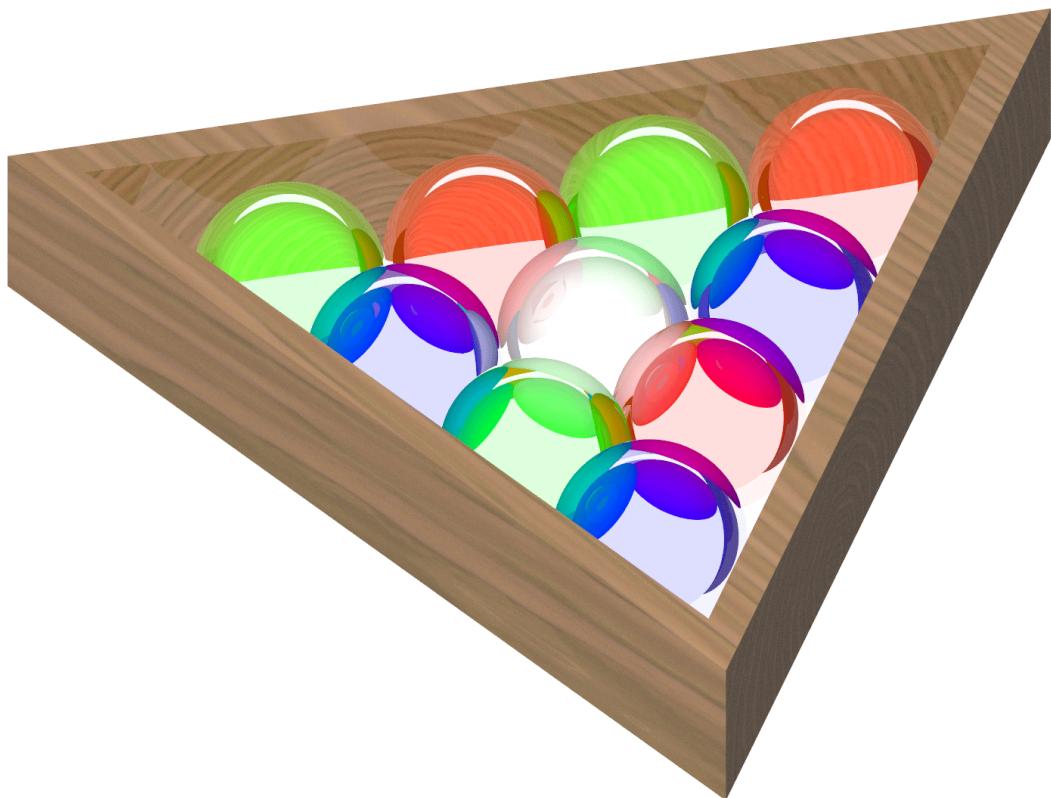


PARTICLES & WIZARDS

INTRODUCTION TO QUANTUM FIELD THEORY



A VERY SHORT PROTO-BOOK BY
N. BOOKER

To my parents

Contents

1 Instead of a foreword	5
1.1 How to use this book (5)	
1.2 Acknowledgements (6)	
1.3 References (7)	
I Preliminaries	8
2 Classical field theory	9
2.1 Action principle (9)	
2.2 Equations of motion (13)	
2.3 Noether's theorem (15)	
2.4 Poincaré transformations (16)	
3 Non-relativistic quantum mechanics	19
3.1 Quantum mechanics of mixed states (19)	
3.2 Time evolution and dynamical pictures (22)	
II Canonical quantisation	26
4 Free fields	27
4.1 Klein-Gordon equation and its demise (27)	
4.2 Fock space (29)	
4.3 Quantisation of the Klein-Gordon field (31)	
5 Interacting fields I: ϕ^4 theory	35
5.1 Self-interaction (35)	
5.2 Scattering matrix (38)	
5.3 Feynman diagrams (39)	
5.4 ϕ^4 theory Feynman rules (41)	
6 Interacting fields II: QED	46
6.1 Dirac equation (46)	
6.2 Quantisation of the Dirac field (51)	
6.3 Story of a spinor (53)	
6.4 Spin, helicity and chirality (54)	
6.5 CPT theorem (60)	
6.6 Maxwell's equations (63)	
6.7 Quantisation of the photon field (67)	
6.8 QED Feynman rules (68)	
III Path integrals	72
7 Free fields	73
7.1 Path integrals in non-relativistic quantum mechanics (73)	
7.2 Generating functional (76)	
7.3 Green's function (78)	
8 Interacting fields I: ϕ^4 theory	80
8.1 Generating functional (80)	
8.2 Perturbative expansion (81)	
8.3 Effective action (83)	
8.4 Dyson-Schwinger equations (86)	
8.5 Global symmetry breaking (89)	
9 Interacting fields II: QED	93
9.1 Fermion propagator (93)	
9.2 Photon field propagator (94)	
9.3 Vertex (98)	

IV Renormalisation and regularisation	100
10 Regularisation	101
10.1 Motivation (101) 10.2 Mathematical toolkit (103) 10.3 Cutoff regularisation (105) 10.4 Pauli-Villars regularisation (105) 10.5 Dimensional regularisation (106)	
11 Renormalisation I: ϕ^4 theory	109
11.1 Emergence of renormalisation (109) 11.2 Renormalisability (110) 11.3 Counterterms (111) 11.4 Renormalisation schemes (113) 11.5 1-loop renormalisation of ϕ^4 theory (114) 11.6 Callan-Symanzik equation (116) 11.7 Renormalisation group (118) 11.8 Källén-Lehmann spectral representation (119)	
12 Renormalisation II: QED	121
12.1 Ward-Takahashi identity (121) 12.2 Ward identity (123) 12.3 Renormalisation constants (125) 12.4 Tensor decomposition (126) 12.5 1-loop renormalisation of QED (128) 12.6 Beyond perturbation theory (132)	
V Standard model	133
13 Non-Abelian gauge theories	134
13.1 Generalisation of the Abelian gauge (134) 13.2 Gauge fixing and Faddeev-Popov ghosts (139) 13.3 Quantisation (140) 13.4 BRST symmetry (144) 13.5 Quantum chromodynamics (146)	
14 Electroweak theory	151
14.1 Quantum flavourdynamics (151) 14.2 Lie theory (152) 14.3 Spinors (155) 14.4 Local gauge symmetry breaking (156) 14.5 Higgs mechanism (158) 14.6 Quark and neutrino mixing (161) 14.7 Feynman rules (163)	
15 Experimental high energy physics	167
15.1 Probability amplitude (167) 15.2 $2 \rightarrow 2$ processes (170) 15.3 Rise of quarks (175) 15.4 Electron-quark scattering (176) 15.5 Feynman diagrams (179) 15.6 Detectors (179)	
16 Instead of a postscript	182
16.1 Beyond the standard model (182) 16.2 Where do we go from here? (183)	

Chapter 1

Instead of a foreword

Quote 1.1 But QFT is not conceptually difficult. It's just classical field theory with commutators, at least if one stays clear of more esoteric things like supersymmetry or string theory. If you already have experience in CFT, QFT should pose no difficulties to you at all.

Paulina Schlachter, 29 September 2024

Quote 1.2 When conquering QFT you'll find yourself in a similar position as Napoleon with Russia.

Felix Halbwedl, 20 October 2024

1.1 How to use this book

Quantum field theory is the unification of quantum mechanics and special relativity. It is not a theory of quantum gravity because it still operates within the completely flat Minkowski space. While the development of high-energy physics preceded that of QFT historically, QFT actually forms the theoretical basis of HEP.

Our ultimate objective in QFT is thus to calculate the final state from some initial state. This is accomplished by applying the scattering matrix S_{fi} to the initial state. The interaction-relevant part of S_{fi} is another matrix \mathcal{M}_{fi} known as the transition amplitude, which is significant in its own right.

The construction \mathcal{M}_{fi} is made easy thanks to Feynman diagrams. As it turns out, every process in high energy physics can be represented by a Feynman diagram, which consists of three types of objects: lines, vertices and loops, which are nothing but lines coming back to themselves¹. The mathematical expression of the Feynman diagram, which happens to be \mathcal{M}_{fi} , can then be constructed by the Feynman rules of the field theory of interest, which assign each type of object to a mathematical expression and multiply them together.

To derive these rules is the heart of quantum field theory. We start with the Lagrangian of our field theory of interest, which can always be split into a free part and an interacting part. The free part corresponds to lines, while the interacting part corresponds to vertices. There are two ways to go from the Lagrangian to the Feynman rules: the canonical quantisation formulation and the path integral formulation.

This book is structured as follows:

- While knowledge of special relativity is assumed, we provide a brief overview of classical field theory and non-relativistic quantum mechanics in Part I. The two topics directly precede the formulation of QFT and usually constitute the beginning of a standard ‘Quantum Field Theory I’ course in most universities. An experienced reader should have no problem reading this part like a novel.
- Part II covers canonical quantisation, in which fields are quantised. Historically, it is also called ‘second quantisation’ as it builds on ‘first quantisation’ in quantum mechanics, where physical quantities are quantised. For this reason, it is the more intuitive approach for quantum electrodynamics

¹Mathematically, this means we integrate the expression of a line over the internal loop momenta.

(QED), and the canonical quantisation of QED is the staple of a standard ‘Quantum Field Theory I’ course in most universities.

- Part III covers the path integral formulation, in which, from the action S of a system, we derive a generating functional $Z[J]$ that takes the form of a path integral. Propagators are then derived by taking functional derivatives of $Z[J]$. This approach is somewhat easier but less intuitive. As such, it is usually at the intersection between ‘Quantum Field Theory I’ and ‘Quantum Field Theory II’ courses in most universities.
- As it turns out, both formulations, which yield the same Feynman rules, fail in structures beyond the tree level of Feynman diagrams due to the emergence of infinities. The elimination of these unphysical infinities requires the theory to be renormalised, which can only be done by first performing the nice mathematical trick of regularisation. The two topics, which are covered in Part IV, are also usually at the intersection between ‘Quantum Field Theory I’ and ‘Quantum Field Theory II’ courses in most universities.
- Now that QED has been formulated and renormalised up to the first loop order, we will complete the standard model in Part V by looking at quantum chromodynamics (QCD) and the weak force, both of which are non-Abelian gauge theories, unlike QED, which is Abelian. The weak force also combines with QED to form electroweak theory, where a spontaneous breaking of the gauge symmetry gives rise to the well-known Higgs mechanism. These topics lie firmly in ‘Quantum Field Theory II’ courses in most universities.

Many parts of the book require knowledge of spinors, symmetries and Lie theory. These topics will not be covered in this book. Rather, the reader is encouraged to consult the companion book *Spinors & Symmetries* or other standard Lie theory texts.

This proto-book is licensed under [CC BY-NC-SA 4.0](#). For any comments, suggestions or typos, please ‘shoot’ an e-mail to

`neil(dot)booker(at)ucl.ac(dot)uk`

Quote 1.3 🐾, aber sicher doch. Ich mache auch Fehler.

Felix Halbwedl, 8 February 2025

1.2 Acknowledgements

Quote 1.4 Credit for those who work hard, I just happen to know some things.

Felix Halbwedl, in his infinite humility, 22 December 2024

Quote 1.5 Not a word of Halbwedl and not a thought of Booker.

The Author, on this book, 11 April 2025

Most of the work on this book was carried out during the 2024-25 winter and summer semesters at University College London. The initial work was based on the lecture notes on Quantum Field Theory and Standard Model courses, authored by Prof. Alessio Serafini² and Prof. Gavin Hesketh respectively. While the book eventually acquired a range of references, I would like to thank them for teaching the courses and for the physical discussions I had with them.

Furthermore, I want to extend my gratitude to Felix Halbwedl, who stimulated many physical discussions on various topics in HEP, contributed countless quotes to the book and offered much advice on its contents and the formatting. I am thankful to Abhijeet Vats, under whose guidance I was able to develop my L^AT_EX skills to a satisfactory level. I am also grateful to Francisco Silva, who also contributed to physical discussions and provided several references.

I would like to thank Felix Halbwedl, Paul Kothgasser, Robert Schwarzl and other members of the Basisgruppe NAWI Physik (BaGru) at the Technische Universität Graz for ‘adopting’ me into their

²Known lovingly as the ‘Wizard’ due to his character appearing as a wizard in the [UCL PandA Day](#) plays.

student community during the writing of this book.

Lastly and most importantly, I dedicate this book to my parents, whose immense love and support throughout my life I would never be able to repay.

1.3 References

- *Introduction to Gauge Field Theory* by David Bailin (University of Sussex) and Alexander Love (University of Sussex)
- *Quantum Field Theory I* by Niklas Beisert (ETH Zürich)
- *Quantum Field Theory II* by Niklas Beisert (ETH Zürich)
- *Quantum Fields* by Nikolay Bogoliubov (JINR) and Dmitry Shirkov (JINR)
- *Quantum Field Theory* by Gernot Eichmann (Technische Universität Graz)
- *Quantenfeldtheorie* by Matthias Gaberdiel (ETH Zürich)
- *Quantum Field Theory II* by Matthias Gaberdiel (ETH Zürich) and Aude Gehrmann-De Ridder (ETH Zürich)
- *Quantum Field Theory I* by Axel Maas (Technische Universität Graz)
- *Quantum Field Theory II: Gauge Theories* by Axel Maas (Technische Universität Graz)
- *Notes on Quantum Field Theory I* by Marco Serone (SISSA)
- An Introduction to Quantum Field Theory by Michael E. Peskin (SLAC) and Daniel V. Schroeder (Weber State University)

Part I

Preliminaries

Chapter 2

Classical field theory

Up to the mid-20th century, significant equations of motion had been usually derived by intuition, and this (understandably) carries over when these equations are introduced in undergraduate physics. We will begin with the action principle, which is possibly the most important idea in the entirety of physics from which the Euler-Lagrange equations then follow as a result of the vanishing of the boundary term. Then, it can be seen that the equations of motion of any area in physics can be derived systematically by inserting its Lagrangian¹ into the *Euler-Lagrange equations*.

2.1 Action principle

The two elements leading to the action principle are the action itself and the concept of symmetries. We begin by transitioning from classical mechanics to classical field theory and work slowly towards the action.

Note 2.1 (Metric signature) Unlike GR, convention dictates that the Minkowski 4-metric in HEP has the signature $(+, -, -, -)$. That is, the line element has the form

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \quad (2.1)$$

We now briefly discuss the *variational formalism*. Two equivalent formulations of the variational formalism exist - Lagrangian and Hamiltonian mechanics. In classical mechanics, the central quantities are the 4-position x (or often q) and momentum p .

Note 2.2 (Reference frames) We can choose certain frames that simplify calculations:

- For a spacelike separation $(x - y)^2 < 0$, one can always, without loss of generality, choose a frame to set $(x^0 - y^0) = 0$.
- For a timelike separation $(x - y)^2 > 0$, one can always, without loss of generality, choose a frame to set $(\vec{x} - \vec{y}) = 0$.

We now take the monumental step of actually understanding what a field theory actually means. The central point is migrating from a coordinate-centric system we have seen up to this point to a field-centric system. In field theories, the 4-position x is replaced with a 4-field $\phi(x) = (\phi_0, \phi_1, \phi_2, \phi_3)$. ϕ_1, ϕ_2 and ϕ_3 are simply the spatial components of the corresponding 3-field, while ϕ_0 is a *scalar* or *time-like* component of the 4-field². For example, we consider the *Lagrangian density*³ with only one *kinetic term*. Previously, this would merely be the kinetic energy:

$$\mathcal{L} = \frac{mv^2}{2} \quad (2.2)$$

In a field theory, we turn the velocity to derivative over 4-coordinates or the *4-derivatives* of the field and absorb the coupling-like mass m to a normalised value of 1. This then constructs the simplest field

¹Which, regrettably, are ultimately formulated via intuition as there is no other way to do so.

²For example, in the electromagnetic 4-potential, ϕ_0 is the electric scalar potential.

³Often also simply called the *Lagrangian*, although you will be able to tell the difference by looking at the notation.

theory Lagrangian, which is that of a free massless scalar field, which can be used to model particles like massless scalar bosons⁴:

Definition 2.1 (Free massless scalar field Lagrangian)

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi \partial^\mu \phi \quad (2.3)$$

The sole term is the *kinetic energy density*, which arises from the variation of the field ϕ over the 4-coordinates.

The only term in this Lagrangian is a so-called kinetic term, which we can now properly define as a term consisting of field derivatives. In analogy to classical mechanics, the rest of the terms make up the *potential*.

One way we can introduce a potential to the Lagrangian is to make the free field massive instead of massless. Note that this does not mean that the field itself is massive (which makes little sense) but rather that the particle that generates the field is massive. This gives rise to a mass coupling⁵ term, and the Lagrangian becomes the standard Lagrangian of a free massive scalar field, which we use to derive the Klein-Gordon equation we will see later:

Definition 2.2 (Free massive scalar field Lagrangian)

$$\mathcal{L} = \frac{1}{2} \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} m^2 \phi^2 \quad (2.4)$$

where the second term is the *potential energy density*.

Let us compare this CFT result with its classical mechanics analogue, which is the Lagrangian of a harmonic oscillator:

$$\mathcal{L} = \frac{1}{2} m \dot{x}^2 - \frac{1}{2} k x^2 \quad (2.5)$$

Essentially, $\frac{1}{2} m^2 |\phi|^2$ represents the field-theory version of Hooke's law. It is a potential energy term that determines how the field oscillates about its vacuum. The mass term m^2 replaces the spring constant k and sets the curvature (strength) of the potential.

From the Lagrangian, we can formulate how quantities related to it are defined in CFT. In classical mechanics, the *canonical momentum* is defined as

$$p = \frac{d\mathcal{L}}{dq} \quad (2.6)$$

where q is the generalised 4-coordinates. In CFT, this becomes

Definition 2.3 (Canonical momentum)

$$\pi = \frac{\partial \mathcal{L}}{\partial(\partial_t \phi)} \quad (2.7)$$

The action in CFT is unchanged from its CM counterpart, but uses the CFT Lagrangian instead of the CM Lagrangian:

Definition 2.4 (Action) For a set of fields ϕ_i with the 4-position x^i , the *action* is defined as

$$S = \int \mathcal{L}(\phi, \partial_\mu \phi) d^4x = \int L dx^0 = \int L dt \quad (2.8)$$

where L is the Lagrangian and \mathcal{L} is the Lagrangian density.

We are now in a position to discuss symmetries. Physicists use the word 'symmetry' as a more sophisticated way of saying that a physical quantity we are interested in stays invariant under the change of

⁴We see it more often in approximate models as fundamental massless scalar particles with zero mass are rare.

⁵We will often see the word 'coupling'. Sometimes, it is the short form of the *coupling constant*, which denotes any physical constant that a term in the Lagrangian may have. We then speak of the field *coupling* to whatever the physical quantity this coupling constant represents. For example, in a mass term, the coupling constant is the mass, and we say that in the term, the field couples to mass.

some other ‘background’ quantity. If this physical quantity of interest instead changes with the background quantity, we then sadly say that symmetry is ‘broken’. Broadly speaking, we are interested in the following symmetries:

- **Global and local gauge symmetries:**

- *Global symmetries* have transformations that are identical everywhere in spacetime. They lead to conserved quantities (like energy, momentum, charge) and usually reflect real physical invariances.
- *Local gauge symmetries* have transformations that can vary from point to point in spacetime. They are not symmetries of nature but symmetries (or rather redundancies) in our description.

- **External and internal symmetries:**

- *External symmetries* are transformations that involve changes to the spacetime coordinates themselves. One example is Poincaré symmetry.
- *Internal symmetries* are transformations that act on internal degrees of freedom of fields (e.g. charge, spin, etc.), leaving spacetime coordinates unchanged.

- **Discrete and continuous symmetries:**

- *Continuous symmetries* are governed by transformation parameters that can admit a continuous range of values. One example is Poincaré symmetry.
- *Discrete symmetries* involve transformations that take on only specific values. Examples are the C, P and T symmetries.

So far, we have heard about the layman’s version of Noether’s theorem ‘All symmetries lead to conservation laws’. We have made a conceptual overview of symmetries, but what are their mathematical implications? As it turns out, symmetries are defined with respect to an action principle. Consider an *infinitesimal coordinate transformation*

$$x^\mu \rightarrow x'^\mu = x^\mu + \epsilon^\mu \quad (2.9)$$

Up to the first order expansion, a generic field in x^μ undergoes the corresponding transformation:

$$\phi(x) \rightarrow \phi'(x') = \phi(x) - \epsilon^\mu \partial_\mu \phi(x) \quad (2.10)$$

The variation of the field can then be written as

$$\delta\phi(x) = \phi'(x') - \phi(x) = -\epsilon^\mu \partial_\mu \phi(x) \quad (2.11)$$

An important point of note concerns the Lagrangian (density). While it has rank 0, the Lagrangian \mathcal{L} in some theories might not transform as a scalar. Rather, its variation takes the general form of the total derivative of a current-like vector field K^μ . Without considering fields, the variation of the Lagrangian *arising from a symmetry transformation* has the general form

$$\delta\mathcal{L} = \partial_\mu K^\mu(\phi, \partial_\mu \phi) \quad (2.12)$$

where:

- Physically, $K^\mu(\phi, \partial_\mu \phi)$ is the measure of the failure of \mathcal{L} to transform as a scalar.
- $\partial_\mu K^\mu(\phi, \partial_\mu \phi)$ is then a ‘total derivative’ of $K^\mu(\phi, \partial_\mu \phi)$, which depends on both the field ϕ and the field’s 4-derivative $\partial_\mu \phi$.

Derivation 2.1 (Action principle) We can solve for the variation of the action by integrating the variation of the Lagrangian (2.12), which gives:

$$\delta S = \int d^4x \delta\mathcal{L} = \int d^4x \partial_\mu K^\mu(\phi, \partial_\mu \phi) \quad (2.13)$$

Using the divergence theorem, this integral can be converted into a surface integral over the boundary

$\mathcal{S} = \partial V$ of the spacetime region V :

$$\delta S = \int_{\partial V} d^3x K^\mu(\phi, \partial_\mu \phi) n_\mu \quad (2.14)$$

where n_μ is the normal vector to the boundary. We are left with a *boundary term* that is exactly $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ where n_μ is a directional 4-vector.

There are two scenarios in which this boundary term can be ignored:

- The variation $\delta\phi$ (and with that, ϕ and its derivative) vanishes on the boundary^a.
- The boundary extends into infinity.

The important step now is to *assume the first point*, which can be justified if we impose boundary conditions. Depending on the physical scenario, we usually use one of the two main boundary conditions.

Definition 2.5 (Dirichlet boundary condition) The *Dirichlet boundary condition* or the *boundary condition of the first type* sets the position to be time-invariant at the boundary:

$$\dot{q}|_{\partial V} = 0 \quad \partial_\mu \phi|_{\partial V} = 0 \quad (2.15)$$

Definition 2.6 (Neumann boundary condition) The *Neumann boundary condition* or the *boundary condition of the second type* sets the momentum is time-invariant at the boundary:

$$\frac{\partial L}{\partial \dot{q}} \Big|_{\partial V} = 0 \quad \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} \Big|_{\partial V} = 0 \quad (2.16)$$

where we see that this boundary condition is realised by exploiting Hamilton's equations^a.

^aWe only derive Hamilton's equations later on. However, our current lack of knowledge of Hamilton's equations does not prevent us from imposing the Neumann boundary condition. Rather, it only temporarily prevents us from interpreting the LHS as the momentum time derivative - which we have revealed as a spoiler anyway.

Either way, the term $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ then vanishes due to its dependency on both ϕ and $\partial_\mu \phi$:

$$\int_{\mathcal{S}} d^3x K^\mu(\phi, \partial_\mu \phi) n_\mu = 0 \quad (2.17)$$

In fact, assuming that we choose suitable boundary conditions, all total derivative terms of the form $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ are boundary terms, and hence identically vanish.

Note 2.3 Boundary terms do not contribute to the equations of motion. As such, we can always add or subtract boundary terms to/from the Lagrangian arbitrarily^a.

^aThe same applies to zero terms for the simple reason that they are zero

From this, we see that the action is invariant under the symmetry:

Theorem 2.1 (Action principle)

$$\delta S = 0 \quad (2.18)$$

This is the almighty *action principle*.

^aThis is typically justified in physical field theories where fields and their variations vanish at spatial or temporal infinity.

Remark 2.1 The *action principle* is simply another name for the *principle of stationary action*, which is itself often erroneously known as the *principle of least action*⁶. This simply means that the time derivative of the action of an isolated system is zero. As the principle can be used for action generated by *any* field, it is often mentioned as '*an* action principle' instead of '*the* action principle'.

⁶This is because the principle states that instead of at a minimum, action tends to stay *stationary*, be it a maximum, a minimum or a saddle point.

2.2 Equations of motion

The physical significance of our previous derivation is not immediately obvious, especially with respect to how the boundary term vanishes by dint of the Dirichlet and Neumann boundary condition. One can lift this shroud of confusion by realising that the action principle allows us to derive the equations of motion. Let us show this with the general example in classical mechanics.

Derivation 2.2 (Euler-Lagrange equations) Now that we have assumed the vanishing of the boundary term, let us evaluate the Lagrangian variation $\delta\mathcal{L}(\phi, \partial_\mu\phi)$ explicitly. We consider an infinitesimal variation of the field:

$$\phi(x) \rightarrow \phi(x) + \delta\phi(x) \quad (2.19)$$

The variation of the Lagrangian is exactly analogous to differentiating a function of two variables:

$$f(x, y) \rightarrow f(x + \delta x, y + \delta y) \rightarrow \delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y \quad (2.20)$$

As \mathcal{L} has two variables, the field ϕ and the field derivative ∂_μ , its total variation is given by

$$\delta\mathcal{L} = \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \right) \quad (2.21)$$

Using the product rule gives

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi} \delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta(\partial_\mu\phi) \quad (2.22)$$

Noting that $\delta(\partial_\mu\phi) = \partial_\mu(\delta\phi)$, we can write

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi} \delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial_\mu(\delta\phi) \quad (2.23)$$

Inserting this result into (2.13) gives

$$\delta S = \int d^4x \left[\frac{\partial\mathcal{L}}{\partial\phi} \delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial_\mu(\delta\phi) \right] \quad (2.24)$$

We can apply integration by parts to the second term and apply Gauss's law^a. This gives

$$\delta S = \int d^4x \frac{\partial\mathcal{L}}{\partial\phi} \delta\phi + \underbrace{\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi}_{\text{boundary term}} \Big|_{\partial V} - \int d^4x \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \right) \delta\phi \quad (2.25)$$

where we recall that ∂V denotes the boundary.

Through this process, we have explicitly exposed the so-called boundary term. But there is something very peculiar about this boundary term, isn't there? Let us inspect the two components:

- $\delta\phi$ is nothing but the time derivative (variation) of the position (field), which is zero at the boundary under the Dirichlet boundary condition (2.15).
- $\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}$ is nothing but the time derivative of the momentum, which is zero at the boundary under the Neumann boundary condition (2.16).

So you see, no matter which boundary condition we take, this boundary term is ultimately nothing but zero. The important conclusion you should yield from our discussion is the following:

Note 2.4 Due to its vanishing, the boundary term *does not contribute to the equations of motion*.

In contrast, the rest of the expression, which contribute to the equations of motion, is known as the *bulk term*^b.

The final step we have left is to simply apply the action principle, which then leads to the famous Euler-Lagrange equations:

Theorem 2.2 (Euler-Lagrange equations)

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial (\partial_\mu \phi)} = 0 \quad (2.26)$$

^aAs such, the boundary term is also called the *surface term*.

^bThis expression is rarely used, and almost always in gravitation.

Before proceeding, we make some comments on the bulk and boundary terms:

- The bulk term is so-called because it integrates over the entire volume of spacetime. It is the term that contributes to the action. When an action principle is imposed, the integrand of the bulk term vanishes, as seen in (2.26).
- The boundary term reflects the influence of boundary conditions for the action. It is the term that does not contribute to the action. By imposing an action principle, we have also assumed that $\delta\phi = 0$ on the boundary - a boundary condition.

Like in CM, the *Hamiltonian* is essentially a Legendre transformation of the Lagrangian:

Definition 2.7 (Hamiltonian and Hamiltonian density) The *Hamiltonian* H is

$$H = \int \mathcal{H}(\phi, \pi, \partial_\mu \phi) d^3x \quad (2.27)$$

which is the volume integral of the *Hamilton density* \mathcal{H} . Also simply called the *Hamiltonian*, it is given by

$$\mathcal{H}(\phi_i, \pi_i, t) = \sum_i \pi_i \dot{\phi}_i(\phi_j, \pi_j) - \mathcal{L}(\dot{x}_k(x_j, p_j), x_k, t) \quad (2.28)$$

where ϕ is the field and π is the canonical momentum.

Remark 2.2 As it turns out, the quantity which we have been led to believe to be the Hamiltonian as undergrads is actually the Hamiltonian density \mathcal{H} .

Derivation 2.3 (Hamilton's equations) By taking the variation of (2.28), one finds

$$\delta\mathcal{H} = \sum_i \delta\pi_i \dot{\phi}^i(\phi_j, \pi_j) - \delta\pi^i \frac{\partial \mathcal{L}}{\partial \pi^i} = \sum_i \delta\pi_i \dot{\phi}^i(\phi_j, \pi_j) - \sum_i \delta\phi^i(\phi_j, \pi_j) \dot{\pi}_i \quad (2.29)$$

Now compare this against the general variation:

$$\delta\mathcal{H} = \delta\phi^i(\phi_j, \pi_j) \frac{\partial \mathcal{H}}{\partial \phi^i(\phi_j, \pi_j)} + \delta\pi_i \frac{\partial \mathcal{H}}{\partial \pi_i} \quad (2.30)$$

By equating the two expressions for $\delta\mathcal{H}$, we recover the so-called *Hamilton's equations*:

Theorem 2.3 (Hamilton's equations)

$$\frac{\partial \mathcal{H}}{\partial \phi_i} = -\dot{\pi}_i \quad \frac{\partial \mathcal{H}}{\partial \pi_i} = \dot{\phi}_i \quad (2.31)$$

Defining the Poisson bracket $\{f, g\}$ of some two quantities f and g as

Definition 2.8 (Poisson bracket)

$$\{f, g\} = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right) \quad (2.32)$$

we can rewrite Hamilton's equations as

$$\dot{x}_i = \{x_i, H\} \quad \dot{p}_i = \{p_i, H\} \quad (2.33)$$

As we will see later, the quantum version of the first equation is simply the Heisenberg equation, or the Schrödinger equation in the Heisenberg picture.

2.3 Noether's theorem

We can now finally derive Noether's theorem. The first step is to realise that the variation of the Lagrangian has been written in two formulations (2.21) and (2.12) respectively. Combining them yields

$$\delta\mathcal{L} = \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \right) = \partial_\mu K^\mu \rightarrow \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \right) - \partial_\mu K^\mu = 0 \quad (2.34)$$

Now let us define another current-like quantity called the *Noether current*:

Definition 2.9 (Noether current) The 4-vector equivalent of the probability density ϱ is the so-called Noether current, *probability 4-current* or *conserved current* J^μ . Its zeroth component is simply the good ol' probability density, and its 3 other (spatial) components are the *probability (3-)current* J^i .

From the Noether current, one can derive a charge-like quantity representing the total probability called the *Noether charge*, the *probability charge* or the *conserved charge* as it is conserved with respect to time:

Definition 2.10 (Probability charge)

$$Q = \int d^3x J^0 \quad (2.35)$$

As probability is conserved, J^μ is Lorentz-invariant and satisfies the *continuity equation*:

Theorem 2.4 (Continuity equation)

$$\partial_\mu J^\mu = 0 \quad (2.36)$$

Rather cheatingly, we can now equate (2.34) and (2.36):

$$\partial_\mu J^\mu = \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \right) - \partial_\mu K^\mu = 0 \quad (2.37)$$

By removing the partial derivatives, we recover the expression for the Noether current, known as Noether's theorem:

$$J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi - K^\mu \quad (2.38)$$

In the most common case, the Lagrangian is strictly invariant, which leads to the vanishing of the vector field K^μ . We then have:

Theorem 2.5 (Noether's theorem)

$$J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \delta\phi \quad (2.39)$$

The implications are twofold:

- In principle, this is used to calculate (or ‘read off’ if the reader is experienced) the 4-current (or Noether current) for any field that transforms under a global symmetry.
- In practice, these fields are, among others, the Klein-Gordon and Dirac scalar fields.

Quote 2.1 Yes, but it is the same sound

— David Steiner, comparing the ‘oe’ in ‘Noether’ with ‘ö’, 21 November 2024

Note 2.5 Noether's theorem implies the conservation of the charge associated with the probability current:

$$\frac{dQ}{dt} = \int d^3x \partial_0 J^0 = 0 \quad (2.40)$$

Finally, we can directly relate the Noether current and the action, from (2.25) and (2.26), one can see that

$$\frac{\partial\mathcal{L}}{\partial\phi} = \partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = \frac{\delta S}{\delta\phi} \quad (2.41)$$

which, by insertion into (2.37), gives

$$\partial_\mu J^\mu = \delta\phi \frac{\delta S}{\delta\phi} = \frac{\delta\phi}{\delta\theta} \frac{\delta S}{\delta\phi} \quad (2.42)$$

where θ is the parameter for an infinitesimal transformation satisfying

$$x \rightarrow x' = x + \theta\delta x \quad (2.43)$$

Exercise 2.1 Let ϕ be a free scalar field obeying the Klein-Gordon equation (4.2), and let J^μ be the associated density and current 4-vector. Derive the continuity equation (2.36).

Remark 2.3 One essential type of symmetry in QFT is the so-called *gauge symmetries*, which are both internal symmetries and continuous symmetries. Gauge symmetries are governed by Lie groups we have seen in *Spinors & Symmetry*.

2.4 Poincaré transformations

As an example, we now derive the Noether current under Poincaré transformations, which, as seen in *Spinors & Symmetries*, includes translations and Lorentz transformations.

Derivation 2.4 (Translation) Let us assume the same transformations as (2.9) and (2.10). The field variation is then shown in (2.11), which we substitute into the transformation of \mathcal{L} in (2.21). We obtain

$$\delta\mathcal{L} = -\epsilon^\nu \left(\frac{\partial\mathcal{L}}{\partial\phi} \partial_\nu\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial_\mu\partial_\nu\phi \right) \quad (2.44)$$

Now substitute this into (2.13):

$$\delta S = -\epsilon^\nu \int d^4x \left(\frac{\partial\mathcal{L}}{\partial\phi} \partial_\nu\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial_\mu\partial_\nu\phi \right) \quad (2.45)$$

By inserting the equivalence from the Euler-Lagrange equations, we can rewrite the integral as

$$\delta S = -\epsilon^\nu \int d^4x \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial_\nu\phi - \delta_\nu^\mu \mathcal{L} \right) \quad (2.46)$$

where the terms within the bracket is the canonical stress-energy tensor. A simple shift of indices gives its contravariant form:

Definition 2.11 (Stress-energy tensor)

$$T^{\mu\nu} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} \partial^\nu\phi - g^{\mu\nu} \mathcal{L} \quad (2.47)$$

Importantly, by comparing (2.13) and (2.46), one can identify

$$\delta\mathcal{L} = \partial_\mu(\epsilon_\nu T^{\mu\nu}) = \partial_\mu K^\mu \quad (2.48)$$

In other words, our new friend $\epsilon_\nu T^{\mu\nu}$ resembles the previously seen boundary/surface term K^μ . Finally, if one takes a partial derivative of the stress-energy tensor and applies the Euler-Lagrange equations, they will find that

$$\partial_\mu T^{\mu\nu} = 0 \quad (2.49)$$

This shows the well-known conservation of matter-energy content, and follows directly from Noether's theorem applied to spacetime translations.

Note 2.6 (2.49) merely states that $T^{\mu\nu}$ is invariant when differentiated^a over 4-coordinates. As such, it describes the conservation of energy-momentum *density* as opposed to energy and momentum themselves. In cosmology, where inflation exists, the total energy of the universe is *not* conserved. Rather, as inflation gives rise to the expansion of spacetime, the energy *increases* to preserve the invariance of

energy-momentum density.

^aIn curved space, the partial derivative becomes the covariant derivative in GR.

Derivation 2.5 (Lorentz transformation) A Lorentz transformation can be either a rotation or a Lorentz boost. Assuming the parameterisation

$$\Lambda_\nu^\mu = \delta_\nu^\mu + \omega_\nu^\mu \quad \omega^{\mu\nu} = -\omega^{\nu\mu} \quad (2.50)$$

where $\omega^{\nu\mu}$ is some parameter, the coordinate and field transformations are

$$x^\mu \rightarrow x'^\mu = x^\mu + \Lambda_\nu^\mu x^\nu \quad \phi(x) \rightarrow \phi'(x) = \phi(x) + \frac{1}{2}\omega^{\rho\sigma}\Sigma_{\rho\sigma}\phi \quad (2.51)$$

where $\Sigma_{\rho\sigma}$ are the generators of the representations of the Lorentz group abstract elements corresponding to ϕ (e.g., for scalars $\Sigma_{\rho\sigma} = 0$, for vectors $\Sigma_{\rho\sigma}$ corresponds to antisymmetric tensors, and so on). Again, by using Noether's theorem, we can find that the contribution to the Noether current is the total angular momentum, which includes the orbital angular momentum and spin:

$$M^{\mu\rho\sigma} = x^\rho T^{\mu\sigma} - x^\sigma T^{\mu\rho} + \frac{\partial \mathcal{L}}{\partial(\partial_\mu\phi)}\Sigma^{\rho\sigma}\phi \quad (2.52)$$

For Lorentz transformations, $K^\mu = \partial_\nu(x^\nu J^\mu - x^\mu J^\nu)$, and the Noether current becomes:

$$J^\mu = \omega_{\rho\sigma} M^{\mu\rho\sigma} \quad (2.53)$$

where the nature of $\omega_{\rho\sigma} M^{\mu\rho\sigma}$ as a boundary term is easily seen.

Hence, combining both types of symmetries, the most general form of the Noether current under Poincaré transformations can be written as:

$$J^\mu = \epsilon_\nu T^{\mu\nu} + \frac{1}{2}\omega_{\rho\sigma} M^{\mu\rho\sigma} \quad (2.54)$$

where $T^{\mu\nu}$ represents energy-momentum contributions, and $M^{\mu\rho\sigma}$ represents both orbital and intrinsic angular momentum contributions.

We will now discuss an important point that will pop up over and over again in the rest of the book. As should be well understood at this point, QFT is a quantum theory that incorporates SR, which is based on Lorentz transforms:

- A quantity is *Lorentz-covariant* if it transforms under the Lorentz group representation corresponding to its type (scalar, vector, axial vector, spinor, rank-2 tensor, etc.).
- A quantity is *Lorentz-invariant* if it is invariant under Lorentz transformations.
- A equation is Lorentz-covariant/invariant if all its quantities are Lorentz-covariant/invariant.

While Lorentz invariance is intuitive, we find it prudent to define Lorentz covariance rigorously:

Theorem 2.6 (Lorentz covariance) A field $\phi(x)$ transforms under a Lorentz transformation Λ as:

$$\phi'(x') = \rho(\Lambda)\phi(x)\rho(\Lambda)^{-1} \quad (2.55)$$

where the dash represents the new field/coordinate and $\rho(\Lambda)$ is a *representation* of the Lorentz group ($\text{SO}(1, 3)$) acting on the field.

As seen in *Spinors & Symmetries*, this representation differs:

- A scalar field $\phi(x)$ transforms trivially:

$$\phi'(x') = \phi(x) \quad (2.56)$$

- A vector field $A^\mu(x)$, transforms as:

$$A'^\mu(x') = \Lambda^\mu_\nu A^\nu(x) \quad (2.57)$$

- A spinor field $\psi(x)$, transforms as

$$\psi'(x') = T(\Lambda)\psi(x) \quad (2.58)$$

where $T(\Lambda)$ is a spinorial representation we will derive in Derivation 6.3.

We end with some physical comments. Consider a quantum field theory *within the standard model*:

- Fields are Lorentz-covariant because they transform under representations of the Lorentz group.
- Equations of motion and physical laws are Lorentz-covariant as they involve only Lorentz-covariant quantities. This Lorentz covariance is effectively enforced due to SR.
- The Lagrangian is Lorentz-invariant, ensuring that the theory as a whole respects Lorentz symmetry.

A more complete overview of the representation theory of the Lorentz group can be seen in *Spinors & Symmetries*.

Chapter 3

Non-relativistic quantum mechanics

Quote 3.1 Fortunately, quantum mechanics is easy and can be summarised in a few lines.

Alessio Serafini

The other prerequisite of QFT is non-relativistic quantum mechanics, which is actually a special case of QFT where there is only one dimension - time.

3.1 Quantum mechanics of mixed states

A quantum state can always be represented by a Hermitian, positive semi-definite operator with trace 1 ρ^1 (i.e. all eigenvalues of ρ are positive semi-definite and add up to 1).

Definition 3.1 (Positive definiteness and positive semi-definiteness) A positive definite operator ρ always yields a positive expectation value

$$\langle \psi | \rho | \psi \rangle > 0 \quad (3.1)$$

A positive semi-definite operator ρ always yields a non-negative expectation value

$$\langle \psi | \rho | \psi \rangle \leq 0 \quad (3.2)$$

We now introduce the so-called *Sylvester's criterion*.

Definition 3.2 (Minor) A *minor* of some matrix is the determinant of the resultant matrix after deleting an arbitrary number of rows and columns from the initial matrix. For a square matrix, a minor is called a *principal minor* when the indices of the deleted rows and those of the deleted columns are *identical*^a.

^aThis is significant in that if one deletes, say, the 3rd row and the 4th column, the resulting matrix is a minor but not a principal minor.

Theorem 3.1 (Sylvester's criterion) One can use minors of a matrix to test positive definiteness and positive semi-definiteness. For a Hermitian $n \times n$ matrix:

- Positive definiteness holds if all the *leading* principal minors are positive. i.e. if the determinants of the top-left $1 \times 1, \dots, n \times n$ sub-matrices are positive.
- Positive semi-definiteness holds if *all* principal minors are non-negative.

While a pure quantum state is simply a bra or a ket, the bra-ket notation is insufficient for a *mixed state*, which is described by a *density matrix* or a *density operator*. Unlike a state vector, which is an element of the Hilbert space, the density matrix is an operator *on* the Hilbert space.

Physically, a mixed state is a statistical mixture (i.e. ensemble) of i different pure states $|\Psi_i\rangle$ with

¹Note that this might not be the *density operator* ρ .

probabilities p_i , such that:

$$\rho = \sum_i p_i |\Psi_i\rangle\langle\Psi_i| \quad (3.3)$$

In this way, the density matrix extends the concept of quantum state to mixed states, systems where we do not have complete knowledge.

Derivation 3.1 (Von Neumann equation) Let us first derive the *theorem of Liouville*^a in classical field theory, which states that the phase space distribution function $\rho(p, q, t)$ is constant along the trajectories of the system:

$$\frac{d\rho}{dt} = 0 \quad (3.4)$$

This expression is not useful on its own. Rather, we write down the total derivative of ρ against t explicitly

$$\frac{d\rho}{dt} = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial q_i}{\partial t} - \frac{\partial f}{\partial p_i} \frac{\partial p_i}{\partial t} \right) + \frac{\partial \rho}{\partial t} \quad (3.5)$$

Using Hamilton's equations (2.31), we can rewrite the total derivative as

$$\frac{d\rho}{dt} = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} \right) + \frac{\partial \rho}{\partial t} \quad (3.6)$$

We can absorb the first term on the RHS into the Poisson bracket. This gives us

Theorem 3.2 (Theorem of Liouville)

$$\frac{d\rho}{dt} = \{\rho, \mathcal{H}\} + \frac{\partial \rho}{\partial t} = 0 \quad (3.7)$$

where the final RHS term vanishes when

When ρ has no *explicit* time dependence (like in Hamiltonian systems), the term $\frac{\partial \rho}{\partial t}$ vanishes, and we simply have

$$\{\rho, \mathcal{H}\} = 0 \quad (3.8)$$

In QFT, the density operator ρ is functionally equivalent to the phase space distribution function. Hence we have an equivalent for the theorem of Liouville, which, importantly, does *not* generally give $\frac{d\rho}{dt} = 0$:

Theorem 3.3 (Von Neumann equation)

$$i \frac{d\rho}{dt} = [H, \rho] + i \frac{\partial \rho}{\partial t} \quad (3.9)$$

where, again, the last RHS term vanishes when ρ has no *explicit* time dependence.

Due to the aforementioned equivalence, the equation is also called the *quantum Liouville equation* or the *Liouville-von Neumann equation*. It is also the mixed state equivalent of the Schrödinger equation, which deals with pure states.

Remark 3.1 In QM, the Hamiltonian H is an operator, and the Hamiltonian density does not exist.

^aOr more commonly in the English-speaking world, *Liouville's theorem*.

Derivation 3.2 (Reduction to pure state) To prove the last statement, we consider a pure state which has no explicit dependence on time and where we only have one possible i . The density matrix is

$$\rho = |\Psi\rangle\langle\Psi| \quad (3.10)$$

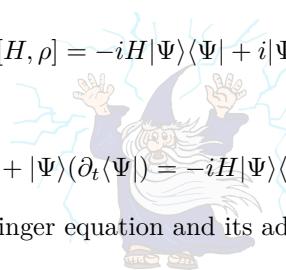
Substituting this into the LHS of the von Neumann equation gives

$$\partial_t \rho = (\partial_t |\Psi\rangle)\langle\Psi| + |\Psi\rangle(\partial_t \langle\Psi|) \quad (3.11)$$

Substituting into the RHS yields

$$-i[H, \rho] = -iH|\Psi\rangle\langle\Psi| + i|\Psi\rangle\langle\Psi|H \quad (3.12)$$

We now equate the two sides:



$$(\partial_t|\Psi\rangle)\langle\Psi| + |\Psi\rangle(\partial_t\langle\Psi|) = -iH|\Psi\rangle\langle\Psi| + i|\Psi\rangle\langle\Psi|H \quad (3.13)$$

where we have recovered the Schrödinger equation and its adjoint form

$$\partial_t|\Psi\rangle = -iH|\Psi\rangle \quad \partial_t\langle\Psi| = i\langle\Psi|H \quad (3.14)$$

Now we remind ourselves of how measurements work. We have already seen the so-called *projection-valued measures* or *PVMs*. Previously, they have been known to us as *projectors*.

Definition 3.3 (Projection-valued measure) A projection-valued measure P_i is a linear and positive semi-definite operator that satisfies, for the density operator ρ :

- Normalisation condition:

$$\sum_i P_i = \mathbb{I} \quad \text{or} \quad \sum_i \text{Tr}(\rho P_i) = 1 \quad (3.15)$$

where \mathbb{I} is the unit matrix of appropriate dimension.

- Orthogonality condition:

$$P_i P_j = \delta_{ij} P_i \quad \text{or} \quad P_i^2 = P_i \quad (3.16)$$

Functionally, it maps a quantity in a vector space V into a subspace $W \subset V$. Each P_i corresponds to an eigenvalue of an observable (e.g. position or spin).

Theorem 3.4 (Born rule) The probability of obtaining the outcome i is, for a mixed state:

$$p_j = \text{Tr}(\rho P_j) = \text{Tr}(|j\rangle\langle\phi|k\rangle\langle j|) = \langle j|\phi\rangle\langle\phi|j\rangle = |\langle j|\phi\rangle|^2 \quad (3.17)$$

For a pure state, this simplifies to

$$p_i = \langle\Psi|P_i|\Psi\rangle \quad (3.18)$$

Remark 3.2 Here we see the significance of (3.15), which is that all probabilities expectedly sum up to 1. So far, we have been working with PVMs, which are utterly perfect, innocent and idealised measurements. In real life, measurement devices are not ideal, giving rise to noise. These imperfect (or rather *generalised*) measurements are described by *positive operator-valued measures* or *POVMs*.

Quote 3.2 ‘Positive Operator Valued Measure’, an acronym fabricated by mathematical physicists to scare all others away.

Alessio Serafini

Definition 3.4 (Positive operator-valued measure) A positive operator-valued measure \prod_i is a linear and positive semi-definite operator that satisfies the normalisation condition only:

$$\sum_i \prod_i = \mathbb{I} \quad \text{or} \quad \sum_i \text{Tr}(\rho \prod_i) = 1 \quad (3.19)$$

By imposing the condition

$$\text{Tr}\left[\prod_i \prod_j\right] = \delta_{ij} \quad (3.20)$$

POVMs reduce to PVMs/projectors.

Remark 3.3 The Born rule is the same as for PVMs, save for the nominal replacement of P_i s by \prod_i s.

3.2 Time evolution and dynamical pictures

In innocent non-relativistic quantum mechanics, the generic Hamiltonian consists of the kinetic energy and the potential:

$$H = \frac{p^2}{2m} + V(\mathbf{x}) \quad (3.21)$$

In field theories, we can generalise this. The kinetic energy, which represents a free particle, becomes the *free Hamiltonian* H_0 while the potential energy, which represents particle interactions, becomes the *interaction Hamiltonian* H_{int} . The total Hamiltonian hence reads

$$H = H_0 + H_{\text{int}} \quad (3.22)$$

The splitting of our beloved Hamiltonian into so-called free and interacting parts gives rise to some convenient tricks in calculations. This is possible because our interpretation of which quantity is evolving in time in quantum mechanics is ultimately philosophically arbitrary. Depending on which term(s) we believe to be actually evolving with time, there exist three *dynamical pictures* or *representations*: the *Schrödinger picture*, the *Heisenberg picture* and *interaction picture*. We consider:

- A series of ‘in’ states $|\psi, \text{in}\rangle$ which denote *prepared* or *incoming* particles at some initial time t_0 .
- A series of ‘out’ states $\langle \alpha, \text{out}|$ which denote *detected* or *outgoing* particles at current time t . These states are the same ‘in’ states, but *after* undergoing time evolution.
- Finally, for simplicity, we assume that the only operation that has happened between the ‘in’ and ‘out’ states is unitary time evolution whose operator is $U(t, t_0)$.

If $t = t_0$, time evolution has not happened, and the three pictures are identical.

Definition 3.5 (Schrödinger picture) The representation we have used throughout undergrad QM is the Schrödinger picture, where the unitary time evolution operator in natural units is

$$U(t, t_0) = e^{-iH(t-t_0)} \quad (3.23)$$

Time evolution is represented as follows:

- Operators are time-invariant.
- ‘in’ states evolve under the *total Hamiltonian* H :

$$|\psi, t\rangle = e^{-iH(t-t_0)} |\psi, \text{in}\rangle = U(t, t_0) |\psi, \text{in}\rangle \quad (3.24)$$

- ‘out’ states are time-invariant. This is because they are actually ‘in’ states that have *already* undergone time evolution in disguise. That is to say, $U(t, t_0)$ is absorbed into the ‘out’ state by definition:

$$\langle \alpha, \text{out}| = \langle \psi, \text{in}| e^{iH(t-t_0)} = \langle \psi, \text{in}| U^\dagger(t, t_0) \quad (3.25)$$

As we all know, the Schrödinger picture equation of motion is simply the *Schrödinger equation*. By differentiating ψ by time and accounting for $U(t, t_0)$, we naturally find

Theorem 3.5 (Schrödinger equation)

$$i \frac{d\psi}{dt} = H\psi \quad (3.26)$$

Definition 3.6 (Heisenberg picture) The *Heisenberg picture* is the opposite of the Schrödinger picture. While $U(t, t_0)$ has the same form, it is the operators that evolve instead of states. Time evolution is represented as follows:

- Operators evolve under the *total Hamiltonian* H :

$$O_H = e^{iH(t-t_0)} O e^{-iH(t-t_0)} = U^\dagger(t, t_0) O U(t, t_0) \quad (3.27)$$

- ‘in’ states are time-invariant.

- ‘out’ states evolve under the *total Hamiltonian* H :

$$_H\langle \alpha, t | = \langle \alpha, \text{out} | e^{-iH(t-t_0)} = \langle \alpha, \text{out} | U(t, t_0) \quad (3.28)$$

Note 3.1 (‘Out’ state as half an operator) There is an important subtlety here which only emerges when we inspect the Heisenberg picture. Even though our ‘out’ state is nominally a state, it is effectively regarded as ‘half of a density operator’ instead of a state:

$$P_\alpha = |\alpha, \text{out}\rangle \langle \alpha, \text{out}| \quad (3.29)$$

Derivation 3.3 (Heisenberg equation) The Heisenberg picture equivalent of the Schrödinger equation is the so-called *Heisenberg equation*. Instead of differentiating ψ by time, we differentiate O_H since it is now the time-dependent quantity.

$$\frac{\partial O_H}{\partial t} = \frac{\partial U^\dagger}{\partial t} O_S U + U^\dagger O_S \frac{\partial U}{\partial t} + \frac{\partial O_H}{\partial t} = U^\dagger H O_S U - U^\dagger O_S H U + \frac{\partial O_H}{\partial t} = U^\dagger [H, O_S] U + \frac{\partial O_H}{\partial t} \quad (3.30)$$

Reabsorbing U and U^\dagger gives us

Theorem 3.6 (Heisenberg equation)

$$i \frac{dO_H}{dt} = [O_H, H] + \frac{\partial O_H}{\partial t} \quad (3.31)$$

where, again, the last RHS term vanishes when O_H has no explicit dependence on time.

Finally, we note that this equation then corresponds well with (3.7) and (3.9).

Remark 3.4 Expectedly, the Heisenberg and Schrödinger equations are identical at $t = t_0$.

The *interaction picture* lies between the Schrödinger and Heisenberg pictures. The central idea is to decompose the Hamiltonian into two parts, the *free Hamiltonian* H_0 and the *interacting Hamiltonian* H_I . The total Hamiltonian H is then

$$H = H_0 + H_I \quad (3.32)$$

Note 3.2 (Determining H_0 and H_I) H_0 is usually number operators like $a^\dagger a$, $b^\dagger b$, etc. (representing the energy of independent modes) while H_I is usually mixed terms like $a^\dagger b$, ab^\dagger , $a^\dagger ab^\dagger b$, etc. (representing products of operators from different modes or nonlinearities) and external driving terms $a^\dagger e^{-i\omega t} + \text{h.c.}$ (representing interaction with classical fields). This will make more sense when you finish reading this part.

Definition 3.7 (Interaction picture) The evolution operator in the interaction picture is

$$U(t, t_0) = e^{iH_0(t-t_0)} e^{-iH(t-t_0)} \quad (3.33)$$

Time evolution is represented as follows:

- Operators evolve under the *free Hamiltonian* H_0 :

$$O_H = e^{iH_0(t-t_0)} O e^{-iH_0(t-t_0)} \quad (3.34)$$

Importantly, however, H_0 itself is always time-independent even though it is an operator. This is because the interaction picture is built *around* the dynamics of H_0 .

- ‘in’ states evolve under both the *total* and *free Hamiltonians*:

$$|\psi, t\rangle = e^{iH_0(t-t_0)} e^{-iH(t-t_0)} |\psi, \text{in}\rangle \quad (3.35)$$

where we often label *interaction picture evolution operator* $U(t, t_0) = e^{iH_0(t-t_0)} e^{-iH(t-t_0)}$.

- ‘out’ states evolve under the *free Hamiltonian* H_0 :

$${}_I\langle \alpha, t | = \langle \alpha, \text{out} | e^{-iH_0(t-t_0)} \quad (3.36)$$

Note 3.3 As noted previously, the three pictures are functionally identical. To verify this, simply acknowledge that:

- If we write the ‘out’ states as-is, there is a ‘conservation of time evolution terms’: Multiplying all 3 time-evolved terms should leave only a prefactor of $e^{-iH(t-t_0)}$, where H is expectedly the *total Hamiltonian*.
- If we write out the ‘out’ states as a time evolution of ‘in’ states, then there is a prefactor of 1.

While the Heisenberg picture generally simplifies calculations, the interaction picture is advantageous when the Hamiltonian includes an interaction term.

It is now possible to define *symmetries of concrete systems* in all three dynamical pictures.

Theorem 3.7 (Conserved quantity) An observable represented by some operator \hat{A} is said to be a conserved quantity if the time derivative of its expectation value is zero:

$$\frac{d}{dt} \langle \hat{A} \rangle = 0 \quad \text{or} \quad \partial_t \hat{A} = 0 \quad (3.37)$$

This has two equivalent statements in the form of commutators:

- In the Schrödinger and Heisenberg pictures:

$$[\hat{A}, \hat{H}] = 0 \quad (3.38)$$

- In the interaction picture:

$$[\hat{A}_I(t), \hat{H}_0] = 0 \quad (3.39)$$

Derivation 3.4 (Time ordering) By differentiating (3.33) against t , one can derive an alternate expression of $U(t, t_0)$ in terms of itself at a different point in time:

$$U(t, t_0) = \mathbb{I} - i \int_{t_0}^t dt_1 H_{\text{int}, I}(t_1) U(t_1, t_0) \quad (3.40)$$

By inserting $U(t, t_0)$ into $U(t_1, t_0)$ over and over and over again, one yields the so-called *Dyson series*:

$$U(t, t_0) = \mathbb{I} + \sum_{j=1}^{\infty} (-i)^j \int_{t_0}^t dt_j \cdots \int_{t_0}^t dt_1 H_{\text{int}, I}(t_j) \cdots H_{\text{int}, I}(t_1) \quad (3.41)$$

To ensure that the integrations are performed at the correct temporal order, we introduce the so-called *time ordering symbol*:

Definition 3.8 (Time ordering symbol) Consider a series operators $A_1(x_1) \cdots A_n(x_n)$, each of which can be represented in the form of creation and annihilation operators like $A_n(x_n) = A_n^+(x_n) + A_n^-(x_n)$. The time ordering symbol is a loosely defined convenience which reorders the operators according to their 4-position:

$$T[A_1(t_1) \cdots A_n(t_n)] = (-1)^p A_{i_1}(t_{i_1}) \cdots A_{i_n}(t_{i_n}) \quad \text{for } x_{i_1} \leq \cdots \leq x_{i_n} \quad (3.42)$$

where p , the parity, can be represented by Grassmann parity^a:

$$p = \sum_{i < j} \pi(A_i) \pi(A_j) \mod 2 \quad (3.43)$$

where the Grassmann parity $\pi(A_i)$ observes

- $\pi(A_i) = 0$ for bosonic operators, which commute as they are Grassmann-even.
- $\pi(A_i) = 1$ for fermionic operators, which anticommute as they are Grassmann-odd.

^aSee *Spinors & Symmetries*.

Remark 3.5 One can also define this without Grassmann mathematics, albeit less satisfyingly:

- Bosonic-bosonic swaps do not contribute to p . No sign change occurs as bosonic operators commute.
- Bosonic-fermionic swaps do not contribute to p . No sign change occurs as bosonic and fermionic operators commute.
- Due to anticommutation, fermionic-fermionic swaps contribute to p by the following:

$$p = \text{num. of fermionic-fermionic swaps} \pmod{2} \quad (3.44)$$

Remark 3.6 When a system is entirely comprised of bosons or fermions, this simplifies:

- For bosonic operators, $p = 0$.
- For fermionic operators, $p = 0$ if the number of swaps is even and $p = 1$ if the number of swaps is odd.

Noting that the Dyson series can be represented compactly via an exponential, we represent $U(t, t_0)$ as a time-ordered exponential:

$$U(t, t_0) = T \left[\exp \left(-i \int_{t_0}^t dt' H_{\text{int}, I}(t') \right) \right] \quad (3.45)$$

In simplified scenarios, T can also be represented mathematically with the *Heaviside step function* $\theta(t)$:

$$T[A_1(t_1)A_2(t_2)] = \theta(t_1 - t_2)A_1(t_1)A_2(t_2) \pm \theta(t_2 - t_1)A_2(t_2)A_1(t_1) \quad (3.46)$$

where:

Definition 3.9 (Heaviside step function)

$$\theta(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases} \quad (3.47)$$

Theorem 3.8 (Heaviside step function properties)

$$\partial_t \theta(t) = \delta(t) \quad (3.48)$$

$$(\partial_t \theta(t))\phi(t) = -\delta(t)(\partial_t \phi(t)) \quad (3.49)$$

Part II

Canonical quantisation

Chapter 4

Free fields

Quote 4.1 What could possibly go wrong?

Alessio Serafini, 16 January 2024

In this chapter, we make an overview of the so-called Klein-Gordon equation, an intuitive attempt at developing a special relativistic quantum theory, as well as its failures. By quantising the Klein-Gordon field, we will get a first taste of canonical quantisation.

4.1 Klein-Gordon equation and its demise

In quantum mechanics, there is only one dimension, which is time. This becomes a problem in high-energy physics, where we take into account the movement of the particles. Also, we must note that particle interactions in high-energy physics often take place in the relativistic limit. As such, any candidates for a theory of HEP must satisfy the following:

- The theory must be in 4D (3 spatial dimensions and 1 temporal dimension).
- The theory must incorporate special relativity.

An earlier attempt at constructing such a theory is the infamous *Klein-Gordon equation*. At that point, QFT was still not developed, and it was naively assumed that the Schrödinger equation could be modified to be Lorentz-invariant by simply applying the wavefunction ϕ to both sides of the equivalence

Theorem 4.1 (On-shell condition)

$$E_p^2 = m^2 + |p|^2 \quad \text{or} \quad p^\mu p_\mu = m^2 \quad (4.1)$$

where, for convenience, we have set $c = \hbar = 1$.

This is known as the *on-shell condition*¹. Intuitively, this gives

Theorem 4.2 (Klein-Gordon equation) The Klein-Gordon equation describes scalar (spin-0) particles in a relativistic framework:

$$(\square + m^2)\phi = 0 \quad (4.2)$$

where $(\square + m^2)$ is known as the *Klein-Gordon operator*.

Remark 4.1 Here we make a historical comment. The Klein-Gordon equation (and the Dirac equation, as we will see later) was historically derived by intuition, like we have done here. However, the modern standard derivation involves inserting the free massive scalar field Lagrangian (2.4) (which we ultimately resign to formulating by intuition) into the Euler-Lagrange equations.

Remark 4.2 As free fields are solutions to the Klein-Gordon equation, a Klein-Gordon operator acting on a free field always gives 0.

¹So-called in reference to the so-called *mass shell*. Essentially, *on-shell* refers to what satisfies the classical equations of motion, whereas *off-shell* refers to those that do not.

Exercise 4.1 Show that the Klein-Gordon equation can be recovered by using an action principle on the free massive scalar field Lagrangian (2.4).

The Klein-Gordon equation has a plane wave general solution

$$\phi(x) = Ne^{-iE_p t - p \cdot x} \quad (4.3)$$

where N is a normalisation constant.

Derivation 4.1 (Demise) Consider a simple 1D potential barrier of the form

$$V(x) = \begin{cases} 0 & x < 0 \\ V & x \geq 0 \end{cases} \quad (4.4)$$

According to the Klein-Gordon equation, the simplest solution would be

$$\phi(t, x) = \begin{cases} e^{-i(E_p t - px)} + ae^{-i(E_p t + px)} & x < 0 \\ be^{-i(E_p t + kx)} & x \geq 0 \end{cases} \quad (4.5)$$

where p is the momentum, $k = \sqrt{(E_p - V_0)^2 - m^2}$, and

- $e^{-i(E_p t - px)}$ is the part of the field travelling at the +ve x -direction that has not yet reached the potential barrier.
- $ae^{-i(E_p t + px)}$ is the part of the field reflected at the barrier travelling at the -ve x -direction.
- $be^{-i(E_p t + kx)}$ is the part of the field transmitted through the barrier travelling at the +ve x -direction.

Intuitively, both $\phi(x)$ and $\partial_x \phi(x)$ are continuous at $x = 0$, from which we find the parameters

$$a = \frac{p - k}{p + k} \quad b = \frac{2p}{p + k} \quad (4.6)$$

By inserting the $x < 0$ solution into the Klein-Gordon equation, we find a dispersion relation

$$p = \pm \sqrt{E_p^2 + m^2} \quad (4.7)$$

To reflect the forward-travelling nature of $e^{-i(E_p t - px)}$, the group velocity $v_g = \partial_p E_p$ must be positive. This forces us to adopt the positive ($x \geq 0$) solution, which we insert into the Klein-Gordon equation. Due to the non-zero potential V , we replace the partial derivatives with covariant derivatives^a:

Definition 4.1 (Scalar field covariant derivative)

$$i\partial_t \rightarrow i\partial_t - V \quad \partial_t \rightarrow \partial_t + iV \quad (4.8)$$

This gives

$$k = \mp \sqrt{(E_p - V)^2 - m^2} \quad (4.9)$$

Again, to reflect the forward-travelling nature of $be^{-i(E_p t + kx)}$, the group velocity or its inverse $\frac{1}{v_g} = \frac{\partial k}{\partial E_p} = \mp \frac{E_p - V}{|k|}$ must be positive. Now consider the case $V > E_p$. From the group velocity condition, we are forced to adopt the negative solution.

One can find a negative energy solution for each positive energy solution. However, this can be handwaved, as we will see much, much later, as antimatter. The real problem lies with the probability density, which we recall to be the 0th component of the conserved current:

$$\varrho = i(\phi^*(\partial_t + iV)\phi - \phi(i\partial_t - V)\phi^*) \quad (4.10)$$

which, in this case, is simply

$$\varrho = 2b^2(E_p - V) \quad (4.11)$$

For $E < V$, this probability density is always negative.

Remark 4.3 Wait, what?

^aWe will justify this when we reach (6.24).

A negative probability density is always unphysical, which can be resolved by turning the (classical) Klein-Gordon field into an operator. This is the beginning of quantum field theory.

Historically, the quantisation of fields/operators as the so-called *field operator* is known as *second quantisation*, in contrast to quantised particles, which was known as the *first quantisation*². Today, we call both *canonical quantisation* as canonical commutation relations are utilised in both quantisation processes.

4.2 Fock space

A generic state in QFT is essentially a linear combination of k particle states for some arbitrary k . This is significant in that k is not fixed - particles might be created and annihilated. Hilbert spaces, which have a fixed number of particles, fail to describe QFT. Rather, the vector space QFT lies in is known as a *Fock space*:

Definition 4.2 (Fock space) The Fock^a space $\mathcal{F}(\mathcal{H}_1)$ is the direct sum of all n -particle Hilbert spaces:

$$\mathcal{F}(\mathcal{H}_1) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n \quad n \in \mathbb{Z} \quad (4.12)$$

where $\mathcal{H}_0 = \mathbb{C}$ is the vacuum state space (zero particles), and $\mathcal{H}_n = \mathcal{H}_1^{\otimes n}$ is the n -particle Hilbert space.

^aNamed after Vladimir Fock, or *Fok* in scientific transliteration.

Remark 4.4 In mathematics, this decomposition is known as a \mathbb{Z} -grading as $n \in \mathbb{Z}$.

Definition 4.3 (Creation and annihilation operators) Fock spaces are equipped with creation and annihilation operators a^\dagger and a that adds a particle to the final state and removes a particle from the initial state respectively:

$$a^\dagger |n\rangle = |n+1\rangle \quad a|n\rangle = |n-1\rangle \quad (4.13)$$

Remark 4.5 Bosonic creation and annihilation operators are near-identical the good ol' ladder operators in QM as they obey the same commutation relations. Fermionic creation and annihilation operators observe anticommutation relations, which are powered by Grassmann mathematics covered in *Spinors & Symmetries*.

Theorem 4.3 (Bosonic and fermionic operator commutations)

- Bosonic operators commute as they are Grassmann-even:

$$[B_i, B_j] = B_i B_j - B_j B_i = 0 \quad (4.14)$$

- Fermionic operators anticommute^a as they are Grassmann-odd:

$$\{F_i, F_j\} = F_i F_j + F_j F_i = 0 \quad (4.15)$$

- A bosonic operator commutes with a fermionic operator:

$$[B, F] = BF - FB = 0 \quad (4.16)$$

^aThe physical justification is that if they commute, the Hamiltonian would be unbounded from below and causality would be violated.

Let us revise the properties of commutators and anticommutators.

Theorem 4.4 (Commutator and anticommutator properties) Commutators:

- **Bilinearity:** The commutator is linear in both the first and second arguments

$$[\alpha A + \beta C, B] = \alpha[A, B] + \beta[C, B] \quad [A, \alpha B + \beta C] = \alpha[A, B] + \beta[A, C] \quad (4.17)$$

²In fact, second quantisation is a slight misnomer as quantising operators is also possible (although unnecessary) in QM.

- **Jacobi identity:** The cyclic sum of nested commutators vanishes

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0 \quad (4.18)$$

- **Product rules:** These are Leibniz-type identities

$$[A, BC] = [A, B]C + B[A, C] \quad [AB, C] = A[B, C] + [A, C]B \quad (4.19)$$

Anticommutators:

- **Bilinearity:** The anticommutator is linear in both the first and second arguments

$$\{\alpha A + \beta C, B\} = \alpha\{A, B\} + \beta\{C, B\} \quad \{A, \alpha B + \beta C\} = \alpha\{A, B\} + \beta\{A, C\} \quad (4.20)$$

- **Product rules:** They look a bit different than for commutators

$$\{A, BC\} = \{A, B\}C - B[A, C] \quad \{AB, C\} = A\{B, C\} - [A, C]B \quad (4.21)$$

So far we have been working in position space, which is, informally speaking, the collection of all possible positional vectors. However, it is often convenient to employ *momentum space* in canonical quantisation. This is because, as we will see:

- The free field Hamiltonian becomes diagonal.
- Expressions for propagators are simplified.

A momentum space is a generalisation of the *reciprocal space* or *wavevector space* you may have seen before in crystallography. The momentum and position (or physical) spaces are Fourier transforms of each other.

Definition 4.4 (Fourier transform)

$$\tilde{f}(p) = \int d^4x e^{ipx} f(x) \quad (4.22)$$

At this point, one should familiarise themselves with three implications. The first two stem from the so-called *Dirac delta*:

Derivation 4.2 (Dirac delta properties)

Definition 4.5 (Dirac delta) The Dirac delta has two equivalent definitions, which basically say that the Dirac delta has unit area (i.e. an area of 1):

$$\delta(x)dx = \begin{cases} \infty & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases} \quad \int_{x_1}^{x_2} \delta(x)dx = \begin{cases} 1 & \text{if } x_1 < 0 < x_2 \\ 0 & \text{otherwise} \end{cases} \quad (4.23)$$

We see that

- The first definition has $\delta(x)$ at infinity (∞) at an infinitesimal range ($1/\infty$). Hence multiplying gives $\infty/\infty = 1$, which is the unit area of 1 we mentioned^a.
- When $x \neq 0$, one has $\delta = 0$. Hence, for meaningful calculations, x must be fixed to 0.

^aThis just shows how unrigorously defined it is.

Any function that satisfies these conditions can be a Dirac delta, like an infinitely narrow Gaussian

$$\delta(x) = \lim_{\sigma \rightarrow 0} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (4.24)$$

The Dirac delta has two interesting properties. One emerges when we consider the Dirac delta of a function $f(x)$. We set a x_0 such that $f(x_0) = 0$. Then, by performing the integral $\int_{x_1}^{x_2} \delta(f(x)) \frac{df}{dx} dx$, we find that

Theorem 4.5 (Dirac delta as a functional)

$$\delta(f(x)) = \left| \frac{\partial f}{\partial x} \right|^{-1} \delta(x) \quad (4.25)$$

The second property is the *sifting property*:

Theorem 4.6 (Sifting property of the Dirac delta)

$$\int d^n x \delta^n(x - y) f(x) = f(y) \quad (4.26)$$

The so-called sifting property effectively allows us to eliminate integrals, just like how the Kronecker delta effectively eliminates one of its 2 indices. Hence, the Dirac delta can be thought of as a continuous form of our good friend, the discrete *Kronecker delta*.

The sifting property is very useful because we then have the first implication of the Fourier transform:

Theorem 4.7 (Elimination of integral via the Fourier transformation)

$$\int d^m y \int d^n x \delta^n(x - y) f(x) = \int d^m y f(y) \quad (4.27)$$

Now let us set $y = 0$. We then have

$$\int d^n x f(x) \delta^n(x) = f(0) \quad (4.28)$$

which gives the second implication

Theorem 4.8 (Fourier transform and exponential function)

$$\int d^n x e^{\pm i k x} = (2\pi)^n \delta^n(k) \quad (4.29)$$

The final implication concerns the 4-derivative ∂_μ . Under a Fourier transform, we have

$$\partial_\mu f(x) \rightarrow \int d^4 x e^{ip \cdot x} \partial_\mu f(x) = -ip_\mu \tilde{f}(p) \quad (4.30)$$

if one removes the functions themselves, one finds

Theorem 4.9 (Fourier transform of 4-derivative)

$$\partial_\mu \rightarrow -ip_\mu \quad (4.31)$$

For example, the momentum space Klein-Gordon equation is

$$(p^2 + m^2)\phi = 0 \quad (4.32)$$

In the following derivation, even though the field and canonical momentum operators are still in position space, their expressions are in momentum space, which we Fourier-transform back to position space.

4.3 Quantisation of the Klein-Gordon field

Derivation 4.3 (Field operator) The Klein-Gordon equation general solution (4.3) can be rewritten to account for negative energy solutions:

$$\phi(x, t) = \int d^3 p N_p \left(f_p e^{-i(E_p t - p \cdot x)} + f_p^* e^{i(E_p t - p \cdot x)} \right) \quad (4.33)$$

where N_p , a real function of p , is the previously seen normalisation factor and f_p is a complex function of p (and hence based in momentum space). Importantly, as the field operator $\phi(x, t)$ is based in

position space, we must perform a Fourier transform $\int d^3p N_p$ to convert f_p from momentum space to position space.

To ensure that the resultant quantised Hamiltonian will evolve with time in the same way, we replace f_p and f_p^* with the annihilation and creation operators a_p and a_p^\dagger . As the Klein-Gordon field is a scalar field (spin-0), it is a bosonic field and its components commute under field quantisation:

Theorem 4.10 (Bosonic creation and annihilation operator commutations) For two arbitrary momenta p and q in bosonic fields, their creation and annihilation operators a_p , a_p^\dagger , a_q and a_q^\dagger observe:

$$[a_p, a_q^\dagger] = (2\pi)^3 \delta^3(p - q) \quad (4.34)$$

$$[a_p, a_q] = [a_p^\dagger, a_q^\dagger] = 0 \quad (4.35)$$

Recalling that special relativity must be observed, we must choose an N_p that makes $\phi(x)$ Lorentz-invariant. The Lorentz-invariant phase space volume element for a single particle is given by

Definition 4.6 (Lorentz-invariant phase space volume element)

$$dV = \frac{dp^3}{(2\pi)^3 2E_p} \quad (4.36)$$

For the field operator, we can remove a factor of $1/\sqrt{2E_p}$ to this volume element so that no factors of E_p emerge in the field and momentum operator commutation relations^a. Taking $N_p = 1/((2\pi)^3 \sqrt{2E_p})$ the Klein-Gordon field operator is then written as

Definition 4.7 (Klein-Gordon field operator)

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3 \sqrt{2E_p}} (a_p e^{ip \cdot x} + a_p^\dagger e^{-ip \cdot x}) \quad (4.37)$$

^aLorentz invariance is preserved as the factor $1/\sqrt{2E_p}$ is not really gone, but rather absorbed into the field amplitude.

Note 4.1 (Two formalisms of $\phi(p)$) Intuitively, the field operator in momentum space is the Fourier transform of its position space counterpart

$$\phi(p) = \int d^3x e^{-ip \cdot x} \phi(x) \quad (4.38)$$

However, the notation $\phi(p)$ is overloaded: a second formalism exists in some literature where $\phi(p)$ denotes the *momentum space contribution* to the field operator (4.37):

$$\phi(p)_{\text{alt}} = \frac{1}{\sqrt{2E_p}} (a_p e^{ip \cdot x} + a_p^\dagger e^{-ip \cdot x}) \quad (4.39)$$

which gives rise the (equally correct) formula for the position space field operator

$$\phi(x) = \int \frac{d^3p}{(2\pi)^3} \phi(p)_{\text{alt}} \quad (4.40)$$

In this book, we use the first formalism.

The momentum operator can then be written by adding a factor of iE_p and changing the sign on the first term

Definition 4.8 (Klein-Gordon momentum operator)

$$\pi(x) = - \int \frac{d^3p}{(2\pi)^3 \sqrt{2}} (-a_p e^{ip \cdot x} + a_p^\dagger e^{-ip \cdot x}) \quad (4.41)$$

The commutation relations for bosonic fields are then

Theorem 4.11 (Bosonic field and momentum operator commutations) For some arbitrary spacetime coordinates x and y

$$[\phi(x), \pi(y)] = i\delta^3(p - q) \quad (4.42)$$

$$[\phi(x), \phi(y)] = [\pi(x), \pi(y)] = [\phi^*(x), \phi(y)] = [\pi^*(x), \pi(y)] = 0 \quad (4.43)$$

Remark 4.6 The vanishing of the commutator between ϕ and π with their complex conjugates is a result of the Grassmann-evenness of bosonic fields³. The implication of this is that ϕ and π are treated as independent from their complex conjugates.

By recalling the definition of the Hamiltonian, we can derive the Klein-Gordon Hamiltonian density from the Klein-Gordon Lagrangian density as

$$\mathcal{H} = \frac{1}{2}(\pi^2 + |\nabla\phi|^2 + m^2\phi^2) \quad (4.44)$$

We then find the (time-independent) Hamiltonian operator:

$$H = \int \frac{d^3p}{(2\pi)^3 2E_p} E_p^2 (a_p^\dagger a_p + a_p a_p^\dagger) = \underbrace{\int \frac{d^3p}{(2\pi)^3} E_p a_p^\dagger a_p}_{\textcircled{1}} + \underbrace{\int \frac{d^3p}{(2\pi)^3} \frac{E_p}{2} [a_p, a_p^\dagger]}_{\textcircled{2}} \quad (4.45)$$

However, as $[a_p, a_p^\dagger]$ is proportional to the identity, so is the term $\textcircled{2}$, and we can neglect it for convenience. Due to the ordering of its creation and annihilation operators, the remaining term $\textcircled{1}$ is known as the *normal-ordered* or *Wick-ordered* Hamiltonian, named after Gian Carlo Wick:

Definition 4.9 (Normal-ordered Klein-Gordon Hamiltonian)

$$:H: = \int \frac{d^3p}{(2\pi)^3} E_p a_p^\dagger a_p \quad (4.46)$$

Here we have introduced the concept of *normal ordering*:

Definition 4.10 (Normal ordering) For some operator O that can be expressed as a polynomial of creation and annihilation operators, its normal ordering or *Wick ordering* $:O:$ is the same polynomial, but with all creation operators to the left of all annihilation operators. This is often necessary in that it eliminates unphysicality created by vacuum fluctuations like $a_p^\dagger a_p$.

Remark 4.7 For example, given some $O = a_p a_q a_p^\dagger$, we have $:O: = a_p^\dagger a_p a_q$.

By considering (4.46) as the entire Hamiltonian⁴, we are able to easily diagonalise it. Therefore, we can find that $:H:$ admits the eigenvectors $|n\rangle$ or $a_p |n\rangle$, with the eigenvalues being

$$:H:a_p |n\rangle = (-n - E_p) a_p |n\rangle \quad (4.47)$$

Remark 4.8 Conversely, *antinormal ordering* places creation operators to the right instead.

Theorem 4.12 (Vacuum expectation value) The vacuum expectation value of any normal-ordered expression yields zero.

We are now in a position to consider the physicality of our free field, Klein-Gordon theory. This involves distinguishing between free field theories, of which our Klein-Gordon theory is one, and interacting theories, in which every physical theory in the standard model lies.

Definition 4.11 (Free and interacting field theories)

- *Free terms* in the Lagrangian are terms which, upon insertion into the Euler-Lagrange equations, yield a linear equation of motion (like the Klein-Gordon equation we have seen and the Dirac equation we will see).
- *Interacting terms* in the Lagrangian are terms which, upon insertion into the Euler-Lagrange

³In other words, they are *enforced* by the Grassmann-evenness of the fields.

⁴An advantage of considering a quantity's normal-ordered counterpart as itself is that we eliminate any uninteresting constant terms. This often simplifies calculations.

equations, yield a nonlinear equation of motion (like the full ϕ^4 theory and QED equations of motion, which are not very interesting for us except for completeness).

We will elaborate on these two statements in the next chapter. For now, it suffices to understand that the Klein-Gordon theory Lagrangian consists purely of free terms due to the linearity of the Klein-Gordon equation, which makes it clearly unphysical:

- As you will see in the next chapter, Feynman diagrams include propagators, vertices and loops. A free field theory can only have propagators: As there are no interactions, vertices do not exist.
- For this reason, loops do not exist either, except for standalone ones.
- The propagator goes on forever in time in the absence of interactions, giving rise to a scattering matrix that is always trivial (simply the unit matrix).
- The cross-section for any scattering process (where next to nothing happens) is then always zero. It should then be obvious that this field theory is not very interesting.

Before we get into any physical theories (which has the problem of gauge fixing), we want to look a toy model that is interacting to get the full grasp of the building block of QFT while not having to deal with those pesky gauges. In most literature, this toy model is ϕ^4 theory. This is nothing but Klein-Gordon theory with an added quartic interacting term which gives rise to the vertex. As the free terms are identical, the propagator stays the same. For this reason, we will derive the free field propagator not in this chapter, but rather the next one where, in ϕ^4 theory, its significance can be fully appreciated.

Chapter 5

Interacting fields I: ϕ^4 theory

Quote 5.1 Tbh, hand-drawn sketches have their own Charme.

Felix Halbwedel, 27 August 2024

We now investigate particle interactions as well as the two matrices S_{fi} and \mathcal{M}_{fi} , the first of which physically governs the probability a certain interaction will take place. The Feynman rules, which we use to build the formula for calculating elements of \mathcal{M}_{fi} , are then derived for ϕ^4 theory, a simple toy model. Some concluding discussions on their use in high energy physics are then made.

5.1 Self-interaction

We now consider an interacting field theory where a field interacts with its environment. This comes into play by introducing interacting terms in the Lagrangian. One type of interaction is *self-interaction*, which is the interaction between a particle and its own field. A good toy model is the so-called ϕ^4 theory¹, which gets rid of the problematic² 3rd-order term, adding only a quartic interaction term $-\frac{\lambda}{4!}\phi^4$ to (2.4):

Definition 5.1 (ϕ^4 theory Lagrangian)

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4 \quad (5.1)$$

where λ is some dimensionless coupling constant that determines the strength of the interaction, and the $1/4!$ is a combinatorial normalisation factor that compensates for the fact that there are $4!$ ways to permute the (4 identical) fields in ϕ^4 theory.

ϕ^4 theory is among a family of theories known as ϕ^n theories. It is superior to all other ϕ^n theories for two reasons:

- **Energetic stability:** ϕ^4 theory has the second-simplest interaction term that respects the symmetry $\phi \rightarrow -\phi$ ³:
 - The general form for the potential energy in ϕ^3 theory is $V(\phi) = m^2\phi^2 + g\phi^3$. Due to the odd exponential in $g\phi^3$, one can get a negative Hamiltonian expectation for a large, negative ϕ , even if the coupling constant g is positive.
 - In contrast, the $\lambda\phi^4$ term ϕ^4 theory is positive as long as λ is positive, ensuring a stable minimum, which is crucial for physical systems.
- **Renormalisability:** In certain dimensions, ϕ^3 theory is non-renormalisable, unlike ϕ^4 theory.

¹Even though its only physical application is describing the self-interaction term of the Higgs field, it has become a staple of QFT books.

²In that it makes the theory non-renormalisable. You will see what this means in Part IV.

³ ϕ^2 theory has the simplest interaction term. But this term is simply the previously seen mass term $\frac{1}{2}m^2\phi^2$.

Note that this is still a bosonic field theory that satisfies the Klein-Gordon equation. Before moving onto scattering, we need introduce one last mathematical tool, which is the *Wick contraction*.

Definition 5.2 (Wick contraction) For operators A and B , the Wick contraction^a or simply *contraction* is simply its vacuum expectation value

$$\overline{AB} = \langle 0|AB|0\rangle \quad (5.2)$$

^aSo-called as just like the contraction of indices in GR (which, in the case of a rank-2 tensor, starts with two indices and ends with a scalar), it starts with two operators and ends with a number.

Derivation 5.1 (Alternative forms of the Wick contraction) From the definition of Wick contractions, we can find several equivalences:

- The product of two operators can always be split into a part that contributes to the vacuum expectation value and a part that does not, which is actually the normal ordering^a $:AB:$.

Theorem 5.1 (Wick's first theorem)

$$AB = \langle 0|:AB|0\rangle + :AB: = \overline{AB} + :AB: \quad (5.3)$$

In some literature, a rearranged version of *Wick's first theorem* is actually used as a less intuitive definition of Wick contractions.

$$\overline{AB} = AB - :AB: \quad (5.4)$$

- The vacuum expectation value is inherently time-ordered, so if A and B are the fields $\phi(x)$ and $\phi(y)$, we can even make the equivalence

$$\overline{\phi(x)\phi(y)} = \langle 0|\phi(x)\phi(y)|0\rangle = \langle 0|T[\phi(x)\phi(y)]|0\rangle = D_F(x-y) \quad (5.5)$$

where $D_F(x-y)$ is the Feynman propagator.

^aAs normal-ordered operators always have creation operators before annihilation operators, their vacuum expectation value $\langle 0|:AB|0\rangle$ is always zero.

Remark 5.1 In other words, the Wick contraction of operators returns their ‘nontrivial’ part. This part, which represents quantum fluctuations and interactions, contributes to the vacuum expectation value, whereas the normal-ordered part does not.

Often, it is more convenient to use a very similar operation called the *time-ordered pairing* instead of Wick contractions.

Definition 5.3 (Time-ordered pairing)

$$\underline{A(x)B(y)} = \begin{cases} \overline{A(x)B(y)} & x^0 > y^0 \\ (-1)^p \overline{B(x)A(y)} & y^0 > x^0 \end{cases} \quad (5.6)$$

where p , last seen in Definition 3.8, is our good friend, the parity.

Quote 5.2 In more accurate books like Bogoliubov’s, the time dependent contractions are written down with bottom brackets.

Felix Halbwedel, 22 December 2024

Derivation 5.2 (Wick's second theorem for two fields) A field can be split into creation and annihilation parts:

$$\phi(x) = \phi^+(x) + \phi^-(x) \quad (5.7)$$

where

$$\phi^+(x) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} a_p e^{-ip \cdot x} \quad \phi^-(x) = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{\sqrt{2E_p}} a_p^\dagger e^{+ip \cdot x} \quad (5.8)$$

Let us evaluate $T[\phi(x)\phi(y)]$ under this decomposition. When $x^0 > y^0$, we have, by time-ordering

$$T[\phi(x)\phi(y)] = \phi(x)\phi(y)$$

Now substitute (5.7)

$$\phi(x)\phi(y) = [\phi^+(x) + \phi^-(x)][\phi^+(y) + \phi^-(y)] \quad (5.9)$$

We then rewrite this sum by grouping into the normal-ordered piece plus the piece that arises from swapping one creation past one annihilation:

$$\phi(x)\phi(y) = : \phi(x)\phi(y) : + [\phi^+(x), \phi^-(y)] \quad (5.10)$$

But using the mode expansions, one can directly see that

$$[\phi^+(x), \phi^-(y)] = \int \frac{d^3 p}{(2\pi)^3} \frac{1}{2E_p} e^{-ip \cdot (x-y)} = D_F(x-y) \quad (5.11)$$

where the Feynman propagator is exactly the contraction of two fields. Thus for $x^0 > y^0$, we have

$$T[\phi(x)\phi(y)] = : \phi(x)\phi(y) : + D_F(x-y) \quad (5.12)$$

By the same argument, this is also true for $y^0 > x^0$. (5.12) is known as Wick's second theorem for two fields.

Now we generalise this to an arbitrary number of fields.

Theorem 5.2 (Wick's second theorem) The time-ordering $T[A_1 A_2 A_3 A_4 A_5 A_6 \dots]$, where all operators are made up of creation and annihilation operators like $A_i = A_i^+ + A_i^-$, can be expressed in terms of time-ordered pairings:

$$T[A_1 A_2 A_3 A_4 A_5 A_6 \dots] = :A_1 A_2 A_3 A_4 A_5 A_6 \dots: + \underbrace{\sum_{\text{single}} :A_1 \underline{A_2} A_3 A_4 A_5 A_6 \dots:}_{\textcircled{1}} + \underbrace{\sum_{\text{double}} :A_1 \underline{A_2} \underline{A_3} A_4 A_5 A_6 \dots:}_{\textcircled{2}} + \dots \quad (5.13)$$

where:

- (1) denotes the sum of all the possible results of $A_1 A_2 A_3 A_4 A_5 A_6 \dots$ undergoing one Wick contraction somewhere in the expression:

$$\sum_{\text{single}} :A_1 \underline{A_2} A_3 A_4 A_5 A_6 \dots: = :A_1 \underline{A_2} A_3 A_4 A_5 A_6 \dots: + :A_1 \underline{A_3} A_2 A_4 A_5 A_6 \dots: + \dots \quad (5.14)$$

- (2) denotes the sum of all the possible results of $A_1 A_2 A_3 A_4 A_5 A_6 \dots$ undergoing two Wick contractions somewhere in the expression:

$$\sum_{\text{double}} :A_1 \underline{A_2} \underline{A_3} A_4 A_5 A_6 \dots: = :A_1 \underline{A_2} \underline{A_3} A_4 A_5 A_6 \dots: + :A_1 \underline{A_3} \underline{A_2} A_4 A_5 A_6 \dots: + \dots \quad (5.15)$$

- ...and so on.

An alternative, non-time-ordered version of *Wick's second theorem* is

$$A_1 A_2 A_3 A_4 A_5 A_6 \dots = :A_1 A_2 A_3 A_4 A_5 A_6 \dots: + \sum_{\text{single}} :\overline{A_1} \underline{A_2} A_3 A_4 A_5 A_6 \dots: + \sum_{\text{double}} :\overline{A_1} \underline{A_2} \overline{A_3} \underline{A_4} A_5 A_6 \dots: + \dots \quad (5.16)$$

where we simply use the normal contraction.

Finally, a nice trick is the so-called *Wick's third theorem* or *Wick's theorem for vacuum expectation values*:

Theorem 5.3 (Wick's third theorem) For operators A, B_1, \dots, B_n , the following is observed:

$$\langle 0 | T[AB_1 \cdots B_n] | 0 \rangle = \sum_i \langle 0 | T[\overbrace{AB_1 \cdots B_i}^{\square} \cdots B_n] | 0 \rangle \quad (5.17)$$

5.2 Scattering matrix

An interacting field theory gives rise to *scattering*, where we have the *S-matrix* or the *scattering matrix* that encodes all the information about the probabilities of different scattering processes. It can be derived by taking limits of $U(t, t_0)$, where time evolution covers the entire history of the system, from the infinite past to the infinite future:

$$S = \lim_{t \rightarrow \infty} \lim_{t_0 \rightarrow -\infty} U(t, t_0) \quad (5.18)$$

Definition 5.4 (S-matrix operator) For the initial/incoming state $|\psi, \text{in}\rangle$ and the final/outgoing state $\langle \alpha, \text{out}|$, the S-matrix element S_{fi} represents the probability amplitude that $|\psi, \text{in}\rangle$ evolves into $\langle \alpha, \text{out}|$. It can be represented via the S-matrix operator S :

$$S_{fi} = \langle \alpha, \text{out} | S | \psi, \text{in} \rangle \quad (5.19)$$

An important property is the so-called *cluster decomposition*, which states that an experiment cannot influence another experiment carried out far away from it, and vice versa. Its implication in QFT is as follows:

Theorem 5.4 (Cluster decomposition) Consider two experiments (or rather *clusters*) 1 and 2. An initial state α which includes parts in both clusters α_1 and α_2 evolves into a final state β which likewise includes parts in both clusters β_1 and β_2 . The scattering matrix can always be decomposed to

$$S_{\beta\alpha} = S_{\beta_1\alpha_1} S_{\beta_2\alpha_2} \quad (5.20)$$

where $S_{\beta_1\alpha_1}$ and $S_{\beta_2\alpha_2}$ are ‘parts’ of $S_{\beta\alpha}$ in 1 and 2.

Now consider a highly idealised system of n particles which are prepared with momenta $\{p_i, i = 1, \dots, n\}$ at time $t_0 \rightarrow -\infty$. They interact with (i.e. scatter in) a perturbed Hamiltonian $H = H_0 + H_{\text{int}}$, reaching a set of final momenta $\{q_i, i = 1, \dots, n\}$ time $t \rightarrow \infty$. One can represent the initial and final states with

$$|p, \text{in}\rangle = \prod_{i=1}^n a_{p_i}^\dagger |0\rangle \quad \langle q, \text{out}| = \langle 0| \prod_{i=1}^n a_{q_i}^\dagger \quad (5.21)$$

where $|0\rangle$ is the ground state. In this scenario, the S-matrix elements are

$$S_{qp} = \langle q, \text{out} | e^{-iH(t-t_0)} | p, \text{in} \rangle \quad (5.22)$$

To evaluate the elements of S_{qp} , we need to diagonalise the full Hamiltonian H . As this is typically impossible, we must use a perturbative approach to deal with H_{int} , where we assume that the interaction Hamiltonian $H_{\text{int}}(t)$ is zero at $t \rightarrow -\infty$ and $t \rightarrow \infty$.

Derivation 5.3 (LSZ reduction formula) Let us now consider scattering in the Heisenberg picture. We apply the Heisenberg equation to a_p :

$$\dot{a}_p = i[H, a_p] \rightarrow \dot{a}_p = iE_p a_p \quad (5.23)$$

Plugging this result into the field operator yields

$$\phi(x)_H = \int \frac{d^3 p}{(2\pi)^3 \sqrt{2E_p}} (a_p e^{-ip \cdot x} + a_p^\dagger e^{ip \cdot x}) \quad (5.24)$$

From this, one can represent the creation and annihilation operators in terms of the wavefunctions.

$$ia_{p,H}^\dagger = \int \frac{d^3 x}{\sqrt{2E_p}} (e^{-ip \cdot x} (\partial_0 \varphi_H(x)) - \varphi_H(x) (\partial_0 e^{-ip \cdot x})) \quad (5.25)$$

In the Heisenberg picture, the S -matrix element is

Definition 5.5 (Heisenberg picture S -matrix element)

$$S_{qp,H} = \langle 0 | a_{q_1,H} \cdots a_{q_m,H} a_{p_1,H}^\dagger \cdots a_{p_n,H}^\dagger | 0 \rangle \quad (5.26)$$

We can repeatedly insert (5.25), which yields the so-called *LSZ formula* or the *LSZ reduction formula* for S -matrix elements, named after Harry Lehmann, Kurt Symanzik and Wolfhart Zimmermann. By defining a so-called *vacuum expectation value*

Definition 5.6 (Vacuum expectation value) The expectation value of time-ordered field operators evaluated at $m+n$ spacetime points is labelled G_{m+n} as the time ordering symbol is a $(m+n)$ -point *Green's function*:

$$G_{m+n} = \langle 0 | T[\phi_H(x_1) \cdots \phi_H(x_{m+n})] | 0 \rangle \quad (5.27)$$

We can write the LSZ formula as

Theorem 5.5 (LSZ reduction formula for S -matrix elements) For a series of spacetime coordinates x_i , the scattering matrix elements can be represented by

$$S_{qp,H} = \int \frac{d^4x_1}{\sqrt{2E_{p_1}}} \cdots \int \frac{d^4x_{m+n}}{\sqrt{2E_{q_m}}} e^{-i \sum_{j=1}^n p_j \cdot x_j + i \sum_{j=1}^m q_j \cdot x_{n+j}} \prod_{j=1}^{m+n} (\partial_{x_j}^2 + m^2) G_{m+n} \quad (5.28)$$

where the Klein-Gordon operators $(\partial_{x_j}^2 + m^2)$ enforce that the external particles are on-shell.

Remark 5.2 Here, we see that all our field quantisation in Fock space and derivation of (anti)commutation relations have paid off. Exactly because of canonical quantisation, we were able to construct the LSZ formula, which is essential for constructing the S -matrix. We thus see the significance of canonical quantisation, even though, as we will discover, our end result actually does *not* make use of field operators whatsoever.

Now we would like to migrate to the interaction picture (denoted by the subscript I) due to its simplicity⁴. Our vacuum expectation value/Green's function then becomes the *interacting time-ordered propagator*:

Theorem 5.6 (Interacting time-ordered propagator)

$$G_{m+n,I} = \frac{\langle 0 | T[\phi_I(x_1) \cdots \phi_I(x_{m+n})] S | 0 \rangle}{\langle 0 | S | 0 \rangle} \quad (5.29)$$

It is also called the *interacting^a Green's function*, the $(m+n)$ -point *Green's function* or the *correlation function^b*.

^aThe ‘interacting’ does not refer to the dynamical picture. Instead, it is used to describe the presence of an interaction term in the Hamiltonian.

^bSo-called as it is used to study correlations between field operators at different spacetime points in the interacting vacuum.

As is easily seen, the LSZ formula is very cumbersome. As such, while it is theoretically significant, we usually do not actually use it as-is when deriving S_{fi} . Instead, we utilise the *Feynman rules*, which are the practical implementation of what the LSZ formula theoretically justifies. The next two sections will be dedicated to deriving the Feynman rules and elucidating their physical and practical meaning.

5.3 Feynman diagrams

Now we will begin our long journey to the Feynman rules which we use in lieu of the LSZ reduction formula. One can construct \mathcal{M}_{fi} (and by that, S_{fi}) of a given field theory through its Feynman rules. These rules make use of the so-called *Feynman diagrams*, which is effectively the graphical representation of a specific process:

⁴Also because we have not derived an expression for the Heisenberg picture fields ϕ_H anyway.

Definition 5.7 (Feynman diagram elements) A Feynman diagram includes propagators, vertices, and loops^a:

- **Propagators:** They correspond to moving free particles and correspond to the free part of the Lagrangian.
 - The propagators representing the initial and final particle states are represented by *external propagators*, *external legs* or *external lines*, whose ends are *external points*:
 - * An *incoming external leg* is an initial state (i.e. incoming) particle, typically on the left side.
 - * An *outgoing external leg* is a final state (i.e. outgoing) particle, typically on the right side.

For a total number of n such external points, one has an n -point Feynman diagram.

- The propagators that are intermediate lines between vertices and represent virtual particles are *internal propagators* or *internal lines*.

Remark 5.3 External and internal propagators have the identical mathematical formulation. However, external propagators do not contribute to the scattering matrix or the transition amplitude as are a part of the *initial & final states*, while internal propagators contribute to both as they are a part of the transition/scattering process.

Vertices: Denoted by visible round dots. The number of vertices, known as the *order*, corresponds to the order of the coupling constant of the field theory. Vertices correspond to the interacting part of the Lagrangian. Importantly, as we will soon demonstrate:

- The number of particles/fields/lines that connect to the propagator of a field theory is strictly fixed.
- As mentioned previously, vertices cannot be derived exactly. Rather, we must employ perturbation theory in both the canonical quantisation and path integral formulations.

Internal loops: Internal loops can be loosely thought of as an internal line coming back to itself. Mathematically, we write it as a loop integral of the (internal) line propagator. We speak of a *tree-level diagram* if the diagram in question has no loops. Loops do not correspond to any part of the Lagrangian. Instead, they arise from combining interaction and free parts in higher-order (loop) perturbation theory^b.

^aAs ϕ^4 theory is a purely self-interacting toy model and does not involve any specific particles, *which is not the case in real life*. We will use plain lines for all particles and propagators.

^bAs we will soon see, internal loops integrate over internal momenta. This often creates divergences, which is a hallmark of loop-order QFT that must be treated by *renormalisation*.

A Feynman diagram of the 0th order has no vertices:

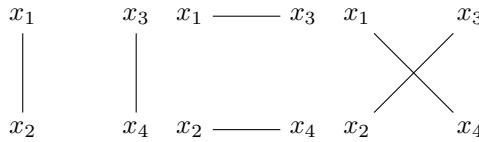


Figure 5.1: 0th-order Feynman diagrams

A Feynman diagram of the 1st order has a single vertex:

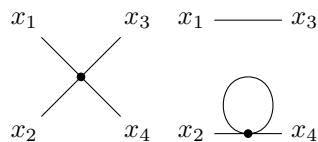


Figure 5.2: 1st-order Feynman diagrams

Note the existence of the loop in the diagram on the right. Functionally, this diagram is identical to the middle entry in the 0th-order diagrams, with the only difference being the removal of the loop. Hence, it is not the simplest expression this particular interaction can take. The middle 0th-order diagram is then known as the *leading order* Feynman diagram with respect to the interaction it represents as it is the most reduced form of the interaction.

A Feynman diagram of the 2nd order has two vertices⁵:

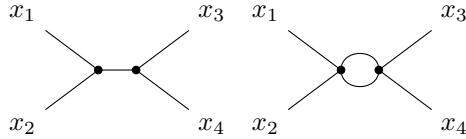


Figure 5.3: 2nd-order Feynman diagrams

All of the diagrams above are 4-point Feynman diagrams.

Quote 5.3 You can still insert some hand drawings. Hand drawings are better than no drawings, and if you don't do it now, it eventually never happens.

Felix Halbwedel, encouraging the author to finish the Feynman diagram illustrations, 22 November 2024

One importance of Feynman diagrams is that it gives more intuitive physical meanings to G_{m+n} and S_{fi} :

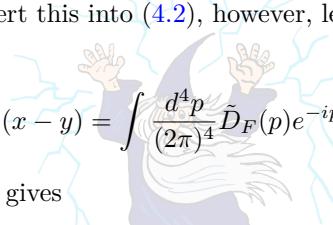
- G_{m+n} sums over all possible Feynman diagrams (both connected and disconnected) with $m + n$ external points. That is to say, it encodes the *sum* of probability *amplitudes* over all possible processes allowed by the theory of interest for fields to be created or annihilated at those points, including all quantum fluctuations consistent with the field dynamics.
- S_{fi} is the probability *itself*, and encodes a specific the scattering process (i.e. only a specific Feynman diagram). So in a way, G_{m+n} is the full ‘storyboards’ of diagrams, and S_{fi} is the final edited ‘scene’ of a specific storyboard that makes it to the screen.

5.4 ϕ^4 theory Feynman rules

The first derivation we shall make is that of the propagator which we, in our infinite wisdom, have deferred up to this point. At this point, we can already evaluate the multi-point Green’s function by using Wick’s second theorem. This may look like a tedious process. Luckily for us, in doing so, many terms cancel out, and (5.29) reduces to a series of two-point Green’s functions or the *Feynman propagators* of some spacetime coordinates x and y , labelled $D_F(x - y)$:

$$D_F(x - y) = \langle 0 | T[\phi_I(x)\phi_I(y)] | 0 \rangle \quad (5.30)$$

Derivation 5.4 (Feynman propagator) Now we want to solve for the exact form of this propagator. The central idea which we shall utilise is the fact that $D_F(x - y)$ is a solution of the Klein-Gordon equation. Before we mindlessly insert this into (4.2), however, let us perform a mathematical trick by performing a Fourier transform:



$$D_F(x - y) = \int \frac{d^4 p}{(2\pi)^4} \tilde{D}_F(p) e^{-ip \cdot (x-y)} \quad (5.31)$$

Substituting this into the equation gives

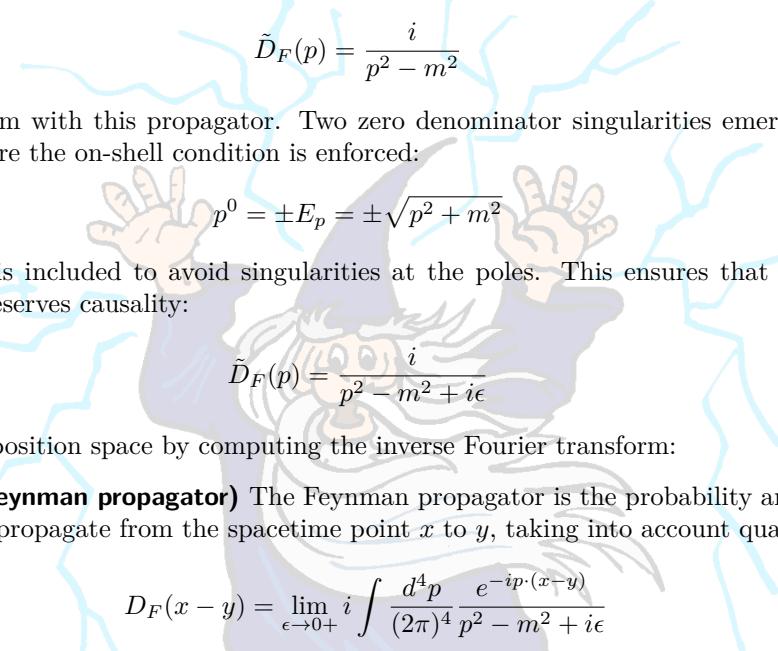
$$\int \frac{d^4 p}{(2\pi)^4} \tilde{D}_F(p) e^{-ip \cdot (x-y)} (p^2 - m^2) = \delta^4(x - y) \quad (5.32)$$

⁵At this point, it should be clear that the Feynman diagrams we have drawn are nothing but ‘toy’ Feynman diagrams. Recall that a 3-point vertex is impossible in ϕ^4 theory.

This implies that $\tilde{D}_F(p)$ must satisfy, for the momentum space Klein-Gordon operator ($p^2 - m^2$)

$$(p^2 - m^2)\tilde{D}_F(p) = i \quad (5.33)$$

Thus, the propagator in momentum space is

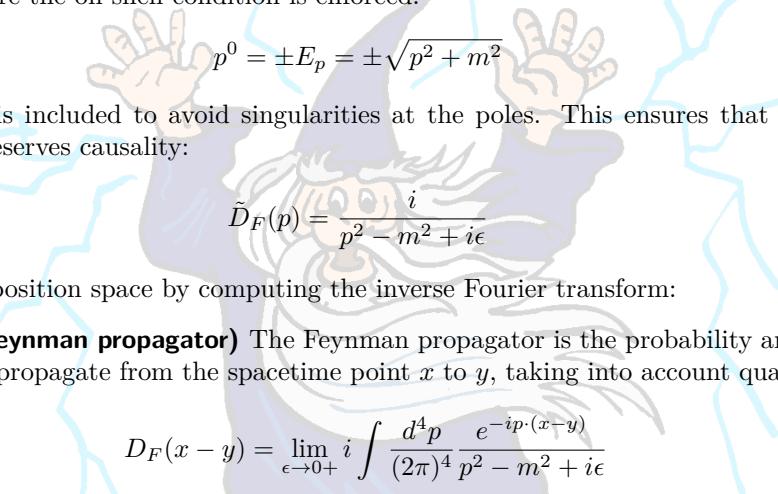


$$\tilde{D}_F(p) = \frac{i}{p^2 - m^2} \quad (5.34)$$

There is one problem with this propagator. Two zero denominator singularities emerge at so-called *poles*, positions where the on-shell condition is enforced:

$$p^0 = \pm E_p = \pm \sqrt{p^2 + m^2} \quad (5.35)$$

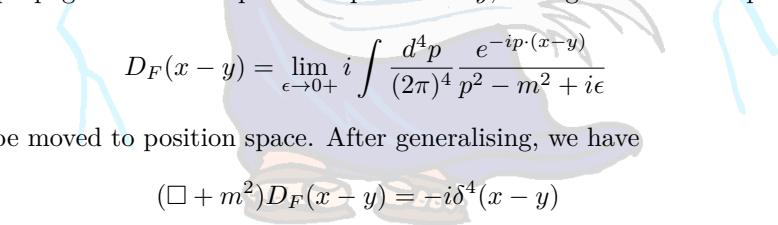
As such a *smol ie* is included to avoid singularities at the poles. This ensures that our integral is well-defined and preserves causality:



$$\tilde{D}_F(p) = \frac{i}{p^2 - m^2 + i\epsilon} \quad (5.36)$$

Now, we return to position space by computing the inverse Fourier transform:

Definition 5.8 (Feynman propagator) The Feynman propagator is the probability amplitude for a scalar particle to propagate from the spacetime point x to y , taking into account quantum fluctuations:



$$D_F(x - y) = \lim_{\epsilon \rightarrow 0+} i \int \frac{d^4 p}{(2\pi)^4} \frac{e^{-ip \cdot (x-y)}}{p^2 - m^2 + i\epsilon} \quad (5.37)$$

(5.33) can likewise be moved to position space. After generalising, we have



$$(\square + m^2)D_F(x - y) = -i\delta^4(x - y) \quad (5.38)$$

where, importantly, we see that the Feynman propagator is a Green's function related to the Klein-Gordon operator. It represents the probability amplitude for a particle to propagate from spacetime point y to x . The equation above simply shows that the Klein-Gordon operator governs this propagation.

Note 5.1 For convenience, we often suppress (i.e. conveniently forget writing) $\lim_{\epsilon \rightarrow 0+}$ or both $\lim_{\epsilon \rightarrow 0+}$ and $i\epsilon$ for brevity, in which case their existence is assumed.

This is the easy part. Next, we will see how a perturbative expansion can allow us to read off the propagator. First, we show that (5.29) is a Taylor expansion of λ in the ϕ^4 theory. A Feynman diagram with m incoming external legs and n outgoing external legs is represented a $(m+n)$ -point Green's function, which itself is made up of Feynman propagators. As the Feynman propagator has a S_{fi} term, it can be written as a perturbative expansion like S_{fi} :

$$G_{m+n} = \sum_k G_{m+n}^{(k)} \quad (5.39)$$

Each $G_{m+n}^{(k)}$ corresponds to a class of (various possible) Feynman diagrams with k vertices⁶, where the coupling constant λ is of the order k . For example:

- The 0^{th} term represents the so-called *free propagator* where no interactions happen. It has only external leg propagators.
- For $k > 0$, the k^{th} term represents the Feynman propagator where k interaction happens (represented by a Feynman diagram of order k). It has both external and internal propagators.

Derivation 5.5 ($2 \rightarrow 2$ processes)



⁶That is to say, each interaction vertex arises from an insertion of the interaction term from the Lagrangian.

Quote 5.4 It's like an electric board with four sockets, and each field is a plug.

Alessio Serafini, on 2 → 2 Feynman diagrams, 27 February 2025

Putting it all together, we now look at the example of a $(2+2)$ -point Green's function with incoming external legs x_1 and x_2 , outgoing external legs x_3 and x_4 and no interaction in the middle of the Feynman diagram. From Wick's third theorem (5.17), we see that the Green's function can be represented entirely in terms of Feynman propagators as

$$G_{2+2}^{(0)} = D_F(x_1 - x_2)D_F(x_3 - x_4) + D_F(x_1 - x_3)D_F(x_2 - x_4) + D_F(x_1 - x_4)D_F(x_2 - x_3) \quad (5.40)$$

If 1 interaction happens at point y , we have the extra term

$$G_{2+2}^{(1)} = -i \frac{\lambda}{4!} \frac{4!}{1!} \int d^4y D_F(x_1 - y)D_F(x_2 - y)D_F(x_3 - y)D_F(x_4 - y) \quad (5.41)$$

where λ and $1/4!$ are the coupling constant and the normalisation term seen in (5.1). The factor $4!$ that follows accounts for the fact that each vertex has four ϕ -fields, and there are $4!$ ways to contract these (four) fields with external legs or propagators^a. The factor $1!$ accounts for the first-order perturbation. If 2 interactions happen at points y_1 and y_2 on the left and right sides respectively, we have the extra term, where the factors λ , $1/4!$ and $4!$ are squared:

$$G_{2+2}^{(2)} = - \left(\frac{\lambda}{4!} \right)^2 \frac{4!^2}{2!} \int d^4y_1 \int d^4y_2 D_F(x_1 - y_1)D_F(x_2 - y_1)D_F(x_3 - y_2)D_F(x_4 - y_2) \quad (5.42)$$

where the factor $1!$ accounts for the second-order perturbation, and so on in higher orders.

^aThis is the multiplicity C which we will see later.

We also note that the same Feynman diagram can be yielded from different ways of contracting operators, which may arise multiple times in the series expansion. This is reflected in the Feynman diagram by the multiplicity C . We thus define:

Definition 5.9 (Multiplicity) The *multiplicity* or *multiplicity factor* C with respect to a specific Feynman diagram is the number of possible Wick contractions that result in the specific diagram.

This is a natural result from Wick's second theorem, where we have summed over all possible contractions of the field operators, many of which are actually identical.

Quote 5.5 It is not trivial.

Paulina Schlachter, on calculating C , 25 February 2025

Finally, we are in a position to discuss the *Feynman rules*. The Feynman rules do not produce S_{fi} itself, but rather a quantity called the *scattering amplitude* or *transition amplitude* \mathcal{M}_{fi} , a matrix related to the *S-matrix*. The advantage of the Feynman rules is that they are set up specifically to make constructing \mathcal{M}_{fi} fast and intuitive. But first, we need to provide a definition for \mathcal{M}_{fi} :

Derivation 5.6 (Scattering amplitude) Let us begin with (5.28), where we have m incoming external legs and n outgoing external legs. At this point, two simplifications can be made. The first is *amputating* our poor propagators (also known as *truncating* our poor external legs):

Definition 5.10 (Amputated propagator) It is convenient to write all 'internal' or 'interior' parts of the propagator as a so-called *amputated propagator* $\bar{G}(y_1, \dots, y_l)$:

$$G_{m+n} = D_F(x_1 - y_1) \cdots D_F(x_{n+m} - y_l) \times \bar{G}(y_1, \dots, y_l) \quad (5.43)$$

where we have $l \leq m + n$ since more than one leg may couple to the same vertex.

Remark 5.4 As the amputated propagator is effectively the whole propagator with all external leg propagators removed^a, it is known to be 'amputated'.

Quote 5.6 Arm raised, but not too high.

Felix Halbwedl, 30 March 2025

Now the question arises: Why do we do this? Let us inspect the RHS of (5.28). On the far right, we have the two factors $\prod_{j=1}^{m+n} (\partial_{x_j}^2 + m^2) G_{m+n}$. Importantly, we see that $\prod_{j=1}^{m+n} (\partial_{x_j}^2 + m^2)$ are effectively the inverses of the external leg Feynman propagators, which cancel out with the external leg Feynman propagators themselves in G_{m+n} . As such, the two factors together is effectively the amputated propagator (5.43).

The second simplification is the Fourier transformation, through which we eliminate the rather pesky 4-integrals. We can then define the scattering amplitude \mathcal{M}_{fi} as

Definition 5.11 (Transition amplitude)

$$S_{fi} = \delta_{fi} + i(2\pi)^4 \delta^4(p_f - p_i) \mathcal{M}_{fi} \quad (5.44)$$

where m and n are indices for initial and final particles, $p_i = \sum_{j=1}^n p_j$ and $p_f = \sum_{j=1}^n q_j$, and $\delta^4(p_i - p_f)$ enforces momentum conservation.

Remark 5.5 The physical significance of \mathcal{M}_{fi} can be quickly found: The Kronecker delta δ_{fi} is essentially the identity that accounts for the case where there is no interaction^b. Hence, \mathcal{M}_{fi} is a rescaled version of the interaction-dependent part of S_{fi} . Crudely speaking, we can reduce the relationship to

$$\mathcal{M}_{fi} = \langle f | S_{fi} - 1 | i \rangle \quad (5.45)$$

^aNote however that external *points* which represent initial and final states are preserved. This is trivial in ϕ^4 theory as all external points are 1, but become important in QED where external points represent states, which are 4-spinors or 4-vectors.

^bThis is often omitted as we assume that there *is* an interaction.

The Feynman rules of ϕ^4 theory are thus as follows:

Theorem 5.7 (ϕ^4 theory Feynman rules) For a given Feynman diagram in ϕ^4 theory, the scattering amplitude matrix elements with a complex factor $i\mathcal{M}_{fi}$ are constructed as follows:

ϕ^4 THEORY FEYNMAN RULES (PARTIAL)	
For each	Add to expression
Incoming/outgoing scalar particle	1^{a}
Internal line	$\frac{i}{k_j^2 - m^2}$
Internal loop	$\int \frac{d^4 k_j}{(2\pi)^4} \frac{i}{k_j^2 - m^2}$
Vertex	$-i \frac{\lambda}{4!} (2\pi)^4 \delta^4 \sum_j p_j^{\text{b}}$

where k_j is the propagator momenta, p_j is the incoming momenta and q_j is the outgoing momenta. We also perform the following steps:

- Introduce the multiplicity C for the number of contractions leading to the same diagram.
- Include the $1/k!$ factor from the Taylor expansion, where k is the perturbative order.
- Remove a factor of $(2\pi)^4 \delta^3(p - q)$, where p and q are the total initial and detected momenta, as it is usually already accounted for in the definition of the cross-section or decay rate in terms of the matrix elements.

^aUnlike spinor and vector fields in QED, scalar fields in our good ol' harmless ϕ^4 theory manifests in a mere number (i.e. scalar) which, in our case, just so happens to be unity.

^bThe sum is over all lines exiting the vertex and forces four-momentum conservation at the vertex.

Remark 5.6 For $2 \rightarrow 2$ interactions ($m = n = 2$), \mathcal{M}_{fi} reduces very nicely to $-i\lambda$. Here the nature of ϕ^4 theory as a useful toy model becomes clear.

Note 5.2 (Sneak peek of renormalisation) It should be obvious that, to make sure that our results are physical, the scattering amplitude must observe

$$0 \leq |\mathcal{M}|^2 \leq 1 \quad (5.46)$$

At the tree level, this is always observed. However, the introduction of loops will often bring about divergences (often into infinity) that violate this condition. These divergences suggest that our theory is not entirely physical beyond the tree level. To kill the divergences, we must *renormalise* our field theory of interest. How we do this will be at the heart of Part IV.

Happily, we have now constructed $i\mathcal{M}_{fi}$ from the Feynman rules. By removing the complex factor and inserting \mathcal{M}_{fi} into (5.44), we can solve for S_{fi} , which is much easier than using the LSZ formula. Finally, we note that:

- External propagators/legs/lines have the same expression as internal lines, which is $i(k_j^2 - m^2)^{-1}$. However, they are not included in the Feynman rules as the external legs are already truncated in \mathcal{M}_{fi} .
- In the same vein, external legs are truncated in S_{fi} . Both are a result of the cancelling out of external legs in (5.28). However, this does not mean that external legs are physically meaningless, as they still appear in Feynman diagrams.

Chapter 6

Interacting fields II: QED

A generalised and actually physical version of the Klein-Gordon equation is the Dirac equation. The Dirac field, which represents fermions (i.e. electron and positron), then accounts for one-half of QED. The other half is the photon field, which is governed by Maxwell's equations and is the half of QED that makes it a gauge theory. Using canonical quantisation, we will quantise the Dirac and photon (Maxwell) fields and develop a Lagrangian for quantum electrodynamics, which accounts for fields generated by both electrons/positrons and photons. Both fields are a departure from our innocent ϕ^4 fields: the Dirac field is a fermion field instead of a boson field, and the photon field, while a boson field, is a gauge field that requires gauge fixing.

6.1 Dirac equation

The ill-fated Klein-Gordon equation, which we have found to be *kaputt*, is a Lorentz-invariant 2nd-order DE. We now propose a better candidate in the form of a Lorentz-invariant 1st-order DE, whose most general form would be known as the *Dirac equation*.

Theorem 6.1 (Dirac equation)

$$(i\gamma^\mu \partial_\mu - m)\phi = (i\cancel{a} - m)\psi = 0 \quad (6.1)$$

γ^μ is a yet undetermined 4-vector, and we have defined the so-called *Feynman slash notation* for some four-vector a

$$\cancel{a} = \gamma^\mu a_\mu \quad (6.2)$$

Quote 6.1 The equation was more intelligent than its author.

Paul Dirac, on his equation^a (disputed)

^aThe rationale behind the quote, as per Victor Weisskopf, was that 'A great deal more was hidden in the Dirac equation than the author had expected when he wrote it down in 1928'.

Remark 6.1 Again, this formulation stems from intuition. The modern standard derivation involves inserting the Dirac Lagrangian (13.22) (which we ultimately construct by intuition) into the Euler-Lagrange equations.

Derivation 6.1 (γ^μ matrices) The Dirac equation does not directly conflict with the Klein-Gordon equation:

- As the Klein-Gordon equation must still be satisfied to fulfil the SR energy-momentum relation, every solution to the Dirac equation is also a solution to the Klein-Gordon equation^a.
- The reverse is not true, as the spinorial nature of the solution means that the probability $\rho = \psi^\dagger \psi$ will always be non-negative, and *all negative probability states are excluded*.

The first observation then inspires us to investigate the nature of the γ^μ matrices by considering what happens in the Klein-Gordon limit. We impose the differential operator $i\gamma^\mu \partial_\mu + m$ to the Dirac

equation^b and equating it with the Klein-Gordon equation

$$\underbrace{-(\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu - m^2)\psi}_{\text{application of } i\gamma^\mu \partial_\mu + m} = \underbrace{-(\partial^\mu \partial_\mu - m^2)\psi}_{\text{Klein-Gordon equation}} = 0 \quad (6.3)$$

Immediately, we see that the term $\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu$ must be equal to $\partial^\mu \partial_\mu$. Effectively, through $\gamma^\mu \gamma^\nu$, one index was shifted up. Hence, a sensible guess of $\gamma^\mu \gamma^\nu$ would be the metric:

$$\gamma^\mu \gamma^\nu = g^{\mu\nu} \quad (6.4)$$

However, this is wrong for the reason that the off-diagonal components of the 4-metric are zero, thus implying

$$\gamma^0 \gamma^1 = 0 \quad \text{and} \quad (\gamma^0)^2 = -(\gamma^1)^2 = \mathbb{I}_4 \quad (6.5)$$

at the same time. In fact, such conditions can never be satisfied as long as the components of γ^μ are mere numbers. However, if one switches the indices on the LHS of (6.3) and adds this otherwise identical expression to (6.3), they will find

$$\gamma^\mu \gamma^\nu \partial_\mu \partial_\nu + \gamma^\nu \gamma^\mu \partial_\nu \partial_\mu = 2\partial^\mu \partial_\mu \quad (6.6)$$

This becomes an anticommutation relation in the form of

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}\mathbb{I}_4 \quad (6.7)$$

One can see from (6.7) that γ^μ are elements of a *Clifford algebra*^c, from the definition of which it is clear that elements of γ^μ must be a set of 4 matrices.

Definition 6.1 (γ^μ matrices in the Dirac basis) In the *Dirac basis*, the γ^μ matrices are

$$\gamma^0 = \begin{pmatrix} \mathbb{I}_2 & 0 \\ 0 & \mathbb{I}_2 \end{pmatrix} \quad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix} \quad (6.8)$$

where $j \in \{1, 2, 3\}$ (sometimes also labelled as i), and σ_j are the Pauli matrices we know and love.

^aIn fact, the Dirac equation becomes the Klein-Gordon equation when one takes its time derivative and enforces the γ^μ properties we will see later.

^bNote the + sign of the second term of this operator!

^cThis poses no conflict with Grassmann algebras that define the fields because no single algebra can be said to formulate QED. Rather, both do so simultaneously.

Note 6.1 (Alternative formulation of the γ^μ matrices) It is often convenient to redefine γ^0 and γ^i . We define β and α^i :

$$\beta := \gamma^0 \quad \gamma^i = \beta \alpha^i \quad (6.9)$$

While β is nothing but a relabelling of γ^0 meant to confuse people, α^i actually has its merits, as we will see later.

Interestingly, the Pauli matrices satisfy the following relation:

Theorem 6.2 (Pauli matrices properties)

$$\sigma_j \sigma_k = i\epsilon_{jkl} \sigma_l + \delta_{jk} \mathbb{I}_2 \quad (6.10)$$

$$[\sigma_j, \sigma_k] = 2i\epsilon_{jkl} \sigma_l \quad (6.11)$$

where we have once again encountered our good friends, the Kronecker delta δ_{jk} and the Levi-Civita symbol ϵ_{jkl} .

Finally, we conclude with a list of useful formulae for γ^μ matrices:

Theorem 6.3 (Commonly used γ^μ matrix formulae I)

- γ^0 is nothing but $\text{diag}(\mathbb{I}_2, -\mathbb{I}_2)$ and satisfies:

$$\gamma^{0\dagger} = \gamma^0 \quad \gamma^{0\dagger}\gamma^0 = \gamma^0\gamma^0 = \mathbb{I}_4 \quad \gamma^0\gamma^\mu\gamma^0 = \gamma^{\mu\dagger} \quad \gamma^{0\dagger}\gamma^\mu\gamma^0 = -\gamma^{\mu T} \quad (6.12)$$

- γ^j , where $j = \{1, 2, 3\}$, satisfy:

$$\gamma^{j\dagger} = -\gamma^j \quad \gamma^j\gamma^j = -\mathbb{I}_4 \quad \gamma^{j\dagger}\gamma^j = -(-\mathbb{I}_4) = \mathbb{I}_4 \quad (6.13)$$

- Due to the peculiar anticommutation relation among γ^μ matrices

$$\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}\mathbb{I}_4 \quad (6.14)$$

we can swap the order of matrices γ^μ and γ^ν and introduce a sign flip, so long as $\mu \neq \nu$ ^a.

- A crucial identity involving the γ^2 matrix we use to prove the charge conjugation^b identity (last identity in (6.12)) is

$$\gamma^2\gamma^\mu\gamma^2 = \gamma^{\mu T} \quad (6.15)$$

^aIf $\mu = \nu$, then it is intuitive that the swap does not need a sign flip.

^bWe will soon see what this means.

Theorem 6.4 (Commonly used γ^μ matrix formulae II)

$$\text{Tr}(\gamma^\mu\gamma^\nu) = 4g^{\mu\nu} \quad (6.16)$$

$$\begin{aligned} \text{Tr}(\gamma^{\mu_1} \dots \gamma^{\mu_{2n}}) &= g^{\mu_1\mu_2} \text{Tr}(\gamma^{\mu_3} \dots \gamma^{\mu_{2n}}) - g^{\mu_1\mu_3} \text{Tr}(\gamma^{\mu_2}\gamma^{\mu_4} \dots \gamma^{\mu_{2n}}) + \dots + \\ &\quad g^{\mu_1\mu_n} \text{Tr}(\gamma^{\mu_2} \dots \gamma^{\mu_{2n-1}}) \end{aligned} \quad (6.17)$$

$$\text{Tr}(\gamma^{\mu_1} \dots \gamma^{\mu_{2n+1}}) = 0 \quad (6.18)$$

$$\text{Tr}(\not{a}\not{b}) = 4a \cdot b \quad (6.19)$$

$$\text{Tr}(\not{a}\not{b}\not{c}\not{d}) = 4(a \cdot bc \cdot d - a \cdot cb \cdot d + a \cdot db \cdot c) \quad (6.20)$$

$$\gamma^\alpha\gamma^\mu\gamma_\alpha = -2\gamma^\mu \quad (6.21)$$

$$\gamma^\alpha\gamma^\mu\gamma^\nu\gamma_\alpha = 4g^{\mu\nu} \quad (6.22)$$

$$\gamma^\alpha\gamma^\mu\gamma^\nu\gamma^\rho\gamma_\alpha = -2\gamma^\rho\gamma^\nu\gamma^\mu \quad (6.23)$$

Remark 6.2 As the γ^μ matrices are 4×4 , the wave solution of the Dirac equation ϕ has 4 components. However, it is important to note that this ψ is *not* a vector due to it not transforming under general coordinate transformations. Rather, it is a *spinor* which one can better understand by reading the companion book *Spinors & Symmetries*. While it is possible to simply regard ψ as a 4-vector in some regards, doing so would be quite morally questionable.

Before we recover the Dirac Lagrangian, there remains one loose end to take care of. As we will soon see, this innocent Dirac field is not alone. Rather, it is coupled to a (scalar) photon field A^μ that we will see later. Here, a problem has arisen in the partial derivative field $\partial_\mu A^\mu$, which does not transform covariantly. To compensate for this failure, we augment the partial derivative by introducing a correction term. This new derivative is called the *covariant derivative*, which, in a general gauge theory with a vector field A^μ and coupling constant g , is given by

Definition 6.2 (Abelian covariant derivative)

$$D^\mu = \partial^\mu + igA^\mu \quad (6.24)$$

From this, we can construct the field strength tensor for some vector field A^μ , which has the form

Definition 6.3 (Field strength tensor)

$$F^{\mu\nu} = i[D^\mu, D^\nu] = \partial^\mu A^\nu - \partial^\nu A^\mu - i[A_\mu, A_\nu] \quad (6.25)$$

Remark 6.3 But this feels a bit familiar, doesn't it?

As it turns out, we can make an analogy between gauge theories and good ol' GR:

- The general coordinate transformations we see in GR are analogous to the gauge transformations

in gauge theories.

- In GR, the covariant derivative is introduced to preserve general coordinate transformations. In gauge theories, the covariant derivative is introduced to preserve Noether's theorem in local phase (i.e. gauge) transformations.
- The Christoffel symbols $\Gamma_{\mu\nu}^\sigma$ we use in GR covariant derivatives are analogous to the gauge field A^μ , and both represent a ‘correction’ to their respective transformation¹.
- The Ricci tensor $R_{\mu\nu}$ is analogous to the field strength $F^{\mu\nu}$.
- ...and so on.

The implication of this for QED is then intuitive: In the presence of this scalar field, Our innocent partial derivative ∂ must be rewritten as D . Its second term is modified from that of the generalised covariant derivative by the charge e , which is the coupling constant of QED:

$$D_\mu = \partial_\mu + ieA_\mu \quad (6.26)$$

We can now write down the Lagrangian and Hamiltonian densities for the (free field) Dirac equation for Abelian gauge theories:

$$\mathcal{L} = \bar{\psi}(i\cancel{D} - m)\psi \quad (6.27)$$

which, by noting the Dirac equation itself, is always zero. The bar on top denotes the *Dirac adjoint*:

Definition 6.4 (Dirac adjoint) The Dirac adjoint $\bar{\psi}$ of some ψ is

$$\bar{\psi} = \psi^\dagger \gamma^0 \quad (6.28)$$

$\bar{\psi}$ is not a new field or, indeed, anything physical. It is a mathematically altered ψ that ensures the Lorentz invariance of the Dirac Lagrangian.

Let us write out the covariant derivative explicitly. The Lagrangian is then

Definition 6.5 (Dirac Lagrangian)

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu + ieA_\mu) - m)\psi = \underbrace{\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi}_{\text{Dirac field}} - \underbrace{e\bar{\psi}\gamma^\mu\psi A_\mu}_{\text{interaction}} \quad (6.29)$$

The interaction term denotes interaction with the photon field A^μ . It replaces the interaction $-J^\mu A_\mu$ in the classical Maxwell Lagrangian we will see in (6.121). If further fields (e.g. proton fields) are added, this changes to

$$\mathcal{L} = \sum_f \bar{\psi}(i\gamma^\mu(\partial_\mu + ie_f A_\mu) - m_f)\psi \quad (6.30)$$

where f is an index representing the range of fermions we are concerned with.

Sticking to the simpler case, it then follows that the canonical momentum reads

Definition 6.6 (Dirac canonical momentum)

$$\pi = i\bar{\psi}\gamma^0 = i\psi^\dagger\gamma^0\gamma^0 = i\psi^\dagger \quad (6.31)$$

The Hamiltonian is then

Definition 6.7 (Dirac Hamiltonian)

$$\mathcal{H} = i\psi^\dagger\partial_0\psi - \mathcal{L} = i\psi^\dagger\partial_0\psi = \bar{\psi}(-i\gamma^j\partial_j + m)\psi \quad (6.32)$$

Remark 6.4 The canonical momentum is $\pi = i\psi^\dagger$ by dint of $\gamma^0\gamma^0 = \mathbb{I}_4$.

Derivation 6.2 (Dirac equation general solution) Despite ψ being a spinor^a, a general plane wave solution is again of the form $\psi = ue^{-ip\cdot x}$. From this and the Klein-Gordon on-shell condition, the

¹The gauge field is a connection on the *principal fibre bundle*, while the Christoffels are connections on the *tangent bundles*.

eigenvalue equation is

$$(\not{p} - m\mathbb{I}_4)u = 0 \quad (6.33)$$

The slashed momentum matrix is of the form

$$\not{p} = \begin{pmatrix} p^0\mathbb{I}_2 & p \cdot \sigma \\ p \cdot \sigma & -p^0\mathbb{I}_2 \end{pmatrix} \quad (6.34)$$

where p is the 3-vector and σ is a 3D vector whose elements are the Pauli matrices^b.

Essentially, $(\not{p} - m\mathbb{I}_4)$ is a matrix while the *Dirac spinors* u are 4-spinors that also happen to be the eigenspinors of our matrix $(\not{p} - m\mathbb{I}_4)$. One can solve them as

Theorem 6.5 (Dirac spinor)

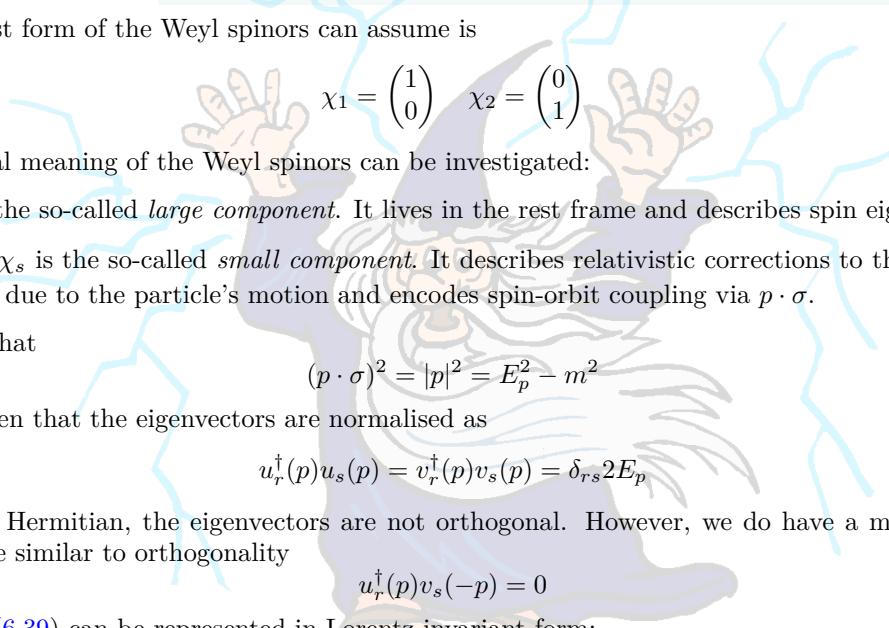
$$u_s(p) = \sqrt{E_p + m} \begin{pmatrix} \chi_s \\ \frac{p \cdot \sigma}{E_p + m} \chi_s \end{pmatrix} \quad v_s(p) = \sqrt{E_p + m} \begin{pmatrix} \frac{p \cdot \sigma}{E_p + m} \chi_s \\ \chi_s \end{pmatrix} \quad (6.35)$$

for $s = 1, 2$. χ_s are 2-component spinors, or so-called *Weyl spinors*.

Quote 6.2 This is a four-dimensional beast.

Alessio Serafini, on the Dirac spinor, 13 March 2025

The simplest form of the Weyl spinors can assume is



$$\chi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \chi_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (6.36)$$

The physical meaning of the Weyl spinors can be investigated:

- χ_s is the so-called *large component*. It lives in the rest frame and describes spin eigenstates.
- $\frac{p \cdot \sigma}{E_p + m} \chi_s$ is the so-called *small component*. It describes relativistic corrections to the spin eigenstates due to the particle's motion and encodes spin-orbit coupling via $p \cdot \sigma$.

By noting that

$$(p \cdot \sigma)^2 = |p|^2 = E_p^2 - m^2 \quad (6.37)$$

it can be seen that the eigenvectors are normalised as

$$u_r^\dagger(p)u_s(p) = v_r^\dagger(p)v_s(p) = \delta_{rs}2E_p \quad (6.38)$$

As \not{p} is not Hermitian, the eigenvectors are not orthogonal. However, we do have a metric relation that is quite similar to orthogonality

$$u_r^\dagger(p)v_s(-p) = 0 \quad (6.39)$$

(6.38) and (6.39) can be represented in Lorentz-invariant form:

$$\bar{u}_r(p)u_s(p) = -\bar{v}_r(p)\bar{v}_s(p) = \delta_{rs}2m \quad \bar{u}_r(p)v_s(p) = 0 \quad (6.40)$$

Note that our Dirac spinors are nothing but 4-spinors. They are *not* the Dirac equation general solution, which is a *wave solution*. However, there exists an intuitive relationship

$$\psi(x) = u(p)e^{-ip \cdot x} \quad (6.41)$$

This essentially moves us back to position space. Finally, we are in a position to write down the full general solution of the Dirac equation, which represents fermions like electrons and positrons:

Theorem 6.6 (Dirac equation general solution) The general solution of the Dirac equation is a spinor field known as the *Dirac field*:

$$\psi = \int d^3p \sum_{s=1}^2 (b_s(p)u_s(p)e^{-ip \cdot x} + d_s(p)v_s(p)e^{ip \cdot x}) \quad (6.42)$$

where $b_s(p)$ and $d_s(p)$ are 4 constants.

The Dirac field is a so-called *Grassmann-valued field*, which is a field whose components take values in a Grassmann algebra. Recalling Grassmann mathematics from *Spinors & Symmetries*, one can realise that the components of Grassmann fields anticommute rather than commute.

From their Dirac equation, spinors $u_s(p)$ and $v_s(p)$ further satisfy

$$\sum_s u_s(p) \bar{u}_s(p) = \not{p} + m \quad \sum_s v_s(p) \bar{v}_s(p) = \not{p} - m \quad (6.43)$$

The on-shell condition is enforced by $p^0 = E_p$. The $u_s(p)$ term accounts for positive energy solutions which represent matter, while the $v_s(p)$ term accounts for negative energy solutions which represent antimatter.

^aThe horror!

^bNote that we are still living in momentum space.

Let us continue with two physical discussions. The first concerns how the negative energy solutions gave rise to antimatter:

- Like the infamous Klein-Gordon equation, the Dirac equation also has negative energy solutions. However, rather than dismiss them as unphysical, Dirac proposed that all negative energy states are filled in a kind of vacuum state, which is now known as the *Dirac sea*.
- A ‘hole’ in this sea - i.e. a missing negative-energy electron - would appear as a positive-energy, positive-charge particle. This is the positron which was discovered by Carl Anderson in 1932.
- The formulation of the positron was later expanded, which gave rise to antimatter as antiparticles of matter. The emergence of antiparticles has thus justified the negative energy solutions as physical. This is known as the *Feynman-Stückelberg interpretation*.

Definition 6.8 (Feynman-Stückelberg interpretation) We consider the exponential term of the wave solution^a which has the following equivalence:

$$e^{-i(-E)(-t)} = e^{-iEt} \quad (6.44)$$

As such, we can understand negative energy solutions in two equally valid ways:

- A positive energy antiparticle travelling forward in time. This implies antimatter and is the physically useful interpretation.
- A negative energy particle travelling backwards in time. This is the practically useful interpretation.

^aFor our intuitive convenience, we only look at the zeroth components of the inner product, but the principle applies to all components.

The second is to do with the nature of the Dirac equation. The Dirac equation fails to describe a single particle and requires a many-particle treatment like QFT for two reasons:

- For a single particle, the Dirac equation implies unbounded negative energy. This is solved by the Dirac sea and antimatter, which inherently implies many particles.
- Even if we remove this historical convenience, the Dirac equation still implies particle creation and annihilation, which is impossible for a single particle.

6.2 Quantisation of the Dirac field

Having acquired the the Dirac field as the general solution of the Dirac equation in (6.42), we now attempt to quantise it, which allows us to quantise several significant quantities using it. In quantising the Klein-Gordon equation, we replaced f_p and f_p^* with creation and annihilation operators (see Derivation 4.3). Here we start less ambitiously. We order that $b_s(p)$ become an operator. For $d_s(p)$, we do the same but replace it with $d_s^\dagger(p)$. The general solution then reads

$$\psi = \int \frac{d^3 p}{(2\pi)^3 \sqrt{2E_p}} \sum_{s=1}^2 (b_s(p) u_s(p) e^{-ip \cdot x} + d_s^\dagger(p) v_s(p) e^{ip \cdot x}) \quad (6.45)$$

where again, for the sake of convenient normalisations, we have added a normalisation factor of $\frac{1}{(2\pi)^3 \sqrt{2E_p}}$ previously seen in (4.36). The canonical momentum can likewise be found:

$$\pi = \int \frac{d^3 p}{(2\pi)^3 \sqrt{2E_p}} \sum_{s=1}^2 (b_s^\dagger(p) u_s^\dagger(p) e^{ip \cdot x} + d_s(p) v_s^\dagger(p) e^{-ip \cdot x}) \quad (6.46)$$

One can then represent the fermionic field creation and annihilation operators $b_s(p)$ and $d_s^\dagger(p)$ in terms of wavefunctions, much like the interacting field creation and annihilation operators (5.25):

$$-id_s^\dagger(p) = \int \frac{d^3 x}{\sqrt{2E_p}} \frac{\bar{v}_s}{2m} (e^{-ip \cdot x} \partial_0 \psi - \psi \partial_0 e^{-ip \cdot x}) \quad ib_s^\dagger(p) = \int \frac{d^3 x}{\sqrt{2E_p}} (e^{-ip \cdot x} \partial_0 \bar{\psi} - \bar{\psi} \partial_0 e^{-ip \cdot x}) \frac{u_s}{2m} \quad (6.47)$$

where the factor of $1/2m$ arises from the normalisation condition (6.41).

Before mindlessly assigning the bosonic commutation relations to $b_s(p)$ and $d_s(p)$, we stop for a moment and realise that the Dirac field is not actually a bosonic field. The solution to the Dirac equation is not a scalar, but a spinor. Hence, it describes spin- $\frac{1}{2}$ particles² and is a fermionic field instead. As such, we instead impose the following *fermionic anticommutation relations* for some indices i and j :

Theorem 6.7 (Fermionic creation and annihilation operator anticommutations)

$$\{b_i(p), b_j^\dagger(q)\} = \{d_i(p), d_j^\dagger(q)\} = \delta_{ij} (2\pi)^2 \delta^3(p - q) \quad (6.48)$$

$$\{b_i(p), b_j(q)\} = \{d_i(p), d_j(q)\} = 0 \quad (6.49)$$

Operators for different particle species (electrons and positrons) anticommute to zero:

$$\{b_i(p), d_j(q)\} = \{b_i^\dagger(p), d_j(q)\} = \{b_i(p), d_j^\dagger(q)\} = \{b_i^\dagger(p), d_j^\dagger(q)\} = 0 \quad (6.50)$$

Suddenly recalling the useful relation (6.43) for no reason whatsoever, we can rewrite it using the definition of the Dirac adjoint (6.28) as

$$\sum_s u_s(p)_\alpha u_s^\dagger(p)_\beta = (E_p + m) \delta_{\alpha\beta} - (p \cdot \gamma \gamma^0)_{\alpha\beta} \quad \sum_s v_s(p)_\alpha v_s^\dagger(p)_\beta = (E_p - m) \delta_{\alpha\beta} - (p \cdot \gamma \gamma^0)_{\alpha\beta} \quad (6.51)$$

Putting (6.46), (6.48), (6.49) and (6.51) all together, we can solve for the commutator $[\phi_\alpha(x), \pi_\beta(y)]$ and find the following anticommutation relation

Theorem 6.8 (Fermionic field and momentum operator anticommutations)

$$\{\psi_\alpha(x), \psi_\beta(y)\} = \{\pi_\alpha(x), \pi_\beta(y)\} = \{\pi_\alpha(x), \pi_\beta^\dagger(y)\} = 0 \quad (6.52)$$

$$\{\psi_\alpha^\dagger(x), \psi_\beta(y)\} = \delta_{\alpha\beta} \delta^3(x - y) \quad (6.53)$$

Importantly, this anticommutator, like its boson counterpart $[\phi, \phi^*]$, is not so much derived as it is enforced, in this case by the fact that fermionic fields are Grassmann-odd. The implication is that ψ^\dagger , and by extension, $\bar{\psi}$ ^a, are not independent from ψ .

Recalling the result from (6.31), we have

$$\{\psi_\alpha(x), \pi_\beta(y)\} = i \delta_{\alpha\beta} \delta^3(x - y) \quad (6.54)$$

^aThis justifies our comment following (6.28).

By integrating (6.32) and then using (6.38), we can find the normal-ordered Hamiltonian

Definition 6.9 (Dirac equation normal-ordered Hamiltonian)

$$H = \int \frac{d^3 p}{(2\pi)^3} E_p \sum_{s=1}^2 (b_s^\dagger(p) b_s(p) + d_s^\dagger(p) d_s(p)) \quad (6.55)$$

We can likewise find the charge by integrating the 0th component 4-current $J^0 = \psi^\dagger \psi$:

²Here we see why a spinor is often called a ‘rank-half tensor’.

Definition 6.10 (Electric charge)

$$Q = e \int d^3x : \psi^\dagger \psi : = e \int \frac{d^3p}{(2\pi)^3} \sum_{s=1}^2 (b_s^\dagger(p)b_s(p) - d_s^\dagger(p)d_s(p)) \quad (6.56)$$

where the normal ordering is used to ensure the result is physical. This is simply the electric charge, from which one can verify a particle and its antiparticle carry equal but opposite charges.

6.3 Story of a spinor

Our quantisation of the Dirac fermion field was easy, or at least no harder than quantising the ϕ^4 field. Before we proceed to quantising the electromagnetic field, which is vastly harder, let us sit on the Dirac equation for a bit and consider its implications, especially with respect to symmetries. This will be useful because, as we will see in Part V, scalar fermions as described by the Dirac equation exist not only in QED but also in all three forces covered by the standard model. One is then motivated to study it more carefully. This is our goal for the next three sections.

Previously, we have said that spinors, unlike tensors, do not undergo general coordinate transformations. Instead, they undergo rotation-like transformations defined by Lie groups. With respect to Dirac spinors, we consider several Lie groups. The starting point is the *homogeneous Lorentz group* $\mathrm{SO}^+(1, 3)$, which is a *connected* group that governs all (continuous) spacetime boosts and rotations³:

- If we add three more components (parity P , time reversal T and PT), we get the *full lorentz group* or the *inhomogeneous Lorentz group* $\mathrm{O}(1, 3)$, a *disconnected* group with four *smoothly separated* components:

$$\mathrm{O}(1, 3) \cong \mathrm{SO}^+(1, 3) \cup P \cdot \mathrm{SO}^+(1, 3) \cup T \cdot \mathrm{SO}^+(1, 3) \cup PT \cdot \mathrm{SO}^+(1, 3) \quad (6.57)$$

- $\mathrm{SO}^+(1, 3)$ is a double cover of $\mathrm{SL}(2, \mathbb{C})$. This is how we go from the 4D Dirac spinor to the 2D Weyl spinor⁴.
- The rotation subgroup of $\mathrm{SO}^+(1, 3)$ is $\mathrm{SO}(3)$, which itself is famously a double cover of $\mathrm{SU}(2)$.
- $\mathrm{SU}(2)$ is fundamental for the concept of spin and all that follows.

Definition 6.11 (Proper and improper Lorentz transformations) There exists two kinds of Lorentz transformations:

- *Proper Lorentz transformations* have matrices with determinant +1. These are rotations and boosts that make up $\mathrm{SO}^+(1, 3)$
- *Improper Lorentz transformations* have matrices with determinant -1. These are the P , T and PT transformations that make up the rest of $\mathrm{O}(1, 3)$.

Let us first consider $\mathrm{SO}^+(1, 3)$.

Derivation 6.3 (Lorentz group generators) Given a Lorentz transformation $x \rightarrow x' = \Lambda x$, we expect the Dirac equation to be Lorentz-invariant:

$$(i\gamma^\mu \partial_\mu - m)\psi(x) \rightarrow (i\gamma^\mu \partial'_\mu - m)\psi'(x') \quad (6.58)$$

where the wavefunction transforms according to the (internal) spinor transformation $T(\Lambda)$, defined by

$$\psi'(x') = T(\Lambda)\psi(x) = T(\Lambda)\psi(\Lambda^{-1}x') \quad (6.59)$$

Now we try to determine $T(\Lambda)$. One can write out the transformation in terms of indices:

$$x'^\mu = \Lambda_\mu^\nu x^\mu \quad \partial_\mu \Lambda_\mu^\nu \partial'_\nu \quad (6.60)$$

³This is a spacetime symmetry that transforms the spinors and is unrelated to the U(1) internal symmetry we will see later.

⁴Recall that the transformations of spinors are ultimately governed by Lie groups.

The original, untransformed and final, transformed equations can then be written as

$$(i\gamma^\nu \Lambda_\nu^\mu \partial'_\nu - m)\psi(\Lambda^{-1}x') = 0 \quad T\Lambda^{-1}(i\gamma^\mu \partial'_\mu - m)T(\Lambda)\psi(\Lambda^{-1}x') = 0 \quad (6.61)$$

By equating the two^a, one finds

$$T(\Lambda)^{-1}\gamma^\mu T(\Lambda) = \lambda_\nu^\mu \gamma^\nu \quad (6.62)$$

Using the parameterisation of generators in *Spinors & Symmetries*, an infinitesimal Lorentz transformation may be parameterised as

$$\Lambda_\nu^\mu = \delta_\nu^\mu - \omega_{\rho\sigma}(g^{\rho\nu}\delta_\nu^\sigma - g^{\sigma\mu}\delta_\nu^\rho) + O(\omega^2) \quad (6.63)$$

and a transformation may be given as

$$T(\Lambda) = e^{i\omega_{\rho\sigma}s^{\rho\sigma}} = \mathbb{I}_4 + i\omega_{\rho\sigma}s^{\rho\sigma} + O(\omega^2) \quad (6.64)$$

where $s^{\rho\sigma}$ are the all-too-familiar generators that we have parameterised via $\omega_{\rho\sigma}$ in *Spinors & Symmetries*. Plugging the two expressions into (6.62) gives

$$i[s^{\rho\sigma}, \gamma^\mu] = g^{\rho\mu}\gamma^\sigma - g^{\sigma\mu}\gamma^\rho \quad (6.65)$$

which surprisingly reduces, via (anti)commutation relations, to the simple expression

Definition 6.12 (Lorentz group generators) The following generators define the spinor representation of the (proper) Lorentz group

$$s^{\rho\sigma} = \frac{i}{4}[\gamma^\rho, \gamma^\sigma] \quad (6.66)$$

We conclude by saying that the transformations are the representations $T(\Lambda) = e^{i\omega_{\rho\sigma}s^{\rho\sigma}}$.

^aOne can do so as both equations apply for all ψ s.

6.4 Spin, helicity and chirality

The hard work in the last section is not pointless. We know already that Lorentz covariance and invariance are important in a field theory and that the Dirac fermion field will return again and again. This motivates us to introduce the so-called *Lorentz group bilinears* or simply *bilinears*, which resemble archetypical terms one would commonly find in gauge theory Lagrangians that:

- Involve the γ^μ matrices.
- Are Lorentz-covariant and thus preserve SR.

Recall from *Spinors & Symmetries* that mathematically, a bilinear is a map that is linear in both its arguments. For Lorentz groups, these 2 arguments are ψ and $\bar{\psi}$. In our context, a bilinear is essentially a Lorentz-covariant object that sandwiches a 4×4 matrix between $\bar{\psi}$ and ψ .

Definition 6.13 (Bilinears) From (6.7), we know that there are a total of 16 possible 4×4 matrix constructions from the anticommutation $\{\gamma^\mu, \gamma^\nu\}$. With this, we have a total of 16 Lorentz group bilinears. Depending on the quantity we sandwich with $\bar{\psi}$ and ψ , we can split them into five categories according to their transformation properties:

- **Scalar:**

$$\bar{\psi}\mathbb{I}_4\psi \quad (6.67)$$

In a typical Lagrangian, we can see that this becomes a mass term if we attach $\frac{1}{2}m^2$.

- **Pseudoscalar:**

$$\bar{\psi}\gamma^5\psi \quad (6.68)$$

where $\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3$. It is associated with chirality.

- **Vector:**

$$\bar{\psi}\gamma^\mu\psi \rightarrow \Lambda_\nu^\mu\bar{\psi}\gamma^\mu\psi \quad (6.69)$$

It can represent conserved currents^a and obeys the continuity equation to enforce charge conservation:

$$\partial_\mu(\bar{\psi}\gamma^\mu\psi) = 0 \quad (6.70)$$

- **Pseudovector:**

$$\bar{\psi}\gamma^\mu\gamma^5\psi \rightarrow \det(\Lambda)\Lambda_\nu^\mu\bar{\psi}\gamma^\mu\gamma^5\psi \quad (6.71)$$

It appears in theories involving axial currents^b and chiral symmetry.

- **Rank-2 tensor:**

$$\bar{\psi}\sigma^{\mu\nu}\psi \rightarrow \Lambda_\lambda^\nu\Lambda_\sigma^\mu\bar{\psi}\sigma^{\lambda\sigma}\psi \quad (6.72)$$

where the $\sigma_{\mu\nu}$ is defined as the commutatation of two γ^μ matrices up to a factor

$$\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu\gamma^\nu] \quad (6.73)$$

^aIn fact, the electromagnetic 4-current can be written as $J^\mu = \bar{\psi}\gamma^\mu\psi$.

^bAs such, it is also called a *axial vector*. For example, the axial current in weak interactions is $\bar{\psi}\gamma^\mu\gamma^5\psi$ and plays a role in describing the *handedness* of particles.

Some comments are in order:

- By looking at the 4D spacetime index, we see that the scalar, pseudoscalar, vector, pseudovector and rank-2 tensor have 1, 1, 4, 4 and 6⁵ incarnations respectively.
- We can have higher-rank terms that are Lorentz-covariant, but they will always be combinations of the bilinears we defined above. As such, our 16 bilinears are actually the entire range of *independent* bilinears.
- Each type of bilinear is so-called not because they are scalars, vectors, etc. (they are all 4×4 matrices), but rather because of how they behave under Lorentz transformations: The scalar and the pseudoscalar are invariant (i.e. transform as scalars). The vector and the pseudovector transform as vectors. The rank-2 tensor transforms as a rank-2 tensor.
- We attach ‘pseudo-’ prefix before the pseudoscalar and the pseudovector because they change sign under the parity transformation. This, as we have seen, results from the flipping of the chirality of the γ^5 matrix.
- A hidden significance of these bilinears will be seen by the end of Part V.

The most interesting object here is the so-called γ^5 *matrix*, which is actually a combination of our previous γ^μ matrices but is nonetheless important enough to merit its own symbol.

Definition 6.14 (γ^5 matrix)

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad (6.74)$$

This is a γ^μ matrix in name only which has nothing to do with the Lorentz group generators⁶. That being said, the γ^5 matrix is not a mathematical convenience, and its significance is not restricted to the bilinears. We will see this when it comes back to annoy us over and over again throughout the rest of the book.

Theorem 6.9 (γ^5 matrix properties)

- γ^5 is unitary:

$$\gamma^{5\dagger} = \gamma^5 \quad (\gamma^5)^2 = \mathbb{I}_4 \quad (6.75)$$

- γ^5 anticommutes with γ^μ matrices:

$$\{\gamma^5, \gamma^\mu\} = 0 \rightarrow \gamma^5\gamma^\mu = -\gamma^\mu\gamma^5 \quad (6.76)$$

⁵While there are a total of 16 possible combinations, they reduce to $\frac{4 \times (4-1)}{2} - 6$ due to the antisymmetry of $\sigma^{\mu\nu}$.

⁶In fact, the notation γ^5 is nothing but a historical artefact: In the old days, indices we call 0, 1, 2 and 3 were labelled 1, 2, 3 and 4, and what we today call the γ^5 matrix became so-named due to the superficial similarity between it and the γ^μ matrices observed by 1930s physicists.

To even begin to understand the significance of γ^5 , we need to first return to quantum mechanics. Spin concerns the rotation of Dirac equation solutions, which are Dirac spinors. We already know that the entirety of Lorentz transformations - translations and rotations - are governed by $\text{SO}^+(1, 3)$, which is where we will start:

- If we remove translations, we get the 3D rotation group $\text{SO}(3)$.
- Now consider which double cover this group forms (i.e. go from Dirac spinors to Weyl spinors), and we get $\text{SU}(2)$, which is intuitively the Lie group that governs spin.
- The generators of $\text{SU}(2)$ are the Pauli matrices.

However, at this point, we have to go back from Weyl spinors to Dirac spinors. This results in the so-called *spin operator* \mathbf{S} , which is actually a ‘vector of matrices’. To see what this means, we write down the spin operator in its individual components S_i as well as its entirety:

Definition 6.15 (Spin angular momentum operator)

$$\hat{S}_i = \frac{1}{2}\Sigma_i = \frac{1}{2} \begin{pmatrix} \sigma_i & 0 \\ 0 & \sigma_i \end{pmatrix} \quad \mathbf{S} = \frac{1}{2}\boldsymbol{\Sigma} = \frac{1}{2} \begin{pmatrix} \boldsymbol{\sigma} & 0 \\ 0 & \boldsymbol{\sigma} \end{pmatrix} \quad (6.77)$$

where we have defined

$$\Sigma_i = \frac{i}{2}\epsilon_{ijk}\gamma^j\gamma^k = \frac{i}{2}\epsilon_{ijk}\alpha^j\alpha^k \quad (6.78)$$

Essentially, \mathbf{S} is a 3-vector whose components are operators.

From quantum mechanics, one recalls that spin has no classical analogue. It contributes, along with the *orbital angular momentum* which is carried over from classical mechanics, to the total angular momentum, which is ultimately conserved. We will prove this now.

Derivation 6.4 (Conservation of angular momentum) The total angular momentum operator \mathbf{J} is the sum of the orbit and spin angular momentum operators. This is known as *spin-orbit coupling*:

$$\mathbf{J} = \mathbf{L} + \mathbf{S} \quad (6.79)$$

To prove that it is a conserved quantity, all we have to do is to show that it satisfies (3.38). This can be done by evaluating $[\hat{L}_i, H]$ and $[\hat{S}_i, H]$ individually. Let us first recall the definition of \hat{L}_i :

Definition 6.16 (Orbital angular momentum operator)

$$\hat{L}_i = \epsilon_{ijk}x^j p^k \quad \mathbf{L} = \mathbf{x} \times \mathbf{p} \quad (6.80)$$

We now calculate the Dirac Hamiltonian (density) by inserting *the free part of* (13.22) into (2.7) and then both into (2.28):

$$H = \boldsymbol{\alpha} \cdot \mathbf{p} + \beta m \quad (6.81)$$

Now take the commutator of (6.80) and this. We get

$$[\hat{L}_i, H] = i\epsilon_{ijk}\gamma^j p^k \quad [\mathbf{L}, H] = i\boldsymbol{\gamma} \times \mathbf{p} \quad (6.82)$$

Doing the same for (6.79) gives

$$[\hat{S}_i, H] = -i\epsilon_{ijk}\gamma^j p^k \quad [\mathbf{S}, H] = -i\boldsymbol{\gamma} \times \mathbf{p} \quad (6.83)$$

Indeed, we are left with

$$[\hat{L}_i + \hat{S}_i, H] = 0 \quad (6.84)$$

from which we recover the conservation of the total angular momentum of \mathbf{J} . We then know that \mathbf{J} is also Lorentz-invariant.

Having finished our little recap, we can discuss some concepts relevant in HEP.

Definition 6.17 (Helicity operator) The *helicity operator* is the projection of spin in the direction of the momentum. In index and vector notations, it is written as^a

$$\hat{h} = \frac{\Sigma_i p^i}{|p_i|} = \frac{\gamma^5 \gamma^0 \gamma_\mu p^\mu}{|p_\mu|} \quad \hat{h} = \frac{\mathbf{\Sigma} \cdot \mathbf{p}}{|\mathbf{p}|} = \frac{\gamma^5 \gamma^0 \boldsymbol{\gamma} \cdot \mathbf{p}}{|\mathbf{p}|} \quad (6.85)$$

Just like how the spin s of a 4-spinor is the eigenvalue of the spin operator, the *helicity* h of a 4-spinor is then its eigenvalue under the helicity operator.

^aSome alternative definitions exist up to a rescaling factor.

However, somewhat confusingly, we also speak of *left-handed* and *right-handed* helicity:

- When $h = +1$, one may speak of the helicity as right-handed. As the helicity is positive, the particle's spin direction is the same as its direction of motion. This orientation is so-called as it observes the right-hand rule: Align the right thumb in the momentum's direction, and the curled fingers should align with the spin direction.
- When $h = -1$, one may speak of the helicity as left-handed. As the helicity is negative, the particle's spin direction is opposite to its direction of motion. This orientation is so-called as it observes the left-hand rule: Align the left thumb in the momentum's direction, and the curled fingers should align with the spin direction.

Now we want to solve for Dirac equation eigenspinors that are also helicity eigenstates. This means that the Weyl spinors have more complex structures compared to (6.36). The approach is simple. We take the Weyl spinors of each possible 4-spinor as an unknown and solve them by applying the helicity operator. The most convenient basis is the so-called *helicity basis*, where Dirac spinors are nothing but eigenspinors of the helicity operator with eigenvalues $h = \pm 1$:

$$\hat{h}\psi = \lambda\psi = \pm\psi \quad (6.86)$$

The eigenstates are then

Theorem 6.10 (Helicity eigenstates) For massive spin- $\frac{1}{2}$ free fermions, the helicity basis helicity eigenstates of matter propagating in the (θ, ϕ) direction are

$$u_\uparrow = \sqrt{E+m} \begin{pmatrix} \cos(\theta/2) \\ e^{i\phi} \sin(\theta/2) \\ \frac{|\vec{p}|}{E+m} \cos(\theta/2) \\ \frac{|\vec{p}|}{E+m} e^{i\phi} \sin(\theta/2) \end{pmatrix} \quad u_\downarrow = \sqrt{E+m} \begin{pmatrix} -\sin(\theta/2) \\ e^{i\phi} \cos(\theta/2) \\ \frac{|\vec{p}|}{E+m} \sin(\theta/2) \\ -\frac{|\vec{p}|}{E+m} e^{i\phi} \cos(\theta/2) \end{pmatrix} \quad (6.87)$$

where θ is the polar angle (deviation from the z -axis) and ϕ is the azimuthal angle, whose term $e^{i\phi}$ represents the rotation effect by the xy plane.

The states observe

$$\hat{h}u_\uparrow = +u_\uparrow \quad \hat{h}u_\downarrow = -u_\downarrow \quad (6.88)$$

For antimatter, the helicity operator $\hat{h}^v = -\hat{h}$ is the negative of its matter counterpart, and one has the eigenstates

$$v_\uparrow = \sqrt{E+m} \begin{pmatrix} \frac{|\vec{p}|}{E+m} \sin(\theta/2) \\ -\frac{|\vec{p}|}{E+m} e^{i\phi} \cos(\theta/2) \\ -\sin(\theta/2) \\ e^{i\phi} \cos(\theta/2) \end{pmatrix} \quad v_\downarrow = \sqrt{E+m} \begin{pmatrix} \frac{|\vec{p}|}{E+m} \cos(\theta/2) \\ \frac{|\vec{p}|}{E+m} e^{i\phi} \sin(\theta/2) \\ \cos(\theta/2) \\ e^{i\phi} \sin(\theta/2) \end{pmatrix} \quad (6.89)$$

which observe

$$\hat{h}^v v_\uparrow = +v_\uparrow \quad \hat{h}^v v_\downarrow = -v_\downarrow \quad (6.90)$$

Alas, we have been going too far. If we go back to (6.85), we will see that γ^5 has reemerged in the helicity operator. That is to say, the ghost of γ^5 past has yet again come back to haunt us. This forces us to consider that this nasty little matrix may have additional physical meanings. To accomplish this, we

introduce the concept of *chirality* (from ‘hand’ in Greek), which denotes the ‘handedness’ of a particle. The so-called *chiral transformation* is given by

Definition 6.18 (Chiral transformation) The chiral transformation \mathcal{C} is defined with respect to the transformation $e^{i\alpha\gamma^5} \in U(1)_A$ in the *axial U(1) group* labelled $U(1)_A$, which is mathematically identical but physically distinct from the $U(1)$ group in QED:

$$\mathcal{C}\psi = e^{i\alpha\gamma^5}\psi \quad (6.91)$$

where α is the rotation angle. The eigenvalue of a 4-vector under a chiral transformation is then known as *chirality*.

Yet this is not the end of the story. Let us consider the specific case of $\alpha = \pi$. From Euler’s formula, this gives

$$\mathcal{C}\psi = \gamma^5\psi \quad (6.92)$$

This is known as *flipping the helicity*, and γ^5 is thus known as the *helicity operator*. But we have avoided discussing why we connected chirality with handedness. This shroud of mystery lifts itself when we define two, so-called, *chiral projectors*:

Definition 6.19 (Chiral projectors) The left-handed and right-handed chiral projectors are defined as

$$P_L = \frac{\mathbb{I}_4 - \gamma^5}{2} \quad P_R = \frac{\mathbb{I}_4 + \gamma^5}{2} \quad (6.93)$$

Through them, we acquire the left-chiral and right-chiral spinors:

$$\psi_L = P_L\psi = \frac{\mathbb{I}_4 - \gamma^5}{2}\psi \quad \psi_R = P_R\psi = \frac{\mathbb{I}_4 + \gamma^5}{2}\psi \quad (6.94)$$

where

$$\psi = \psi_L + \psi_R \quad (6.95)$$

and, for matter and antimatter

$$\hat{P}_R u_R = u_R \quad \hat{P}_L u_R = 0 \quad \hat{P}_R u_L = 0 \quad \hat{P}_L u_L = u_L \quad (6.96)$$

$$\hat{P}_R v_R = 0 \quad \hat{P}_L v_R = v_R \quad \hat{P}_R v_L = v_L \quad \hat{P}_L v_L = 0 \quad (6.97)$$

Both ψ_L and ψ_R display eigenvalues of ± 1 when γ^5 is imposed:

$$\underbrace{\gamma^5\psi_R = \psi_R}_{\text{matter}} \quad \underbrace{\gamma^5\psi_L = -\psi_L}_{\text{matter}} \quad \underbrace{\gamma^5\psi_R = -\psi_R}_{\text{antimatter}} \quad \underbrace{\gamma^5\psi_L = \psi_L}_{\text{antimatter}} \quad (6.98)$$

That is to say, by using γ^5 , we can acquire the left- and right-handed components of any 4-spinor.

Theorem 6.11 (Conservation of chirality) Chirality is conserved in any interaction which includes currents $\bar{\psi}\gamma^\mu\phi$.

$$\begin{aligned} \bar{\psi}\gamma^\mu\phi &= \bar{\psi}_R\gamma^\mu\phi_R + \bar{\psi}_R\gamma^\mu\phi_L + \bar{\psi}_L\gamma^\mu\phi_R + \bar{\psi}_L\gamma^\mu\phi_L \\ &= \bar{\psi}_R\gamma^\mu\phi_R + 0 + 0 + \bar{\psi}_L\gamma^\mu\phi_L \end{aligned} \quad (6.99)$$

Two implications arise from the conservation of chirality:

- Right-handed particles only interact with right-handed particles, and left-handed particles only interact with left-handed particles.
- It is impossible to turn a left-handed chiral state into a right-handed chiral state through the exchange of a virtual particle, and vice versa⁷.

⁷e.g. The chirality of an electron cannot be changed when it interacts with a photon.

Derivation 6.5 (Weyl basis) So far we have been working in the Dirac basis^a. A second basis is the so-called *Weyl basis* or *chiral basis*, where the γ^0 is slightly changed, while the other γ^μ matrices and the formulas, including the left- and right-hand projections (6.94), stay the same. The Weyl basis has the form

Definition 6.20 (Weyl basis)

$$\gamma_{\text{ch}}^0 = \begin{pmatrix} 0 & \mathbb{I}_2 \\ \mathbb{I}_2 & 0 \end{pmatrix} \quad \gamma_{\text{ch}}^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix} \quad (6.100)$$

A unitary matrix U can be used to transform objects from the Dirac basis to the Weyl basis:

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathcal{I}_2 & \mathcal{I}_2 \\ -\mathcal{I}_2 & \mathcal{I}_2 \end{pmatrix} \quad (6.101)$$

Respectively, a 4-spinor and the γ^μ matrices transform as

$$\psi_{\text{Weyl}} = U\psi_{\text{Dirac}} \quad \gamma_{\text{Weyl}}^\mu = U\gamma_{\text{Dirac}}^\mu U^\dagger \quad (6.102)$$

The Weyl basis is very useful in combination with chiral projectors. This is because in the Weyl basis, the two Weyl spinors of a Dirac spinor are automatically the non-zero halves of the Dirac spinor's left-handed and right-handed components^b. That is to say, one always has, for 2-spinors ξ and η :

$$\psi = \psi_L + \psi_R \quad \psi_L = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \quad \psi_R = \begin{pmatrix} 0 \\ \eta \end{pmatrix} \quad (6.103)$$

As half of the components of ψ_L and ψ_R are zero, only 2 independent degrees of freedom are left. As such, one can effectively discard the zero components. This gives rise to another formalism where ψ_L and ψ_R reduce to 2D and, hence, are regarded as Weyl spinors:

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix} \quad \psi_L = \xi, \quad \psi_R = \eta \quad (6.104)$$

Both formalisms are common. Note that they can only be used in the Weyl basis.

^aLike the majority of the literature, we follow the Dirac basis unless noted.

^bOne might be tempted to ask whether the Weyl basis or the chiral projectors came first, but such a question is as meaningful as asking which came first, the chicken or the egg.

However, the flipping of chirality has a deeper meaning we have not elucidated yet, and we yet again find ourselves terrorised by γ^5 . Here it is necessary to introduce the concept of *symmetry breaking*.

Definition 6.21 (Symmetry breaking) Two types of symmetry breaking exist:

- In *spontaneous symmetry breaking*, the equations of motion are invariant, but the ground state (vacuum) of the system is not^a.
- In *explicit symmetry breaking*, the equations of motion are not invariant.

^aWe do not need to worry about this until Part III.

It may seem surprising at this point, but as we will see, explicit symmetry breaking happens all the time. One example can be seen by placing the helicity-flipped Dirac spinor $\gamma^5\psi$ into the Dirac equation.

Derivation 6.6 ($\gamma^5\psi$ as massless solution) We consider the LHS of the Dirac equation $i\partial_\mu$ is Fourier-transformed with p_μ :

$$(i\gamma^\mu \partial_\mu - m)\gamma^5\psi = i\gamma^\mu \gamma^5 \partial_\mu \psi - m\gamma^5\psi \quad (6.105)$$

Using anticommutation, we move γ^5 past γ^μ and find

$$i\gamma^\mu \gamma^5 \partial_\mu \psi - m\gamma^5\psi = \gamma^5(i\gamma^\mu \partial_\mu + m)\psi \quad (6.106)$$

Removing γ^5 on the RHS gives us $(i\gamma^\mu \partial_\mu \psi + m)\psi$, which is *not* the LHS of the Dirac equation and is hence *not* zero. Hence, the helicity-flipped 4-spinor is *not* a solution to the Dirac equation.

However, there is one funny exception. If we assume the fermions to be massless^a, we then have

$$\gamma^5 \underbrace{(i\gamma^\mu \partial_\mu \psi)}_{\text{Dirac equation}} = 0 \quad (6.107)$$

where we have recovered the Dirac equation. That is to say, $\gamma^5 \psi$ is a solution to the Dirac equation if and only if the fermion is massless.

^aInconceivable!

The failure of the Dirac equation to preserve itself in (6.106) is a case of explicit symmetry breaking. Specifically, the derivative term is invariant under a chirality flip, while the mass term is not. That is to say, the mass term breaks chiral symmetry.

Another implication is that in the massless limit, helicity and chirality are de facto equivalent. This can be shown by taking the second formalism (6.104), which means that we treat ψ_L and ψ_R as Weyl spinors instead of half-zero Dirac spinors.

Derivation 6.7 (Equivalence of helicity and chirality in the massless limit) Inserting (6.104) into the Dirac equation gives

$$i\gamma^\mu \partial_\mu \psi_L - m\psi_R = 0 \quad i\gamma^\mu \partial_\mu \psi_R - m\psi_L = 0 \quad (6.108)$$

The mass term couples left and right chiralities, and one cannot have a mass term for a fermion unless both ψ_L and ψ_R are present. In the ultra-relativistic limit ($E \gg m$), and $m \rightarrow 0$. We then have

Theorem 6.12 (Weyl equations)

$$i\gamma^\mu \partial_\mu \psi_L = 0 \quad i\bar{\gamma}^\mu \partial_\mu \psi_R = 0 \quad (6.109)$$

Now, since the particle is massless, its speed is c in all reference frames. Unlike the massive particle case, it is no longer possible to ‘overtake’ it by a Lorentz boost. That is to say, helicity becomes Lorentz-invariant like chirality, and the two become identical.

Remark 6.5 One important implication of this concerns the ultra-relativistic limit, where $E \ll m$. The speed of the particle is then very close to c , and the previous points likewise apply.

6.5 CPT theorem

This entire mathematical toolkit has led us to a famous theorem, which is the so-called *CPT* (*charge conjugation, parity, time reversal*) theorem, also known with an alternate initial ordering as the *PCT* theorem. All three are symmetries of the Dirac field.

Quote 6.3 Yes, it's the oldschool order.

Felix Halbwedel, on the alternate initial order ‘PCT’, 3 January 2025

Transformations form a connected Lie groups are continuous, while the as-of-yet-undefined transformations P and T are discrete transformations, which involve specific finite changes that cannot be continuously connected to the identity transformation. They are not elements of actually $O(1, 3)$, but rather representations of the elements Λ_P and Λ_T (which are Lorentz transformations) of the group $O(1, 3)$:

$$\Lambda_P = \text{diag}(1, -1, -1, -1) \quad \Lambda_T = \text{diag}(-1, 1, 1, 1) \quad (6.110)$$

A third discrete symmetry we are interested in is the charge conjugation C , which is an outer automorphism⁸ of the $U(1)$ group, even though it is not continuous like $U(1)$ symmetry.

Now, let us define C , P and T explicitly:

Definition 6.22 (Parity operation) P stands for the *parity operation* P , which changes the 3-position to its inverse:

$$P : (t, \mathbf{x}) \rightarrow (t, -\mathbf{x}) \quad (6.111)$$

⁸An automorphism is a symmetry of the symmetry... oh well.

Theorem 6.13 (Transformations under P)

Object	TRANSFORMATIONS UNDER P	
	Behaviour under P^a	Intrinsic parity $P(\text{object})$
Scalar S	Invariant	+1
Pseudoscalar S_p	Sign flip	-1
4-spinor ψ in Dirac basis	$P\psi = \gamma^0\psi$	+1 (fermions), -1 (antifermions)
Vector V^μ	Sign flip on V^i , invariant otherwise	-1 ^b
Pseudovector A^μ	Sign flip on A^0 , invariant otherwise	+1
Metric $g^{\mu\nu}$	Invariant ^c	+1
Stress-energy tensor $T^{\mu\nu}$	Sign flip on T^{0j} , T^{i0} and some of T^{ij} , invariant otherwise	+1
Faraday tensor $F^{\mu\nu}$	Sign flip on F^{0j} , invariant otherwise	None (mixed)

^a $i, j \in [1, 2, 3]$ ^bIf it describes an interacting particle like a photon.^cComponents undergo general coordinate transformations.**Remark 6.6** As $P^2 = \mathbb{I}$ for a Dirac spinor, it is a Z_2 symmetry labelled Z_2^P .**Definition 6.23 (Time reversal)** T stands for the so-called *time reversal* T , which flips the time coordinate:

$$T : (t, \mathbf{x}) \rightarrow (-t, \mathbf{x}) \quad (6.112)$$

Theorem 6.14 (Transformations under T)

Object	TRANSFORMATIONS UNDER T	
	Behaviour under P	
Scalar S	Invariant	
Pseudoscalar S_p	Invariant	
4-spinor ψ in Dirac basis	$T\psi = i\gamma^1\gamma^3\psi^a$	
Vector V^μ	Sign flip on V^i , invariant otherwise	
Pseudovector A^μ	Sign flip on A^0 , invariant otherwise	
Metric $g^{\mu\nu}$	Invariant ^b	
Stress-energy tensor $T^{\mu\nu}$	Sign flip on T^{0j} and T^{i0} , invariant otherwise	
Faraday tensor $F^{\mu\nu}$	Sign flip on F^{ij} , invariant otherwise	

^aThis is a mathematical tool. The actual physical operator has an extra complex conjugation operator K and is thus anti-unitary. It has the form $\mathcal{T} = TK = i\gamma^1\gamma^3K$.^bComponents undergo general coordinate transformations.**Definition 6.24 (Charge conjugation)** C stands for *charge conjugation* C , which is the sign-flip of all charges:

$$C : Q \rightarrow -Q \quad (6.113)$$

Theorem 6.15 (Transformations under C) The action of C on other quantities are simpler, and we summarise them below:

Object	TRANSFORMATIONS UNDER C	Behaviour under C
Scalar S		Sign flip
Pseudoscalar S_p		Invariant
4-spinor ψ in Dirac basis		$C\psi = i\gamma^2\gamma^0\psi$
Vector V^μ		Sign flip
Pseudovector A^μ		Invariant
Metric $g^{\mu\nu}$		Invariant ^a
Stress-energy tensor $T^{\mu\nu}$		Invariant
Faraday tensor $F^{\mu\nu}$		Sign flip

^aApplies even for charged fields. This is because the stress-energy tensor depends quadratically on fields, and the sign changes cancel out.

From the actions of C , P , T on Dirac spinors, it is then easy to infer their actions on Weyl spinors. This brings about some funny effects. For example, Theorem 6.15 implies, for the left- and right-handed Weyl spinors χ_L and χ_R that make up the Dirac spinor:

$$C\chi_L = -i\sigma_2\chi_R^* \quad C\chi_R = -i\bar{\sigma}_2\chi_L^* \quad (6.114)$$

Hence, for a theory to be C invariant, it must contain:

- Left-handed Weyl spinors
- Right-handed complex conjugate of Weyl spinors

However, it is possible to get away with only one handedness. This is the so-called *Majorana*⁹ *spinor*, which is technically a very specific subcategory of the Dirac spinor:

Definition 6.25 (Majorana spinor) A Majorana spinor is a Dirac spinor with the following construction

$$\psi = \begin{pmatrix} \chi_L \\ -i\sigma_2\chi_L^* \end{pmatrix} \quad (6.115)$$

Due to being a Dirac spinor, it transforms as one, with the key distinction of undergoing charge conjugation as

$$C\psi = \psi \quad (6.116)$$

meaning that a *Majorana fermion* is its own antiparticle. As a result, a Majorana particle is always charge-neutral.

Derivation 6.8 (γ^5 matrix) Interestingly, if one consults the parts of Theorem 6.15, Theorem 6.13 and Theorem 6.14 on Dirac spinors, one sees that CPT , put together, acts on a Dirac spinor as

$$CPT\psi = \gamma^5\psi \quad (6.117)$$

Quote 6.4 The γ^5 matrix is the trilobites of high energy physics!

The Author, 21 April 2025

Finally, let us wrap up our discussions on the γ^5 matrix, which is already too long, by consider the symmetry breaking in CPT . These symmetries can be broken to a certain degree: The breaking of individual or two (i.e. C , P , T , CP , CT , PT) symmetries is not prohibited in QFT. However, due to the need for Lorentz invariance, the breaking of CPT symmetry is disallowed in QFT. This is illustrated by the so-called CPT theorem:

Theorem 6.16 (CPT theorem) Any quantum field theory that is Lorentz-invariant and has a well-defined local interaction must respect CPT symmetry. i.e. for some quantity H , the combination of

⁹To this day, his demise remains a historical mystery. For more, see [here](#).

charge conjugation, parity, and time reversal is always a symmetry:

$$(CPT)H(CPT)^{-1} = H \quad (6.118)$$

As mentioned, the CPT theorem ultimately results from Lorentz invariance. Hence, it cannot be spontaneously broken like gauge symmetries (e.g. electroweak symmetry). At the time of writing (2025), experimentalists have yet to observe CPT symmetry breaking, showing how well QFT has withstood the tests of time.

There is, of course, a more terrifying implication if we turn the first statement above backwards: If CPT were spontaneously broken, it would suggest a violation of Lorentz invariance, a common feature of physics beyond the standard model (BSM).

6.6 Maxwell's equations

It is well-known that QED is a gauge theory governed by U(1) symmetry. This gauge symmetry is so-called as it is defined with respect to the Lie group U(1)¹⁰, which is Abelian. In *Spinors & Symmetries*, it is seen that representations of abstract elements of an Abelian Lie group commute.

$$[A_\mu, A_\nu] = 0 \quad (6.119)$$

In our case, this is the gauge (photon) field A^μ , which is the 4-potential, consisting of the electric scalar potential and the magnetic vector potential¹¹. Inserting (6.119) into (6.25) gives us the QED field strength tensor, which is the well-known Faraday tensor:

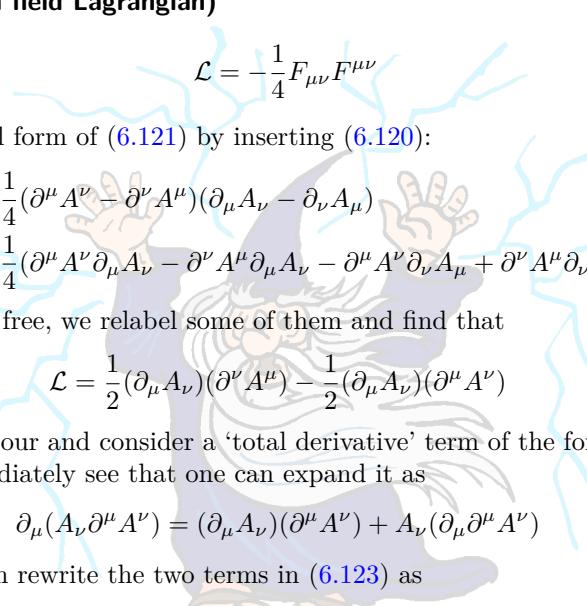
Definition 6.26 (Faraday tensor)

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \quad (6.120)$$

Unsurprisingly, the equations of motion that govern the photon half of QED are Maxwell's equations, which we can derive using the standard method of constructing the Lagrangian and insert it into the Euler-Lagrange equations.

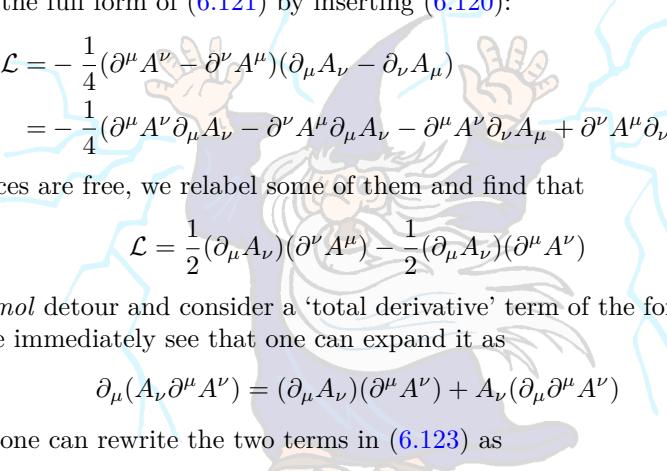
Derivation 6.9 (Recovery of Maxwell's equations) We set the photon field Lagrangian (density) as

Definition 6.27 (Photon field Lagrangian)



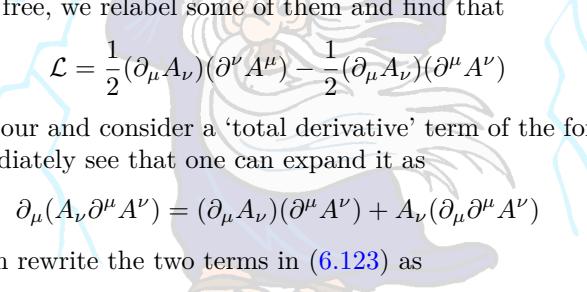
$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (6.121)$$

We can write down the full form of (6.121) by inserting (6.120):



$$\begin{aligned} \mathcal{L} &= -\frac{1}{4} (\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &= -\frac{1}{4} (\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\nu A^\mu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + \partial^\nu A^\mu \partial_\nu A_\mu) \end{aligned} \quad (6.122)$$

Noting that all indices are free, we relabel some of them and find that



$$\mathcal{L} = \frac{1}{2} (\partial_\mu A_\nu)(\partial^\nu A^\mu) - \frac{1}{2} (\partial_\mu A_\nu)(\partial^\mu A^\nu) \quad (6.123)$$

Now let us take a *smol* detour and consider a ‘total derivative’ term of the form $\partial_\mu(A_\nu \partial^\mu A^\nu)$. As per the product rule, we immediately see that one can expand it as

$$\partial_\mu(A_\nu \partial^\mu A^\nu) = (\partial_\mu A_\nu)(\partial^\mu A^\nu) + A_\nu(\partial_\mu \partial^\mu A^\nu) \quad (6.124)$$

Using this relation, one can rewrite the two terms in (6.123) as

$$\frac{1}{2} (\partial_\mu A_\nu)(\partial^\nu A^\mu) = \frac{1}{2} \partial_\mu(A_\nu \partial^\nu A^\mu) - \frac{1}{2} A_\nu \partial^\nu \partial^\mu A_\mu \quad (6.125)$$

$$-\frac{1}{2} (\partial_\mu A_\nu)(\partial^\mu A^\nu) = \frac{1}{2} A_\nu \square A^\nu - \frac{1}{2} \partial_\mu(A_\nu \partial^\mu A^\nu) \quad (6.126)$$

¹⁰We will investigate its gauge transformations very soon.

¹¹We can also use an alternative 4-potential, comprising of the magnetic scalar potential and the electric vector potential. However, this is rarely used due to the absence of observed magnetic monopoles.

But at the same time, the term $\partial_\mu(A_\nu \partial^\mu A^\nu)$ looks a bit familiar, doesn't it? One can see that this so-called term is nothing but our good friend, the boundary term in (2.12). Recalling that zero terms and the boundary terms can be added or removed from the Lagrangian to our liking, (6.123) is then

$$\mathcal{L} = \frac{1}{2} A_\mu (\square A^\mu - \partial^\mu \partial^\nu) A_\nu \quad (6.127)$$

By applying the Euler-Lagrange equations, we find that

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu A^\nu)} \right) - \frac{\partial \mathcal{L}}{\partial A^\nu} = 0 \quad (6.128)$$

where A^ν , which we ultimately recognise as a field variable, replaces ψ .

We then recover Maxwell's equations in index notation

Theorem 6.17 (Maxwell's equations)

$$\partial_\mu F^{\nu\mu} = J^\nu \quad (6.129)$$

The electric and magnetic 3-fields then follow intuitively:

$$E_i = F^{0i} = -\partial_i A^0 + \partial_0 A^i \quad B^i = -\frac{1}{2} \epsilon^{ijk} F_{jk} = \epsilon^{ijk} \partial_j A_k \quad (6.130)$$

Now we investigate the gauge transformation brought about by our so-called U(1) symmetry. In gauge theory, there exist physical observables that are invariant under certain transformations of the potentials. In the case of electromagnetism, this manifests in the invariance of the Faraday tensor $F^{\nu\mu}$ (and hence, the EM fields E and B) under a set of 4-potential gauge transformations that is unsurprisingly called the U(1) transformations:

Definition 6.28 (U(1) transformations)

$$\psi \rightarrow e^{-i\Lambda(x)} \psi \quad \bar{\psi} \rightarrow \bar{\psi} e^{i\Lambda(x)} \quad A_\mu(x) \rightarrow A_\mu(x) + \partial_\mu \Lambda(x) \quad (6.131)$$

where $\Lambda(x)$ is a scalar field and satisfies the wave equation.

The U(1) group has only a single generator, that being the 1×1 identity matrix \mathbb{I}_1 . This sole generator has an 'eigenscalar' with only one component, from which electric charge arises.

But what about the electrical current? If $\Lambda(x)$ is constant, then (6.131) corresponds to a genuine physical symmetry. Thus the current can be derived from (2.39) as

Definition 6.29 (Electrical current)

$$J^\mu = \bar{\psi} \gamma^\mu \psi \quad (6.132)$$

Derivation 6.10 (Invariance of the Faraday tensor) One can verify the invariance of the Faraday tensor by noting that the following conditions simultaneously hold

$$\frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = F^{\mu 0} \quad \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\nu)} = 0 \quad (6.133)$$

where $F^{\mu 0}$ are the canonically conjugate EM fields.

Another way to see this is that, assuming zero 4-current, Maxwell's equations can be written as

$$\partial_\nu F^{\nu\mu} = \square A^\mu - \partial^\mu(\partial_\nu A^\nu) = 0 \quad (6.134)$$

Immediately, we note that the equation does not depend on A^μ itself. Rather, the only dependence lies in $\partial_\nu A^\nu$.

Either way, we see that we can shift A_μ by any gradient $\partial_\mu \Lambda$, as we just did in our U(1) transformation, without affecting the Faraday tensor $F_{\mu\nu}$.

As it turns out, this is expected from the so-called *Noether's second theorem*.

Theorem 6.18 (Noether's second theorem) If a Lagrangian is invariant under an infinite-dimensional local symmetry group (like U(1) symmetry), there exist differential identities that the Euler-Lagrange equations are subject to, known as Noether identities, that make the equations to be not independent from each other. These Noether identities then reflect the redundancy in the equations of motion.

We can verify that such an identity exists for Maxwell's equations.

Derivation 6.11 (Noether identity in Maxwell's equations) The Lagrangian is invariant under our U(1) transformation, and it is possible to write the variation of the action as

$$\delta S = \int d^4x \partial_\mu \Lambda \cdot \left(\frac{\delta \mathcal{L}}{\delta A_\mu} \right) = \int d^4x \Lambda \cdot \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta A_\mu} \right) = 0 \quad (6.135)$$

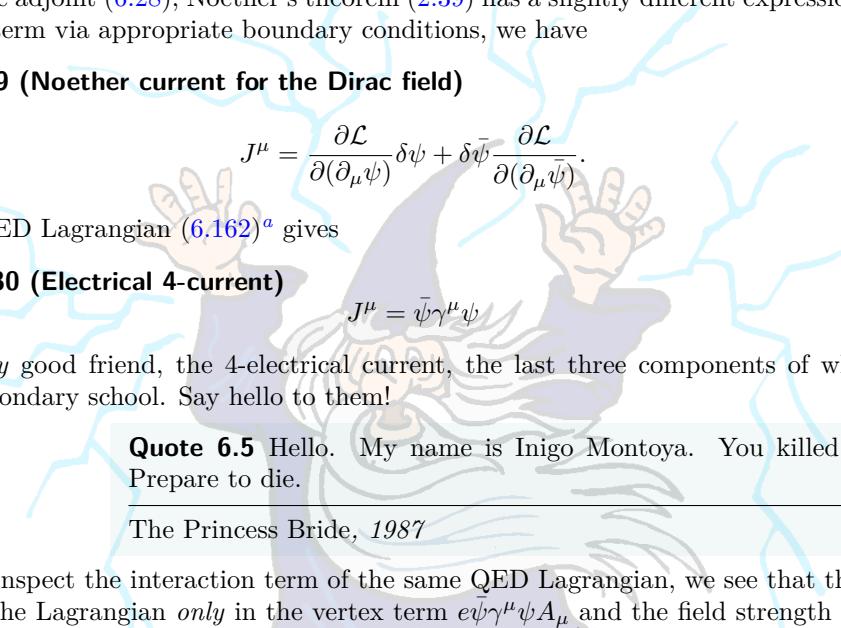
But since $\Lambda(x)$ is arbitrary, we know that, identically:

$$\partial_\mu \left(\frac{\delta \mathcal{L}}{\delta A_\mu} \right) = 0 \quad (6.136)$$

This is not immediately useful. Let us then take a detour and, for no reason whatsoever for the time being, derive the conserved current, which we know to exist due to the invariance of the Lagrangian under the global U(1) transformation (6.131).

Due to the Dirac adjoint (6.28), Noether's theorem (2.39) has a slightly different expression. Assuming zero boundary term via appropriate boundary conditions, we have

Theorem 6.19 (Noether current for the Dirac field)



$$J^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \delta\psi + \delta\bar{\psi} \frac{\partial \mathcal{L}}{\partial(\partial_\mu \bar{\psi})}. \quad (6.137)$$

Inserting the QED Lagrangian (6.162)^a gives

Definition 6.30 (Electrical 4-current)

$$J^\mu = \bar{\psi} \gamma^\mu \psi \quad (6.138)$$

This is our *very* good friend, the 4-electrical current, the last three components of which we have known since secondary school. Say hello to them!

Quote 6.5 Hello. My name is Inigo Montoya. You killed my father. Prepare to die.

The Princess Bride, 1987

However, if we inspect the interaction term of the same QED Lagrangian, we see that the gauge field A_μ appears in the Lagrangian *only* in the vertex term $e\bar{\psi}\gamma^\mu\psi A_\mu$ and the field strength tensor terms. However, if we are to, again for no apparent reason, calculate $\frac{\partial \mathcal{L}}{\partial A_\mu}$, the field strength terms do not contribute as they depend on $\partial_\nu A_\mu$ instead of A_μ itself. Hence, the only relevant term is vertex term^b, and we have

$$\frac{\partial \mathcal{L}}{\partial A_\mu} = -e\bar{\psi}\gamma^\mu\psi \quad (6.139)$$

In natural units, the charge e becomes nothing but unity. Consulting (6.138), we find that, amazingly:

$$J^\mu = -\frac{\partial \mathcal{L}}{\partial A_\mu} \quad (6.140)$$

This is a cute trick, but it quickly rises in significance. We recall that all this has been nothing but a giant detour from deriving the QED Noether identity. Recalling the tensorial form of Maxwell's equations (6.129), we can write

$$-\frac{\delta \mathcal{L}}{\delta A_\mu} = \partial_\nu F^{\nu\mu} \quad (6.141)$$

Now let us take a second 4-derivative on both sides. Recalling (6.136) which we derived earlier, one can conclude that the LHS is zero. This yields our Noether identity

Theorem 6.20 (QED Noether identity)

$$\partial_\mu \partial_\nu F^{\nu\mu} = 0 \quad (6.142)$$

which holds identically due to the antisymmetry of $F^{\nu\mu}$.

^aWhile we introduce it later on, this is purely a choice of pacing. Without gauge fixing, it is nothing but the Dirac and EM Lagrangians added together.

^bThis is known as *minimal coupling*.

The takeaway from this gauge invariance of $F^{\nu\mu}$ is that not all degrees of freedom in A_μ are physical. The unphysical degrees of freedom correspond to pure gauge and are thus called *gauge freedoms* which, if not eliminated, will lead us to mistakenly count multiple configurations of A^μ s as distinct and give rise to erroneous results.

Derivation 6.12 (Fixing the gauge) The standard procedure to eliminate gauge freedoms is *gauge fixing*. In classical EM, we attempt to eliminate this gauge freedom in (6.131) by the Lorentz-invariant Lorenz gauge.

Quote 6.6 Amazingly, the missing “t” is not a typo here.

Alessio Serafini

$$\partial_\mu A^\mu = 0 \quad (6.143)$$

whence Maxwell's equations reduce to the d'Alembertian (wave) equation

$$\square A^\mu = 0 \quad (6.144)$$

The mathematical justification of this gauge fixing is known as the *Gupta-Bleuler formalism*^a. We will not justify it fully, but the main ideas are as follows:

- So far our Fock space has been positive-definite. However, when $A^\mu(x)$ is quantised, we have one timelike component which leads to negative norm states. We then construct a so-called *Krein space*, which is a Fock space that includes these negative norms.
- The negative norm is clearly unphysical. So we impose the Lorenz gauge. In doing so, we have reduced our Krein space into a physical subspace.

The quantum analogue of the Lorenz gauge is the R_ξ gauge. We begin with the photon field Lagrangian (6.121). The R_ξ gauge adds a gauge-fixing term to the Lagrangian:

Theorem 6.21 (R_ξ gauge)

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2\xi} (\partial_\mu A^\mu)^2 \quad (6.145)$$

where ξ is a parameter.

Like in (9.2), we can rewrite the gauge-fixing term and recognise the vanishing of the ‘total derivative’ boundary term:

$$\begin{aligned} \mathcal{L}_{\text{gauge}} &= -\frac{1}{2\xi} (\partial^\mu A_\mu \partial^\nu A_\nu) \\ &= -\frac{1}{2\xi} \partial_\mu (A^\mu \partial^\nu A_\nu) + \frac{1}{2\xi} A^\mu \partial_\mu \partial^\nu A_\nu \\ &= \frac{1}{2\xi} A^\mu \partial_\mu \partial_\nu A^\nu \end{aligned} \quad (6.146)$$

The addition of this gauge-fixing term is justified because it is the simplest local, renormalisable term one can introduce to fix the gauge. We will prove this by using the Faddeev-Popov method in Part III.

^aIn non-Abelian gauge theories, it is superceded by the BRST formalism.

The simplest R_ξ gauge is the Feynman-'t Hooft gauge, which is used in most QFT calculations. In this

gauge, one has $\xi = 1$, and the Lagrangian becomes

Theorem 6.22 (Feynman-'t Hooft gauge)

$$\mathcal{L}_{\text{gauge}} = -\frac{1}{2}(\partial_\mu A^\mu)^2 \quad (6.147)$$

We can now quantise the 4-potential, which is a real, massless Klein-Gordon (scalar) field. The classical wave solution is

$$A^\mu(x) = \int d^3p \sum_{\lambda=0}^3 (\epsilon_\lambda^\mu(p) f_\lambda(p) e^{-ip \cdot x} + \epsilon_\lambda^{\mu*}(p) f_\lambda^*(p) e^{ip \cdot x}) \quad (6.148)$$

where ϵ^μ is a *polarisation vector*, a 4-versor¹². Significantly, it holds the following summation property:

Theorem 6.23 (Polarisation vector property)

$$\sum_{\lambda=0}^d \epsilon_\lambda^{\nu*}(p) \epsilon_\lambda^\mu(p) = -g^{\mu\nu} \quad (6.149)$$

where the index λ labels the polarisation states of the photon.

One notes this to quite resemble the *tetrad fields* in general relativity.

Derivation 6.13 (Polarisation and the Lorentz gauge) For a photon, there are four possible indices, but not all are physical:

- $\lambda = 0$: This is a *longitudinal polarisation* that is often unphysical.
- $\lambda = 1, 2$: They are the two physical *transverse polarisations*^a of the photon.
- $\lambda = 3$: This is a *scalar polarisation* that is also often unphysical.

As photons in QED are gauge bosons, the choice of polarisation vectors is not unique. This gauge freedom can be removed by applying Lorenz gauge:

$$p_\mu \epsilon_\lambda^\mu = 0 \quad (6.150)$$

where p^μ is the photon's 4-momentum.

In doing so, the two unphysical components of the 4-versor have been eliminated due to them being unphysical under the gauge, and only the two transverse polarisation states remain physical for photons. This means that $k_\mu \epsilon^\mu(\lambda, k) = 0$ for transverse polarisations $\lambda = 1, 2$, but not necessarily for $\lambda = 0, 3$.

^ai.e. they are perpendicular to the direction of propagation and to each other.

6.7 Quantisation of the photon field

Using the same procedure we have done before, we insert the normalisation factor into (6.151) and replace $f_\lambda(p)$ and $f_\lambda^*(p)$ with creation and annihilation operators. This gives the photon field, which is actually the 4-potential we see in classical electromagnetism.

Definition 6.31 (Photon field)

$$A^\mu(x) = \int \frac{d^3p}{(2\pi^3)\sqrt{2E_p}} \sum_{\lambda=0}^3 (\epsilon_\lambda^\mu(p) a_\lambda(p) e^{-ip \cdot x} + \epsilon_\lambda^{\mu*}(p) a_\lambda^\dagger(p) e^{ip \cdot x}) \quad (6.151)$$

where the creation and annihilation operators have the form

$$ia_\lambda^\dagger(p) = \int \frac{d^3x}{\sqrt{2E_p}} g_{\lambda\lambda} \epsilon_\lambda^\mu(p) (e^{-ip \cdot x} \partial_0 A_\mu - A_\mu \partial_0 e^{-ip \cdot x}) \quad (6.152)$$

¹²One can recall from *Spinors & Symmetries* that a versor is simply a unit quaternion (i.e. it has norm 1). In simpler terms, a versor is simply a vector whose magnitude is unity (i.e. 1).

Note 6.2 (Gauge field normalisation) So far, we have been using *relativistic normalisation*, which is given by (4.34). For gauge fields, a different normalisation convention is used instead. The key difference is in how the one-particle states are defined:

$$[a_p, a_q^\dagger] = 2E(2\pi)^3 \delta^3(p - q) \quad (6.153)$$

To account for (or rather *absorb*) this extra factor of $2E$, (5.44) becomes

$$S_{fi} = \mathcal{M}_{fi} i \prod_{j=1}^n \frac{1}{2E_{f_j}} \prod_{j=1}^m \frac{1}{2E_{i_j}} (2\pi)^4 \delta^4(p_i - p_f) \quad (6.154)$$

In this gauge theory normalisation convention, (15.14) takes the form

$$\text{LIPS}(m) \equiv (2\pi)^4 \delta^4(p_i - p_f) \prod_{k=1}^m \frac{d^3 q_k}{(2\pi)^3} \quad (6.155)$$

They observe the commutation relation

$$[a_\lambda(p), a_{\lambda'}^\dagger(q)] = -g_{\lambda\lambda'}(2\pi)^3 2E \delta^3(p - q) \quad (6.156)$$

We then derive the canonical momentum of a photon field by inserting (6.121):

$$\pi^\mu = \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = \partial_0 A^\mu \quad (6.157)$$

Plugging in (6.151) gives

Definition 6.32 (Photon field canonical momentum)

$$\pi^\mu = - \int \frac{d^3 p}{(2\pi^3) \sqrt{2E_p}} \sum_{\lambda=0}^3 (\epsilon_\lambda^{\mu*}(p) a_\lambda^\dagger(p) e^{ip \cdot x} - \epsilon_\lambda^\mu(p) a_\lambda(p) e^{-ip \cdot x}) \quad (6.158)$$

Again we consider the nature of the photon. It has spin 1, and is thus a boson. The standard bosonic commutations thus apply:

$$[A^\mu(x), \pi^\nu(u)] = -ig^{\mu\nu} \delta^3(x - u) \quad (6.159)$$

$$[A^\mu(x), A^\nu(u)] = [\pi^\mu(x), \pi^\nu(u)] = 0 \quad (6.160)$$

One final loose end is the Hamiltonian. By inserting (6.157) into (2.28) (where, notably, the field is A^μ instead of ϕ), the photon field Hamiltonian reads

Definition 6.33 (Photon field Hamiltonian)

$$\mathcal{H} = \frac{1}{2} \dot{A}^\nu \dot{A}_\nu + \frac{1}{2} D A^\nu D A_\nu \quad (6.161)$$

6.8 QED Feynman rules

The final term we need to add to the QED Lagrangian is the interaction term that represents the vertex, which is the coupling of two fermion fields to a photon field. This allows us to write the QED Lagrangian as

Definition 6.34 (QED Lagrangian)

$$\mathcal{L} = \bar{\psi}(iD - m)\psi - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \quad (6.162)$$

We now include the gauge-fixing term and write down the full QED Lagrangian by consulting (6.26), (9.2) and (6.146). The fully simplified, gauge-fixed QED Lagrangian is then

$$\mathcal{L}_{\text{QED}} = \underbrace{\bar{\psi}(i\gamma^\mu \partial - m)\psi}_{\text{free fermion}} - \underbrace{e\bar{\psi}\gamma^\mu A_\mu\psi}_{\text{vertex}} - \underbrace{\frac{1}{2} A^\mu (\square g_{\mu\nu} - \partial_\mu \partial_\nu) A^\nu}_{\text{free photon}} + \underbrace{\frac{1}{2\xi} A^\mu \partial_\mu \partial_\nu A^\nu}_{R_\xi \text{ gauge}} \quad (6.163)$$

We now calculate the QED propagators.

Note 6.3 (Propagator Feynman rule) Generally, the Feynman rule of a generic propagator is derived as follows:

- Fourier-transform the Lagrangian to momentum space.
- Recognise that the Lagrangian corresponding to the propagator is nothing but an operator acting on our field of interest.
- Strip off the field, leaving only the operator itself.
- This is then a so-called *kernel* which is the reverse of the propagator.
- Solve for the propagator by dint of the product of kernel and the propagator being the identity.

Derivation 6.14 (Fermionic propagator) As the fermionic propagator is fermionic^a, we denote it distinctly as S_F :

$$S_F(x - y) = \langle 0 | T[\psi(x)\bar{\psi}(y)] | 0 \rangle \quad (6.164)$$

This is the only useful fermionic propagator, as anticommutation implies that

$$\langle 0 | T[\psi_\alpha(x)\psi_\beta(y)] | 0 \rangle = \langle 0 | T[\bar{\psi}_\alpha(x)\bar{\psi}_\beta(y)] | 0 \rangle = 0 \quad (6.165)$$

By definition, it is a Green's function of the Dirac operator:

$$(i\cancel{D}_x - m)S_F(x - y) = \delta^4(x - y) \quad (6.166)$$

Now we want to convert this expression to momentum space. We begin by Fourier-transforming (6.1) into momentum space. The field transformations are

$$\psi(x) = \int \frac{d^4 p}{(2\pi)^4} \psi(p) e^{-ip \cdot x} \quad \bar{\psi}(x) = \int \frac{d^4 p}{(2\pi)^4} \bar{\psi}(p) e^{ip \cdot x} \quad (6.167)$$

Using this and (4.31), we have

$$(\cancel{p} - m)\tilde{\psi}(p) = 0 \quad (6.168)$$

The second element to be Fourier-transformed is the propagator itself:

$$S_F(p) = \int d^4 x e^{ip \cdot x} S_F(x) \quad (6.169)$$

Putting it all together, we find

$$(\cancel{p} - m)S_F(p) = i\mathbb{I} \rightarrow S_F(p) = \frac{i}{\cancel{p} - m} = \frac{i(\cancel{p} + m)}{p^2 - m^2 + i\epsilon} \quad (6.170)$$

where we added a *smol*, pole-eliminating term in the last step. Returning to position space gives us

Definition 6.35 (Fermionic propagator)

$$S_F(x - y) = -i \lim_{\epsilon \rightarrow 0+} \int \frac{d^4 p}{(2\pi)^4} \frac{i(\cancel{p} + m)e^{-ip \cdot (x-y)}}{p^2 - m^2 + i\epsilon} \quad (6.171)$$

^aWho could've guessed?

Derivation 6.15 (Photon field propagator) Now we attack the photon field propagator. As it is bosonic, we denote it as $D_F^{\mu\nu}$, identical to the Feynman propagator save for the tensorial indices:

$$D_F^{\mu\nu}(x - y) = \langle 0 | T[A^\mu(x)A^\nu(y)] | 0 \rangle \quad (6.172)$$

The derivation is conceptually similar to that of the Feynman and fermion propagators. While we have not derived the (gauge) photon field equations of motion before, we can do so now by inserting

the gauge EM Lagrangian (6.147) into (2.26).

$$\partial_\mu F^{\mu\nu} + \frac{1}{\xi} \partial^\nu (\partial_\mu A^\mu) = 0 \quad (6.173)$$

Expanding $F^{\mu\nu}$ and simplifying yields

$$\square A^\nu - \partial^\nu (\partial_\mu A^\mu) + \frac{1}{\xi} \partial^\nu (\partial_\mu A^\mu) = 0 \quad (6.174)$$

Now we move this to momentum space. The Fourier transform of the photon field is

$$A_\mu(x) = \int \frac{d^4 k}{(2\pi)^4} e^{-ik \cdot x} \tilde{A}_\mu(k) \quad (6.175)$$

Using this and (4.31), the equations of motion become

$$-k^2 \tilde{A}^\nu(k) + \left(1 - \frac{1}{\xi}\right) k^\nu (k_\mu \tilde{A}^\mu(k)) = 0 \quad (6.176)$$

where k^μ is the 4-momentum. We define the aforementioned kernel, which is the inverse of the photon propagator:

$$D_{\mu\nu}^{-1}(k) \tilde{A}^\nu(k) = 0 \rightarrow D_{\mu\nu}^{-1}(k) = -k^2 g_{\mu\nu} + \left(1 - \frac{1}{\xi}\right) k_\mu k_\nu \quad (6.177)$$

It should not be surprising that the product of the propagator and the kernel must be the Kronecker tensor^a δ_μ^ρ :

$$D_{\mu\nu}^{-1}(k) D^{\nu\rho}(k) = \delta_\mu^\rho \quad (6.178)$$

We conveniently assume the (tensorial) propagator to be of the form

$$D_{\mu\nu}(k) = A(k^2) g_{\mu\nu} + B(k^2) k_\mu k_\nu \quad (6.179)$$

Inserting this into (6.178) yields

$$\left[-k^2 g_{\mu\nu} + \left(1 - \frac{1}{\xi}\right) k_\mu k_\nu \right] [A(k^2) g^{\nu\rho} + B(k^2) k^\nu k^\rho] = \delta_\mu^\rho \quad (6.180)$$

This is a set of tensorial equations which have the solution

$$A(k^2) = -\frac{1}{k^2} \quad B(k^2) = -\frac{\xi - 1}{\xi} \frac{1}{k^4} \quad (6.181)$$

This gives the photon field propagator as

$$D_{\mu\nu}(k) = -\frac{1}{k^2} \left[g_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right] \quad (6.182)$$

This lives in the Landau gauge. In the Feynman gauge, which is $\xi = 1$, it reduces to:

$$D_{\mu\nu}(k) = -\frac{g_{\mu\nu}}{k^2} \quad (6.183)$$

Quote 6.7 Do not forghetti, die Kirche ruft uns arme Seelen zur Führung!
Morgen um 15 Uhr bei da Hans Jesus Kirche

Julia Neumeister, 4 April 2025

Do not forget that we still need to return to position space. Transforming back, we find

Definition 6.36 (Photon field propagator)

$$D_F^{\mu\nu}(x-y) = -i \lim_{\epsilon \rightarrow 0+} \int \frac{d^4 k}{(2\pi)^4} \frac{g^{\mu\nu} e^{-ik \cdot (x-y)}}{p^2 + i\epsilon} \quad (6.184)$$

^aHow rude!

Finally, using Wick's second theorem and taking the Grassmann parity of Fermions into account, we can formulate the QED Feynman rules:

Theorem 6.24 (QED Feynman rules) For a given Feynman diagram in QED, the scattering amplitude matrix elements with a complex factor $i\mathcal{M}_{fi}$ are constructed as follows:

QED FEYNMAN RULES (PARTIAL)	
For each	Add to expression
Incoming and outgoing electron	$\bar{u}_\alpha(s, p)$ and $u_\alpha(s, p)$
Incoming and outgoing positron	$v_\alpha(s, p)$ and $\bar{v}_\alpha(s, p)$
Incoming and outgoing photon	$\epsilon^{*\mu}(\lambda, p)$ and $\epsilon^\mu(\lambda, p)$
Internal photon line	$\frac{-ig^{\mu\nu}}{p^2}$
Internal fermion line	$\frac{i(p+m)e^{-ip \cdot (x-y)}}{p^2 - m^2}$
Internal loop	$\int d^4 l_n \frac{1}{(2\pi)^4}$ over corresponding line
Vertex	$-ie\gamma^\mu_{\alpha\beta}$
Vertex ^a	$(2\pi)^4\delta^3(k_i - k_f)$

The indices are tricky. First, we note that γ^μ is actually an array, with each element being a 4-matrix:

- The incoming and outgoing photon indices are μ and ν . They govern the metric and γ^μ as an array.
- The incoming and outgoing fermion indices are α and β . They govern the fermion and photon 4-spinors and the specific elements of each 4×4 matrix in γ^μ .

The incoming and outgoing 4-momenta are k_i and k_f , and each internal loop has a so-called *internal momentum* l_n .

Finally, before taking a well-deserved break, remove a factor of $(2\pi)^4\delta^3(p - q)$, where p and q are the *total* initial and detected momenta.

^aThis term enforces 4-momentum conservation.

One should note the following points:

- The symbols s and λ seen in the electrons, positrons and photons are actually indices, which are put into the bracket purely for ease of viewing.
- In QED, the previously seen factor C observes $C = k!$, and C and $1/k!$ cancel out.
- When the scattering involves more than one diagram, Wick's theorem will be used, in which the exchange of two fermion operators changes the sign of the expression as per the Grassman parity.

We can now draw QED Feynman diagrams. Unlike ϕ^4 theory, QED is not a toy model and concerns itself with real particles - in specific, fermions and photons. One can represent them in a Feynman diagram as follows:

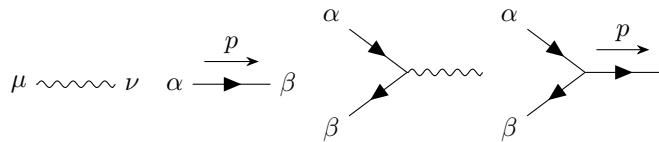


Figure 6.1: QED Feynman diagram elements. L-R: photon propagator, fermion propagator, vertex leading to photon propagator, vertex leading to fermion propagator

Thus, we have finished the canonical quantisation derivation of QED.

Part III

Path integrals

Chapter 7

Free fields

Quote 7.1 Juice, juice, orange juice...

Richard Feynman, playing the bongos, September 1981

While canonical quantisation is the more intuitive approach to developing field theories, it suffers from certain drawbacks, especially with respect to extra degrees of gauge freedom (as we have seen in deriving the QED Feynman rules). In developing more complicated field theories like QCD, we will make use of the other formalism: path integrals.

7.1 Path integrals in non-relativistic quantum mechanics

Unsurprisingly, the end goal in path integrals is the same as that in canonical quantisation, which is to derive the S -matrix S_{fi} , the transition amplitude \mathcal{M}_{fi} and the Feynman rules of a given field theory that constructs \mathcal{M}_{fi} .

It is useful to recognise the relationship between S_{fi} and the unitary time evolution operator $U(t, t_0)$. When t is the final time t_f of some process and t_0 is the initial time t_i , one recovers the scattering matrix:

$$S_{fi} = U(t_f, t_i) \quad (7.1)$$

We can see why this reasoning is useful by investigating path integrals in innocent non-relativistic quantum mechanics. The central idea is that a particle in motion can and *will* take every possible trajectory or *path*. We postulate that each path contributes a factor of e^{iS} to the $U(t, t_0)$:

$$U(t, t_0) = \sum_{\text{all paths}} e^{iS} \quad (7.2)$$

where S is the action.

As is well known, in non-relativistic quantum mechanics, $U(t, t_0)$ is given by

$$U(t, t_0) = \langle f | e^{-iHT} | 1 \rangle \quad (7.3)$$

where 1 is the initial state¹ and $T = t - t_0$ is the time interval.

Derivation 7.1 (Time-slicing) One can solve the contribution to $U(t, t_0)$ of a certain path by slicing the path's time interval into *smol* time steps of ϵ . We use the *Lie product formula*:

Theorem 7.1 (Lie product formula) For any operators or square matrices \hat{A} and \hat{B} , one has

$$e^{\hat{A}+\hat{B}} = \lim_{N \rightarrow \infty} \left(e^{\hat{A}/N} e^{\hat{B}/N} \right)^N = \lim_{N \rightarrow \infty} \left(e^{\hat{B}/N} e^{\hat{A}/N} \right)^N \quad (7.4)$$

where N is the so-called *Trotter number*.

¹We have avoided writing i to prevent confusion with indices that will appear later.

Practically, this has an alternative formulation. Suppose that, instead of solving directly for $e^{\hat{A}+\hat{B}}$, we solve it segment by segment, solving first a $e^{\epsilon(\hat{A}+\hat{B})}$ for a *smol* ϵ and then calculate $(e^{\epsilon(\hat{A}+\hat{B})})^{1/\epsilon}$, where the segment $e^{\epsilon(\hat{A}+\hat{B})}$ has the form

$$e^{\epsilon(\hat{A}+\hat{B})} = e^{\epsilon\hat{A}}e^{\epsilon\hat{B}} + O(\epsilon^2) \quad (7.5)$$

Our good friend, the unitary time evolution operator, can then be approximated as

$$e^{-iHT} = (e^{-iH\epsilon})^N \quad (7.6)$$

where $N = T/\epsilon$ is again the Trotter number.

From this, we are in a position to construct the generic time evolution operator $U(t, t_0)$. We write over a series of N steps:

$$\begin{aligned} U(t, t_0) &= \langle f, t_N | e^{-iH\epsilon} | N-1, t_{N-1} \rangle \cdots \langle 1, t_1 | e^{-iH\epsilon} | i, t_0 \rangle \\ &= \int dx_{N-1} \cdots dx_1 \langle f, t_N | e^{-iH\epsilon} | x_{N-1}, t_{N-1} \rangle \cdots \langle x_1, t_1 | e^{-iH\epsilon} | x_i, t_0 \rangle \end{aligned} \quad (7.7)$$

where we have integrated over all intermediate positions $dx_1 \cdots dx_{N-1}$ ^a and $\epsilon = T/N$ is again a single time step.

Remark 7.1 Note that the second line is not a single integral but $N - 1$ integrals.

^aThe initial and final positons x_0 and x_N are not integrated as they are fixed - remember that the path integral is integrating over a range of *possible* positions!

The Hamiltonian can be decomposed as

$$H = \frac{1}{2}p_i^2 + V(q_i) \quad (7.8)$$

where we recall p_i and q_i to be generalised momenta and coordinates.

Derivation 7.2 (One slice in phase space) We now want to evaluate one slice, represented by $\langle q_{i+1}, t_{i+1} | e^{-iH\epsilon} | q_i, t_i \rangle$. Importantly, as neither p_i nor q_i are scalars, we *cannot* simply say that $e^{-iH\epsilon} = e^{-i\epsilon(p_i^2/2+V(q_i))} = e^{-i\epsilon p_i^2/2} e^{-i\epsilon V(q_i)}$. A trick must be used to solve for the decomposed result:

Theorem 7.2 (Baker-Campbell-Hausdorff formula) Suppose one has the known matrices X and Y and the unknown matrix Z which satisfy $e^X e^Y = e^Z$. Z can be solved by

$$Z = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}[X, [X, Y]] - \frac{1}{12}[Y, [X, Y]] + \dots \quad (7.9)$$

where square brackets are commutators.

Amazingly, however, as ϵ is *smol*, (7.9) yields convenient approximation

$$e^{-iH\epsilon} \approx e^{-i\epsilon p_i^2/2} e^{-i\epsilon V(q_i)} \quad (7.10)$$

which is *almost* identical to the result if p_i and q_i were scalars. We can then say that, for some arbitrary step i , that

$$\langle q_{i+1}, t_{i+1} | e^{-iH\epsilon} | q_i, t_i \rangle \approx \langle q_{i+1}, t_{i+1} | e^{-i\epsilon p_i^2/2} e^{-i\epsilon V(q_i)} | q_i, t_i \rangle \quad (7.11)$$

We exploit the following identity, which is a complete set of momentum eigenstates:

$$\int \frac{dp}{2\pi} |p\rangle \langle p| = 1 \quad (7.12)$$

Insertion gives

$$\langle q_{i+1} | e^{-i\epsilon p_i^2/2} e^{-i\epsilon V(q)} | q_i \rangle = \int \frac{dp}{2\pi} \langle q_{i+1} | e^{-i\epsilon p_i^2/2} | p \rangle \langle p | e^{-i\epsilon V(q)} | q_i \rangle \quad (7.13)$$

Both $V(q)$ and p^2 act diagonally in position space, and we find

$$\langle p | e^{-i\epsilon V(q)} | q_i \rangle = e^{-i\epsilon V(q_i)} \langle p | q_i \rangle = e^{-i\epsilon V(q_i)} \frac{e^{-ipq_i}}{\sqrt{2\pi}} \quad \langle q_{i+1} | e^{-i\epsilon \frac{p^2}{2}} | p \rangle = e^{-i\epsilon \frac{p^2}{2}} \langle q_{i+1} | p \rangle = e^{-i\epsilon \frac{p^2}{2}} \frac{e^{ipq_{i+1}}}{\sqrt{2\pi}} \quad (7.14)$$

Putting it all together, and considering multiple degrees of freedom labeled by $j = 1, \dots, M$, we have

$$\langle q_{i+1}, t_{i+1} | e^{-iH\epsilon} | q_i, t_i \rangle = e^{-iV(q_i)} \prod_j^M \frac{dp_j^i}{2\pi} \exp \left[i\epsilon \left(p_j \frac{q_j^{i+1} - q_j^i}{\epsilon} - \frac{p_j^{i2}}{2} \right) \right] \quad (7.15)$$

where M is the degree of freedom (i.e. the number of dimensions/coordinates).

This is as far as we can go in phase space, and we are forced to move to momentum space.

Derivation 7.3 (Many slices in configuration space) Let us now go back to many slices so that we can recover the time evolution operator. Say that we have N slices. The integrals are then defined be over all intermediate positions and momenta:

$$\langle q_f, t_f | q_0, t_0 \rangle = \int \prod_{i=1}^{N-1} dq_i \prod_{i=0}^{N-1} \left(\frac{dp_i}{2\pi} \right) \exp \left[i\epsilon \sum_{i=0}^{N-1} \left(p_i \frac{q_{i+1} - q_i}{\epsilon} - \frac{p_i^2}{2} - V(q_i) \right) \right] \quad (7.16)$$

The momenta p_i can now be integrated out. This is, up to phase, standard Gaussian integral, which can be evaluated as

$$\int_{-\infty}^{\infty} dp e^{-ap^2/2+bp+c} = \sqrt{2\pi/a} e^{b^2/2a+c} \quad (7.17)$$

In our case, this is

$$\langle q_f, t_f | q_0, t_0 \rangle = \int \prod_{i=1}^{N-1} dq_i \sum_{i=0}^N \epsilon \left(\frac{1}{2} \dot{q}_i^2 - V(q_i) \right) = \int \prod_{i=1}^{N-1} dq_i \sum_{i=0}^N \epsilon L(q_i, \dot{q}_i) \quad (7.18)$$

We now make a shorthand that represents the integration over *all possible intermediate configurations* of the path $q(t)$ between the endpoints q_0 and q_f - i.e. the ‘sum over all paths’. This is sometimes called the *integration measure*:

Definition 7.1 (Integration measure)

$$\mathcal{D}q = \prod_{i=1}^{N-1} dq_i \quad (7.19)$$

where N is once again the number of steps.

Thus, the transition amplitude becomes

$$\langle q_f, t_f | q_0, t_0 \rangle = \int \mathcal{D}q e^{i \sum_{i=0}^{N-1} \epsilon L(q_i, \dot{q}_i)} \quad (7.20)$$

In the continuum limit, we send N to infinity and find

$$\langle q_f, t_f | q_0, t_0 \rangle = \int \mathcal{D}q e^{iS} = \int \mathcal{D}q e^{i \int_{t_0}^{t_f} dt L(q, \dot{q})} \quad (7.21)$$

where S is the action. This verifies our postulate (7.2)^a.

^aImportantly, time ordering is not a concern as the integration the Lagrangian naturally preserves the time order. The same can be seen in the time steps (7.7).

Theorem 7.3 (Integration measure properties)

$$\int \mathcal{D}\phi = \phi(x) \quad (7.22)$$

$$\int \mathcal{D}\phi\phi = \frac{\phi^2(x)}{2} \quad (7.23)$$

Quote 7.2 The strict built-in ordering of times makes commutation relations between the operators irrelevant.

Niklas Beisert, 2017

Now we turn this from quantum mechanics to a field theory:

- The Lagrangian L is replaced by the Lagrangian density \mathcal{L} .
- The coordinates q are replaced with the fields ϕ .
- From the last point, $\mathcal{D}q$ becomes $\mathcal{D}\phi$.

which gives, after introducing sensible limits

Theorem 7.4 (Time evolution of fields)

$$U(t, t_0) = \int_{\phi(t_0)}^{\phi(t)} \mathcal{D}\phi e^{i \int d^4x \mathcal{L}} \quad (7.24)$$

7.2 Generating functional

Quote 7.3 As any reader of Dirac knows, it is sometimes convenient to speak of a distribution as if it were a function.

Sidney Coleman and Jeffrey Mandula, in ‘All Possible Symmetries of the S Matrix’, 16 March 1967

In the path integral formulation, we introduce, for every field $\phi(x)$, a classical external field $J(x)$ called a *source* that couples linearly to the field. This so-called source is more of a mathematical convenience than a physical entity, and it is significant in that it allows the creation of Green’s functions.

Before introducing the main concepts of this section in formulae, we summarise the main ideas in text:

- The Lagrangian density (and by this, the action) is modified by adding a source term $J(x)\phi(x)$.
- The time evolution operator, which is an exponential of the Hamiltonian, is generalised as the almighty $Z(J(x))$ *generating functional* which, as we will see later, is so-called because it is used to generate Green’s functions.
- Loosely speaking, the Green’s function can be derived by differentiating the *generating functional* by the source and setting $J(x) = 0$ afterwards².

Definition 7.2 (Generating functional) The generating function is an exponential of the now-modified action:

$$Z[J] = \int \mathcal{D}\phi e^{i \int d^4x (\mathcal{L}(\phi(x)) + J(x)\phi(x))} \quad (7.25)$$

Remark 7.2 But this looks a bit familiar, doesn’t it?

If you share this opinion, you’d be right. As $Z(J(x))$ integrates over all possible field configurations, it is the quantum analogue of the well-known *partition function* in statistical physics. In fact, $Z(J(x))$ is simply called the *partition function* in some literature.

So far, this generating functional diverges into infinity. We thus introduce the *normalised generating functional* $Z_0[J]$:

²This is where one can appreciate the nature of the source as a mathematical convenience. Ultimately, QFT describes free or interacting fields instead of external influences. Hence, the source to zero as this external influence would have remained otherwise.

Definition 7.3 (Normalised generating functional)

$$Z_0[J] = \frac{Z[J]}{Z[0]} \quad (7.26)$$

Derivation 7.4 (Klein-Gordon field) We are now in a position to derive the (renormalised) generating functional of the Klein-Gordon field. Consider the Klein-Gordon Lagrangian (2.4):

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2$$

The generating functional is hence

$$Z[J] = \int \mathcal{D}\phi e^{i \int d^4x \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + i \int d^4x J(x)\phi(x)} \quad (7.27)$$

Let us analyse the exponential

$$i \int d^4x \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 + i \int d^4x J(x)\phi(x) \quad (7.28)$$

Assuming fields vanish at infinite, we can employ integration by parts. This gives

$$\frac{1}{2}i \int d^4x \phi(-\partial^2 - m^2)\phi + i \int d^4x J(x)\phi(x) \quad (7.29)$$

Importantly, the field $\phi(x)$ can be decomposed into two parts, a classical (background) part and a quantum fluctuation part:

$$\phi(x) = \phi_c(x) + \phi_q(x) \quad (7.30)$$

where:

- $\phi_c(x)$ is chosen to satisfy the classical equation of motion involving the source

$$(\partial^2 + m^2)\phi_c(x) = -J(x) \quad (7.31)$$

- $\phi_q(x)$ is the quantum fluctuation around the classical solution.

This is the essence of the path integral method: quantum effects come from summing over fluctuations around classical paths. From (7.31), the exponential reduces to

$$-\frac{1}{2}i \int d^4x \phi_q(\partial^2 + m^2)\phi_q + i \int d^4x J(x)\phi_c(x) \quad (7.32)$$

At the same time, it is easy to rewrite (7.31) as

$$\phi_c(x) = \int d^4y D_F(x-y)J(y) \quad (7.33)$$

where, in our utter ignorance, we have accidentally rederived the Feynman propagator $D_F(x-y)$ in the path integral formulation.

Hence:

$$i \int d^4x J(x)\phi_c(x) = i \int d^4x d^4y J(x)D_F(x-y)J(y) \quad (7.34)$$

Turning back to the generating functional, we now have

$$Z[J] = \int \mathcal{D}\phi e^{i \int d^4x \phi_q(\partial^2 + m^2)\phi_q + i \int d^4x d^4y J(x)D_F(x-y)J(y)} \quad (7.35)$$

However, we are still integrating over the whole field ϕ . As elucidated earlier, ϕ_c is nothing but a measly classical background, and the real object of interest is ϕ_q . That is to say, it is ϕ_q instead of the

whole ϕ that we now want to integrate over. This is a linear change of variables, so the integration measure reduces simply:

$$\mathcal{D}\phi = \mathcal{D}(\phi_c + \phi_q) = \mathcal{D}\phi_q \quad (7.36)$$

Quote 7.4 Hal!

Nikos Stavros, in Red Alert 1

The generating functional is then

$$Z[J] = \int \mathcal{D}\phi_q e^{i \int d^4x \phi_q (\partial^2 + m^2) \phi_q + i \int d^4x d^4y J(x) D_F(x-y) J(y)} \quad (7.37)$$

Now consider the case where the source is zero. As the source is the cause of the classical field ϕ_c , $\phi_c(x) = 0$ when $J = 0$, and we can effectively rewrite $\phi(x)$ as $\phi_q(x)$. Hence, we can write

$$Z[0] = \int \mathcal{D}\phi_q e^{i \int d^4x \phi_q (\partial^2 + m^2) \phi_q} \quad (7.38)$$

and

$$Z[J] = Z[0] e^{i \int d^4x d^4y J(x) D_F(x-y) J(y)} \quad (7.39)$$

We now derive, for the Klein-Gordon field, the expression of the normalised generating function $Z_0[J]$ defined in (7.26). Amazingly, this is already implied in (7.39):

$$Z_0[J(x)] = e^{i \int d^4x d^4y J(x) D_F(x-y) J(y)} \quad (7.40)$$

By the definition of the exponential function, we can easily see that $Z_0[0]$ is always 1. This verifies that our result is indeed normalised.

7.3 Green's function

We are now almost in a position to derive the full Green's function from the generating function we have just derived. This involves differentiating $Z_0[J]$, which is a *functional*, by sources $J(x_n)$, which are *functions* themselves. Hence, standard derivatives fail, and we need to introduce the so-called *functional derivative* δ , which is the analogue of a normal derivative for a function.

Derivation 7.5 (Motivation of the functional derivative) Suppose we have a functional

$$J[f] = \int_a^b L(x, f(x), f'(x)) dx \quad (7.41)$$

where we add a *smol* δf to $f(x)$. L then becomes

$$L(x, f(x), f'(x)) \rightarrow L(x, f + \delta f, f' + \delta f') \quad (7.42)$$

where $\delta f'$ is the *variation* of the *derivative*. The change in J , due to this variation, is then

$$\begin{aligned} \delta J &= \int_a^b \left(\frac{\partial L}{\partial f} \delta f(x) + \frac{\partial L}{\partial f'} \frac{d}{dx} \delta f(x) \right) dx \\ &= \int_a^b \left(\frac{\partial L}{\partial f} - \frac{d}{dx} \frac{\partial L}{\partial f'} \right) \delta f(x) dx + \frac{\partial L}{\partial f'}(b) \delta f(b) - \frac{\partial L}{\partial f'}(a) \delta f(a) \end{aligned} \quad (7.43)$$

where, importantly, $(\delta f)'$ is the *derivative* of the *variation*.

The functional derivative satisfies the following properties:

Theorem 7.5 (Functional derivative properties)

$$\frac{\delta\phi(y)}{\delta\phi(x)} = \delta(x - y) \quad (7.44)$$

$$\frac{\delta}{\delta\phi(x)}(\alpha(y)\beta(z)) = \frac{\delta\alpha(y)}{\delta\phi(x)}\beta(z) + \alpha(x)\frac{\delta\beta(z)}{\delta\phi(x)} \quad (7.45)$$

Derivation 7.6 (Green's function) We again begin with the (unnormalised) generating functional for a (free or interacting) scalar field theory:

$$Z(J) = \int \mathcal{D}\phi \exp \left(iS[\phi] + i \int d^4x J(x)\phi(x) \right) \quad (7.46)$$

Taking the functional derivative of $Z(J)$ with respect to the source $J(x)$ brings down factors of $\phi(x)$ from the exponential:

$$\frac{\delta Z(J)}{\delta J(x)} = i \int \mathcal{D}\phi \phi(x) \exp \left(iS[\phi] + i \int d^4x J(x)\phi(x) \right) \quad (7.47)$$

Now, for a system of n 4-positions, we have an n -point function, and hence n sources. Repeating the functional derivative for every source gives us

$$\frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)} = i^n \int \mathcal{D}\phi \phi(x_1) \cdots \phi(x_n) \exp \left(iS[\phi] + i \int d^4x J(x)\phi(x) \right) \quad (7.48)$$

Recall that it is the Green's function we want. This is the *vacuum* expectation value of the time-ordered product. We hence set the source as zero:

$$\frac{\delta^n Z(J)}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0} = i^n \int \mathcal{D}\phi \phi(x_1) \cdots \phi(x_n) e^{iS[\phi]} \quad (7.49)$$

Normalising the generating functional gives

$$\frac{1}{i^n} \frac{\delta^n Z_0(J)}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0} = \frac{1}{Z[0]} \int \mathcal{D}\phi \phi(x_1) \cdots \phi(x_n) e^{iS[\phi]} \quad (7.50)$$

Now let us connect this to the vacuum expectation value. In (classical) probability theory, the definition of the expectation value is well-known as

$$\langle f \rangle = \frac{1}{Z} \int f(x) e^{-S(x)} dx \quad (7.51)$$

From (7.24), we see that time evolution and expectation values are calculated as weighted sums over all field configurations, with weight $e^{iS[\phi]}$. The VEV of a time-ordered product of field operators is a probabilistic average over all possible quantum fluctuations of the vacuum state fields. Hence, like (7.51), we can construct the expression

$$\langle 0 | T[\phi(x_1) \cdots \phi(x_n)] | 0 \rangle = \frac{\int \mathcal{D}\phi \phi(x_1) \cdots \phi(x_n) e^{iS[\phi]}}{\int \mathcal{D}\phi e^{iS[\phi]}} = \frac{1}{Z[0]} \int \mathcal{D}\phi \phi(x_1) \cdots \phi(x_n) e^{iS[\phi]} \quad (7.52)$$

where, again, $Z[0]$ normalises the expression.

Hence, we finally see that (7.50) is exactly the Green's function in disguise:

Theorem 7.6 (Propagator-generating functional relation)

$$G_0^{(n)} = \langle 0 | T[\phi(x_1) \cdots \phi(x_n)] | 0 \rangle = \frac{1}{i^n} \frac{\delta^n Z_0[J]}{\delta J(x_1) \cdots \delta J(x_n)} \Big|_{J=0} \quad (7.53)$$

This then allows us to reconstruct the generating functional in terms of the propagator:

$$Z[J] = \sum_{n=0}^{\infty} \int d^d x_1 \cdots d^d x_n \langle 0 | T[\phi(x_1) \cdots \phi(x_n)] | 0 \rangle J(x_1) \cdots J(x_n) \quad (7.54)$$

Chapter 8

Interacting fields I: ϕ^4 theory

In the last chapter, we left off from (7.53), which can be written explicitly as¹:

$$\langle T[\phi_1 \cdots \phi_n] \rangle = \left. \frac{\int \mathcal{D}\phi \phi_1 \cdots \phi_n e^{iS[\phi, J]}}{\int \mathcal{D}\phi e^{iS[\phi, J]}} \right|_{J=0} \quad (8.1)$$

This chapter will see us tackling this expression.

8.1 Generating functional

While innocent-looking, the expression of our so-called generating functional is actually quite unwieldy. Similar to what we did in canonical quantisation, we split the Lagrangian in the numerator (i.e. the generating functional $Z[J]$) into two parts, the (quadratic) free part \mathcal{L}_F which we can single out from the integration measure and the interacting part \mathcal{L}_I with a dependence on ϕ .

$$Z[J] = \int \mathcal{D}\phi e^{i \int d^d x (\mathcal{L}_F + \mathcal{L}_I + J\phi)} = \int \mathcal{D}\phi e^{i \int d^d x (\mathcal{L}_F + J\phi)} e^{i \int d^d x \mathcal{L}_I} \quad (8.2)$$

where we have a third implicit term $\mathcal{L}_S = J(x)\phi(x)$, which is the source part.

Now we evaluate this expression. The free part of the Lagrangian is nothing but the Klein-Gordon Lagrangian, whose generating functional we have already derived in (7.40).

Derivation 8.1 (Interaction part) For the interaction term, we rewrite it in terms of functional derivatives,

$$e^{iS_I[\phi]} = e^{i \int d^d x \mathcal{L}_I[\phi]} \quad (8.3)$$

Here a cute trick can be used. Consider the expression $\frac{\delta}{\delta J(x)} e^{i \int d^d x \mathcal{L}_I[\phi]}$. By evaluating the partial derivative, we can write

$$\frac{\delta}{\delta J(x)} e^{i \int d^d x \mathcal{L}_I[\phi]} = i\phi(x) e^{i \int d^d x \mathcal{L}_I[\phi]} \quad (8.4)$$

If we eventually set the source to zero (which, as we have seen, we will), the exponential term is nothing but unity. Hence, assuming that the source is ultimately set to zero, we can always make the substitution

$$\phi(x) = -\frac{\delta}{\delta J(x)} \quad (8.5)$$

Thus, the interaction part becomes

$$e^{i \int d^d x \mathcal{L}_I[\phi]} = e^{i \int d^d x \mathcal{L}_I\left[-\frac{\delta}{\delta J(x)}\right]} \quad (8.6)$$

which we can single out from the integral due to the lack of dependence on ϕ .

¹At first glance this might seem slightly confusing. Note that the indexless ϕ is not standalone but is to be read as a part of the integration measure $\mathcal{D}\phi$.

Hence, the generating functional is

$$Z[J] = e^{i \int d^d x \mathcal{L}_I \left[\frac{\delta}{i \delta J} \right]} e^{-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y)} \quad (8.7)$$

Compare this with (7.39), we can set the source to zero and write

$$Z[J] = Z[0] e^{-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y)} \quad (8.8)$$

The free part of the ϕ^4 Lagrangian is identical to that of the Klein-Gordon Lagrangian, and this is why we have already re-derived our good friend, the Feynman propagator in the last chapter. However, it remains for us to derive the vertex which arises from the new ϕ^4 quartic interacting term. One strength of the path integral formulation lies in the fact that one can directly read off the vertex term from the Lagrangian. We will do so now:

Derivation 8.2 (Lazy vertex) Let us take ϕ^4 theory as an example, where \mathcal{L}_I is actually $-\lambda\phi^4/4!$. Immediately, we can see that two terms have nothing to do with the vertex:

- ϕ^4 denote field self-interactions.
- $4!$ is the combinatorial factor.

We then have $-\lambda$ left. However, there is also a prefactor i brought about by the generating functional. If we combine the two, the resulting expression $-i\lambda$ is exactly the vertex factor in the Feynman rules.

However, this is not very rigorous, and we will proceed with the full derivation in the next section.

8.2 Perturbative expansion

Let us again consider a $2 \rightarrow 2$ process in ϕ^4 theory. The compact form of the Green's function reads

$$\langle T[\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)] \rangle = \frac{\int \mathcal{D}\phi \phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4) e^{iS[\phi,J]}}{\int \mathcal{D}\phi e^{iS[\phi,J]}} \Big|_{J=0} \quad (8.9)$$

The denominator is merely the generating functional $Z[J]$ (in this case $Z[0]$), while the numerator can be realised as the generating functional differentiated with respect to x_1, \dots, x_4 .

$$\left\langle T \left[\prod_i^4 \phi(x_i) \right] \right\rangle = \frac{1}{Z[0]} \frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} \int \mathcal{D}\phi e^{iS[\phi,J]} \Big|_{J=0} \quad (8.10)$$

Note that while $\int \mathcal{D}\phi e^{iS[\phi,J]}$ is essentially $Z[J]$, we cannot cancel it with $1/Z[0]$ due to the existence of the differential operators $\delta^4/\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)$. Instead, we decompose the action as per (8.2). Substituting the integration results from (8.7):

$$\left\langle T \left[\prod_i^4 \phi(x_i) \right] \right\rangle = \frac{1}{Z[0]} \frac{\delta^4}{\delta J(x_1)\delta J(x_2)\delta J(x_3)\delta J(x_4)} e^{i \int d^d x \mathcal{L}_I \left[\frac{\delta}{i \delta J} \right]} e^{-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y)} \Big|_{J=0} \quad (8.11)$$

where d is the number of dimensions in the spacetime.

Derivation 8.3 (2 → 2 processes) We cannot calculate the Green's function with this result alone. Rather, we perform a perturbative expansion of the exponential. We start with $e^{i \int d^d x \mathcal{L}_I \left[\frac{\delta}{i \delta J} \right]} e^{-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y)}$, which is the evaluated form of $Z[J]$.

Recall from Part II that the interacting Lagrangian in ϕ^4 theory is

$$\mathcal{L}_I = -\frac{\lambda}{4!} \phi^4 \quad (8.12)$$

where, importantly, λ is the almighty coupling constant. The two exponentials then each expand to a series of polynomials:

$$\begin{aligned} \left\langle T \left[\prod_i^4 \phi(x_i) \right] \right\rangle &= \frac{1}{Z[0]} \frac{\delta^4}{\delta J(x_1) \delta J(x_2) \delta J(x_3) \delta J(x_4)} \left(\sum_n \frac{1}{n!} \left(\frac{\lambda}{4!} i \int d^d y \frac{\delta^4}{i \delta J(y)^4} \right)^n \right) \times \\ &\quad \left. \left(\sum_m \frac{1}{m!} \left(-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y) \right)^m \right) \right|_{J=0} \end{aligned} \quad (8.13)$$

where m and n are the powers.

Both indices m and n go up to infinity. However, we are actually not too interested in them on their own. Rather, we inspect the order of λ which, like in canonical quantisation, denotes the number of vertices. This puts us in a position to return to (5.39).

Merely at first order, we have the terrible-looking expression

$$\begin{aligned} \left\langle T \left[\prod_i^4 \phi(x_i) \right] \right\rangle &= \frac{Z_2[0]}{Z[0]} \frac{\delta^4}{\delta J(x_1) \delta J(x_2) \delta J(x_3) \delta J(x_4)} \left(\frac{1}{2!} \left(-\frac{i}{2} \int d^d x d^d y J(x) D_F(x-y) J(y) \right)^2 \right. \\ &\quad \left. - \frac{i\lambda}{4!} \int d^d z \frac{1}{4!} \frac{\delta^4}{\delta J(z)^4} \left(-\frac{i}{2} \int d^d x d^d y J(z) D_F(x-y) J(y) \right)^4 + \mathcal{O}(\lambda^2) \right) \end{aligned} \quad (8.14)$$

This will produce a series of derivatives, many of which are identical. In the interest of brevity, we will jump through this hard part and arrive at the conclusion that one can write

$$\frac{1}{Z[0]} = \frac{1}{1 + \lambda D} \quad \text{the rest} = A + \lambda(B + C + AD) + \mathcal{O}(\lambda^2) \quad (8.15)$$

With some effort, A , B , C and D can be solved. We first look at A :

$$A = G_{2+2}^{(0)} = D_F(x_1 - x_2) D_F(x_3 - x_4) + D_F(x_1 - x_3) D_F(x_2 - x_4) + D_F(x_1 - x_4) D_F(x_2 - x_3) \quad (8.16)$$

This is *exactly* what we got in Part I using Wick's third theorem! Now let us investigate the other terms:

$$\lambda B = -i\lambda \int d^d x D_F(x - x_1) D_F(x - x_2) D_F(x - x_3) D_F(x - x_4) \quad (8.17)$$

$$\lambda C = -\frac{i\lambda}{2} \sum_{P(ijkl)} D_F(x_i - x_j) \int d^d x D_F(x - x) D_F(x - x_k) D_F(x - x_l) \quad (8.18)$$

where $P(ijkl)$ permutes over all possible indices i , j , k and l (i.e. 1 and 2).

$$\lambda AD = -\frac{i\lambda}{8} \int d^d x D_F(x - x) D_F(x - x) \sum_{P(ijkl)} D_F(x_i - x_j) D_F(x_k - x_l) \quad (8.19)$$

Again, these results agree with what we got from canonical quantisation. Putting it all together, we have

$$\langle T[\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)] \rangle = \frac{A + \lambda(B + C + AD)}{1 + \lambda D} \quad (8.20)$$

One can expand the RHS factorial, yielding

$$\langle T[\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)] \rangle = (A + \lambda(B + C + AD))(1 - \lambda D) + \mathcal{O}(\lambda^2) = A + \lambda(B + C) + \mathcal{O}(\lambda^2) \quad (8.21)$$

We hence recover

$$G_{2+2}^{(1)} = \lambda(B + C) \quad (8.22)$$

If one is masochistic, it is possible to write out the full form of the propagator via this expansion and set the sources to zero. In principle, this recovers the very same ϕ^4 Feynman rules as we have seen before.

Already, we could give some comments comparing path integrals and canonical quantisation:

- So far, we have categorically avoided using Wick's theorem, and we will not use it even if we write out the full form of the propagator. As we have seen in the $2 \rightarrow 2$ process, all possible contractions are ‘automatically’ generated (i.e. arise naturally) through differentiating the (Gaussian²) path integral. Here we see the advantage of the path integral formulation.
- We note that the expansion is very unwieldy in position space, and the expansion is again better carried out in momentum space. Here we see the disadvantage of the path integral formulation.

8.3 Effective action

Our previous encounter with amputated propagators was brief. Now, with path integrals in our hands, we can investigate it and its related concepts more sophisticatedly. Starting from the full propagator or the *dressed propagator*, one can increasingly simplify it:

- If we preserve the connected parts of the Feynman diagram only, the Green's functions reduce to *connected Green's functions*, which is generated by the so-called *free energy* $W[J]$.
- If we further remove (i.e. amputate) the external (leg) propagators³, we are left with the previously seen amputated propagators. This extracts the core interaction structure.
- If we remove even the reducible parts of the Feynman diagram, we are left with the *one-particle-irreducible (1PI or OPI in short) Green's functions*, which describe fundamental interaction vertices beyond the classical action (e.g. loops).
- Conversely, if we remove all OPI Green's functions from the full propagator, we find the *bare propagator* G_0 .

Quote 8.1 Help me, OPI-Wan. You're my only hope.

Star Wars, 1977

Dear reader (yes, you), chapters ago, you learned about the amputated propagators (5.43). Now they beg you to help them in their struggle against path integrals.

We first define the *free energy*, which is simply the generating functional of connected Green's functions.

Definition 8.1 (Free energy)

$$W[J] = -i \ln(Z[J]) \quad (8.23)$$

Expanding $W[J]$ in terms of $J(x)$ provides the connected n -point Green's functions G_c

Definition 8.2 (Connected Green's function)

$$G_c = \langle T[\phi(x_1) \cdots \phi(x_n)] \rangle_c = \left. \frac{\delta^n W[J]}{\delta J(x_1) \cdots \delta J(x_n)} \right|_{J=0} \quad (8.24)$$

Normal Green's functions can be expressed in terms of connected ones via *combinatorics*:

$$G^{(2)} = G_c^{(2)} \quad (8.25)$$

$$G^{(4)} = G_c^{(4)} + G_c^{(2)} G_c^{(2)} \quad (8.26)$$

$$G^{(6)} = G_c^{(6)} + \textcircled{1} + G_c^{(2)} G_c^{(2)} G_c^{(2)} \quad (8.27)$$

where $\textcircled{1}$ denotes the products of 4-point and 2-point connected functions.

Note 8.1 (Plot twist) A sinister plot twist is as follows: The Green's function we saw in the LSZ reduction formula (5.28) is actually the *connected Green's function*, not the Green's function itself. In a 2-point function, this is not a problem because, as we saw in (8.25), G and G_c are equivalent, but it

²Gaussian integrals inherently sum over all possible contractions.

³Whether one does so in the full Green's function or the connected Green's function makes no difference, and both result in the amputated propagator.

is something to be noted in more complicated processes.

But why is this? The reason is twofold:

- **Vacuum bubbles:** They are so-called due to being disconnected diagrams that have no external legs at all. They are written as an overall factor $\langle 0|0 \rangle_J$, the vacuum-to-vacuum amplitude under a source J .
- **Separate scattering events:** They are processes distinct from, and for our purposes, essentially far apart in spacetime from whatever our process of interest is. They are not included in the LSZ formula because they do not represent the process we are trying to measure.

We continue by introducing the so-called *classical field* φ . It is defined as the expectation value of the quantum field, or essentially the mean field over all configurations, in the presence of the source. φ is given by taking the functional derivative of W with respect to J :

Definition 8.3 (Classical field)

$$\varphi = \langle \phi(x) \rangle = \frac{\delta W[J]}{\delta J(x)} \quad (8.28)$$

This so-called classical field is distinct from the classical field ϕ_c from the last section, and the two converge only when there are no quantum corrections.

Before we proceed, a few comments are in order:

- Our classical field contains all loop corrections as well as the source.
- If we set the source to zero, our classical field φ reduces to the vacuum expectation value $\langle \phi \rangle$.
- At tree level, our classical field becomes the classical part of the field ϕ_c that we get when we decompose the field like in (7.30).
- At both zero source and tree level, the classical field is zero⁴, with the exception of spontaneously symmetry-breaking theories we will see by the end of this chapter, where it could take values of equal magnitude and opposite signs.

The *effective action* is then the Legendre transform of $W[J]$:

Definition 8.4 (Effective action)

$$\Gamma[\varphi] = W[J] - \int d^d x J(x) \varphi(x) \quad (8.29)$$

where $J(x)$ is understood as a functional of $\varphi(x)$ through the inversion of $\varphi(x) = \delta W / \delta J$.

From the effective action, one can derive a galaxy of useful quantities. For example, if one is to assume a constant classical field φ_{const} , one gets the sign-flipped *effective potential* V_{eff} at that constant field, multiplied by the spacetime volume:

Definition 8.5 (Effective potential) The effective potential is simply the renormalised potential at a constant field, given by

$$\Gamma[\varphi_{\text{const}}] = -V_{\text{eff}} \int d^d x \quad (8.30)$$

where the superscript d is the spacetime dimension.

Now consider derivatives against the classical field φ . The first functional derivative of $\Gamma[\varphi]$ gives the source. This is nothing but the equation of motion for our so-called classical field⁵:

$$\frac{\delta \Gamma}{\delta \varphi} = -J \quad (8.31)$$

⁴Recall that when we *now* set the source to zero, we have $\phi_G[A_\mu] = 0$.

⁵This also justified why our so-called classical field is so-named. Its behaviour is governed by an action principle due to lacking any quantum corrections.

The second functional derivative of $\Gamma[\varphi]$ gives the *inverse propagator*:

$$\frac{\delta^2 \Gamma[\varphi]}{\delta \varphi(x) \delta \varphi(y)} = (G_{\text{full}})^{-1}(x, y) \quad (8.32)$$

where $G_{\text{full}}(x, y)$ is the full propagator or the *dressed propagator* including quantum corrections. Any higher-order derivatives of $\Gamma[\varphi]$ yield the OPI Green's functions G_{OPI} , sometimes also denoted as γ :

Definition 8.6 (OPI Green's function)

$$G_{\text{OPI}} = \left. \frac{\delta^n \Gamma[\varphi]}{\delta J(x_1) \cdots \delta J(x_n)} \right|_{J=0} \quad (8.33)$$

In other words, $\Gamma[\varphi]$ is the generating functional of OPI Green's functions. For example, $\Gamma^{(4)}$ represents the four-point interaction vertex.

Remark 8.1 The amputated Green's function essentially represents the sum of OPI diagrams that make up the full vertex function.

Remark 8.2 Importantly, the term ‘one-particle-irreducible’ does not describe interactions involving only one particle. Rather, a Feynman diagram is called OPI if it cannot be split into two separate diagrams by cutting a single internal propagator. This means:

- The diagram remains connected if any single internal propagator is removed.
- It does not factorise into two separate diagrams through a single-particle exchange.

This is in contrast to one-particle-reducible (OPR) diagrams, which can be split by cutting a single propagator, meaning they contain a disconnected propagator that acts as a bridge.

Let us summarise what we have so far:

- We have 4 main quantities of which we can take functional derivatives - the classical action and the 3 we have seen in this section.
- They are functionals of either the source $J(x)$ or the field $\phi(x)$.
- In an n-point QFT process, one has a series of positions $x_1 \cdots x_n$.

We go over them one by one:

- **Classical action $S[\phi]$:** Functional derivatives are taken w.r.t. $\phi(x)$.
 - The first derivative simply generates the classical equations of motion.
 - Higher derivatives against the n points generate vertices in the classical theory.
- **Generating functional $Z[J]$:** Functional derivatives are taken w.r.t. $J(x)$ and generate vacuum expectation values (i.e. correlation functions or Green's functions).
- **Free energy $W[J]$:** Functional derivatives are taken w.r.t. $J(x)$ and generate connected Green's functions.
- **Effective action $\Gamma[\phi]$:** Functional derivatives are taken w.r.t. $\phi(x)$.
 - The first derivative generates the master Dyson-Schwinger equation (8.41) which incorporate loop corrections⁶.
 - The second derivative generates the inverse propagator.
 - Higher derivatives against the n points generate the OPI Green's functions.

One final, albeit important, remark concerns the so-called *self-energy* Σ , which we have seen to be instrumental in the Dyson equation we will use in renormalisation.

Definition 8.7 (Self-energy) We define the self-energy of a particle as the energy that a particle has due to its interaction and its environment^a. Mathematically, it is the sum of all its OPI *two-point*

⁶Hence the name ‘effective action’ is justified.

diagrams:

$$\Sigma(p) := \sum G_{\text{OPI}}(p) \quad (8.34)$$

^ai.e. the part of the total energy that relates the particle back to itself

Remark 8.3 It is worth noting that self-energy is not actually energy. Rather, it represents corrections to the full propagator arising from the interaction between a particle and its environment. This is analogous to the original concept of self-energy in electromagnetism, which refers to the energy required to assemble a charged particle due to its own electromagnetic field.

So far, we have seen quite a few gimmicks that we can exploit from the effective action formalism. However, this is not the end of the story. Rather, as we will see in the rest of this chapter, the effective action gives rise to two fundamental results.

8.4 Dyson-Schwinger equations

An interesting analogy exists between the action S and the generating functional Z . Consider a *smol* variation of the field (i.e. a gauge transformation). As this is nothing a change of variables in the path integral, the generating functional (7.25) is invariant:

$$\delta Z = \int \mathcal{D}\phi \frac{\delta}{\delta \phi(x)} e^{i \int d^4x (\mathcal{L}(\phi(x)) + J(x)\phi(x))} = 0 \quad (8.35)$$

This is analogous to the action principle, which states that the action is invariant under a *smol* variation of coordinates⁷. Also by analogy to the action, we can, using the gauge invariance of the generating functional, derive a series of equation that look like equations of motion.

Derivation 8.4 (Dyson-Schwinger equation) We first evaluate the functional derivative inside the integral

$$\frac{\delta}{\delta \phi(x)} e^{i \int d^4y (\mathcal{L}(\phi(y)) + J(y)\phi(y))} = \left(i \frac{\delta \mathcal{L}}{\delta \phi(x)} + iJ(x) \right) e^{i \int d^4y (\mathcal{L} + J\phi)} \quad (8.36)$$

This gives

$$\delta Z = \int \mathcal{D}\phi \left(i \frac{\delta \mathcal{L}}{\delta \phi(x)} + iJ(x) \right) e^{i \int d^4y (\mathcal{L} + J\phi)} = 0 \quad (8.37)$$

By realising that the action is nothing but the 4-integral of the Lagrangian (density), we can write

$$\frac{\delta \mathcal{L}}{\delta \phi(x)} = \frac{\delta S}{\delta \phi(x)} \quad (8.38)$$

Thus

$$\delta Z = i \int \mathcal{D}\phi \frac{\delta S}{\delta \phi(x)} e^{iS[\phi] + i \int J\phi} + iJ(x) \int \mathcal{D}\phi e^{iS[\phi] + i \int J\phi} = 0 \quad (8.39)$$

Factoring off i on both sides and recognising that $Z[J] = \int \mathcal{D}\phi e^{iS+i \int J\phi}$, this becomes

$$\frac{\delta S}{\delta \phi(x)} [\phi] Z[J] = -J(x)Z[J] \quad (8.40)$$

Recalling the cute trick that was (8.5), we replace ϕ by functional derivatives of J :

Theorem 8.1 (Dyson-Schwinger equation)

$$\frac{\delta S}{\delta \phi} \left(-i \frac{\delta}{\delta J} \right) Z[J] = -J(x)Z[J] \quad (8.41)$$

This is the master equation for the rest of the Dyson-Schwinger equations, which are infinite. The Dyson-Schwinger equations for propagators are derived by performing functional derivatives of the source J .

⁷Remember that in QFT, we have replaced 4-coordinates with 4-fields.

An alternate form of this master equation exists, which involves the quantities we derived in the last section. Recall from (8.28) that the classical field is the expectation value of a field and is related to the free energy W (8.23). By inserting the field expectation/classical field it into (8.41), we can rewrite it in terms of the free energy W :

$$\frac{\delta S}{\delta \phi} \left(\frac{\delta W}{\delta J} + \frac{\delta}{\delta J} \right) = -J(x) \quad (8.42)$$

Using a Legendre transformation and inserting (8.31), a dependence on the effective action Γ (8.29) can also be acquired:

$$\frac{\delta \Gamma}{\delta \varphi(x)} + \frac{\delta S}{\delta \varphi(x)} \left(\varphi(x) + \frac{\delta^2 W}{\delta J(x) \delta J(y)} + \frac{\delta}{\delta J(y)} \right) = 0 \quad (8.43)$$

from which the Dyson-Schwinger equations for propagators are derived by taking derivatives of the classical field φ .

Derivation 8.5 (Dyson equation) We will now derive the Dyson-Schwinger equation of propagators.

We perform a single functional derivative with respect to the source:

$$\begin{aligned} -i \frac{\delta}{\delta J(y)} \left[\frac{\delta S}{\delta \phi(x)} \left[-i \frac{\delta}{\delta J} \right] Z[J] \right] &= -i \frac{\delta}{\delta J(y)} [-J(x)Z[J]] \\ &= -i \left[-\delta(x-y)Z[J] - J(x)(-i) \frac{\delta Z}{\delta J(y)} \right] \end{aligned} \quad (8.44)$$

Now we set the source to zero:

- The LHS becomes a time-ordered two-point function.
- The second LHS term vanishes.

Hence:

$$\langle T\phi(x)\phi(y) \rangle = -i \frac{\delta}{\delta J(y)} [-J(x)Z] \Big|_{J=0} \quad (8.45)$$

$$\frac{1}{Z[0]} \left(-i \frac{\delta}{\delta J(x)} \right) \left(-i \frac{\delta}{\delta J(y)} \right) Z[J] \Big|_{J=0} = -i [-\delta(x-y)Z[0]] \quad (8.46)$$

$$Z[0] \left\langle \frac{\delta S}{\delta \phi(x)} \phi(y) \right\rangle = i\delta(x-y)Z[0] \quad (8.47)$$

Finally, we see that the two-point Dyson-Schwinger equation or the Dyson-Schwinger equation for propagators is

Theorem 8.2 (Dyson-Schwinger equation for propagators)

$$\left\langle \frac{\delta S}{\delta \phi(x)} \phi(y) \right\rangle = i\delta(x-y) \quad (8.48)$$

As an example, let us see what this implies in ϕ^4 theory. Let us insert the ϕ^4 Lagrangian (5.1) and recognise that $\delta(x-y)$ becomes $\delta^4(x-y)$ due to dimensionality:

$$\langle (-\partial_x^2 - m^2) \phi(x) \phi(y) \rangle - \frac{\lambda}{3!} \langle \phi(x)^3 \phi(y) \rangle = i\delta^4(x-y) \quad (8.49)$$

We now rewrite this expression by introducing two objects:

- The (now full) propagator $G(x,y) \equiv \langle T[\phi(x)\phi(y)] \rangle$
- The OPI insertion Σ , given by

$$\langle \phi(x)^3 \phi(y) \rangle_{\text{connected}} \rightarrow 3 \int d^4z d^4w G(x,z) \Sigma(z,w) G(w,y) \quad (8.50)$$

where the factor of 3 comes from choosing one of the three $\phi(x)$ lines to hook into the OPI two-point subdiagram.

The equation then reads

$$(-\partial_x^2 - m^2) G(x, y) - \int d^4 z \Sigma(x, z) G(z, y) = i\delta^4(x - y) \quad (8.51)$$

By recognising that the bare propagator $G_0(x, y)$, which is the Feynman propagator (5.37), satisfies

$$G_0^{-1}(x, y) = (-\partial_x^2 - m^2) \delta(x - y) \quad (8.52)$$

We can again rewrite the equation as

$$(G_0^{-1} - \Sigma) G = i \quad (8.53)$$

which, after a Fourier transformation to momentum space, is

$$(G_0^{-1}(p) - \Sigma(p)) G(p) = i \quad (8.54)$$

Rearranging gives us the so-called *Dyson equation*:

Theorem 8.3 (Dyson equation)

$$G_{\text{full}} = G_0 + \Sigma = \frac{i}{p^2 - m^2 - \Sigma(p) + i\epsilon} \quad (8.55)$$

where we have clearly labelled G as the full propagator G_{full} .

If we are to interpret this equation as a geometric series (which we will), it essentially decomposes the full propagator into the bare and dressed propagators. Equivalently, one can see this as summing all insertions of the OPI self-energy. The expansion reads:

$$D_F(p) = \underbrace{\frac{i}{p^2 - m_0^2 + i\epsilon}}_{0 \text{ loops}} + \underbrace{\frac{i}{p^2 - m_0^2 + i\epsilon} (-i\Sigma(p^2)) \frac{i}{p^2 - m_0^2 + i\epsilon}}_{1 \text{ loop}} + \underbrace{\frac{i}{p^2 - m_0^2 + i\epsilon} (-i\Sigma(p^2)) \frac{i}{p^2 - m_0^2 + i\epsilon} (-i\Sigma(p^2)) \frac{i}{p^2 - m_0^2 + i\epsilon}}_{2 \text{ loops}} + \dots \quad (8.56)$$

where the terms go up to infinity.

But what does this physically mean? The key to interpreting this result lies in the realisation that our assumption in Part I that a propagator is nothing but a single line has been a lie all along. The single-line propagator we have grown to know and love is nothing but the *bare propagator* - that is to say, the simplest possible version of it. In contrast, when we go to higher loop orders, we have so-called *dressed propagators*, which is the result when one attaches internal loops to a bare propagator. That is to say, like the vertex, the propagator *cannot* be exactly solved either - it, too, must be derived perturbatively.

With this realisation, we can make sense of our result:

- The n^{th} term has n loops sandwiched among (i.e. multiplied by) $n + 1$ internal lines.
- The LHS is the dressed Feynman propagator, and the first RHS term is the bare Feynman propagator.
- The rest of the RHS are OPI Feynman propagators, collectively known as self-energy, relate the dressed and bare Feynman propagators.

The usefulness of this section and the previous one will not be immediately obvious. However, they will prove essential in Part IV.

8.5 Global symmetry breaking

Quote 8.2 Flatfish evolved to survive on ocean bed. And basically, they moved their eyes to one side of the body. (...) Halibuts have their eyes on the right side of the body, while flounders have their eyes on the left side of the body. Why is that? We don't know. But they are symmetric.

Emilia Szymańska, 13 September 2024

Outside of the convenient tricks we introduced in the last section, another important implication of the effective action formalism is the spontaneous symmetry breaking. In the scope of this book, we are interested in two types of spontaneous symmetry breaking:

- Global symmetry breaking
- Local gauge symmetry breaking

In this section, the former is covered. We will do this via the motivating example of a modified ϕ^4 theory. If one replaces the mass coupling m^2 in the Lagrangian with some negative parameter $-\mu^2$, we find

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}\mu^2\phi^2 - \frac{\lambda}{4!}\phi^4 \quad (8.57)$$

Unlike standard ϕ^4 theory, this model exhibits a funny effect upon closer inspection. As we have learned in Part I, a Lagrangian can be split into kinetic terms and the potential, which is

$$V = -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4 \quad (8.58)$$

It is intuitive to see that this potential has two minima $\pm v$, where v is the vacuum expectation value of ϕ :

$$v = \pm\sqrt{\frac{6}{\lambda}}\mu \quad (8.59)$$

This minimum is completely fixed in that it has no dependence on the 4 coordinates. It is then possible to decompose the field as

$$\phi(x) = v + \sigma(x) \quad (8.60)$$

Let us consider the case where the field lives near the VEV. Inserting this decomposition then gives

$$\mathcal{L} = \frac{1}{2}\partial_\mu\sigma\partial^\mu\sigma - 2\mu^2\sigma^2 - \sqrt{\frac{6}{\lambda}}\mu\sigma^3 - \frac{\lambda}{4!}\sigma^4 \quad (8.61)$$

We can immediately see that:

- The field perturbation $\sigma(x)$ is massive as we have a mass-like quadratic term $-2\mu^2\sigma^2$.
- Symmetry is no longer preserved when we flip the sign of ϕ . This is due to the sign also flipping in the third-order term $-\sqrt{\frac{6}{\lambda}}\mu\sigma^3$.

What is the physical implication of the second point? Well, nothing happens when the energy is well above the VEV, where the ϕ - V function still looks like a single trough. However, as we approach the VEV, the shape of the function begins to resemble a so-called *sombrero*. That is to say, instead of having one minimum, we have two minima along with a third unstable stationary point in between. Importantly, this means that a minimum can be reached at two different points of ϕ .



Figure 8.1: Various sombreros at the Museo de Arte Popular, CDMX. Credit: Alejandro Linares García

The breaking of this spontaneous symmetry is discrete. However, just like how we have discrete and continuous symmetries, there exist discrete and continuous spontaneous symmetry breakings, the latter of which gives rise to the so-called *Goldstone's theorem*.

Consider the so-called *linear sigma model* or the *linear σ model*. Superficially, it seems identical to ϕ^4 theory, of which it is a generalised version. The key lies in the fact that instead of a *single* scalar field, the linear σ model has a series of N scalar fields with identical couplings, denoted (i.e. summed over) by index $i = 1, \dots, N$.

Definition 8.8 (Linear σ Lagrangian)

$$\mathcal{L} = \frac{1}{2} \partial_\mu \phi_i \partial^\mu \phi_i + \frac{1}{2} m^2 \phi_i^2 + \frac{\lambda}{4!} \phi_i^4 \quad (8.62)$$

The important realisation here is to imagine the collection of scalar fields as a vector, albeit not in the same way as a vector field. This requires us to realise the fact that the number of our scalar fields N can be different from the spacetime dimensions:

- The components of this so-called vector are full-fledged scalar fields determined by the 4-coordinates in Minkowski space instead of components of a vector field.
- Rather, the vector components together make up a ‘field space’ of N dimensions, distinct from our Minkowski spacetime.

Now consider a good ol’ orthogonal group of rotations denoted by $O(N)$. As we know, the representations of this group encompass all $N \times N$ orthogonal matrices R_{ij} . The linear σ Lagrangian is then invariant under the transformation

$$\phi_j = R_j^i \phi^i \quad (8.63)$$

This transformation, which is *continuous*, is analogous to our previous *discrete* sign-flip.

Again we solve for our minima, which reads

$$(\phi_{i,0})^2 = \frac{\mu^2}{\lambda} \quad (8.64)$$

This only poses a constraint on the magnitude of $\phi_{i,0}$. For our convenience, we can set our so-called ‘field space’ coordinates such that the only non-zero component of $\phi_{i,0}$ lies in the N^{th} direction:

$$\phi_{i,0} = (0, \dots, 0, v) \quad v = \frac{\mu}{\sqrt{\lambda}} \quad (8.65)$$

The shifted fields then read, for some index $k = 1, \dots, N - 1$:

$$\phi_i(x) = (\pi_k(x), v + \sigma(x)) \quad (8.66)$$

where $\pi_k(x)$ is the shift of the (previously zero) fields of index $1, \dots, N - 1$, and $\sigma(x)$ is the shift of our (previously non-zero) field of index N .

Our good friend, the Lagrangian, is then

$$\mathcal{L} = \frac{1}{2} \partial_\mu \pi_k \partial^\mu \pi_k + \frac{1}{2} \partial_\mu \sigma \partial^\mu \sigma - \mu^2 \sigma^2 - \sqrt{\lambda} \mu \sigma^3 - \sqrt{\lambda} \mu \pi_k^2 \sigma - \frac{\lambda}{4} \pi_k^2 \sigma^2 - \frac{\lambda}{4} \sigma^4 - \frac{\lambda}{4} \pi_k^4 \quad (8.67)$$

where Greek indices indicate spacetime components, and Latin indices indicate ‘field space’ components.

Quote 8.3 Because it is a Griechenlandflagge.

Felix Halbwedl, erroneously asserting that the Scottish flag in the BaGru room is a Greek flag, 26 September 2024

Three physical comments are in order:

- Again, we see that σ is massive due to the existence of the quadratic mass-like term, but the fields π_k are massless.
- There exists a broken continuous symmetry due to the existence of the terms $-\sqrt{\lambda} \mu \sigma^3 - \sqrt{\lambda} \mu \pi_k^2 \sigma$.
- While we no longer have $O(N)$ symmetry, a $O(N - 1)$ symmetry corresponding to the index k is preserved.

So what does the example tell us?

Derivation 8.6 (Goldstone’s theorem) Let us generalise our linear σ result to an arbitrary field theory with several fields $\lambda_i(x)$. Like before, we can split the Lagrangian into the kinetic terms and the potential:

$$\mathcal{L} = T + V \quad (8.68)$$

where, for a minimum in V , the fields find themselves at $\lambda_{i,0}$.

Expanding the potential about $\lambda_{i,0}$ gives

$$V(\phi) = V(\phi_0) + \frac{1}{2} (\phi - \phi_0)_i (\phi - \phi_0)_j \left(\frac{\partial^2}{\partial \phi_i \partial \phi_j} V \right)_{\phi_0} + \dots \quad (8.69)$$

As before, we recognise that the coefficient of the quadratic (mass-like) term is nothing but the mass coupling, or rather the *mass couplings*, which are now a matrix:

$$(m^2)_{ab} = \left(\frac{\partial^2}{\partial \phi_i \partial \phi_j} V \right)_{\phi_0} \quad (8.70)$$

Let us now impose a continuous transformation, which looks like

$$\phi_i \rightarrow \phi_i + \alpha \chi_i(\phi_i) \quad (8.71)$$

Now suppose that V is invariant under this *smol* transformation, we then have

$$V(\phi_i + \alpha \chi_i(\phi_i)) = V(\phi_i) \quad (8.72)$$

We can rephrase this as

$$\frac{\partial}{\partial \alpha} V(\phi + \alpha \chi(\phi)) \Big|_{\alpha=0} = 0 \quad (8.73)$$

By the chain rule,

$$\frac{\partial V}{\partial \phi_i} \chi_i(\phi) = 0 \quad (8.74)$$

At the minimum $\phi = \phi_0$, we have $\partial_i V(\phi_0) = 0$ by definition of a stationary point, and this expression is trivially satisfied. We must then impose a second partial derivative with respect to some ϕ_j and then evaluate at ϕ_0 :

$$0 = \frac{\partial}{\partial \phi_j} \left(\frac{\partial V}{\partial \phi_i} \chi_i(\phi) \right) \Big|_{\phi_0} = \frac{\partial^2 V}{\partial \phi_j \partial \phi_i} \Big|_{\phi_0} \chi_i(\phi_0) + \frac{\partial V}{\partial \phi_i} \Big|_{\phi_0} \frac{\partial \chi_i}{\partial \phi_j} \Big|_{\phi_0} \quad (8.75)$$

But $\partial_i V(\phi_0) = 0$, so the second term vanishes and we get

$$\frac{\partial^2 V}{\partial \phi_j \partial \phi_i} \Big|_{\phi_0} \chi_i(\phi_0) = 0 \quad (8.76)$$

$$(m^2)_{ji} \chi_i(\phi_0) = 0 \quad (8.77)$$

Again we have two cases:

- $\chi_i(\phi_0) = 0$: This means that symmetry is preserved at the VEV (ground state), which makes this case uninteresting to look at.
- $\chi_i(\phi_0) \neq 0$: This is the interesting case as symmetry is broken by the non-zero transformation. However, as the product is zero, we now must have $m^2 = 0$.

The second point yields Goldstone's theorem:

Theorem 8.4 (Goldstone's theorem) If a global continuous symmetry is spontaneously broken, the field theory must contain massless scalar particles, or so-called *Goldstone bosons*, which corresponds to the broken symmetry generators.

However, the reverse is not true. Massless scalar particles do not always give rise to the spontaneous breaking of global symmetries.

But what about renormalisation? This is actually a good question, and using the effective action formalism, we can see that Goldstone's theorem is satisfied in all loop orders⁸.

Derivation 8.7 (Goldstone's theorem for effective potentials) Suppose $V_{\text{eff}}(\varphi)$ has a non-trivial minimum φ_0 . The stationary point condition then reads

$$\frac{\partial V_{\text{eff}}}{\partial \varphi} \Big|_{\varphi_0} = 0 \quad (8.78)$$

The mass-squared matrix of small fluctuations $\varphi_i = \varphi_0 + \eta_i$ is

$$(m^2)_{ij} = \frac{\partial^2 V_{\text{eff}}}{\partial \varphi \partial \varphi} \Big|_{\varphi_0} \quad (8.79)$$

We can see that the variation of the effective action is zero:

$$\delta \Gamma[\varphi] = \int d^4x \frac{\delta \Gamma}{\delta \varphi_i(x)} (T^a)_{ij} \varphi_j(x) = 0 \quad (8.80)$$

Hence yielding the condition

$$(T^a)_{ij} \varphi \frac{\partial V_{\text{eff}}}{\partial \varphi} = 0 \quad (8.81)$$

Again, this relationship is at the first derivative order and is trivial. Taking a second derivative gives

$$(T^a)_{ik} \delta_{ji} \varphi \frac{\partial V_{\text{eff}}}{\partial \varphi} + (T^a)_{ik} \varphi \frac{\partial^2 V_{\text{eff}}}{\partial \varphi \partial \varphi} = 0 \quad (8.82)$$

Now we set the field to the minimum φ_0 . The first term vanishes by (8.78), and one has

$$(m^2)_{ij} (T^a)_{jk} \varphi_0 = 0 \quad (8.83)$$

Hence, each broken generator T^a with $T^a \varphi \neq 0$ is an eigenvector of m^2 with eigenvalue zero. That is to say, our (massless) Goldstone bosons remain massless after renormalisation.

⁸You may be tempted to skip the rest of the section and only return to it after finishing Chapter 11.

Chapter 9

Interacting fields II: QED

As we already saw, QED is a gauge theory, which means that we need gauge fixing to quantise it. In this chapter, we will see that gauge fixing in path integrals is slightly different from canonical quantisation due to the structure of the integral and the presence of functionals.

9.1 Fermion propagator

As seen in (13.22), the Lagrangian for the Dirac fermion (without any further fermions) is given by

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu \partial_\mu - m)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu$$

The generating functional is then

Definition 9.1 (Dirac equation generating functional)

$$Z[J] = \int \mathcal{D}\bar{\psi} \mathcal{D}\psi e^{i \int d^d x (\bar{\psi}(i\cancel{\partial} - m)\psi + \bar{\eta}\psi + \bar{\psi}\eta)} \quad (9.1)$$

where η is the (Grassmann) source for the Dirac field that importantly *anticommutes* like the fermion field.

From this, we can derive the fermion propagator.

Derivation 9.1 (Fermion propagator) Once again, we can decompose the field into two parts:

$$\psi = \psi_c + \psi_q \quad (9.2)$$

where the classical-like ψ_c and satisfies the sourceful Dirac equation

$$(i\cancel{\partial} - m)\psi_c = -\eta \quad (9.3)$$

and is given by

$$\psi_c = -D^{-1}\eta \quad (9.4)$$

where we have used a shorthand by defining the so-called *Dirac operator*, which is, in some sense, an extended derivative^a:

Definition 9.2 (Dirac operator)

$$D = i\cancel{\partial} - m \quad (9.5)$$

Essentially, we have shifted fields to absorb the linear source terms:

$$\psi = \psi_q - D^{-1}\eta \quad \bar{\psi} = \bar{\psi}_q - \bar{\eta}D^{-1} \quad (9.6)$$

such that our Lagrangian reads

$$\bar{\psi}(i\cancel{\partial} - m)\psi + \bar{\eta}\psi + \bar{\psi}\eta = \bar{\psi}D\psi + \bar{\eta}\psi + \bar{\psi}\eta = \bar{\psi}_qD\psi_q - \bar{\eta}D^{-1}\eta \quad (9.7)$$

Again, as we have nothing but a linear change of variables, the integration measure shifts as

$$\mathcal{D}\phi = \mathcal{D}(\psi_c + \psi_q) = \mathcal{D}\psi_q \quad (9.8)$$

and the generating functional reads

$$Z[\eta, \bar{\eta}] = \int \mathcal{D}\psi_q \mathcal{D}\bar{\psi}_q e^{i \int \bar{\psi}_q D\psi_q} e^{-i \int d^d x d^d y \bar{\eta}(x) D^{-1}(x,y) \eta(y)} \quad (9.9)$$

Like in the case of the Feynman propagator, if one sets the source to zero, they find

$$Z[0, 0] = \int \mathcal{D}\psi_q \mathcal{D}\bar{\psi}_q e^{i \int \bar{\psi}_q D\psi_q} \quad (9.10)$$

As such, (9.9) is normalised via the relation

$$Z_0[\eta, \bar{\eta}] = \frac{Z[\eta, \bar{\eta}]}{Z[0, 0]} = e^{-i \int d^d x d^d y \bar{\eta}(x) D^{-1}(x,y) \eta(y)} \quad (9.11)$$

By consulting (7.53), we know that the fermion two-point function is simply the generating functional after taking the functional derivative twice by the source. That is to say:

$$\langle 0 | T[\psi(x)\bar{\psi}(y)] | 0 \rangle = - \left. \frac{\delta Z_0[\eta, \bar{\eta}]}{\delta \bar{\eta}(x) \delta \eta(y)} \right|_{\eta=\bar{\eta}=0} \quad (9.12)$$

Differentiating the exponent $-i \int \bar{\eta} D^{-1} \eta$ twice gives exactly our fermion propagator:

$$\langle 0 | T\psi(x)\bar{\psi}(y) | 0 \rangle = D^{-1}(x, y) = S_F(x - y) \quad (9.13)$$

By consulting the standard Fourier transform result (4.31), it is easy to see that the momentum space Dirac operator is $D = \not{p} - m$. Hence, its inverse is

$$S_F(p) = \frac{i}{\not{p} - m + i\epsilon} = i \frac{\not{p} + m}{p^2 - m^2 + i\epsilon} \quad (9.14)$$

Making the Fourier transform back to position space, we finally rederive the (tree-level) fermion propagator as

$$S_F(x - y) = \int \frac{d^d p}{(2\pi)^d} i \frac{\not{p} + m}{p^2 - m^2 + i\epsilon} e^{-ip(x-y)} \quad (9.15)$$

^aBut *not* a covariant derivative as no gauges are involved.

This concludes the easy half.

9.2 Photon field propagator

Now we arrive at the hard part. Recall the photon field Lagrangian from (). Adding a source term yields the generating functional

Definition 9.3 (Photon field generating functional)

$$Z[J_\mu] = \int \mathcal{D}A_\mu e^{i \int d^d x \left(\frac{1}{2} A^\mu (\square g_{\mu\nu} - \partial_\mu \partial_\nu) A^\nu + J^\mu A_\mu \right)} \quad (9.16)$$

Note 9.1 Here, J_μ remains an arbitrary source instead of the 4-current. This is because we have not included an interaction term (where the 4-current J^μ is present) in the (free) photon field Lagrangian.

Again, gauge freedom complicates many things. We show this by first mindlessly evaluating the path integral without considering them.

Derivation 9.2 (The propagator goes kaputt) From experience, one can reasonably convince themselves that

$$i(\square g_{\mu\nu} - \partial_\mu \partial_\nu) = \delta^4(x - y) D_{\mu\nu}(x - y) \quad (9.17)$$

That is to say, we must calculate the inverse of $\square g_{\mu\nu} - \partial_\mu \partial_\nu$ in order to solve for our beloved photon propagator. We begin by converting it to momentum space:

$$\square g_{\mu\nu} - \partial_\mu \partial_\nu \rightarrow p^2 g_{\mu\nu} - p_\mu p_\nu \quad (9.18)$$

If we contract the RHS with k^ν , we obtain

$$(p^2 g_{\mu\nu} - p_\mu p_\nu) p^\nu = g_{\mu\nu} p^2 p^\nu - p_\mu (p \cdot p) = p^2 p_\mu - p^2 p_\mu = 0 \quad (9.19)$$

i.e. $p^2 g_{\mu\nu} - p_\mu p_\nu$ maps any vector proportional to p_μ to zero and is hence not invertible.

Like in canonical quantisation, extra degrees of freedom must be eliminated by gauges. The way to do this for Abelian gauge theories is known as the so-called *Faddeev-Popov method*:

Derivation 9.3 (Fixing the gauge with the Faddeev-Popov method) As we know, in the case of electromagnetism or indeed, non-Abelian gauge theories in general, the gauge fields A_μ are not uniquely defined because they transform under a gauge transformation as:

$$A_\mu \rightarrow A_\mu^\alpha = A_\mu + \partial_\mu \alpha(x) \quad (9.20)$$

where A_μ^α is the field after it is gauge-transformed by some arbitrary function $\alpha(x)$.

In the case of the Lorenz gauge (6.143), the condition we want to impose is

$$\partial^\mu A_\mu^\alpha = 0 \quad (9.21)$$

which we can expand as

$$\partial^\mu (A_\mu + \partial_\mu \alpha) = \partial^\mu A_\mu + \partial^\mu \partial_\mu \alpha = 0 \quad (9.22)$$

This is a differential equation for α , which can always be solved, ensuring that we can always reach the Lorenz gauge by an appropriate choice of α .

The heart of the Faddeev-Popov method is to write our gauge condition of interest as a functional^a $G[A_\mu]$ satisfying

$$G[A_\mu] = 0 \quad (9.23)$$

Given any A_μ , we can always find a suitable gauge function $\alpha(x)$ such that the transformed field A_μ^α satisfies the gauge condition (9.23). Let us now make use of this in the photon field.

This involves the functional $\Delta[A_\mu]$, which is essentially a delta functional. It is the generalisation of the Dirac delta function, which is related to it by

Definition 9.4 (Delta functional)

$$\Delta[A_\mu]^{-1} = \int \mathcal{D}\alpha \delta(G[A_\mu]) \quad (9.24)$$

This functional is gauge-invariant. For a gauge transformation $\alpha \rightarrow \alpha + \alpha'$:

$$\Delta[A_\mu]^{-1} = \Delta[A_\mu^{\alpha'}]^{-1} \quad (9.25)$$

Interestingly, if we multiply both sides of (9.24) by $\Delta[A_\mu]$

$$1 = \Delta[A_\mu] \int \mathcal{D}\alpha \delta(G[A_\mu]) \quad (9.26)$$

which allows us to semi-cheatingly insert this expression as we please. Let us do so for (9.16):

$$Z[J^\mu] = \int \mathcal{D}A_\mu \left(\Delta[A_\mu] \int \mathcal{D}\alpha \delta(G[A_\mu^g]) \right) e^{iS[A_\mu]} \quad (9.27)$$

where we have used the following shorthand to denote the part of the action irrelevant to gauge fixing,

$$S[A_\mu] = i \int d^d x \left(\frac{1}{2} A^\mu (\square g_{\mu\nu} - \partial_\mu \partial_\nu) A^\nu + J^\mu A_\mu \right) \quad (9.28)$$

We now implement the gauge transformation $\alpha \rightarrow \alpha + \alpha'$:

$$Z[J^\mu] = \int \mathcal{D}\alpha \int \mathcal{D}A_\mu^{\alpha'} \Delta[A_\mu^{\alpha'}] \delta(G[A_\mu^{\alpha'}]) e^{iS[A_\mu^{\alpha'}]} \quad (9.29)$$

and change the variables of integration from $A_\mu^{\alpha'}$ to A_μ

$$Z[J^\mu] = \int \mathcal{D}g \int \mathcal{D}A_\mu \Delta[A_\mu] \delta(G[A_\mu]) e^{iS[A_\mu]} \quad (9.30)$$

hence finally eliminating the function $\alpha(x)$ outside of the integration variable.

To proceed from here, we must employ a nice trick on the term $\Delta[A_\mu]$ (9.24), which we will otherwise struggle to integrate. Let us change the variable of integration from g to $G[A_\mu]$:

$$\Delta[A_\mu]^{-1} = \int \mathcal{D}G[A_\mu] \det \left(\frac{\delta G[A_\mu]}{\delta \alpha} \right)^{-1} \delta(G[A_\mu]) = \det \left(\frac{\delta G[A_\mu]}{\delta \alpha} \right)^{-1} \Big|_{G[A_\mu]=0} \quad (9.31)$$

which easily transforms to

$$\Delta[A_\mu] = \det \left(\frac{\delta G[A_\mu](x)}{\delta \alpha(y)} \right) \Big|_{G[A_\mu](x)=0} \quad (9.32)$$

where we have labelled the two distinct 4-coordinates x and y .

Let us now define the so-called *Faddeev-Popov operator* $M(x, y)$ by the relation

$$\det M(x, y) = \det \left(\frac{\delta G[A_\mu]}{\delta \alpha} \right) \Big|_{G[A_\mu]=0} \quad (9.33)$$

$M(x, y)$ itself can be derived using the chain rule:

Definition 9.5 (Faddeev-Popov operator)

$$M(x, y) = -\partial_\mu^y \frac{\delta G[A_\mu](x)}{\delta \alpha(y)} \quad (9.34)$$

where the partial derivative concerns the y coordinate.

Using the Faddeev-Popov method gauge condition $G[A_\mu] = 0$ (9.23), we have

$$M(x, y) = -\partial^2 \delta(x - y) \quad (9.35)$$

which is nothing but a field-independent functional determinant that *does not* introduce interactions. This allows us to replace $\Delta[A_\mu]$ in (9.30) with $\det M$:

$$Z[J^\mu] = \int \mathcal{D}\alpha \int \mathcal{D}A_\mu \det M \delta(G[A_\mu]) e^{iS[A_\mu]} \quad (9.36)$$

Now the important step is to decompose our friend, the functional $G[A_\mu]$:

$$G[A_\mu] = D[A_\mu] + \omega(x) \quad (9.37)$$

where ω is an arbitrary, so-called ‘dummy’ function which we ultimately integrate. The more interesting part is $D[A_\mu, x]$, in which we will now finally introduce the Lorenz gauge explicitly:

$$D[A_\mu, x] = \partial^\mu A_\mu \quad (9.38)$$

From (9.23), we easily see that

$$D[A_\mu, x] = \partial^\mu A_\mu = -\omega(x) \quad (9.39)$$

Using the property of a Gaussian function, we can introduce a Gaussian averaging over the dummy function $\omega(x)$ with weight:

$$\int \mathcal{D}\omega e^{-\frac{i}{2\xi} \int d^d x \omega(x)^2} = \text{arbitrary constant} \quad (9.40)$$

The size of the constant does not affect the final result of our physical quantities^b. For convenience, we simply set the constant to 1, which normalises this Gaussian integral. Again, we cheatingly multiply the generating functional by 1, and with that, our integral:

$$Z[J^\mu] = \int \mathcal{D}\omega e^{-\frac{i}{2\xi} \int d^d x \omega^2} \int \mathcal{D}\alpha \int \mathcal{D}A_\mu \det M \delta(G[A_\mu]) e^{iS[A_\mu]} \quad (9.41)$$

We also recognise that, from the decomposition

$$\delta(G[A_\mu]) = \delta(\partial^\mu A_\mu + \omega(x)) \quad (9.42)$$

Substituting this into the generating functional yields

$$Z[J^\mu] = \int \mathcal{D}\omega e^{-\frac{i}{2\xi} \int d^d x \omega^2} \int \mathcal{D}\alpha \int \mathcal{D}A_\mu \det M \delta(\partial^\mu A_\mu + \omega(x)) e^{iS[A_\mu]} \quad (9.43)$$

The delta function essentially enforces $\partial^\mu A_\mu = -\omega(x)$, a result we have already seen in (9.39). Indeed, using the sifting property (4.26), we can now integrate over $\omega(x)$ as

$$\int \mathcal{D}\omega \delta(\partial^\mu A_\mu + \omega(x)) e^{-\frac{i}{2\xi} \int d^d x \omega^2} = e^{-\frac{i}{2\xi} \int d^d x (\partial^\mu A_\mu)^2} \quad (9.44)$$

This gives

$$Z[J^\mu] = \int \mathcal{D}\alpha \int \mathcal{D}A_\mu \det M e^{iS[A_\mu] - \frac{i}{2\xi} \int d^d x (\partial^\mu A_\mu)^2} \quad (9.45)$$

There remains the problem of the integration variable α . However, we recognise that by construction, the functional measure $\mathcal{D}A_\mu$ and the action $S[A_\mu]$ are gauge-invariant, as is the Faddeev-Popov determinant $\det M$. Therefore, the $\mathcal{D}\alpha$ integral over α just yields an overall constant factor^c. Like the arbitrary size of the Gaussian integral, we can assume this to be 1 for convenience.

As such, only one integration variable remains, which is A_μ

$$Z[J^\mu] = \int \mathcal{D}A_\mu \det M e^{i \int d^d x (S[A_\mu] + \frac{1}{2\xi} (\partial_\mu A^\mu)^2)} \quad (9.46)$$

Finally, we reinsert the full form of the shorthand (9.28), consult (6.146) and acknowledge that, as the photon field is Abelian, the Faddeev-Popov operator term evolves into its fairly harmless form (9.35). This gives our final, well-defined expression

Definition 9.6 (Gauge-fixed photon field generating functional)

$$Z[J^\mu] = \int \mathcal{D}A_\mu e^{i \int d^d x \frac{1}{2} A^\mu (\square g_{\mu\nu} - (1 - \frac{1}{\xi}) \partial_\mu \partial_\nu +) A^\nu + \int d^d x J^\mu A_\mu} \quad (9.47)$$

^aIt is a functional because A_μ ultimately depends on x_μ .

^bRemember that in deriving physical quantities, we take functional derivatives of $Z[A_\mu]$ and divide the result by $Z[0]$ (or rather use the normalised generating functional $Z_0[A_\mu]$), killing the arbitrary constants in both.

^cThis factor is exactly the volume of the gauge group.

Note 9.2 (Faddeev-Popov ghosts) We append this discussion with a taste of things to come. In QED, which is an *Abelian* gauge theory, (9.35) applies. In non-Abelian gauge theories like QCD, however, the Faddeev-Popov operator term in the generating functional

$$\det M = \int \mathcal{D}c \mathcal{D}\bar{c} e^{iS_{\text{ghost}}} \quad (9.48)$$

which contributes to the Lagrangian. c and \bar{c} are unphysical fields known as the *Faddeev-Popov ghost fields*^a or simply *Faddeev-Popov ghosts* which obey Grassmann anticommutations. We will discuss the treatment of Faddeev-Popov Ghosts much later.

^aThe general notion of a *ghost field* denote unphysical fields emerging in the Lagrangian. Specifically, ghost fields in QFT assume the name ‘Faddeev-Popov ghost fields’.

Now that we have finished gauge fixing, the 3-field vertex can be derived.

Note 9.3 (Vertex Feynman rule) We revise the derivation of the Feynman rule of a generic vertex:

- Perform a Fourier transform to send the corresponding Lagrangian to momentum space.
- Integrate the momentum space Lagrangian over the 4-coordinates x to solve for the action.
- Take functional derivatives of the action exponential (generating functional) w.r.t. the fields to solve for the vertex contribution to \mathcal{M}_{fi} .
- However, remember from the last section that this step can be simplified by stripping off the external fields. The remaining part can then be read off as the vertex.

Now let us put this into action.

Derivation 9.4 (Photon propagator) We again begin by renormalising our generating functional (9.47). When the source is zero, one has

$$Z[0] = \int \mathcal{D}A_\mu e^{i \int d^d x \frac{1}{2} A^\mu (\square g_{\mu\nu} - (1 - \frac{1}{\xi}) \partial_\mu \partial_\nu) A^\nu} \quad (9.49)$$

We go momentum space via a Fourier transform:

$$Z[0] = \int \mathcal{D}A_\mu e^{\frac{1}{2} \int \frac{d^4 p}{(2\pi)^4} A^\mu(-p) (-p^2 g_{\mu\nu} + (1 - \frac{1}{\xi}) p_\mu p_\nu) A^\nu(p)} \quad (9.50)$$

Performing the Gaussian integral yields us the normalised generating functional, which reads

$$Z[J] = Z[0] e^{\frac{i}{2} \int \frac{d^d k}{(2\pi)^d} J^\mu(-k) D_{\mu\nu}(k) J^\nu(k)} \quad (9.51)$$

with the propagator $D_{\mu\nu}(k)$ defined by

$$\left(-p^2 g_{\mu\nu} + \left(1 - \frac{1}{\xi} \right) p_\mu p_\nu \right) D^\alpha_\nu(k) = i \delta_{\mu\nu} \quad (9.52)$$

The rest of the derivation is identical to what we did in Part II from (6.178) on.

9.3 Vertex

We once again find ourselves on the doorsteps of greatness as we formulate QED for a second time. Putting it all together, the QED generating functional is

Definition 9.7 (QED generating functional)

$$Z[J^\mu] = \int \mathcal{D}A_\mu \mathcal{D}\bar{\psi} \mathcal{D}\psi e^{i \int d^d x (A^\mu (\square g_{\mu\nu} - (1 - \frac{1}{\xi}) \partial_\mu \partial_\nu) A^\nu + \bar{\eta}\psi + \bar{\psi}\eta + J^\mu A_\mu)} \quad (9.53)$$

where $\bar{\eta}\psi + \bar{\psi}\eta$ is the fermion field source term, and $J^\mu A_\mu$ is the photon (gauge) field source term.

Remark 9.1 Strictly speaking, one can even introduce the previously seen ghost field terms

$$Z[J] = \int \mathcal{D}A_\mu \mathcal{D}\bar{\psi} \mathcal{D}\psi \mathcal{D}\bar{c} \mathcal{D}c e^{i \int d^d x (A^\mu (\square g_{\mu\nu} - (1 - \frac{1}{\xi}) \partial_\mu \partial_\nu) A^\nu + \mathcal{L}_{\text{ghost}} + J^\mu A_\mu + \bar{J}\psi + \bar{\psi}J)} \quad (9.54)$$

which are trivial (i.e. vanish) in QED anyway.

From the interaction term, we can then derive the 3-point function, which is the vertex.

Derivation 9.5 (Vertex) We start with the QED interaction Lagrangian

$$\mathcal{L}_{\text{int}}(x) = -e\bar{\psi}(x)\gamma^\mu\psi(x)A_\mu(x) \quad (9.55)$$

First, we move it to momentum space by Fourier transforms. The fields transform as

$$\psi(x) = \int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot x} \psi(p) \quad \bar{\psi}(x) = \int \frac{d^4 p'}{(2\pi)^4} e^{+ip' \cdot x} \bar{\psi}(p') \quad A_\mu(x) = \int \frac{d^4 k}{(2\pi)^4} e^{-ik \cdot x} A_\mu(k) \quad (9.56)$$

which give the momentum space Lagrangian

$$\mathcal{L}_{\text{int}}(x) = -e \int \frac{d^4 p d^4 p' d^4 k}{(2\pi)^{12}} e^{i(p' - p - k) \cdot x} \bar{\psi}(p') \gamma^\mu \psi(p) A_\mu(k) \quad (9.57)$$

Integrating over the 4-coordinates x , we find the interaction term contribution to the action

$$S_{\text{int}} = \int d^4 x \mathcal{L}_{\text{int}}(x) = -e \int \frac{d^4 p d^4 p' d^4 k}{(2\pi)^8} (2\pi)^4 \delta^{(4)}(p' - p - k) \bar{\psi}(p') \gamma^\mu \psi(p) A_\mu(k) \quad (9.58)$$

where the δ function enforces momentum conservation at the vertex ($p' = p + k^a$).

We recall that path integrals allow us to read off vertices from the Lagrangian: the amputated 3-point kernel is i times the coefficient of $\bar{\psi}\psi A$ in S_{int} . Here, amputating the external fields/legs also mean stripping off the 4-momentum integrals which are a part of the fields in momentum space. As such:

$$\frac{\delta^3(iS_{\text{int}})}{\delta\bar{\psi}(p')\delta\psi(p)\delta A_\nu(k)} = -ie\gamma^\nu(2\pi)^4\delta^{(4)}(p' - p - k) \quad (9.59)$$

which gives the familiar-looking vertex as

$$\boxed{-ie\gamma^\mu} \quad (9.60)$$

where $(2\pi)^4\delta^{(4)}$ that enforces momentum conservation at the vertex (the global δ -function is usually factored out for the whole diagram).

Fourier transforming the momentum space kernel back gives the position space vertex:

$$-ie(\gamma^\mu)_{\alpha\beta}\delta^{(4)}(x_1 - x_2)\delta^{(4)}(x_1 - x_3) \quad (9.61)$$

where the three fields meet at the same spacetime point with the Dirac structure $-ie\gamma^\mu$.

^aIf all momenta are taken as incoming, this is then $p + p' + k = 0$.

Part IV

Renormalisation and regularisation

Chapter 10

Regularisation

Quote 10.1 The sun set in the west on a notion where no man had dared to venture. And beyond that - infinity.

1492: *Conquest of Paradise*

Even though we noted loops in the Feynman rules, everything we have done so far has been restricted to the tree level. This is obviously problematic because loop interactions emerge in real life. We treat this via renormalisation, which itself must only commence after we perform regularisation.

10.1 Motivation

Previously in Part II, we introduced the complete set of Feynman rules and the concept of the scattering amplitude \mathcal{M} . As the objective now is to go beyond the tree level, it is a good idea to review the full form of \mathcal{M} up to the infinite order:

Theorem 10.1 (Full ϕ^4 theory scattering amplitude)

$$i\mathcal{M} = -i\lambda + \frac{(-i\lambda)^2}{2} \sum I_2 + O(\lambda^3) \quad (10.1)$$

where $\sum I_n$ is the sum of all possible *Feynman integrals* of that order n .

This systematic treatment makes use of the so-called Feynman integrals, which are representations of loop contributions to the scattering amplitude. The Feynman integral of a diagram arises naturally from the Feynman rules:

Definition 10.1 (Feynman integral)

$$I = \lim_{\epsilon \rightarrow 0} \int \prod_{k=1}^l \frac{d^d p_k}{i\pi^{d/2}} \prod_{k=1}^n \frac{1}{(p_k^2 - m_1^2 + i\epsilon)} \quad (10.2)$$

where:

- d is the physical dimension of the manifold.
- l is the number of loops.
- n is the number of the internal legs/lines.

Note 10.1 One can think of Feynman integrals I_n as a generalised version of Feynman propagators D_F . Indeed, a Feynman integral is ultimately a combination of Feynman propagators integrated over internal momenta. It corresponds to the probability amplitude for a particle to propagate from a point *back to the same point* some l times and hence represents l loops.

Previously, the reader has been led to believe that this is the end of the story. Beyond the tree level,

this is unfortunately a lie. Uniquely among all Feynman diagram elements, loops are mathematically represented by integrating over internal loop momenta. To be exact, loop integrals.

It is well-known that loop integrals often diverge into infinity, which is indeed the case for our Feynman integrals. Unsurprisingly, physical systems are not infinite, which suggests that our current formulation of QFT fails when we go beyond the tree level. In this chapter, we restrict our discussion to ϕ^4 theory.

Derivation 10.1 (Tadpole diagram) We illustrate this with the simplest possible example, that being a 1-loop, 1-point/2-point function known as the so-called *tadpole diagram*:

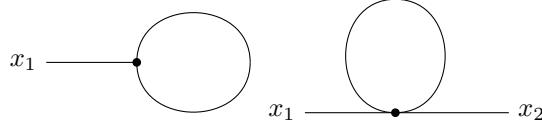


Figure 10.1: Archetypical tadpole diagrams.

The self-energy of a Feynman diagram is directly related to its Feynman integral by a coupling constant (and arbitrary constants). In ϕ^4 theory, the self-energy of an is simply the Feynman integral multiplied by a factor of $\frac{\lambda^n}{2}$ where n is the *number of vertices*. In our case, this is $\frac{\lambda}{2}$. Take the one-point tadpole as an example:

$$\Sigma_{\text{tadpole}} = \frac{\lambda}{2} \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 - m^2 - i\epsilon} \quad (10.3)$$

where, notably, we assume an arbitrary number of dimensions d for the time being. As we will prove later on, this integral evaluates as

$$\Sigma_{\text{tadpole}} = \frac{\lambda}{2} \frac{\pi^{d/2}}{(2\pi)^d} \Gamma\left(1 - \frac{d}{2}\right) m^{d-2} \quad (10.4)$$

where $\Gamma(n)$ is a so-called *gamma function* defined by

Definition 10.2 (Gamma function)

$$\Gamma(n) = \int_0^\infty t^{n-1} e^{-t} dt \quad (10.5)$$

If n is a positive integer, $\Gamma(n)$ reduces to

$$\Gamma(n) = (n-1)! \quad (10.6)$$

Theorem 10.2 (Gamma function properties)

- $\Gamma(n)$ has poles at $n = 0, -1, -2, \dots$
- $\Gamma(n+1) = n\Gamma(n)$
- $\Gamma'(1) = -\gamma \approx -0.5772$ where γ is the Euler-Mascheroni constant.

We now make the arbitrary choice of $d = 2$. The gamma function becomes

$$\Gamma\left(1 - \frac{2}{2}\right) = \int_0^\infty t^{-1} e^{-t} dt = \infty \quad (10.7)$$

Hence, we have $\Sigma_{\text{tadpole}} = \infty$ for $d = 2$.

What is the physical meaning of this? Essentially, when p is *large*, the original integral approaches

$$\Sigma_{\text{tadpole}} = \frac{\lambda}{2(2\pi)^d} \int d^d p \frac{1}{p^2} \quad (10.8)$$

which diverges to infinity for $d \geq 2$. This result is clearly unphysical and is known as the so-called *ultraviolet (UV) singularity*.

Remark 10.1 Another singularity is the so-called *infrared (IR) singularity* which takes occur in low energies/*smol* momenta.

10.2 Mathematical toolkit

The tadpole diagram is a good example of infinities as its integral is already simplified and can be evaluated as-is. This is, regrettably, not the case for almost all other diagrams, for which the elephant in the room remains our inability to integrate over the four-momentum p_μ . Rather, a mathematical trick known as *Feynman parameterisation* is required.

Derivation 10.2 (Feynman parameterisation)

Definition 10.3 (Feynman parameters) In his 1949 paper ‘Space-Time Approach to Quantum Electrodynamics’, Feynman noted the relation

$$\begin{aligned} \frac{1}{ab} &= -\frac{1}{a-b} \left(\frac{1}{a} - \frac{1}{b} \right) = -\frac{1}{a-b} \left[\frac{1}{xa + (1-x)b} \right]^1 \\ &= \int_0^1 dx \frac{1}{(xa + (1-x)b)^2} = \int_0^1 dx \int_0^1 dy \delta(x+y-1) \frac{1}{(xa+yb)^2} \end{aligned} \quad (10.9)$$

where $x, y \in [0, 1]$ are the so-called *Feynman parameters*. One can generalise this relation as

$$\frac{1}{a_1 \cdots a_n} = \int dx_1 \cdots dx_n \delta \left(\sum_{i=1}^n x_i - 1 \right) \frac{(n-1)!}{[\sum_{i=1}^n x_i a_i]^n} \quad (10.10)$$

Now let us apply this in the context of Feynman integrals. Assume the loop momentum k , external momenta p_i , mass m_i and the singularity-avoiding term $i\epsilon$, we define our a_i as

$$a_i = (k + p_i)^2 - m_i^2 + i\epsilon \quad (10.11)$$

The delta function in (10.10) essentially enforces the relation

$$\sum_{i=1}^n x_i - 1 = 0 \rightarrow \sum_{i=1}^n x_i = 1 \quad (10.12)$$

We can thus write, using (10.11) and (10.12):

$$\sum_{i=1}^n x_i a_i = k^2 \sum_{i=1}^n x_i (2kp_i + P_i^2 - m_i^2) + i\epsilon \quad (10.13)$$

We define a rescaled loop momentum l , which is

$$l = k + \sum_{i=1}^n x_i p_i \quad (10.14)$$

This gives rise to the expression

$$\sum_{i=1}^n x_i a_i = l^2 - \left(\sum_{i=1}^n x_i p_i \right)^2 - \sum_{i=1}^n x_i (p_i^2 - m_i^2) + i\epsilon = l^2 - \mu^2 + i\epsilon \quad (10.15)$$

where we have defined a (rescaled) mass parameter μ that is independent of l , satisfying

$$\mu^2 = \left(\sum_{i=1}^n x_i p_i \right)^2 - \sum_{i=1}^n x_i (p_i^2 - m_i^2) \quad (10.16)$$

This is nice, as we have now eliminated the Feynman parameter x from the final expression completely! Here we see why Feynman parameterisation is so-called. Essentially, the new loop momenta and energy are rescaled (and parameterised) using our new friends, the Feynman parameters.

Another nice standard trick one can employ to simplify Feynman integrals is the so-called *Wick rotation*:

Definition 10.4 (Wick rotation) Consider the 4D Minkowski metric

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \quad (10.17)$$

The Wick rotation involves a coordinate rescaling

$$t = -i\tau \quad (10.18)$$

after which the metric reads

$$ds^2 = d\tau^2 + dx^2 + dy^2 + dz^2 \quad (10.19)$$

which is simply the 4D Euclidian metric. The Minkowski space version of our quantity of interest can be recovered by reversing our rescaling (10.18).

The Wick rotation simplifies calculations. One can see this by noting that the path integral becomes, in Euclidian space,

$$e^{iS[\phi]} \rightarrow e^{-S_E[\phi]} \quad (10.20)$$

which converges due to the negative sign.

Note 10.2 In this book, some sacrifices are made for our convenience:

- We will relabel the rescaled l and μ back to p and m after the Feynman parameterisation is complete. This ensures that we see symbols we are familiar with in the final result.
- The same thing will happen w.r.t. Wick rotation. We will announce a Wick rotation has been made, at which point it should be assumed that the physical quantities, which retain their pre-Wick rotation notation, are those after Wick rotation.

Derivation 10.3 (Bubble diagram) We are now in a position to observe the emergence of infinities in a 1-loop, 2-point/4-point function, whose Feynman diagrams are called *bubble diagrams*:

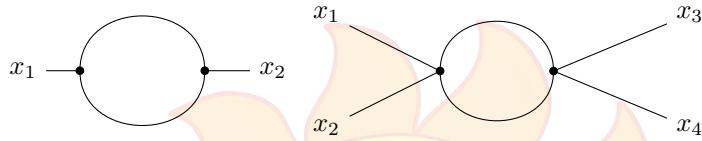


Figure 10.2: Archetypical bubble diagrams.

Let us take the 4-point diagram as an example. We recall that the self-energy of a diagram with n vertices is related to the Feynman integral by $\frac{\lambda^n}{2}$. For the bubble diagram, this is $\frac{\lambda^2}{2}$:

$$\Sigma_{\text{bubble}} = \frac{\lambda^2}{6} \int \frac{d^d p}{(2\pi)^d} \frac{1}{(p^2 - m^2)((q-p)^2 + m^2)} \quad (10.21)$$

where $q = p_1 + p_2 = -p_3 - p_4$. Performing Feynman parameterisation and a Wick rotation yields

$$\Sigma_{\text{bubble}} = -\frac{\lambda^2}{2} \int \frac{d^d p}{(2\pi)^d} \frac{1}{(p^2 - m^2 + i\epsilon)^2} \quad (10.22)$$

where it should be clear that p and m are nothing but the rescaled l and μ which we have relabelled. This simplified integral evaluates as

$$\Sigma_{\text{bubble}} = \frac{\lambda^2}{2} \frac{\pi^{d/2}}{(2\pi)^d} \Gamma\left(2 - \frac{d}{2}\right) \int_0^1 x(1-x)q^2 + m^{2d-2-2} dx \quad (10.23)$$

Here, the UV singularity emerges at $d = 4$. The physical implication is likewise slightly altered: When p is *large*, the original integral approaches

$$\Sigma_{\text{bubble}} = \frac{\lambda^2}{2(2\pi)^d} \int d^d p \frac{1}{p^4} \quad (10.24)$$

which again is a UV singularity where the integral diverges to infinity for $d \geq 4$.

Exercise 10.1 Using Feynman parameterisation, show that a UV singularity emerges in a *sunset diagram*, which looks like

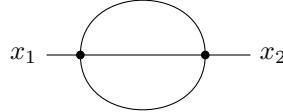


Figure 10.3: Archetypical sunset diagram.

and has the self-energy

$$\Sigma_{\text{sunset}} = \int \frac{d^d k}{(2\pi)^d} \frac{d^d q}{(2\pi)^d} \frac{1}{(k^2 - m^2 + i\epsilon)((p - k - q)^2 - m^2 + i\epsilon)(q^2 - m^2 + i\epsilon)} \quad (10.25)$$

where p is the external momentum, k and a are the loop momenta and m is the mass of the internal propagators.

Quote 10.2 There's potential for a lot of upheaval.

Alessio Serafini, on the sunset diagram, 6 March 2025

In fact, for a sufficiently large d , most Feynman diagrams display UV singularities. They are hence said to be *UV divergent*. One way to treat these infinities is *regularisation*, which, as we will see, is a very hand-wavy way to deal with infinities.

10.3 Cutoff regularisation

Derivation 10.4 (Cutoff regularisation) We return to the tadpole diagram self-energy (10.3). We have written it with an indefinite integral, but it actually spans over positive and negative infinities:

$$\Sigma_{\text{tadpole}} = \frac{\lambda}{2} \int_{-\infty}^{\infty} \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 - m^2 - i\epsilon} \quad (10.26)$$

It has already been established that integrating over the positive infinity is what causes infinities to emerge. So what if we replace the upper limit with a *large* but finite Λ ? The integral becomes

$$\Sigma_{\text{tadpole}} = \frac{i\lambda}{4\pi^2} \left(\Lambda^2 \sqrt{1 + \frac{m^2}{\Lambda^2}} - m^2 \ln \left(\frac{\Lambda + \Lambda \sqrt{1 + \frac{m^2}{\Lambda^2}}}{m} \right) \right) \quad (10.27)$$

where Λ is known as the *regulator*.

This is known as *cutoff regularisation* or the *UV momentum cutoff*, one of the many ways to regularise Feynman integrals. Its physical meaning lies in the fact that we essentially introduce a ceiling of the biggest momentum the system we have, which allows us to avoid the singularity in (10.8).

The original singularity is then recovered when $\Lambda \rightarrow \infty$

The same happens to a bubble diagram, although in this case we use Feynman parameterisation and Wick rotation, given its complexity.

10.4 Pauli-Villars regularisation

The so-called *Pauli-Villars regularisation* or *P-V regularisation* is named after Wolfgang Pauli and Felix Villars, and involves rewriting the first propagator using Feynman parameterisation. In the general case,

the P-V regularisation is accomplished by

$$\frac{1}{k^2 - m^2 + i\epsilon} \rightarrow \frac{1}{k^2 - m^2 + i\epsilon} - \sum_i \frac{a_i}{k^2 - \Lambda_i^2 + i\epsilon} \quad (10.28)$$

for a series of mass parameters Λ_i with indices i .

Derivation 10.5 (Pauli-Villars regularisation) Now consider the bubble diagram. Again, writing out the integration explicitly gives

$$\Sigma_{\text{bubble}} = \frac{(-i\lambda)^2}{2} \int_{-\infty}^{\infty} \frac{d^d p}{(2\pi)^d} \frac{i}{k^2 - m^2 + i\epsilon} \frac{i}{(p - k)^2 + m^2 + i\epsilon} \quad (10.29)$$

where the factor $-i$ accounts for the Wick rotation. We then perform the P-V regularisation:

$$\frac{1}{k^2 - m^2 + i\epsilon} \rightarrow \frac{1}{k^2 - m^2 + i\epsilon} - \underbrace{\frac{1}{k^2 - \Lambda^2 + i\epsilon}}_{\textcircled{1}} = \frac{m^2 - \Lambda^2}{(k^2 - m^2 + i\epsilon)(k^2 - \Lambda^2 + i\epsilon)} \quad (10.30)$$

where λ is again *large* (i.e. $\Lambda \gg m$) but finite, and $\textcircled{1}$ is a pseudo-mass term that represents a fictitious photon mass. The self-energy then reads

$$\Sigma_{\text{bubble}} = -\frac{\lambda^2 \Lambda^2}{2} \int_{-\infty}^{\infty} \frac{d^d p}{(2\pi)^d} \frac{1}{(k^2 - m^2 + i\epsilon)(k^2 - \Lambda^2 + i\epsilon)((p - k)^2 + m^2 + i\epsilon)} \quad (10.31)$$

In regularisation, it is often useful to split the self-energy into finite and infinite (i.e. convergent and divergent) parts so that their physical significance can be better understood:

$$\Sigma(p^2) = \Sigma(0) + \tilde{\Sigma}(p^2) \quad (10.32)$$

In our example, they are

$$\Sigma_{\text{bubble}}(0) = -\frac{\lambda^2 \Lambda^2}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{1}{(k^2 - m^2 + i\epsilon)^2 (k^2 - \Lambda^2 + i\epsilon)} \quad (10.33)$$

$$\tilde{\Sigma}_{\text{bubble}}(p^2) = -\frac{\lambda^2 \Lambda^2}{2} \int \frac{d^4 k}{(2\pi)^4} \frac{2pk - p^2}{(k^2 - m^2)^2 (k^2 - \Lambda^2)((p - k)^2 - m^2)} \quad (10.34)$$

10.5 Dimensional regularisation

The process known as *dimensional regularisation* is arguably the most important method of regularisation in perturbation theory. As we will see later, it can be used in conjunction with the minimal subtraction scheme to perform renormalisation.

Derivation 10.6 (Tadpole diagram) Let us tie up our previous loose end by evaluating the tadpole diagram step by step. The Feynman (loop) integral reads

$$I_2 = \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 - m_0^2 + i\epsilon} \quad (10.35)$$

In this case, we label the regularisation scale M instead of Λ and make the rescaling

$$I_2 = \frac{1}{M^{d-4}} \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 - m_0^2 + i\epsilon} \quad (10.36)$$

We define another parameter $\varepsilon = 4 - d$ (not to be confused with the ϵ in the term $i\epsilon$) and transition to Euclidean space by performing a Wick rotation $k^0 \rightarrow ik^0$, giving:

$$I_2 = iM^\varepsilon \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 + m_0^2} \quad (10.37)$$

which integrates as

$$I_2 = \frac{iM^\varepsilon}{(2\pi)^d} V_S \int_0^\infty d^d p \frac{p^{d-1}}{p^2 + m_0^2} \quad (10.38)$$

where V_S is the surface volume of a unit sphere in d dimensions that has the standard formula $V_S = \frac{2\pi^{d/2}}{\Gamma(d/2)}$.

By making the substitution $x = m_0^2/(p^2 + m_0^2)$, the integral evaluates as

$$I_2 = V_S \frac{m_0^2}{2(2\pi)^d} \left(\frac{M}{m_0}\right)^\varepsilon \int_0^1 (1-x)^{d/2-1} x^{-d/2} \quad (10.39)$$

Very sneakily, the integral $\int_0^1 (1-x)^{d/2-1} x^{-d/2}$ is actually three beta functions in disguise. We know this as it is a standard *Euler beta function*, which has the form

Definition 10.5 (Euler beta function)

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1} (1-x)^{b-1} dx \quad (10.40)$$

We can hence rewrite

$$I_2 = V_S \frac{m_0^2}{2(2\pi)^d} \left(\frac{M}{m_0}\right)^\varepsilon \frac{\Gamma(d/2)\Gamma(1-d/2)}{\Gamma(1)} \quad (10.41)$$

where $\Gamma(1)$ is simply 1. Inserting the so-called surface volume $V_S = \frac{2\pi^{d/2}}{\Gamma(d/2)}$, we get

$$I_2 = \frac{m_0^2}{(2\pi)^{2-\varepsilon/2}} \left(\frac{M}{m_0}\right)^\varepsilon \Gamma(1-d/2) \quad (10.42)$$

We can perform the following expansions

$$\Gamma(1-d/2) = \Gamma(-\varepsilon/2) \approx -\frac{2}{\varepsilon} - \gamma - 1 + O(\varepsilon) \quad (10.43)$$

where γ is again the Euler-Mascheroni constant.

$$\left(\frac{M}{m_0}\right)^{4-d} = 1 + \epsilon \ln\left(\frac{M}{m_0}\right) + \mathcal{O}(\epsilon^2) \quad (10.44)$$

$$(2\pi)^{d/2} = (2\pi)^2 \left(1 - \frac{\epsilon}{2} \ln(2\pi) + \mathcal{O}(\epsilon^2)\right) \quad (10.45)$$

Putting it all together, and treating ε as the parameter of interest, we find

$$I_2 = -\frac{m_0^2}{(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln(4\pi) + \ln \frac{\mu^2}{m_0^2}\right) \quad (10.46)$$

which diverges as $\epsilon \rightarrow 0$ or $d \rightarrow 4$.

As the tadpole diagram has only one measly vertex, its self-energy contribution is related to the Feynman integral by $\frac{\lambda^1}{2}$. Hence, the self-energy is

$$\Sigma_{\text{tadpole}} = -\frac{m_0^2 \lambda}{2(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln(4\pi) + \ln \frac{\mu^2}{m_0^2}\right) \quad (10.47)$$

Derivation 10.7 (Bubble diagram) We know that the loop integral of a bubble diagram is

$$I_4 = \int \frac{d^d k}{(2\pi)^d} \frac{1}{k^2 - m^2 + i\epsilon} \frac{1}{(p-k)^2 - m^2 + i\epsilon} \quad (10.48)$$

where $p = p_1 + p + 2 = p_3 + p_4$ is the external momentum and k is the internal loop momentum. Again, we employ the regularisation scale M and make the rescaling:

$$I_4 = \frac{1}{M^{d-4}} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 - m^2 + i\epsilon)^2} \frac{1}{((p-k)^2 - m^2 + i\epsilon)^2} \quad (10.49)$$

After Feynman parameterisation, we get

$$\begin{aligned} I_4 &= \frac{1}{M^{d-4}} \int \frac{d^d k}{(2\pi)^d} \int_0^1 dx \frac{1}{(x((p-k)^2 - m^2) + (1-x)(k^2 - m^2))^2} \\ &= M^\varepsilon \int \frac{d^d k}{(2\pi)^d} \int_0^1 dx \frac{1}{(k - xp)^2 + x(1+x)p^2 - m^2} \end{aligned} \quad (10.50)$$

By defining $l^\mu = k^\mu - xp^\mu$ and performing a Wick rotation, we can rewrite

$$I_4 = iM^\varepsilon \int \frac{d^d k}{(2\pi)^d} \int_0^1 dx \frac{1}{(l^2 + m^2 - x(1-x)p^2 + i\epsilon)^2} \quad (10.51)$$

Using the properties of the Gaussian function, we have

$$I_4 = \frac{i\pi^{d/2}}{(2\pi)^d} \Gamma(\varepsilon/2) \int_0^1 dx \frac{1}{(m^2 - x(1-x)p^2)^{\varepsilon/2}} \quad (10.52)$$

Let us perform expansions of these terms. Labelling $a^2 = m^2 - x(1-x)p^2$:

$$\Gamma(\varepsilon/2) = \frac{2}{\varepsilon} - \gamma + O(\varepsilon) \quad \frac{1}{(a^2)^{\varepsilon/2}} = 1 - \frac{\varepsilon}{2} \ln a^2 + O(\varepsilon^2) \quad M^\varepsilon = (M^2)^{\varepsilon/2} = 1 + \frac{2}{\varepsilon} \ln M^2 + \dots \quad (10.53)$$

Putting it all together, and treating ε as the parameter of interest, we find

$$I_4 = \frac{i\pi^{d/2}}{(2\pi)^d} \int_0^1 dx \left(\frac{2}{\varepsilon} - \gamma - \ln \frac{m^2 - x(1-x)p^2}{M^2} + O(\varepsilon) \right) \quad (10.54)$$

Using the standard integration techniques, we have

$$I_4 = -\frac{m_0^2}{(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln \frac{m_0^2}{\mu^2} \right) \quad (10.55)$$

which also diverges as $\varepsilon \rightarrow 0$ or $d \rightarrow 4$.

As we have 2 vertices, the one-loop contribution in the Green's function, or rather the self-energy is related to the Feynman integral by $\frac{\lambda^2}{2}$. Hence

$$\Sigma_{\text{bubble}} = \frac{i\lambda^2 \pi^{d/2}}{2(2\pi)^d} \left(\frac{2}{\varepsilon} - \gamma + \ln \frac{m_0^2}{\mu^2} \right) \quad (10.56)$$

where the factor of λ^2 denotes 2 vertices in the bubble diagram.

Chapter 11

Renormalisation I: ϕ^4 theory

Quote 11.1 The infinities are hidden behind the cloud?

Felix Halbwedl, on the watermark of Part IV, 3 March 2025

Fortunately, there are no infinities in real-world physics. Unfortunately, we have not recovered this via regularisation, which only serves as an intermediate step to control divergences by introducing a regulator. To obtain finite, physically meaningful predictions, we must employ *renormalisation*.

11.1 Emergence of renormalisation

The key point lies in realising that quantities depending on λ , including e , m , ψ and A_μ , are actually merely the *tree-level* versions of the terms they *claim* to represent:

- So far, we have mistakenly treated the coupling constant λ as a fixed quantity. This is, in fact, not true: Quantum corrections introduce scale dependence, and λ evolves with energy scales as a result.
- Due to this, quantities we previously employed like the mass m and the charge e , also change in higher orders. This extends to the scalar field ψ , the vector field A_μ and the Green's function G , all of which are dependent on e and m . In higher orders, they are augmented by a series of correction terms, which can be written as a power series of the coupling constant.
- We will now denote the previously seen λ , m , e , ϕ , A and G as λ_0 , m_0 , e_0 , ψ_0 , A_0 and G_0 , which we call the *bare* (read: unrenormalised) quantities.
- The renormalised (read: actual) quantities are then denoted simply as m , e , ψ , A and G , since they are the true quantities.

Remark 11.1 As infinities only emerge in loop integrals, the tree-level part of a propagator is unaffected by renormalisation. At each order, the renormalised counterpart of a bare quantity is different.

Quote 11.2 In practice, we can never measure bare charge or bare mass, because nature always includes the higher order corrections in real world interactions. But the concept of such bare quantities will serve us well in our analyses.

Robert D. Klauber

Importantly, renormalisation does not render regularisation redundant. In fact, the standard way of renormalising a field theory involves regularisation:

- Label the bare quantities with the subscript 0 .
- Identify the divergences (i.e. infinities).
- Regularise the theory¹.

¹We need to do this so that there is actually a way to evaluate our integral

- Replace all instances of the bare coupling constant and mass λ_0 and m_0 with their renormalised counterparts λ and m .
- Remove infinities from the bare quantities using the counterterms. This must be accomplished via specific normalisation schemes.
- Remove the regulator to ‘deregularise’ the theory, e.g.
 - In general, set d to the actual number of physical dimensions.
 - In cutoff regularisation, send $\Lambda \rightarrow \infty$.
 - In dimensional regularisation, send $\varepsilon \rightarrow 0$.

Theorem 11.1 (Renormalisability) If all divergences can be absorbed into a *finite* number of redefined physical parameters^a, the theory is known to be *renormalisable*. If an *infinite* number of counterterms is required, the theory is known to be *non-renormalisable*.

^aEven renormalisable theories may still have infinities at intermediate stages, but these are systematically removed.

11.2 Renormalisability

In addition to our previous loose definition, we want to systematically determine whether a field theory is renormalisable. This can be done in more than one way.

Definition 11.1 (Mass dimension) The so-called *mass dimension* $[Q]$ of some quantity Q is the dimension of mass (or, assuming natural units, energy as well) units in physical units of Q .

The definition above is quite a mouthful, so let us clarify it with a few examples. For a theory in d dimensions:

- The mass dimension $[S]$ of an action S is always zero. As the action is dimensionless, it cannot possibly have units of mass. It can be decomposed to

$$[S] = [\mathcal{L}] + [d^d x] = 0 \quad (11.1)$$

where $[d^d x] = -d$ as $\int d^d x$ integrates over d independent coordinates.

- The mass dimension $[\mathcal{L}]$ of the Lagrangian density \mathcal{L} is always d . This is derived by inserting $[d^d x] = -d$ into (11.1).
- Consider a Lagrangian with only the kinetic term $\mathcal{L} = \frac{\partial^2 \phi^2}{2}$, which gives the expression

$$[\mathcal{L}] = [\partial^2] + [\phi^2] = 2[\partial] + 2[\phi] = 2 + 2[\phi] = d \quad (11.2)$$

The mass dimension $[\phi]$ of a field ϕ is thus

$$[\phi] = \frac{d-2}{2} \quad (11.3)$$

- Now consider a Lagrangian with a series of interacting terms $\mathcal{L} = \lambda_i \phi^{i2}$, which gives the expression

$$[\mathcal{L}] = [\lambda_i] + [\phi^i] = [\lambda_i] + \frac{i(d-2)}{2} = d \quad (11.4)$$

For a specific term in the Lagrangian with i fields, the mass dimension of the term’s corresponding coupling constant λ_i ³ is thus

$$[\lambda] = d - \frac{i(d-2)}{2} \quad (11.5)$$

Take ϕ^4 theory for example. We have $i = 4$ fields. As such, in 4 dimensions:

$$[\lambda] = d - \frac{4(d-2)}{2} = 4 - d = \varepsilon \quad (11.6)$$

where we recall the definition of ε from dimensional regularisation.

²Here, the subscript is an index while the superscript is an exponential.

³This denotes the renormalised coupling constant. As we know, the bare coupling constant is dimensionless. We will return to this later.

Theorem 11.2 (Renormalisability from $[\lambda_i]$) A coupling i in a field theory is renormalisable if $[\lambda_i] \geq 0$. If $[\lambda_i] < 0$, it is non-renormalisable. This means it needs a suppression factor and is treated in an *effective field theory* (EFT) framework.

A second framework is *power counting*, which is a very sophisticated way of saying dimensional analysis:

Definition 11.2 (Superficial degree of divergence) The *superficial degree of divergence* D is a convenient way to determine if a Feynman diagram diverges into infinity. It is the power of the momentum p in the Feynman diagram^a:

$$D = dL - \sum_i (d_i - d)V_i \quad (11.7)$$

where L is the number of (independent) loops (i.e. momentum integrations) in the diagram, V_i is the number of vertices of the vertex type corresponding to the interaction term with i fields, and d_i is the mass dimension of the interaction term associated with vertex type i .

- If $D \geq 0$, the diagram is known to be *superficially divergent*:
 - If $D > 0$, the diagram leads to *logarithmic divergence*.
 - If $D = 0$, the diagram leads to *power law divergence* and requires counterterms.
- If $D < 0$, the diagram is convergent, and no renormalisation is needed.

^ai.e. the momentum power in the numerator minus that in the denominator.

Theorem 11.3 (Renormalisability from power counting) A field theory is renormalisable if $d_i \leq d$:

- Through redefinitions, only a finite number of counterterms are needed.
- The coupling constant is dimensionless.
- Examples are QED and QCD in $d = 4$.

A field theory is super-renormalisable if $d_i < d$:

- This is a subset of renormalisable theories that have only a finite number of divergent Feynman diagrams (usually low-order loops).
- The coupling constant has a positive mass dimension.
- An example is ϕ^3 theory in $d = 6$.

A field theory is non-normalisable if $d_i > d$:

- An infinite number of counterterms are needed.
- As this is not accomplishable, the theory becomes meaningless in high energies (but is acceptable in low energies).
- The coupling constant has a negative mass dimension.
- An example is GR in $d = 4$.

Let us consider this in the context of ϕ^4 theory in 4D, our simplest toy model. In the one-loop order, the 1-point loop (tadpole), the 2-point loop (bubble) and the 4-point loop are all divergent. Conversely, n -point loops with $n > 4$ do not contribute, as they have a negative superficial degree of divergence $D < 0$.

11.3 Counterterms

The bare quantities are related to their renormalised counterparts by the so-called *renormalisation factors* Z_λ , Z_m and Z_ϕ :

Definition 11.3 (ϕ^4 renormalisation factors)

$$\lambda_0 = Z_\lambda \lambda \quad \phi_0 = \sqrt{Z_\phi} \phi \quad m_0^2 = Z_m m^2 \quad (11.8)$$

The ϕ^4 theory Lagrangian (5.1) becomes

$$\mathcal{L} = \frac{Z_\phi}{2} \partial_\mu \phi \partial^\mu \phi + \frac{Z_m Z_\phi}{2} m^2 \phi^2 + \frac{Z_\lambda \lambda Z_\phi^2}{4!} \phi^4 \quad (11.9)$$

The renormalised Feynman propagator, n -point vertex and n -point Green's function are then

$$D_F = \frac{D_{F,0}}{Z_\phi} = \frac{i}{Z_\phi(p^2 - Z_m m^2 + i\epsilon)} \quad \Gamma^{(n)} = Z_\phi^{-n/2} \Gamma_0^{(n)} \quad G^{(n)} = Z_\phi^{-n/2} G_0^{(n)} \quad (11.10)$$

The propagator and vertex can be decomposed. For example, the 2-point propagator function and 4-point vertex function decomposes as

$$\Gamma^{(2)}(p^2) = p^2 - m^2 + \Sigma(p^2) \quad \Gamma^{(4)}(s, t, u) = -i\lambda + \Gamma_{\text{1-loop}}^{(4)}(s, t, u) + \Gamma_{\text{2-loop}}^{(4)}(s, t, u) + \dots \quad (11.11)$$

This decomposition is quite intuitive, as ϕ^4 theory is merely a scalar theory. Its QED counterpart will be much scarier.

Here we acknowledge a certain degree of hand-waviness of the renormalisation parameters:

- All physical quantities are finite.
- All physical quantities are independent of the renormalisation constants Z_i and ϵ (of the $i\epsilon$ fame).

One notable exception to this is the (bare and renormalised) Lagrangian, which is not an observable. With our new friends, the renormalisation factors, the Lagrangian, which we now acknowledge to be actually nothing but the *bare Lagrangian*, can be written as

$$\mathcal{L} = \underbrace{\frac{1}{2} \partial_\mu \phi \partial^\mu \phi - \frac{1}{2} m^2 \phi^2 - \frac{\lambda}{4!} \phi^4}_{\text{renormalised Lagrangian}} + \underbrace{\frac{\delta Z_\phi}{2} \partial_\mu \phi \partial^\mu \phi + \frac{1}{2} \delta m^2 \phi^2 + \frac{\delta \lambda}{4!} \phi^4}_{\text{counterterms}} \quad (11.12)$$

where the δZ_ϕ , δm^2 and $\delta \lambda$ are the so-called *counterterms*⁴:

Definition 11.4 (ϕ^4 counterterms)

$$Z_\phi = 1 + \delta Z_\phi \quad m_0^2 Z_\phi = m^2 + \delta m^2 \quad \lambda_0 Z_\phi^2 = \mu^{-\varepsilon} (\lambda + \delta \lambda) \quad (11.13)$$

Let us justify the counterterms one by one:

- Both Z_ϕ and δZ_ϕ are dimensionless.
- Z_ϕ is dimensionless, while m_0^2 , m^2 and δm^2 have mass dimensions of 2.
- λ_0 is dimensionless, while from (11.6), λ has mass dimension ϵ . To preserve the dimensionlessness of λ_0 , we introduce the term μ^ε , where μ is the so-called *renormalisation scale* and has mass dimension 1.

Thus, we have explicitly established the ultimate goal of renormalisation. To solve for the renormalised quantities or the renormalisation constants, we must calculate the counterterms δ .

Note 11.1 (Running coupling) Importantly, we observe that the renormalised mass and field are independent of the renormalisation scale μ while the renormalised coupling constant is not. The *renormalised* coupling λ is well-known as the *running coupling* as the dependence shows that λ ‘runs’ with μ .

But what does μ physically mean? As it turns out, it is a part of *renormalisation conditions*, which are constraints imposed on the Z_i s. This is actually an umbrella term for many concepts:

⁴We can think of the divergence as rubbish that leaks out during a calculation. We don’t simply throw the rubbish onto the bare parameter directly. Instead, we introduce the counterterm, a trash bin, whose only job is to collect and *cancel* out the garbage. Once that is done, the renormalised result is ‘clean’.

- We choose an arbitrary *renormalisation scale* μ^5 which the counterterms δ are dependent on.
- A corresponding (and likewise arbitrary) *renormalisation point*, which is essentially an artificial constraint we place on μ , is defined.
- Different choices of renormalisation conditions correspond to different *renormalisation schemes*, each of which fixes the counterterms in a specific way.

Some comments on renormalisation schemes should be made:

- Renormalisation schemes are merely mathematical conveniences, and physical variables should not differ among schemes.
- The lack of any technical restrictions on the choice of μ can be exploited: It is often practically convenient to choose μ to be of the same order as the characteristic energy scale of the physical process being studied.
- As different renormalisation schemes are different only by the choice of μ , the different incarnations of a renormalised quantity in different schemes are related to each other by finite constants.
- However, the renormalised Green's function is always the same regardless of the renormalisation scheme. This invariance (i.e. symmetry) gives rise to the *renormalisation group* which we will soon discuss. The renormalisation group also ties up our other loose end, which is the running of the coupling constants.

11.4 Renormalisation schemes

After many, many pages, we are now in a position to actually carry out what we set out to do at the beginning of this chapter - removing the divergence. Now that we have established that our arbitrary choice of μ does not affect the renormalised quantities, we can discuss specific *renormalisation schemes*. In this section, we shall introduce three schemes, which are by design almost always used with dimensional regularisation:

- The *minimal subtraction (MS) scheme*.
- The *modified minimal subtraction (\overline{MS}) scheme*, which is a slightly modified form of the MS scheme.
- The *on-shell (OS) scheme* or the *physical scheme*.

The first question that arises is why dimensional regularisation is the ‘golden boy’ of all three schemes. Previously, we introduced the renormalisation scale μ as a bookkeeping device to make the coupling constants dimensionally correct. The same can be said about the regulator M in dimensional regularisation. As such, we recognise that M is nothing but μ . For a tadpole diagram, this means:

$$I_2 = -\frac{m_0^2}{(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln(4\pi) + \frac{\mu^2}{m_0^2} \right) \quad (11.14)$$

From here on, the three schemes are rather intuitive. The idea is to modify the Feynman integral according to our renormalisation conditions, which ultimately leave the divergence (e.g. $\frac{2}{\varepsilon}$) intact by design. The infinities in the bare term and the counterterm cancel out.

Quote 11.3 -The infinities might cancel each other out.
-Grüß Gott! They just might.

Barclay and Einstein, in ‘The Nth Degree’

Let us first introduce the MS and \overline{MS} schemes, which are highly similar.

Definition 11.5 (MS renormalisation conditions) In the MS scheme, only the divergent part $\frac{2}{\varepsilon}$ (i.e. the infinite pole) is preserved in the counterterm. It is so-called because the minimal possible subtraction is made: *only* the divergent pole $\frac{2}{\varepsilon}$ is deleted from the final renormalised mass.

⁵This is meaningless without the next bullet point.

Definition 11.6 ($\overline{\text{MS}}$ renormalisation conditions) The $\overline{\text{MS}}$ scheme is almost identical to the MS scheme. However, instead of only preserving the infinity in the counterterm, we also preserve the finite constants, removing only the renormalisation scale term $\frac{\mu^2}{m_0^2}$. The $\overline{\text{MS}}$ scheme is often preferred over the MS scheme as the nasty finite constants are also removed.

Now we go onto the OS scheme, which, terrifyingly, does not get rid of the arbitrary renormalisation scale μ from the final result. Instead, all quantities (mass, coupling/charge, etc.) are defined to be the physical observables themselves⁶ through setting the renormalisation point to correspond exactly to the on-shell condition:

Definition 11.7 (OS renormalisation point)

$$p^2 = m^2 \quad (11.15)$$

The implication is that any required scale dependence is effectively hidden inside the physical parameters. Instead of deleting certain terms, we instead define a series of renormalisation conditions around the renormalisation point $p^2 = m^2$:

Definition 11.8 (ϕ^4 theory OS renormalisation conditions)

- **Mass counterterm:** The 2-point propagator vanishes exactly at the physical mass shell

$$\Gamma^{(2)}(p^2 = m^2) = 0 \quad (11.16)$$

This ensures the physical renormalised mass m is the pole of the full propagator, i.e., the physical particle has mass m .

- **Field counterterm:**

$$\left. \frac{d\Gamma^{(2)}(p^2)}{dp^2} \right|_{p^2=m^2} = 1 \quad (11.17)$$

This ensures the propagator has unit residue at the pole, corresponding to a properly normalised one-particle state.

- **Coupling counterterm:** The renormalised 4-point vertex function $\Gamma^{(4)}(p_i)$ is set to equal the physical coupling λ at a particular kinematic configuration. In the on-shell scheme, this is usually:

$$\Gamma^{(4)}(p^2 = m^2) = -i\lambda \quad (11.18)$$

This means that the coupling λ is defined as the physical 2-to-2 scattering amplitude at the symmetric point where all external particles are on-shell and the center-of-mass energy is just at threshold.

The final loose end we need to take care of in the OS scheme is the surviving μ . To tie it up, we once again consult dimensional analysis. For physical intuition, we hence set μ to be simply the mass m , which has the benefit of making logarithm corrections simple:

$$\mu = m \quad (11.19)$$

This brings about another subtlety. Curiously, we note that the renormalised coupling, which depends on μ , now becomes

$$\lambda_0 Z_\phi^2 = m^{-\varepsilon} (\lambda + \delta\lambda) \quad (11.20)$$

which is no longer dependent on μ . The implication is that the coupling $\lambda(\mu)$, which runs with μ in the general case, is no longer running. Instead, it is fixed to m . This is exactly the point of the OS scheme. By making the fixture $\mu = m$, we have now defined the otherwise running coupling to be the actual, physically observed coupling.

11.5 1-loop renormalisation of ϕ^4 theory

Let us get a first taste of what renormalised variables look like:

⁶Hence its alternate name.

Derivation 11.1 (Renormalised field) By applying the renormalisation condition (11.16) to (11.11), the field counterterm can be derived as

$$\delta Z_\phi = - \left. \frac{\partial \Sigma(p^2)}{\partial p^2} \right|_{p^2=m^2} \quad (11.21)$$

where the on-shell condition is accounted for. However, the self-energy is related to the Feynman integral, which is momentum-independent, by $\frac{\lambda}{2}$, and is as such also momentum-independent. This gives

$$\delta Z_\phi = 0 \rightarrow Z_\phi = 1 \quad (11.22)$$

in the tadpole and bubble diagrams. That is to say, in the 1-loop order, the renormalised field is identical to the bare field. This applies to all three schemes. In fact, the simplest correction emerges in the 2-loop order in a sunset diagram.

Derivation 11.2 (Renormalised mass) From (11.22), the mass counterterm relationship for a tadpole diagram reduces nicely to

$$m_0^2 Z_\phi = m_0^2 = m^2 + \delta m^2 \quad (11.23)$$

From mass-energy equivalence (note that momentum is already out of the picture from the renormalised field), the mass counterterm is

$$\delta m^2 = \frac{\lambda}{2} \int \frac{d^d p}{(2\pi)^d} \frac{1}{p^2 - m_0^2 + i\epsilon} + \underbrace{O(\lambda^n)}_{①} \quad (11.24)$$

where only the first-order counterterm is written explicitly, and ① are higher-order counterterms. Treating the first-order counterterm (i.e. tadpole) Feynman integral with dimensional regularisation, we find that it is

$$\delta m^2 = -\frac{m_0^2 \lambda}{2(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln(4\pi) + \frac{\mu^2}{m_0^2} \right) \quad (11.25)$$

- **MS scheme:** The Feynman integral and the first-order counterterm are

$$I_{2,\text{MS}} = \frac{m_0^2}{(4\pi)^{d/2}} \frac{2}{\varepsilon} \quad \delta m_{\text{MS}}^2 = -\frac{m_0^2 \lambda}{(4\pi)^{d/2} \varepsilon} \quad (11.26)$$

Thus, the divergence-free renormalised mass becomes

$$m^2 = m_0^2 + \delta m^2 = m_0^2 - \frac{m_0^2 \lambda}{(4\pi)^{d/2} \varepsilon} \quad (11.27)$$

- **$\overline{\text{MS}}$ scheme:** The Feynman integral and the first-order counterterm are

$$I_{2,\overline{\text{MS}}} = \frac{m_0^2}{(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln 4\pi \right) \quad \delta m_{\overline{\text{MS}}}^2 = -\frac{m_0^2 \lambda}{2(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln 4\pi \right) \quad (11.28)$$

Thus, the renormalised mass is

$$m^2 = m_0^2 + \delta m^2 = m_0^2 - \frac{m_0^2 \lambda}{2(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln 4\pi \right) \quad (11.29)$$

- **OS scheme:** The mass counterterm is fixed to exactly remove the self-energy correction:

$$m_{\text{phys}}^2 = m^2 + \Sigma^{(1)}(m^2) - \delta m^2 = m^2 \quad (11.30)$$

Thus, the Feynman integral and the renormalised mass are

$$I_{2,\text{OS}} = \frac{m_0^2}{(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln 4\pi + \frac{\mu^2}{m_0^2} \right) \quad m^2 = m_0^2 + \delta m^2 = m_0^2 - \frac{m_0^2 \lambda}{2(4\pi)^{d/2}} \left(\frac{2}{\varepsilon} - \gamma + \ln 4\pi + \frac{\mu^2}{m_0^2} \right) \quad (11.31)$$

Derivation 11.3 (Renormalised coupling) Only the 4-point 1-loop (bubble) diagram contributes to the coupling counterterm. We remember that using dimensional regularisation, its integral evaluates as (10.55). The general form of counterterm reads

$$\delta\lambda = \sum_i^{\text{all diagrams}} C_i \text{divergence}_i \quad (11.32)$$

where C is the multiplicity.

At one loop order, the only contributing diagram is the 4-point, 1-loop diagram, which has 3 possible Feynman diagrams corresponding to the s , t and u channels. The counterterm hence reads

$$\delta\lambda = 3I_4 \quad (11.33)$$

- **MS scheme:**

$$\delta\lambda = 3 \frac{i\lambda^2 \pi^{d/2}}{2(2\pi)^d} \frac{2}{\varepsilon} \quad (11.34)$$

- **$\overline{\text{MS}}$ scheme:**

$$\delta\lambda = 3 \frac{i\lambda^2 \pi^{d/2}}{2(2\pi)^d} \left(\frac{2}{\varepsilon} - \gamma \right) \quad (11.35)$$

- **OS scheme:** The condition $\Gamma^{(4)}(p^2 = m^2) = -i\lambda$ makes the fixture $\lambda_{\text{phys}} = \lambda$. Hence

$$\delta\lambda = 3 \frac{i\lambda^2 \pi^{d/2}}{2(2\pi)^d} \left(\frac{2}{\varepsilon} - \gamma + \ln \frac{m_0^2}{\mu^2} \right) \quad (11.36)$$

One problem arises here. We see that different renormalisation schemes give different counterterms:

Aphorism 11.1 (Felix Halbwedi, 22 March 2025) What matters most is the divergent part of the counterterms, they have to agree for all incarnations.

The finite part of the counterterms is not relevant at all, it can be anything finite. The job of the counterterm is to hunt down and kill the divergency inside the bare mass. It cannot retrieve the physical mass held captive by the divergency. There we need to ask Mama nature for a helping hand, and measure the physical mass.

I know, it's mathematically horrible, but in the end it's Mama nature we want to describe.

In this sense, the dependence of the renormalised quantities on the renormalisation scale is also eliminated by measuring quantities physically. This is the central point of the next two sections.

11.6 Callan-Symanzik equation

The creation of our so-called renormalisation scale may not seem immediately satisfying, as we have merely transferred the arbitrariness of δ with an arbitrariness of the renormalisation scheme (i.e. of μ). This apparent contradiction is ultimately reconciled by the so-called *Callan-Symanzik equation*, which makes sure that the theory remains physically meaningful despite the arbitrariness of μ .

Derivation 11.4 (Callan-Symanzik equation)

Quote 11.4 Now we exploit the simple but powerful fact that the bare vertex functions^a $\Gamma_0^{(n)}$ do not know anything about the renormalisation scale μ .

John Cardy

^aFor our example, we use a generic bare Green's function $G_{n,0}$ instead.

As the bare n -point Green's function $G_0^{(n)}$ is independent from the renormalisation scale μ , it is safe to write

$$\mu \partial_\mu G_0^{(n)}(p, \lambda_0) = 0 \quad (11.37)$$

By consulting the part of (11.10) that concerns the Green's function, we can write $G_0^{(n)}$ in terms of Z_ϕ and $G^{(n)}$. Hence, the equation above becomes

$$\mu \partial_\mu (Z_\phi^{n/2}(\lambda, \mu) G^{(n)}(p, \lambda, \mu)) = 0 \quad (11.38)$$

Finally, we use the chain rule:

Theorem 11.4 (Callan-Symanzik equation) To preserve the invariance of physical variables, any *direct* change of Green's function $G_n(p, \lambda, \mu)$ due to the change of μ is compensated by corresponding changes of G_n due to changes of the coupling constant λ and the field ϕ , and the *total* change of G_n arising from μ is zero:

$$\frac{\partial G_n}{\partial \mu} + \frac{\partial \lambda}{\partial \mu} \frac{\partial G_n}{\partial \lambda} + \frac{n}{Z_\phi} \frac{\partial Z_\phi}{\partial \mu} G_n = 0 \quad (11.39)$$

where n is the number of points of the n -point function.

Let us go through the terms one by one:

- The first term is the *direct* dependence of G_n on the renormalisation scale μ .
- The second term is the *implicit* dependence of G_n on μ , but through the running coupling $\lambda(\mu)$, which is ultimately dependent on (or ‘runs’ with) μ . Due to this dependence, $\lambda(\mu)$ is ultimately not a physical quantity^a.
- The third term is a rescaling that accounts for the fact that the field itself is renormalised, and represents the contribution of the field strength renormalisation to the dependence of G_n on μ .
- As the sum of the terms is zero, μ is made irrelevant w.r.t. the physical quantities. We often speak of this as the scale dependence being essentially ‘absorbed’ into physical quantities.

Remark 11.2 As noted in Quote 11.4, it should be immediately obvious that (11.39) can also be derived from $\Gamma^{(n)}$. In fact, the same can be said for the field ϕ . The equations are

$$\frac{\partial \Gamma^{(n)}}{\partial \mu} + \frac{\partial \lambda}{\partial \mu} \frac{\partial \Gamma^{(n)}}{\partial \lambda} - \frac{n}{2Z_\phi} \frac{\partial Z_\phi}{\partial \mu} \Gamma^{(n)} = 0 \quad \frac{\partial \phi}{\partial \mu} + \frac{1}{\phi} \frac{\partial \phi}{\partial \mu} \phi = 0 \quad (11.40)$$

where the running coupling term is missing in the ϕ equation as the field has nothing to do with the coupling constant.

^aRecall from experimental HEP that perturbation theory naturally fails for large coupling constants. Since different choices of μ effectively shift the way we split between ‘low-energy’ and ‘high-energy’ contributions, $\lambda(\mu)$ changes accordingly.

To better illustrate its physical meaning, we can now introduce two functions which are dimensionless and thus depend only on the equally dimensionless λ :

Definition 11.9 (Beta function) The *beta function* $\beta(\lambda)$, which describes how running coupling ‘runs’ with μ :

$$\beta(\lambda) = \frac{\delta \lambda}{\delta \ln \mu} = \mu \frac{\delta \lambda}{\delta \mu} \quad (11.41)$$

where δ is nothing but the functional derivative. This expression is also called the *renormalisation group equations*.

Definition 11.10 (Anomalous dimension) The *anomalous dimension* $\gamma(\lambda)$, which is technically a correction to the *scaling dimension*:

$$\gamma(\lambda) = -\frac{\mu}{Z_\phi} \frac{\delta Z_\phi}{\delta \mu} \quad (11.42)$$

We then have the massless version of (11.39), which reads

$$\left(\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} + n\gamma(\lambda) \right) G_n = 0 \quad (11.43)$$

This establishes a relation between the dependence of G_n on μ and the dependence of G_n on λ .

Note 11.2 (Plot twist) At this point, we note that:

- The total dependence $\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda}$ of the Green's function G_n or the vertex $\Gamma^{(n)}$, explicit *and* implicit, is in fact non-zero and is equal to the field renormalisation term $n\gamma(\lambda)$.
- The presence of the field renormalisation term thus reveals a sinister plot twist. Neither G_n nor $\Gamma^{(n)}$ are physical quantities.

11.7 Renormalisation group

Now that we have eliminated $\lambda(\mu)$ and G_n as unphysical variables, we can finally turn to physical variables.

Derivation 11.5 (Renormalisation group invariance) A physical observable O observes

$$O \sim Z_\phi^{-n} G_n \quad (11.44)$$

where Z_ϕ^{-n} cancels out with the field renormalisation dependence of the Green's function. We thus have

$$\left(\mu \frac{\partial}{\partial \mu} + \beta(\lambda) \frac{\partial}{\partial \lambda} \right) O = 0 \quad (11.45)$$

which implies that the total explicit and implicit dependence of O on μ is zero.

This can be written more concisely. Inserting the full form of the beta function gives

$$\mu \left(\frac{\partial}{\partial \mu} + \frac{d\lambda}{d\mu} \frac{\partial}{\partial \lambda} \right) O = 0 \quad (11.46)$$

Contracting the chain rule, and we have

$$\mu \frac{d}{d\mu} O = 0 \quad (11.47)$$

Finally, we see that a physical observable O is μ -independent.

The μ -invariance illustrated by (11.47) is interesting as it reminds us of symmetries under Lie groups we saw in *Spinors & Symmetries*. For this reason, we often speak of a *renormalisation group*. This is not actually a group, but rather a semigroup as transformations are not necessarily invertible.

Derivation 11.6 (Running coupling) We also want to establish a relation between the dependence of G_n on p and the dependence of G_n on λ . Such a relation can be found in a second alternative form of (11.39), which involves a rescaling of the momentum. Let us begin with the *classical scaling equation*:

$$\left(p \frac{\partial}{\partial p} - n[\phi] + \mu \frac{\partial}{\partial \mu} \right) G_n = 0 \quad (11.48)$$

where we recall that $[\phi]$ is the mass dimension of ϕ .

Subtracting (11.43) from (11.48) gives

$$\left(p \frac{\partial}{\partial p} - \beta(\lambda) \frac{\partial}{\partial \lambda} - n([\phi] + \gamma(\lambda)) \right) G_n = 0 \quad (11.49)$$

This equation can be fitted for our good friend, the running coupling:

$$\left(p \frac{\partial}{\partial p} - \beta(\lambda) \frac{\partial}{\partial \lambda} \right) \lambda(p/\mu) = 0 \quad (11.50)$$

Another note is that this running coupling must be a solution to the beta function equation (11.41),

and an initial condition exists in the form of

$$\lambda(1) = \lambda_0 \quad (11.51)$$

One can thus rewrite the equation as



$$\int_{\lambda_0}^{\lambda(p/\mu)} \frac{\delta\lambda'(p/\mu)}{\beta(\lambda')} = \ln(p/\mu) \quad (11.52)$$

To sum it up, this equation tells us that a change in p inevitably induces a rescaling of λ and a rescaling of ϕ by proxy of $\gamma(\lambda)$:

- For $\beta(\lambda) > 0$, p rises and falls as λ does.
- For $\beta(\lambda) < 0$, p rises as λ falls and falls as λ rises.
- For $\beta(\lambda) = 0$, we label $\lambda|_{\beta(\lambda)=0} := \lambda^*$.

The idea of the *renormalisation group flow* lies in the change of the coupling constant λ due to the change in the beta function $\beta(\lambda)$ ⁷. In other words, it describes how $\beta(\lambda)$ ‘flows’ with λ . This is captured in (11.43), the massless Callan-Symanzik equation. Disregarding the anomalous dimension $\gamma(\lambda)$, we see that

- For $\beta(\lambda) > 0$, the RG flow is towards larger values of λ :
 - The IR singularity happens when $\lambda \rightarrow 0$ and $\beta(\lambda) \rightarrow 0$. Due to the weak λ , perturbation theory is well-suited for the IR region. This is known as an *IR stable fixed point*.
 - The UV singularity happens when λ is *large* and positive. $\beta(\lambda)$ is thus *large* and positive. Due to the weak λ , perturbation theory fails in the UV region.
- For $\beta(\lambda) < 0$, the RG flow is towards smaller values of λ :
 - The UV singularity happens when $\lambda \rightarrow 0$ and $\beta(\lambda) \rightarrow 0$. Due to the weak λ , perturbation theory is well-suited for the UV region. This is known as an *UV stable fixed point*.
 - The IR singularity happens when λ is *large* and positive. $\beta(\lambda)$ is thus *large* and negative. Due to the weak λ , perturbation theory fails in the IR region.
- For $\beta(\lambda) = 0$, we consult (11.49):
 - If $\beta(\lambda)$ goes from negative to positive through λ^* , we will see the momentum approaching zero as $\lambda \rightarrow \lambda^*$:

$$p \rightarrow 0 \quad (11.53)$$

This is an *IR stable zero*.

- If $\beta(\lambda)$ goes from positive to negative through λ^* , we will see the momentum approaching infinity as $\lambda \rightarrow \lambda^*$:

$$p \rightarrow \infty \quad (11.54)$$

This is an *UV stable zero*.

This is an interesting case, as we can extract the behaviour of the Green’s functions by setting $\lambda \rightarrow \lambda^*$ in (11.49).

11.8 Källén-Lehmann spectral representation

Finally, it is useful to reflect upon what we have gone through in a more physical way. The default form of a renormalised tree-level Feynman propagator with mass m in (11.10) should be well-known to the reader at this point. However, we have yet to expand this to higher orders.

As it turns out, there is a default way to represent a propagator of arbitrary order called the *Källén-Lehmann spectral representation*, where the full propagator *in momentum space* is solved via integrating over the mass parameter/renormalisation scale μ :

⁷Which itself dictates how λ evolves with the renormalisation scale μ .

Definition 11.11 (Källén-Lehmann spectral representation)

$$G = i \int_0^\infty d\mu^2 \frac{\rho(\mu^2)}{p^2 - m^2 + i\epsilon} \quad (11.55)$$

Note 11.3 The Källén-Lehmann spectral representation does not actually assist with calculating counterterms. Rather, it shows that interacting propagators can always be expressed as a weighted sum of free propagators, which provides a non-perturbative insight into the structure of the theory.

The central point of this formalism is as follows:

- Effectively, the tree-level propagator can be regarded as a free particle, regardless of whether the theory itself is free or interacting.
- Higher-order propagators, which are effectively quantum corrections, represent the interacting parts of an (interacting) theory.

We can show this formulaically. For a free particle with mass m , the spectral function is

$$\rho(\mu^2) = Z_\phi^{-1} \delta(\mu^2 - m^2) \quad (11.56)$$

This ρ fixes μ to m , and (11.55) reduces to (11.10).

For interacting fields, our propagator under the Källén-Lehmann spectral representation can be rewritten as

$$G = \underbrace{\frac{i}{Z_\phi(p^2 - m^2 + i\epsilon)}}_{\text{free particle term}} + \underbrace{i \int_0^\infty d\mu^2 \frac{\sigma(\mu^2)}{p^2 - m^2 + i\epsilon}}_{\text{continuum terms}} \quad (11.57)$$

where the continuum terms are interacting field contributions from multi-particle states. Hence, the function $\sigma(\mu^2)$ is defined to be explicitly the interacting contribution:

$$\sigma(\mu^2) = \rho(\mu^2) - Z_\phi^{-1} \delta(\mu^2 - m^2) \quad (11.58)$$

Remark 11.3 Hence, the physical meaning of (11.55) and (more obviously) () is seen. It describes the full propagator, representing the high-order quantum corrections that are deviations from the free propagator in (11.10).

We end with some mathematical discussions. Intuitively, both $\rho(\mu^2)$ and $\sigma(\mu^2)$ are positive, the first of which implies that, for some order n of differentiation:

$$\frac{\partial^n D(-p^2)}{\partial (p^2)^n} = (-1)^n i \int_{m_t^2}^\infty d\mu^2 \frac{\rho(\mu^2)}{(p^2 + \mu^2)^n} \quad (11.59)$$

Inserting the decomposed (11.3) into this gives

$$Z_\phi^{-1} + \int_{m_t^2}^\infty d\mu^2 \sigma(\mu^2) = 1 \quad (11.60)$$

As $\sigma(\mu^2) > 0$, we have $0 \leq Z_\phi^{-1} < 1$.

Chapter 12

Renormalisation II: QED

Quote 12.1 Miew. Mowem Purrum.

Felix Halbwedl, 8 March 2025

We are finally in a position to calculate the QED counterterms. Fortunately, the conceptual ideas of renormalisation have already been detailed in the last chapter. The MS, $\overline{\text{MS}}$ and OS schemes we have introduced also apply to QED. All we have to do is to develop a corresponding framework for vector fields.

12.1 Ward-Takahashi identity

With the renormalisation of QED, we are now again in the realm of gauge theories. As it turns out, gauge invariance affects renormalisation factors in some funny ways. This can be verified by going over the Slavnov-Taylor, Ward-Takahashi and Ward identities, which are the final development from applying Noether's theorem in QFT that we will see in a while.

It is expedient to clarify the context that each identity live in, and how the identities relate to each other:

- The Slavnov-Taylor identity applies to the most general case, which is non-Abelian gauge theories like QCD.
- The Ward-Takahashi identity is a reduction of the Slavnov-Taylor identity in the Abelian limit, like QED.
- The Ward identity is a reduction of the Ward-Takahashi identity in the limit when the external fermions are on-shell and when momenta transfer is zero¹.

As we have not reached non-Abelian gauge theories yet, we first derive the Ward-Takahashi identity.

Derivation 12.1 (Ward-Takahashi identity) Beginning with the QED generating functional

$$Z[\bar{J}, J, J^\mu] = \int \mathcal{D}\bar{\psi} \mathcal{D}\psi \mathcal{D}A_\mu \exp \left[i \int d^4x (\mathcal{L}_{\text{QED}} + \bar{J}\psi + \bar{\psi}J + J^\mu A_\mu) \right] \quad (12.1)$$

where we recall J and \bar{J} to be the fermion source terms, J^μ to be the boson source term, and that there exist the shorthands

$$\mathcal{L}_{\text{QED}} = \bar{\psi}(i\gamma^\mu D_\mu - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} \quad D_\mu = \partial_\mu + ieA_\mu$$

It is again the Dyson-Schwinger equations that save the day. We recall that the generating functional is invariant under local gauge transformations, which, in QED, is our good friend, the U(1) symmetry:

$$\psi \rightarrow (1 + i\alpha(x))\psi, \quad \bar{\psi} \rightarrow \bar{\psi}(1 - i\alpha(x)), \quad A_\mu \rightarrow A_\mu - \frac{1}{e}\partial_\mu\alpha(x)$$

¹This is also called the *soft photon limit*.

Let us now solve for the variation of the generating functional. Intuitively, the variations of the fields read

$$\delta\psi(x) = i\alpha(x)\psi(x) \quad \delta\bar{\psi}(x) = -i\alpha(x)\bar{\psi}(x) \quad (12.2)$$

Under this, the action varies as

$$\delta S = \int d^4x \alpha(x) \partial_\mu j^\mu(x) \quad (12.3)$$

and the source terms vary as

$$\delta(\bar{J}\psi + \bar{\psi}J) = i\alpha(x)\bar{J}(x)\psi(x) - i\alpha(x)\bar{\psi}(x)J(x) \quad (12.4)$$

Putting it all together, the total variation of the integrand on the exponential is, to first order in α :

$$\delta \left[S + \int d^4x (\bar{J}\psi + \bar{\psi}J) \right] = \int d^4x \alpha(x) [\partial_\mu j^\mu(x) + i\bar{J}(x)\psi(x) - i\bar{\psi}(x)J(x)] \quad (12.5)$$

Hence, the variation of the generating functional reads

$$\delta Z = i \int \mathcal{D}\bar{\psi} \mathcal{D}\psi e^{iS+i \int \bar{J}\psi + \bar{\psi}J} \int d^4x \alpha(x) [\partial_\mu j^\mu(x) + i\bar{J}(x)\psi(x) - i\bar{\psi}(x)J(x)] \quad (12.6)$$

Again, consulting the Dyson-Schwinger equations, we know that $\delta Z = 0$. We can then divide both sides by $i \int \mathcal{D}\bar{\psi} \mathcal{D}\psi e^{iS+i \int \bar{J}\psi + \bar{\psi}J}$: Factor out the $\alpha(x)$ and write in terms of expectation values:

$$\int d^4x \alpha(x) [\partial_\mu \langle j^\mu(x) \rangle + \bar{J}(x) \langle \psi(x) \rangle - \langle \bar{\psi}(x) \rangle J(x)] = 0 \quad (12.7)$$

Let us consider the simplest vertex, which is a three-point function with two fermions and one photon, where the fermion absorbs/emits a photon of momentum q^a . According to (??), taking functional derivatives of the generating functional with respect to J and \bar{J} gives us the Green's function:

$$\partial_\mu^x G^\mu(x, y, z) = \partial_\mu \langle T[j^\mu(x)\psi(y)\bar{\psi}(z)] \rangle = \delta(x-y) \langle T[\psi(y)\bar{\psi}(z)] \rangle - \delta(x-z) \langle T[\psi(y)\bar{\psi}(z)] \rangle \quad (12.8)$$

where x, y and z are the 4-positions of the three legs.

We take Fourier transforms of both sides. For each component, this works as

$$\begin{aligned} j^\mu(q) &= \int d^4x e^{iq \cdot x} j^\mu(x) \quad \psi(p+q) = \int d^4y e^{i(p+q) \cdot y} \psi(y) \\ \bar{\psi}(p) &= \int d^4z e^{-ip \cdot z} \bar{\psi}(z) \quad \langle T[\psi(y)\bar{\psi}(z)] \rangle = \int \frac{d^4p}{(2\pi)^4} e^{-ip \cdot (y-z)} S_F(p) \end{aligned} \quad (12.9)$$

where p is the total (4-)momentum of the incoming fermion, and q is the momentum of the incoming/outgoing photon^b.

Putting it all together, we see that the delta functions are eliminated:

$$\int d^4x e^{iqx} \partial_\mu \langle T[j^\mu(x)\psi(y)\bar{\psi}(z)] \rangle = e^{iqy} S_F(y-z) - e^{iqz} S_F(y-z) \quad (12.10)$$

This reduces to

$$iq^\mu \langle j^\mu(q)\psi(p+q)\bar{\psi}(p) \rangle = S_F(p) - S_F(p+q) \quad (12.11)$$

where From the OPI decomposition, we can rewrite the LHS in terms of the vertex function $\Gamma^\mu(p+q, p)$:

$$\langle j^\mu(q)\psi(p+q)\bar{\psi}(p) \rangle = S_F(p+q)\Gamma^\mu(p+q, p)S_F(p) \quad (12.12)$$

Inserting this gives

$$iq^\mu S_F(p+q)\Gamma_\mu(p+q, p)S_F(p) = S_F(p) - S_F(p+q) \quad (12.13)$$

Dividing both sides by $S_F^{-1}(p+q)S_F^{-1}(p)$, we find the common form of the Ward-Takahashi identities:

Theorem 12.1 (Ward-Takahashi identity)

$$iq^\mu \Gamma_\mu(p+q, p) = S_F^{-1}(p+q) - S_F^{-1}(p) \quad (12.14)$$

where $\Gamma_\mu(p+q, p)$ is the full (ampurared) vertex function, $S_F(p+q)$ is the full (interacting, renormalised) fermion propagator and $S_F(p)$ is the same propagator but lacking one external photon.

The Ward-Takahashi identity is also called the *rigid identity*, as *rigid symmetries* is an alternative name for global symmetries.

^aNote that this is effectively a momentum transfer by q on the fermion.

^bNote that the sign is automatically taken care of if we know whether it is incoming or outgoing.

Quote 12.2 Because the Ward-Takahashi identity relates the $n + 1$ -point function with the n -point function, which is useful if you want to construct the most general 3-points function.

Felix Halbwedl, on why Feynman diagrams are employed in the Ward-Takahashi identity, 5 February 2025

Some physical comments can be made:

- The identity is a non-perturbative statement as no perturbation theory is used. It holds beyond just tree level.
- While we used the simplest case of the fermion-photon propagator in our derivation, the final identity (12.14) actually applies to all vertices in QED. That is to say, we can use this identity for any vertex in QED that has at least one external photon involved (we will not prove this).
- In this case, we only consider (or ‘zoom into’) the fermion that the photon operates on, which has the incoming and outgoing full propagators $S_F(p)$ and $S_F(p+q)$, and all other fermions or photons in the vertex are ignored. $\Gamma_\mu(p+q, p)$ is still the full vertex, which can now be better written as $\Gamma_\mu(\dots, p+q, p)$ where \dots are the momenta of all other particles we have ignored.

In conclusion, the key takeaway from this identity is twofold:

- Following directly from current conservation and Noether’s theorem, there is, when a fermion emits or absorbs one photon, a correlation between the pre- and post-emission/absorption full fermion propagators with the product of the photon momentum and the full vertex function.
- The longitudinal part of the (3-point) vertex function (i.e. the part proportional to q^μ) is completely determined by the difference in the fermion self-energies (which are 2-point functions).

Note 12.1 (Sneak peek of gauge theory) The Ward-Takahashi identity is actually a specific form of the Slavnov-Taylor identity, which is used in the (most general case of) non-Abelian gauge theories we will soon discuss, like QCD. The Slavnov-Taylor identity is derived from the BRST symmetry of the gauge-fixed Lagrangian.

$$q^\mu \Gamma_\mu(p+q, p) = S_F^{-1}(p+q)G(p+q, p) - G(p+q, p)S_F^{-1}(p) \quad (12.15)$$

where $G(p+q, p)$ is the ghost-fermion scattering kernel. This term disappears in the Abelian limit, and we are left with the Ward-Takahashi identity.

12.2 Ward identity

Let us prove that the Ward-Takahashi identity reduces to the Ward identity in the long wavelength (IR singularity) case, where the vertex function is on-shell and when momentum transfer is zero ($q \rightarrow 0$).

Derivation 12.2 (Ward idnetity) As we are concerned with the $q \rightarrow 0$ limit, it is legal to perform a Taylor expansion of the RHS around $q = 0$

$$S_F^{-1}(p+q) - S_F^{-1}(p) \approx q^\mu \frac{\partial S_F^{-1}(p)}{\partial p^\mu} \quad (12.16)$$

Substituting this into the Ward-Takahashi identity, we find

$$q^\mu \Gamma_\mu(p, p) = q^\mu \frac{\partial S_F^{-1}(p)}{\partial p^\mu} \quad (12.17)$$

Eliminating the common q^μ :

Theorem 12.2 (Ward identity)

$$\Gamma_\mu(p, p) = \frac{\partial S_F^{-1}(p)}{\partial p^\mu} \quad (12.18)$$

This is the generalised form of the Ward identity.

We can derive a few useful variants of the Ward identity:

Derivation 12.3 (Differential form of the Ward idnetity) From the known relation

$$S_F(p) = [S_F^{-1}(p)]^{-1} \quad (12.19)$$

We can use the product rule and find

$$\frac{\partial S_F(p)}{\partial p^\mu} = -S_F(p) \left(\frac{\partial S_F^{-1}(p)}{\partial p^\mu} \right) S_F(p) \quad (12.20)$$

Inserting this into (12.18), we find

$$\frac{\partial S_F(p)}{\partial p^\mu} = -S_F(p) \Gamma_\mu(p) S_F(p) \quad (12.21)$$

which is the differential form of the Ward identity. This is the incarnation of the Ward identity that Yasushi Takahashi started with when he derived the Ward-Takahashi identity in 1957.

Derivation 12.4 (Scattering amplitude form of the Ward idnetity) Consider a process where an *external* photon with momentum q^μ is emitted from an external charged fermion line. The scattering amplitude reads

$$\mathcal{M}^\mu(q) = \bar{u}(p') \Gamma^\mu(p', p) u(p) \quad (12.22)$$

Let us apply the external photon momentum q_μ on both sides:

$$q_\mu \mathcal{M}^\mu(k) = q_\mu \bar{u}(p') \Gamma^\mu(p', p) u(p) \quad (12.23)$$

Using (12.14), we see that

$$q_\mu \mathcal{M}^\mu(k) = \bar{u}(p') [S_F^{-1}(p') - S_F^{-1}(p)] u(p) \quad (12.24)$$

When the external fermions are on-shell, we know that the inverse propagator annihilates on-shell spinors. As such

$$S_F^{-1}(p) u(p) = 0 \quad \bar{u}(p') S_F^{-1}(p') = 0 \quad (12.25)$$

We hence find the all-too-familiar form of the Ward identity

$$q_\mu \mathcal{M}^\mu(q) = 0 \quad (12.26)$$

This essentially reflects current conservation in QED, or equivalently, gauge invariance of the S -matrix. Let us now investigate this external photon. As it is external, there are only two transverse polarisations, whose 4-versors are, by convention:

$$\epsilon_\mu^1 = (0, 1, 0, 0) \quad \epsilon_\mu^2 = (0, 0, 1, 0) \quad (12.27)$$

As the matrix elements are those of the external photon, they project onto the polarisation v兽or states. As such, $\mathcal{M}^\mu(q)$ also has only two non-zero components^a - $\mathcal{M}^1(q)$ and $\mathcal{M}^2(q)$. From (12.18), the momentum hence has

$$q_\mu = (k_0, 0, 0, -k_3) \quad q^\mu = (k_0, 0, 0, k_3) \quad (12.28)$$

where the transverseness of the two polarisations is seen.

^aYes! $\mathcal{M}^\mu(q)$ is a vector! Bet you didn't catch that, did you?

12.3 Renormalisation constants

Let us begin with our good friend, the gauge-fixed Lagrangian (6.163) and make a cosmetic replacement of the gauge 'coupling' ξ in the gauge-fixing term

$$\frac{1}{\xi} = \lambda \quad \frac{1}{2\xi} A^\mu \partial_\mu \partial_\nu A^\nu = \frac{\lambda}{2} A^\mu \partial_\mu \partial_\nu A^\nu \quad (12.29)$$

Again, we acknowledge the fact that the quantities here are actually bare quantities:

$$\mathcal{L}_{\text{QED}} = \bar{\psi}_0 (i\gamma^\mu \partial - m_0) \psi_0 - \frac{1}{2} A_0^\mu (\square g_{\mu\nu} - \partial_\mu \partial_\nu) A_0^\nu - e_0 \bar{\psi}_0 \gamma^\mu A_{\mu,0} \psi_0 + \frac{\lambda_0}{2} A_0^\mu \partial_\mu \partial_\nu A_0^\nu \quad (12.30)$$

We can now write down the conventional renormalisation factors, some of which are defined slightly differently from those in ϕ^4 theory. To begin with, we start with the provisional set of five renormalisation factors

$$\psi_0 = \sqrt{Z_\psi} \psi \quad A_0 = \sqrt{Z_A} A \quad m_0 = Z_m m \quad e_0 = Z_e e \quad \lambda_0 = Z_\lambda \lambda \quad (12.31)$$

Let us define another three provisional, so-called, renormalisation factors:

$$S_{F,0} = Z_S S_F \quad D_{F,0}^{\mu\nu} F = Z_D D_F^{\mu\nu} \quad \Gamma_0 = Z_\Gamma^{-1} \Gamma \quad (12.32)$$

It should be intuitively obvious that we can represent Z_S , Z_D and Z_Γ ² via the original five renormalisation factors, which we will prove now.

Derivation 12.5 (Propagators) The fermion and photon propagators are defined as the following two-point functions

$$S_F(x - y) = \langle 0 | T\{\psi(x)\bar{\psi}(y)\} | 0 \rangle \quad (12.33)$$

$$D_F^{\mu\nu}(x - y) = \langle 0 | T\{A^\mu(x) A^\nu(y)\} | 0 \rangle \quad (12.34)$$

Expressing the renormalised quantities in terms of their bare counterparts and renormalisation factors:

$$S_{F,0}(x - y) = \langle 0 | T\{\psi_0(x)\bar{\psi}_0(y)\} | 0 \rangle = Z_\psi \langle 0 | T\{\psi(x)\bar{\psi}(y)\} | 0 \rangle = Z_\psi S_F(x - y) \quad (12.35)$$

$$D_{F,0}^{\mu\nu}(x - y) = Z_A \langle 0 | T\{A^\mu(x) A^\nu(y)\} | 0 \rangle = Z_A D_F^{\mu\nu}(x - y) \quad (12.36)$$

Hence, we find that

$$Z_S = Z_\psi \quad Z_D = Z_A \quad (12.37)$$

Derivation 12.6 (Vertex) Z_Γ is a bit trickier and requires us to use the Ward-Takahashi identity. Let us begin with the Ward-Takahashi identity with renormalised quantities

$$q_\mu \Gamma^\mu(p + q, p) = S^{-1}(p + q) - S^{-1}(p)$$

Writing the renormalised quantities in their bare counterparts and renormalised factors give

$$q_\mu Z_\Gamma \Gamma_0^\mu(p + q, p) = Z_\psi [S_0^{-1}(p + q) - S_0^{-1}(p)] \quad (12.38)$$

However, we note that the Ward-Takahashi identity also holds for bare quantities:

$$q_\mu \Gamma_0^\mu(p + q, p) = S_0^{-1}(p + q) - S_0^{-1}(p) \quad (12.39)$$

²Which, as seen above, is defined differently from the other two renormalisation factors. This becomes convenient in calculating the counterterms.

Dividing (12.39) by (12.38) gives us

$$Z_\Gamma = Z_\psi \quad (12.40)$$

Remark 12.1 Often, Z_Γ , Z_ψ and Z_A are written as Z_1 , Z_2 and Z_3 due to having been so-denoted in the original 1950s literature.

This is not the end of our troubles, because as it turns out, even the original five renormalisation factors relate to each other.

Derivation 12.7 ($Z_\lambda Z_A = 1$) This can be proven by looking at the gauge fixing term. Writing all the terms explicitly gives

$$\frac{\lambda_0}{2} (\partial_\mu A_0^\mu)^2 = \frac{Z_\lambda \lambda}{2} \left(\partial_\mu \left(Z_A^{1/2} A^\mu \right) \right)^2 \quad (12.41)$$

This is simply

$$\frac{\lambda_0}{2} (\partial_\mu A_0^\mu)^2 = Z_\lambda Z_A \frac{\lambda}{2} (\partial_\mu A^\mu)^2 \quad (12.42)$$

At this point, we note that the gauge fixing term has a unique property: its bare and renormalised forms are *always* identical. Hence

$$Z_\lambda Z_A \frac{\lambda}{2} (\partial_\mu A^\mu)^2 = \frac{\lambda}{2} (\partial_\mu A^\mu)^2 \quad (12.43)$$

from which we can immediately read off

$$Z_\lambda Z_A = 1 \quad (12.44)$$

Derivation 12.8 ($Z_e Z_A^{1/2} = 1$) This can be proven by inspecting the interaction term. Again, we write all terms explicitly

$$ie_0 \bar{\psi}_0 \gamma^\mu \psi_0 A_{0,\mu} = i(Z_e e)(Z_\psi^{1/2} \psi)^\dagger \gamma^\mu (Z_\psi^{1/2} \psi)(Z_A^{1/2} A_\mu) = Z_e Z_\psi Z_A^{1/2} (ie \bar{\psi} \gamma^\mu \psi A_\mu) \quad (12.45)$$

Recall that the interaction term represents the vertex function. As such, this actually describes vertex renormalisation

$$ie_0 \bar{\psi}_0 \gamma^\mu \psi_0 A_{0,\mu} = Z_\Gamma (ie \bar{\psi} \gamma^\mu \psi A_\mu) = Z_e Z_\psi Z_A^{1/2} (ie \bar{\psi} \gamma^\mu \psi A_\mu) \rightarrow Z_\phi = Z_e Z_\psi Z_A^{1/2} \quad (12.46)$$

from which we can immediately read off

$$Z_e Z_A^{1/2} = 1 \quad (12.47)$$

We are now finally in a position to state all the QED renormalisation factors:

Definition 12.1 (QED renormalisation factors)

- **Variables:**

$$\psi_0 = \sqrt{Z_\psi} \psi \quad A_0 = \sqrt{Z_A} A \quad m_0 = Z_m m \quad e_0 = Z_A^{-1/2} e \quad \lambda_0 = Z_A^{-1} \lambda \quad (12.48)$$

- **Propagators and vertex:**

$$S_{F,0} = Z_\psi S_F \quad D_{F,0}^{\mu\nu} F = Z_A D_F^{\mu\nu} \quad \Gamma_0 = Z_\psi \Gamma \quad (12.49)$$

12.4 Tensor decomposition

In QED, the full forms of the fermion propagator, photon propagator and vertex can still be found by summing the loop contributions. This reads

$$iS^{-1}(p) = iS_0^{-1}(p) + \Sigma(p) \quad i(D^{-1})^{\mu\nu}(q) = i(D_0^{-1})^{\mu\nu}(q) + \Pi^{\mu\nu}(q) \quad \Gamma^\mu(p, q) = \Gamma_0^\mu(p, q) + \Lambda^\mu(p, q) \quad (12.50)$$

where $\Sigma(p)$ is the fermion self-energy, $\Pi^{\mu\nu}(q)$ is the photon vacuum polarisation and $\Lambda^\mu(p, q)$ is the vertex correction. We are now interested in formulating an expression for these three terms, which will involve decomposing them. Unlike in (11.11), the fermion and photon propagators and the vertex in QED all

have tensorial structures, which forces us to make use of the so-called *tensor decomposition*. Let us go through the terms one by one.

Derivation 12.9 (Fermion propagator) We have two terms, the free (bare) propagator and the self-energy. The free propagator, which reads $S_{F,0}(p) = \frac{i}{\not{p} - m + i\epsilon}$, decomposes as

$$iS_{F,0}^{-1}(p) = \not{p} - m \quad (12.51)$$

Now we investigate self-energy. To preserve Lorentz covariance, the only possible tensor structures are \not{p} and \mathbb{I} . The independent variable becomes p^2 , which is Lorentz-invariant:

$$\Sigma(p) = [1 - Z_2^{-1}(p^2)] \not{p} - [Z_2^{-1}(p^2)Z_m(p^2) - 1] m \quad (12.52)$$

For ease of reading, we often conventionally define the *fermion wavefunction renormalisation function* $A(p^2)$ and *fermion mass function* $M(p^2)$, which are nothing but shorthands:

$$\Sigma(p) = A(p^2)\not{p} + B(p^2)m \quad (12.53)$$

Substituting them into (12.50) the full inverse propagator decomposition:

$$iS_F^{-1}(p) = \not{p} - m + A(p^2)\not{p} + B(p^2)m = [1 + A(p^2)] \not{p} - [1 - B(p^2)] m \quad (12.54)$$

$$iS_F^{-1}(p) = A(p^2)(\not{p} - \Sigma_M(p^2)\mathbb{I}) = A(p^2)(\not{p} - \Sigma_M(p^2)) \quad (12.55)$$

Inverting this expression returns us

$$S_F(p) = \frac{i}{[1 + A(p^2)] \not{p} - [1 - B(p^2)] m} \quad (12.56)$$

Conversely, we can define the *wavefunction renormalisation function* $Z(p^2)$ and the *momentum-dependent effective mass* $M(p^2)$:

$$Z(p^2) = \frac{1}{1 + A(p^2)} \quad M(p^2) = \frac{1 - B(p^2)}{1 + A(p^2)} m \quad (12.57)$$

This gives us a form similar to the bare propagator

$$S_F(p) = \frac{iZ(p^2)}{\not{p} - M(p^2)} \quad (12.58)$$

Derivation 12.10 (Photon propagator) The photon propagator has rank 2, and the only possible structures are the metric $g^{\mu\nu}$ and the momenta product $q^\mu q^\nu$.

Let us again investigate the terms one by one. In the Feynman gauge, the free photon propagator is

$$D_0^{\mu\nu}(q) = \frac{-ig^{\mu\nu}}{q^2 + i\epsilon} \quad (12.59)$$

Its inverse with a factor of i is then

$$i(D_0^{-1})^{\mu\nu}(q) = q^2 g^{\mu\nu} \quad (12.60)$$

Previously in the Ward identity, we have seen that $\Pi^{\mu\nu}(q)$ must be transverse. As such, it takes the form

$$\Pi^{\mu\nu}(q) = (q^2 g^{\mu\nu} - q^\mu q^\nu) \Pi(q^2) \quad (12.61)$$

where $\Pi(q^2)$ is a scalar function known as the *vacuum polarisation scalar*.

Substituting into the inverse propagator, we find

$$i(D^{-1})^{\mu\nu}(q) = q^2 g^{\mu\nu} + (q^2 g^{\mu\nu} - q^\mu q^\nu) \Pi(q^2) = q^2 \left[g^{\mu\nu} + \left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2} \right) \Pi(q^2) \right] \quad (12.62)$$

Inverting this again in the Feynman gauge gives

$$D^{\mu\nu}(q) = \frac{-i}{q^2 [1 + \Pi(q^2)]} \left(g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2} \right) + \lambda^{-1} \frac{-iq^\mu q^\nu}{(q^2)^2} \quad (12.63)$$

We can write a shorthand of this by defining the transverse part of the inverse photon propagator:

$$T^{\mu\nu} = g^{\mu\nu} - \frac{q^\mu q^\nu}{q^2} \quad (12.64)$$

which gives the decomposition as

$$D^{\mu\nu}(q) = \frac{-i}{q^2 [1 + \Pi(q^2)]} T^{\mu\nu} + \lambda^{-1} \frac{-iq^\mu q^\nu}{(q^2)^2} \quad (12.65)$$

Derivation 12.11 (Vertex) As per the Feynman rules, the tree-level vertex factor is $-ie\Gamma^\mu = -ie\gamma^\mu$, making the tree-level contribution to the vertex simply the gamma matrices γ^μ . Denoting loop corrections as $\Lambda^\mu(p', p)$, we then have the full, corrected vertex

$$\Gamma^\mu(p', p) = \gamma^\mu + \Lambda_{1\text{-loop}}^\mu(p', p) + \Lambda_{2\text{-loop}}^\mu(p', p) + \dots \quad (12.66)$$

where the 1-loop correction reads, for free indices α and β

$$\Lambda^\mu(p', p) = (-ie)^2 \int \frac{d^4 k}{(2\pi)^4} \gamma^\alpha \frac{i}{k-m} \gamma^\mu \frac{i}{k+q-m} \gamma^\beta \left(\frac{-ig_{\alpha\beta}}{(k-p)^2} \right) \quad (12.67)$$

and so on.

This decomposition is often given a more systematic treatment. Ultimately, we construct Γ^μ from available vectors and matrices: vectors p^μ , $p'^\mu = p^\mu + q^\mu$ and q^μ ; Dirac matrices: γ^μ and $\sigma^{\mu\nu} = \frac{i}{2}[\gamma^\mu, \gamma^\nu]$. Finally, we have scalar functions $F_i(q^2)$ known as *form factors*, which depend only on the Lorentz scalar q^2 .

$$\Gamma^\mu(p, q) = F_1(q^2) \gamma^\mu + F_2(q^2) i \sigma^{\mu\nu} q_\nu + F_3(q^2) q^\mu \quad (12.68)$$

where:

- $F_1(q^2)$ is the *Dirac form factor* concerned with charge and normalisation.
- $F_2(q^2)$ is the *Pauli form factor* concerned with anomalous magnetic moment.
- $F_3(q^2)$ which vanishes due to the Ward identity (12.18).

Due to the vanishing of $F_3(q^2)$, the physically relevant vertex decomposition is

$$\Gamma^\mu(p, q) = F_1(q^2) \gamma^\mu + F_2(q^2) \frac{i}{2m} \sigma^{\mu\nu} q_\nu \quad (12.69)$$

This is identical to (12.66). To see this, we consider the tree-level case, where $F_1(q^2) = 1$ and $F_2(q^2) = 0$. This recovers the tree-level vertex:

$$\Gamma^\mu(p, q)|_{\text{tree}} = \gamma^\mu = \Gamma_0^\mu(p, q) \quad (12.70)$$

The same idea follows at higher orders. The equivalence lies in the fact that the correction $\Lambda^\mu(p', p)$ for each order can be recovered by setting $F_1(q^2)$ and $F_2(q^2)$ to their specific incarnation in that order.

12.5 1-loop renormalisation of QED

Like in ϕ^4 theory, we define the field counterterms w.r.t. the *renormalisation constants*, and the mass and charge counterterms w.r.t. the *physical quantities themselves*. We also define a vertex counterterm δZ_Γ so that the Lagrangian can be written fully in terms of bare and counterterms:

Definition 12.2 (QED counterterms)

$$Z_\Gamma = 1 + \delta Z_\Gamma \quad Z_\psi = 1 + \delta Z_\psi \quad Z_A = 1 + \delta Z_A \quad m_0 Z_\psi = m^2 + \delta m^2 \quad (12.71)$$

From the Ward identity, we know that $Z_\Gamma = Z_\psi$. It is hence easy to see that

$$\delta Z_\Gamma = \delta Z_\psi \quad (12.72)$$

We get two counterterms for the price of one!

The bare Lagrangian then reads

$$\mathcal{L}_{\text{bare}} = \underbrace{-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(iD - m)\psi - e\bar{\psi}\gamma^\mu\psi A_\mu}_{\text{renormalised Lagrangian}} - \underbrace{\frac{1}{4} \delta Z_A F_{\mu\nu} F^{\mu\nu} + \bar{\psi}(i\delta Z_\phi D - \delta m)\psi - e\delta Z_\Gamma \bar{\psi}\gamma^\mu\psi A_\mu}_{\text{counterterms}} \quad (12.73)$$

A convenient renormalisation point is

$$\not{p} = m \quad (12.74)$$

for which we define the following renormalisation conditions:

Definition 12.3 (QED OS renormalisation conditions)

- **Mass counterterm:** The electron propagator has a pole at the physical (measured) mass m

$$\Sigma(\not{p})|_{\not{p}=m} = 0 \quad (12.75)$$

where $\Sigma(\not{p})$ is the electron self-energy.

- **Fermion field counterterm:** The residue of the electron propagator at the physical pole is unity

$$\left. \frac{d}{d\not{p}} [\not{p} - m - \Sigma(\not{p})] \right|_{\not{p}=m} = 1 \rightarrow \left. \frac{d\Sigma(\not{p})}{d\not{p}} \right|_{\not{p}=m} = 0 \quad (12.76)$$

- **Photon field counterterm:** The photon remains massless, i.e., its propagator has a pole at $p^2 = 0$

$$\Pi^{\mu\nu}(q) = (q^\mu q^\nu - q^2 g^{\mu\nu})\Pi(q^2) \quad \text{with} \quad \Pi(0) = 0 \quad (12.77)$$

where $\Pi^{\mu\nu}(q)$ is the photon self-energy (vacuum polarisation).

- **Coupling (i.e. charge) counterterm:**

$$Z_1 \Gamma^\mu(p', p)|_{p=p', q=0} = \gamma^\mu \quad (12.78)$$

This is known as the *Thomson limit* and ensures that the vertex function reduces to the bare interaction in the limit of zero momentum exchange.

From here on, we are in a position to calculate the counterterms.

Derivation 12.12 (Fermion field, vertex and mass) The electron self-energy is the sum of the one-loop self-energy and the counterterms, with the standard formula of

$$-i\Sigma(\not{p}) = -i\Sigma_{\text{1-loop}}(\not{p}) + i\Sigma(\delta Z_\phi \not{p} - \delta m) \quad (12.79)$$

Now we impose the renormalisation conditions. By applying (12.75), we have

$$\Sigma_{\text{1-loop}}(m) = m\delta Z_\psi - \delta m \quad (12.80)$$

Consulting (12.76), we find that

$$\left. \frac{d\Sigma_{\text{1-loop}}(\not{p})}{d\not{p}} \right|_{\not{p}=m} = \delta Z_\psi \quad (12.81)$$

The one-loop self-energy is given by

$$-i\Sigma_{\text{1-loop}}(\not{p}) = (-ie)^2 \int \frac{d^4 k}{(2\pi)^2} \gamma^\mu \frac{i(\not{k} + m)}{k^2 - m^2 + i\varepsilon} \gamma^\nu \frac{-ig_{\mu\nu}}{(p - k)^2 + i\varepsilon} \quad (12.82)$$

Using Feynman parameterisation and dimensional regularisation, this evaluates as

$$-i\Sigma_{\text{1-loop}}(\not{p}) = -ie^2 \frac{\mu^\varepsilon}{(4\pi)^{d/2}} \int_0^1 dx \Gamma(\varepsilon/2) \frac{(4-\varepsilon)m - (2-\varepsilon)x\not{p}}{[(1-x)m^2 - x(1-x)p^2]^{\varepsilon/2}} \quad (12.83)$$

We insert this into (12.81), which gives the rather cumbersome result

$$\delta Z_\phi = \frac{e^2}{(4\pi)^{d/2}} \int_0^1 dx \left(\frac{2}{\varepsilon} - \gamma_E - \ln \left(\frac{(1-x)m^2}{\mu^2} \right) + O(\varepsilon) \right) \times \left(-(2-\varepsilon)x + \left(\frac{\varepsilon}{2} \right) x \frac{[4-2x-\varepsilon(1-x)]}{(1-x)} \right) \quad (12.84)$$

Substituting this result back to (12.79) yields

$$\delta m = m \frac{e^2}{16\pi^2} \int_0^1 dx \left(\frac{2}{\varepsilon} - \gamma_E - \ln \left(\frac{(1-x)m^2}{\mu^2} \right) + O(\varepsilon) \right) \times \frac{(\varepsilon-2)x((1+\varepsilon)-2)[4-2x-\varepsilon(1-x)]}{2(1-x)} \quad (12.85)$$

This suggests that the counterterm δm must be directly proportional to m .

Derivation 12.13 (Photon field) Again, we first acknowledge that the photon self-energy is the sum of the one-loop self-energy and the counterterms

$$i\Pi^{\mu\nu}(q^2) = i\Pi_{\text{1-loop}}^{\mu\nu}(q^2) - i(g^{\mu\nu}q^2q^\mu q^\nu)\delta Z_A \quad (12.86)$$

where the photon one-loop self-energy is given as

$$i\Pi_{\text{loop}}^{\mu\nu} = (-1)(ie)^2 \int \frac{d^4 k}{(2\pi)^2} \text{Tr} \left[\gamma^\mu \frac{i}{\not{k} - m} \gamma^\nu \frac{i}{\not{k} + \not{q} - m} \right] \quad (12.87)$$

where the factor of -1 is due to the presence of one closed fermion loop.

Using Feynman parameterisation and exploiting the γ matrix formulae in Theorem 6.4, This becomes

$$i\Pi_{\text{loop}}^{\mu\nu}(q^2) = -4e^2 \int_0^1 dx \int \frac{d^4 k}{(2\pi)^4} \frac{N^{\mu\nu}}{[x((k+q)^2 - m^2) + (1-x)(k^2 - m^2)]^2} \quad (12.88)$$

where we have defined the shorthand

$$N^{\mu\nu} = k^\mu(k+q)^\nu + k^\nu(k+q)^\mu - g^{\mu\nu}[k \cdot (k+q) - m^2] \quad (12.89)$$

By redefining variables and dimensional regularisation, this becomes

$$i\Pi_{\text{loop}}^{\mu\nu} = -i(g^{\mu\nu}q^2 - q^\mu q^\nu) \frac{8e^2}{(4\pi)^{d/2}} \mu^\varepsilon \int_0^1 x(1-x)dx \frac{\Gamma(\varepsilon/2)}{(a^2)^{\varepsilon/2}} \quad (12.90)$$

Consulting (12.61), we actually recover the vacuum polarisation scalar $\Pi(q^2)$ in the 1-loop order:

$$\Pi_{\text{loop}} = -\frac{8e^2}{(4\pi)^{d/2}} \mu^\varepsilon \int_0^1 x(1-x)dx \frac{\Gamma(\varepsilon/2)}{(a^2)^{\varepsilon/2}} \quad (12.91)$$

Inserting this into (12.86) and imposing the renormalisation condition (12.77), we can derive the counterterm as

$$\delta Z_A = \Pi_{\text{loop}}(0) = -\frac{8e^2}{16\pi^2} \int_0^1 x(1-x)dx \left(\frac{2}{\varepsilon} - \gamma_E - \ln \frac{m^2}{\mu^2} + \dots \right) \quad (12.92)$$

Derivation 12.14 (Running charge) Let us consider the true form of a full photon propagator. It will be the sum of the tree-level photon propagator, the 1-loop correction (i.e. OPI diagram), the 2-loop correction, etc. Using the Feynman rules, it is not hard to see the expression

$$\underbrace{\frac{-ig_{\mu\nu}}{q^2} F(q^2)}_{\text{full photon propagator}} = \underbrace{\frac{-ig_{\mu\nu}}{q^2}}_{\text{tree-level contribution}} + \underbrace{\frac{-ig_{\mu\rho}}{q^2} [i(g^{\rho\sigma}q^2 - q^\rho q^\sigma) \Pi(q^2)] \frac{-ig_{\sigma\nu}}{q^2}}_{\text{1-loop contribution}} + \dots \quad (12.93)$$

where we have conveniently defined the *form factor* $F(q^2)$ that represents the ‘summation’ of the contributions at all levels.

Through a series of manipulations of the tensorial objects in this equation, as well as utilising the relationship

$$\Pi(q) + \Pi^2(q^2) + \dots = \frac{1}{1 - \Pi(q^2)} - 1 \quad (12.94)$$

we find that

$$\frac{-ig_{\mu\nu}}{q^2} F(q^2) e^2 = \frac{-ig_{\mu\rho}}{q^2} e^2 \frac{1}{1 - \Pi(q^2)} \quad (12.95)$$

Recognising that the LHS $F(q^2) e^2$ and the RHS e^2 can be rewritten as the effective coupling constant seen in interactions (which is essentially the (renormalised) running charge) e^2 and the bare charge e_0^2 , we find the relationship

$$e^2(q^2) = \frac{e_0^2}{1 - \Pi(q^2)} \quad (12.96)$$

where $\Pi(q^2)$ is the photon self-energy *scalar*. From a similar train of thought to that of the last derivation, we can see that $\Pi(q^2)$ is rather intuitively given by

$$\Pi(q^2) = \Pi_{\text{loop}}(q^2) - \delta Z_A = -\frac{2e^2}{4\pi^2} \int_0^1 dx x(1-x) \ln \left(\frac{m^2}{m^2 - x(1-x)q^2} \right) \quad (12.97)$$

Two comments are in order:

- $\Pi(q^2)$ is clearly a physical quantity independent of μ .
- e increases with the energy scale due to the log function. This is known as the *screening effect*, a general feature of Abelian gauge theories^a.

We end with a note on HEP. (12.96) can be recast by defining a so-called *fine structure constant*:

Definition 12.4 (Fine structure constant) We define the renormalised and bare fine structure constants as α and α_0 respectively:

$$\alpha_0 = \frac{e_0^2}{4\pi} \quad (12.98)$$

The previous relationship (12.96) then becomes

$$\alpha(q^2) = \frac{\alpha_0}{1 - \Pi(q^2)} \quad (12.99)$$

From meticulous experiments, we know that α takes the infamous value of 1/137.

^aAs we will see later, couplings in non-Abelian gauge theories, on the other hand, decrease with energy scale instead.

Quote 12.3 Weesa free!

Unknown Gungan, in Return of the Jedi

As of the time of writing (2025), QED has been renormalised up to the fifth loop order. Calculations for fourth and fifth loop order renormalisations, pioneered by the much-celebrated Toichiro Kinoshita, were only developed from the 1990s on. QCD has also been renormalised up to the fifth loop order, while electroweak theory has only been fully normalised up to the second loop order. To keep this text short, we will not renormalise QCD or electroweak theory.

12.6 Beyond perturbation theory

While powerful, perturbation theory has limitations. It fails for large coupling constants ($\lambda \ll 1$). In cases where the coupling constant is not strictly *small* or new physics appears at different energy scales, perturbation theory can still be applied in a restricted sense. *Effective field theories* (EFTs) use perturbative methods to focus on low-energy phenomena by integrating out high-energy degrees of freedom. We conclude with some philosophical remarks:

- As we proceeded through Part IV, we saw how everything became increasingly arbitrary and hand-wavy. There is first a breakdown of tree-level QFT at the loop level. However, instead of overthrowing this theory, we introduced renormalisation, which is effectively an ‘extension’ of tree-level QFT that becomes increasingly unwieldy as the diagrams increase in their complexity.
- This is because QFT is actually an EFT³. Unlike fundamental theories, EFTs are valid up to a certain energy scale but are expected to break down beyond that scale, giving rise to new physics. Higher-energy effects then manifest as *suppressed corrections* in the form of higher-dimensional operators. In this sense, all classical physics is nothing but effective theories w.r.t. modern physics. In the same vein, GR and modified theories of gravity are likewise EFTs of a future unified theory.
- As we will soon see, the standard model is effectively a collection of QFTs. Hence, it too is an EFT. Most of its extensions, like supersymmetry (SUSY), are likewise EFTs. In the 1970s, physicists initially thought that supergravity (SUGRA), a gravitational extension of SUSY, might provide a complete quantum theory of gravity by itself. However, it was later realised that supergravity alone is not sufficient to quantise gravity at all energy scales. Instead, string theory emerged as a more complete framework, with supergravity appearing as a low-energy effective theory of string theory. Ultimately, the goal of physics is then to construct a single fundamental theory.

³This idea, considered revolutionary when it was created in the 1960s, was formulated by Ken Wilson, whose general ideas are often known as *Wilsonian renormalisation*.

Part V

Standard model

Chapter 13

Non-Abelian gauge theories

Quote 13.1 Repeated pseudo-breakdowns in the realm of HEP are bound to happen.

Felix Halbwedl, 28 July 2025

To date, the so-called standard model is well-known to be the best model that describes the three fundamental forces: the electromagnetic, strong and weak forces. While QED is a gauge theory, it belongs to the simpler case of Abelian gauge theories. The strong force is instead governed by QCD. In contrast, there exist non-Abelian gauge theories¹, so-called due to being defined w.r.t. $SU(N)$ groups, where there is a failure for the structure constants to vanish. Specifically, the strong force is governed by QCD, a non-Abelian gauge theory corresponding to the $SU(3)$ group.

13.1 Generalisation of the Abelian gauge

As its name suggests, Abelian and non-Abelian gauge theories are governed by Abelian and non-Abelian groups respectively. We recall from *Spinors & Symmetries* that the commutators of two generators form a third generator with a so-called *structure constant* as a factor²:

$$[T_a, T_b] = i f_{ab}^c T_c \quad (13.1)$$

where a, b and c are generator indices that span over $1, \dots, N^2 - 1$ for the group $SU(N)$.

Remark 13.1 As we know, the structural constants f_{ab}^c are antisymmetric in all indices.

When the structure constants are zero, we are effectively taking the Abelian limit (6.119), where group generators commute. That is to say, Abelian groups are actually a special case of non-Abelian groups.

Now we discuss how this concerns gauge theories. In our previous adventures in QED, we have defined a remarkably successful (albeit Abelian) gauge theory under $U(1)$ symmetry. The process of extending this to non-Abelian gauge theories involves only one key difference:

- The $U(1)$ group is generalised into a generic $SU(N)$ group, short for special unitary (N) group.
- As such, the photon field is generalised into a gauge field³ A_μ^a with an extra generator index a whose significance we will soon see.
- The fermion field stays as is since we still have both bosons and fermions.

Let us now briefly revise $SU(N)$ transformations. As we know, the transformation U of some $SU(N)$ group is given by

$$U^\dagger U = 1 \quad \det U = 1 \quad (13.2)$$

It satisfies, by definition

$$U^\dagger U = 1 \quad \det U = 1 \quad (13.3)$$

¹Also historically called Yang-Mills theories after Frank C. N. Yang and George Mills.

²Recall that the factor of i on the RHS is imposed by the definition conventionally used in physics. In mathematics, this factor is nonexistent.

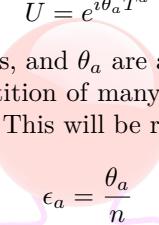
³Even though we only concern ourselves with $SU(N)$ groups, we still call the gauge fields ‘non-Abelian gauge fields’ by convention.

Derivation 13.1 (U as a series of *smol* transformations) Once again, as we know, a transformation U is, by definition

$$U = e^{i\theta_a T^a} \quad (13.4)$$

where T^a are our all-too-familiar generators, and θ_a are a series of parameters.

This transformation can be seen as a repetition of many *smol* transformations: Suppose we divide a transformation angle θ_a into n *smol* steps. This will be reflected in our parameters as



$$\epsilon_a = \frac{\theta_a}{n} \quad (13.5)$$

By a well-known expansion, each *smol* step then looks like

$$U_{\text{smol}} = \mathbb{I} + i\epsilon_a T^a + O(\epsilon^2) \quad (13.6)$$

Repeat this n times, and we find

$$U = (\mathbb{I} + i\epsilon_a T^a + O(\epsilon^2))^n \quad (13.7)$$

If one sends n to infinity (or ϵ to 0), this becomes

$$U = \lim_{n \rightarrow \infty} \left(\mathbb{I} + i \frac{\theta_a}{n} T^a \right)^n = e^{i\theta_a T^a} \quad (13.8)$$

whence (13.4) is recovered. i.e. the matrix exponential precisely encodes the accumulation of infinitely many infinitesimal steps. The significance of this will soon become apparent.

Again, as we know, U are matrices that are representations of abstract elements of the $SU(N)$ group. Two representations exist: the fundamental and adjoint presentations. Both ultimately satisfy the commutation relation in (13.1).

Definition 13.1 (Fundamental representation) The fundamental representation acts on an N -dimensional complex vector space Ψ , whose basis is Ψ_i :

$$\Psi_i = \begin{pmatrix} \Psi_1 \\ \vdots \\ \Psi_N \end{pmatrix} \quad (13.9)$$

We see that:

- The generators T^a themselves are the representations and hence define transformations^a.
- The generators/representations are nothing but $N \times N$ matrices that are Hermitian and traceless.
- There are $N^2 - 1$ generators in total.

This is the representation that we use for matter (Dirac fermion) fields. It corresponds to the basic charge of the theory, like colour charge in QCD.

^aThis makes the fundamental representation conceptually intuitive, which is why it is more prominent in pure mathematics.

Quote 13.2 No frills

Felix Halbwedl, 9 November 2024

We can then write down A generic field ψ is then transformed as, by definition of the fundamental representation

$$\psi_i \rightarrow \psi'_i = U_i^j \psi_j \quad (13.10)$$

The other implication is the invariance of the product between a field and its complex conjugate:

$$\psi_i^* \psi^i \rightarrow \psi'^*_i \psi'_i = U_i^{j*} \psi_j^* U_k^i \psi^k = U_i^{j*} U_k^i \psi_j^* \psi^k = \delta_k^j \psi_j^* \psi^k = \psi_i^* \psi^i \quad (13.11)$$

where we have exploited the fact that all indices are free.

Definition 13.2 (Adjoint representation) The vector space which the adjoint representation acts on is *not generated by the number N itself*. Rather, it is spanned by the *range of generators T^a* , which acts as its basis:

$$T^a = (T^1, \dots, T^{N^2-1}) \quad (13.12)$$

The representations, which we confusingly label $(T^a)_{bc}$, are, importantly, *not the generators themselves*. Rather, it is defined in relation to the structural constants^a:

$$(T^a)_{bc} = i f_{bc}^a \quad (13.13)$$

- The ‘rescaled’ structural constants $(T^a)_{bc}$ are the representations and hence define transformations.
- The representations are $(N^2 - 1) \times (N^2 - 1)$ matrices.

This is the representation that we use for gauge boson fields.

^aIn some literature, an alternative definition $(T^a)_{bc} = -i f_{bc}^a$ that differs by a sign exists.

As a result of the nonvanishing structural constants, the field strength tensor is a little more complex. To begin with, there is a need to redefine the covariant derivative.

Definition 13.3 (Non-Abelian covariant derivative)

$$D_\mu = \partial_\mu - ig A_\mu^a T_a \quad (13.14)$$

which transforms as

$$D_\mu \rightarrow D'_\mu = U^\dagger D_\mu U \quad (13.15)$$

Note 13.1 (A sinister shorthand) A sinister shorthand appears in most of the standard literature, but is rarely explained. Consider some tensorial quantity N_μ or $R_{\mu\nu}$ in Abelian gauge theories, or more specifically, QED. While we only visually see these quantities themselves, they actually come with the generators attached. i.e.

$$N_\mu \xrightarrow{\text{is actually}} N_\mu \cdot 1 \quad R_{\mu\nu} \xrightarrow{\text{is actually}} R_{\mu\nu} \cdot 1 \quad (13.16)$$

However, we see that the generators of $U(1)$ are unity. Hence, they are utterly trivial and are simply absorbed into the quantities for simplicity. As innocuous as this seems, this shorthand becomes problematic when we generalise to the Abelian limit. In a non-Abelian gauge theory, the generators T^a are non-trivial, and previous quantities must be rewritten to reflect this:

$$N_\mu \cdot 1 \xrightarrow{\text{non-Abelian}} N_\mu^a T_a = N_\mu \quad R_{\mu\nu} \cdot 1 \xrightarrow{\text{non-Abelian}} R_{\mu\nu}^a T_a = R_{\mu\nu} \quad (13.17)$$

where the final terms on both expressions can be seen as the ‘short’ forms where the generator index a is ‘contracted’. This is a sinister shorthand.

From here, it is possible to write down a $SU(N)$ gauge transformation matrix by its components. The important starting point is (13.6).

Definition 13.4 (Infinitesimal transformation matrix components)

- In the fundamental representation, we have

$$U_{ij} = \delta_{ij} + i\theta_a (T^a)_{ij} + O(\theta^2) \quad (13.18)$$

where a runs from 1 to $N^2 - 1$, and i and j runs from 1 to N due to the $N \times N$ nature of the matrix U .

- In the adjoint representation, we have

$$U_{bc} = \delta_{bc} + i\theta_a (T^a)_{bc} + O(\theta^2) = \delta_{bc} - \theta_a f_{bc}^a + O(\theta^2) \quad (13.19)$$

where a runs from 1 to $N^2 - 1$, and b and c also runs from 1 to $N^2 - 1$ due to the $(N^2 - 1) \times (N^2 - 1)$ nature of the matrix U .

Let us summarise our findings.

Definition 13.5 (SU(N) transformations of fields) Up to the first order θ_a :

- The scalar (fermion) field lives in the fundamental representation. From (13.18), it transforms as

$$\psi(x) \rightarrow \psi'(x) = (1 + i\theta_a(x)T^a)\psi(x) \quad (13.20)$$

- The vector (gauge) field lives in the adjoint representation. From (13.19), it transforms as

$$A_\mu^a \rightarrow A'_\mu^a = \underbrace{A_\mu^a + g^{-1}\partial_\mu\theta^a}_{(1)} + \underbrace{f_{bc}^a A_\mu^b \theta^c}_{(2)} \quad (13.21)$$

where:

- (1) is the same as Abelian gauge theories (e.g. QED).
- (2) is the term brought about by the non-zeroness of the structural constants^a.

^aWhich itself is because SU(N) groups are non-Abelian.

We are now in a position to think about building our non-Abelian gauge theory Lagrangian. As noted, we have two fields, the Dirac fermion field and the non-Abelian gauge field, both of which undergo some change due to the redefinition of the covariant derivative.

$$\mathcal{L} = \bar{\psi}(i\cancel{D} - m)\psi$$

The fermion field lives in the fundamental representation, so the entire term is

Definition 13.6 (Non-Abelian Dirac Lagrangian)

$$\mathcal{L} = \bar{\psi}(i\gamma^\mu(\partial_\mu - igA_\mu^a T_a) - m)\psi = \underbrace{\bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi}_{\text{Dirac field}} - \underbrace{g\bar{\psi}\gamma^\mu\psi A_\mu^a T_a}_{\text{interaction}} \quad (13.22)$$

where, as before, the first term encodes the free fermion propagator while the second term encodes the 3-vertex between one fermion and two gauge bosons. Inserting (13.14) into (6.25) yields the non-Abelian form of our good friend, the field strength tensor $F^{\mu\nu}$:

Definition 13.7 (Non-Abelian field strength tensor)

$$F_{\mu\nu}^a = \partial_\mu A_\nu^a - \partial_\nu A_\mu^a + g f_{bc}^a A_\mu^b A_\nu^c \quad (13.23)$$

Now let us construct the field strength tensor term. In effect, it is the generalised version of the free photon term or the photon field Lagrangian (6.121). Unfortunately, we cannot simply copy the structure of the Abelian free Lagrangian. This is because the non-Abelian field strength tensor changes under the gauge transformation:

$$F_{\mu\nu}^a \rightarrow F'_{\mu\nu}^a = F_{\mu\nu}^a + \delta F_{\mu\nu}^a = F_{\mu\nu}^a - g f_{bc}^a \alpha^b F_{\mu\nu}^c \quad (13.24)$$

That is to say, it is *not* gauge-invariant. However, the trace of $F^{\mu\nu}$ squared *is*:

$$\delta(F_{\mu\nu}^a F_a^{\mu\nu}) = 2\delta(F_{\mu\nu}^a)F_a^{\mu\nu} = -2g f_{bc}^a \alpha^b F_{\mu\nu}^c F_a^{\mu\nu} = 0 \quad (13.25)$$

To preserve the gauge invariance of the Lagrangian, we thus encode the gauge field kinetic term, the analogue to the photon kinetic term in QED, as⁴

Definition 13.8 (Guage field kinetic term)

$$\mathcal{L} = -\frac{1}{4} \text{Tr}(F^{\mu\nu} F_{\mu\nu}) \quad (13.26)$$

⁴In some older literature we redefine the gauge field as $A_\mu^a \rightarrow gA_\mu^a$. Hence the prefactor is redefined as $\frac{1}{4} \rightarrow \frac{1}{2g^2}$.

A more familiar-looking form can be derived by ‘uncontracting’ the generator indices. Doing so reinserts our good friends, the generators T^a :

$$\mathcal{L} = -\frac{1}{4} \text{Tr}(F^{\mu\nu a} T_a F_{\mu\nu}^a T_a) = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a \text{Tr}(T^a T_a) \quad (13.27)$$

By convention, we conveniently set the normalisation condition as⁵

$$\text{Tr}(T^a T_a) = 1 \quad (13.28)$$

Thus, our kinetic term is conveniently

$$\mathcal{L} = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a \quad (13.29)$$

Note 13.2 (An aside) Recall that the action is the 4-integral (which has mass dimension -4) of the Lagrangian density and has zero mass dimension. Hence, the Lagrangian has a mass dimension of 4. Hence, as long as they have mass dimensions of 4 and are Lorentz-invariant, alternate candidates for the field strength term exist, but cannot be admitted due to being unphysical:

- $\epsilon^{\alpha\beta\mu\nu} F_{\alpha\beta} F_{\mu\nu}$: This term violates P- and T-invariance and hence the CPT theorem.
- $k(F_{\mu\nu} F^{\mu\nu})^2$ or similar terms, where k is a coupling: As the non-coupling part of this term has a mass dimension greater than 4, the coupling must have negative mass dimensions which is forbidden by renormalisability requirements.

In QED, the field strength tensor (Faraday tensor) encodes only one object: the free photon propagator. *This is not the case in non-Abelian gauge theories.* One can see this by writing out the field strength tensors explicitly:

$$\mathcal{L} = -\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a = -\frac{1}{4} [(\partial_\mu A_\nu^a - \partial_\nu A_\mu^a) + g f_{bc}^a A_\mu^b A_\nu^c] [(\partial^\mu A^{a\nu} - \partial^\nu A^{a\mu}) + g f_{b'c'}^a A^{b'\mu} A^{c'\nu}] \quad (13.30)$$

Expanding this gives three objects, all of which are Lorentz invariant and respect gauge symmetry:

$$\mathcal{L} = \underbrace{-\frac{1}{4} (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a)^2}_{\textcircled{1}} - \underbrace{\frac{1}{2} g f_{bc}^a (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a) A^{b\mu} A^{c\nu}}_{\textcircled{2}} - \underbrace{\frac{1}{4} g^2 f_{bc}^a f_{b'c'}^a A_\mu^b A_\nu^c A^{b'\mu} A^{c'\nu}}_{\textcircled{3}} \quad (13.31)$$

where:

- ① encodes our gool ol’ free gauge boson propagator.
- ② encodes the 3-gauge boson vertex as it is cubic in A_μ^a .
- ③ encodes the 4-gauge boson vertex as it is quartic in A_μ^a .

The important realisation is that unlike in QED, the gauge field kinetic term is no longer entirely free. Instead, it includes two vertices that represent self-interaction among gauge bosons. Finally, before we consider gauge fixing, we can write down the entire Lagrangian:

Definition 13.9 (Non-Abelian gauge theory Lagrangian)

$$\mathcal{L} = \underbrace{-\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a}_{\text{gauge field kinetic}} + \underbrace{\bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi}_{\text{fermion kinetic}} - \underbrace{\bar{\psi} \gamma^\mu \psi A_\mu^a T_a}_{\text{boson-fermion vertex}} \quad (13.32)$$

where the second term is free, the third term is interacting, and the first term is a mix of free and interacting parts.

⁵We are allowed to do this as any arbitrary factor is cancelled out when normalising the generating functional.

13.2 Gauge fixing and Faddeev-Popov ghosts

The idea here is highly similar to the Faddeev-Popov method, with some differences due to the emergence of generators and generator indices. Firstly, the U(1) gauge transformation is replaced with a generic SU(N) gauge transformation, which we now label $\beta(x)$ instead of $\alpha(x)$.

$$A_\mu \rightarrow (A_\mu^\beta)^a = A_\mu^a + \frac{1}{g} \partial_\mu \beta^a + f_{bc}^a A_\mu^b = A_\mu^a + \frac{1}{g} (D_\mu \beta)^a \quad (13.33)$$

This is nothing but a copy of (13.21) where we have relabelled the parameters θ^a to β^a and ‘unabsorbed’ the coupling g . Before proceeding, let us summarise the (pseudo-)indices:

- μ is the spacetime index.
- a, b and c are the generator indices (in Latin letters).
- Like α in Part III, β is a (pseudo)index denoting that the 4-potential A_μ has undergone the gauge transformation.

In this sense, $\beta^a(x)$ is not a single function but a series of functions, each of which corresponds to a specific generator.

From (13.13), we can rewrite the covariant derivative (13.14), which lives in the adjoint representation, as

Definition 13.10 (Adjoint representation non-Abelian covariant derivative)

$$(D_\mu \beta)^a = \partial_\mu \beta^a + g f_{bc}^a A_\mu^b \beta^c \quad (13.34)$$

Let us note that the gauge condition itself is now endowed with a generator index. Hence, it now looks like

$$G^a[A^\beta] = \partial^\mu (A_\mu^\beta)^a - w^a(x) \quad (13.35)$$

On the surface, this seems like nothing but its Abelian counterpart with a generator index. However, a problem arises, again due to the failure of the structural constant to vanish, if we calculate the non-Abelian Faddeev-Popov operator.

Derivation 13.2 (Nontrivialness of the Faddeev-Popov operator) Let us substitute (13.33) into (13.35). This yields

$$G^a[A^\beta] = \partial^\mu \left(A_\mu^a + \frac{1}{g} (D_\mu \beta)^a \right) - w^a(x) = \partial^\mu A_\mu^a + \frac{1}{g} \partial^\mu (D_\mu \beta)^a - w^a(x) = G^a[A] + \frac{1}{g} \partial^\mu (D_\mu \beta)^a \quad (13.36)$$

The Faddeev-Popov operator (9.34) is defined as the variation of the gauge condition under an infinitesimal gauge transformation. Here, it takes the form

$$M^{ab}(x, y) = \frac{\delta G^a[A^\beta](x)}{\delta \beta^b(y)} = \frac{1}{g} \frac{\delta \partial_y^\mu (D_\mu \beta)^a(x)}{\delta \beta^b(y)} \quad (13.37)$$

where we again stress that the y in ∂_y^μ is nothing but a pseudoindex.

Our task is hence nothing but to compute the derivative of the covariant derivative^a. This can be done by inserting (13.34):

$$M^{ab}(x, y) = \frac{1}{g} \partial^\mu (\delta^{ab} \partial_\mu \delta(x - y) + g f_c^{ab} A_\mu^c(x) \delta(x - y)) \quad (13.38)$$

Exploiting the Dirac delta gives us

$$M^{ab} = \frac{1}{g} \partial^\mu (\delta^{ab} \partial_\mu + g f_c^{ab} A_\mu^c) \quad (13.39)$$

Finally, we can rewrite this via the differential operator form by consulting (13.34)

$$M = \frac{1}{g} \partial^\mu D_\mu \quad (13.40)$$

^aWhat a mouthful!

This is bad in the sense that our beloved Faddeev-Popov operation turns out to be non-trivial. That is to say, more than simply its harmless Dirac delta form in (9.35). However, the real problem lies in the nontrivialness of the Faddeev-Popov determinant $\det M$.

We can determine $\det M$ using an identity for the determinant of some generic operator A :

$$\det A = \int \mathcal{D}c \mathcal{D}\bar{c} e^i \int d^4x \bar{c} A c \quad (13.41)$$

where $c(x)$ and $\bar{c}(x)$ are anticommuting scalar fields⁶.

We now specify A as the Faddeev-Popov operator as M . As the Faddeev-Popov matrix has dimension N , and by the definition of the determinant, we have

$$\det \left(\frac{1}{g} \partial^\mu D_\mu \right) = \left(\frac{1}{g} \right)^N \det(\partial^\mu D_\mu) \quad (13.42)$$

Now, accounting for the fact that (13.41) is nothing but yet another path integral, we can absorb this factor $\frac{1}{g}$ in normalisation. Hence, we have

Definition 13.11 (Faddeev-Popov ghosts)

$$\det M = \int \mathcal{D}c \mathcal{D}\bar{c} e^i \int d^4x \bar{c} (-\partial_\mu D^\mu) c \quad (13.43)$$

where the factor -1 in $-\partial_\mu D^\mu$ comes from convention, ensuring the kinetic term is positive in Euclidean space and consistent with the operator arising from the gauge fixing.

However, one can soon see that there is something funny going on with the two fields $c(x)$ and $\bar{c}(x)$:

- The fields anticommute, just like fermion fields.
- The fields are scalar fields, just like bosonic fields.

A field clearly cannot be both fermionic and bosonic. Hence, we are forced to conclude that the fields $c(x)$ and $\bar{c}(x)$ are unphysical. i.e. they are *fictitious fields*. $c(x)$ and $\bar{c}(x)$ are then known as the so-called *Faddeev-Popov ghosts* or *Faddeev-Popov ghost fields*⁷.

We are one step away from deriving the full gauge-fixed Lagrangian. From (13.43), the Faddeev-Popov ghost Lagrangian is clearly

$$\mathcal{L}_{\text{ghost}} = \bar{c}^a (-\partial_\mu D_\mu^{ab}) c^b \quad (13.44)$$

Using integration by parts, we can rewrite it as

$$\mathcal{L}_{\text{ghost}} = \partial_\mu \bar{c}^a D_\mu^{ac} c^c \quad (13.45)$$

Hence, we have

Definition 13.12 (Gauge-fixed non-Abelian gauge theory Lagrangian)

$$\mathcal{L} = \underbrace{-\frac{1}{4} F_a^{\mu\nu} F_{\mu\nu}^a}_{\text{gauge field kinetic}} - \underbrace{\frac{1}{2\xi} (\partial_\mu A^\mu)^2}_{\text{gauge-fixing}} + \underbrace{\partial_\mu \bar{c}^a D_\mu^{ac} c^c}_{\text{Faddeev-Popov ghosts}} + \underbrace{\bar{\psi} (i\gamma^\mu \partial_\mu - m) \psi}_{\text{fermion kinetic}} - \underbrace{\bar{\psi} \gamma^\mu \psi A_\mu^a T_a}_{\text{boson-fermion vertex}} \quad (13.46)$$

13.3 Quantisation

Let us compare tree-level Feynman diagram elements in non-Abelian gauge theories with their Abelian counterpart:

- The fermion propagator is identical to its non-Abelian counterpart, save for a redefinition of the covariant derivative.

⁶This will be the subject of another problem on its own.

⁷However, aside from the introduction of the ghosts in the generating functional, gauge fixing is otherwise identical.

- The gauge field propagator is different from its non-Abelian counterpart only by a Kronecker delta δ^{ab} carrying the generator indices a and b .
- The fermion-gauge boson (3-field) vertex is different from its non-Abelian counterpart only by a series of generators T^a .
- There exist two additional vertices due to the Lagrangian field strength term not being completely free, as seen in (13.31).

From the loop order on, everything differs from their Abelian counterparts. A part of this is due to the emergence of ghost field propagators and vertices which appear in renormalisation⁸. We can then summarise what needs to be derived in this section:

- The emergence of δ^{ab} in the gauge field propagator.
- The three- and four-gauge field boson vertex.
- The ghost propagator and vertex.

Derivation 13.3 (Gauge field propagator) The gauge field propagator is not a new object, because the bulk of the work was already done in Derivation 6.15. Still, the propagator is slightly modified. The cause of this lies in the fact that ultimately, the propagator is a two-point Green's function that represents two *different* gauge fields, which means they have distinct indices. Just like how the tensorial indices of the two fields are different in (6.172), so much their generator indices be different. As such, the propagator takes the form

$$D_{\mu\nu}^{ab}(x - y) = \langle 0 | T[A_\mu^a(x) A_\nu^b(y)] | 0 \rangle \quad (13.47)$$

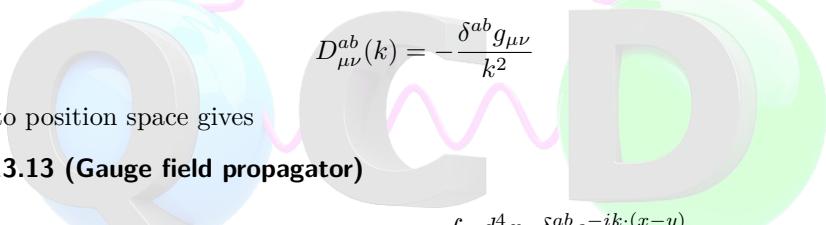
which evaluates as

$$D_{\mu\nu}^{ab}(k) = -\frac{1}{k^2} \delta^{ab} \left[g_{\mu\nu} - (1 - \xi) \frac{k_\mu k_\nu}{k^2} \right] \quad (13.48)$$

As advertised, there is the addition of a Kronecker delta δ^{ab} . This emerges from the fact that:

- In free theory, the term $gf_{bc}^a A_\mu^b A_\nu^c$ is dropped.
- The Lagrangian is hence entirely diagonal in its generator indices^a.
- The Kronecker tensor hence exists to fix one index to another.

In the Feynman gauge ($\xi = 1$), this is



$$D_{\mu\nu}^{ab}(k) = -\frac{\delta^{ab} g_{\mu\nu}}{k^2} \quad (13.49)$$

Moving back to position space gives

Definition 13.13 (Gauge field propagator)

$$D_{\mu\nu}^{ab}(x - y) = -i \lim_{\epsilon \rightarrow 0+} \int \frac{d^4 p}{(2\pi)^4} \frac{\delta^{ab} e^{-ik \cdot (x-y)}}{p^2 + i\epsilon} \quad (13.50)$$

^aIn QCD, where the generator indices are recast into so-called *colour indices*, this carries a physical meaning: If no interactions happen, then there is no mixing of colour indices.

In QCD, this is the so-called *gluon propagator*.

Quote 13.3 Some curly lines here, kinda like my hair.

Felix Halbwedel, on the gluon propagator, 6 July 2025

⁸Even though the ghost fields are utterly unphysical, their Feynman diagram elements arise as mathematical tools in calculations.

Derivation 13.4 (Fermion-gauge boson vertex) This is the easiest derivation, as we do nothing but add a series of generators and replace the charge e with the (so far arbitrary) non-Abelian gauge theory coupling constant g :

Definition 13.14 (Fermion-gauge boson vertex)

$$-ig\gamma^\mu T^a \quad (13.51)$$

Before proceeding, one might want to revise the derivation of the Feynman rule of a generic vertex in Note 9.3.

Derivation 13.5 (Three-gauge boson vertex) We begin by reminding ourselves that the part of the Lagrangian that gives rise to the three-gauge boson vertex is

$$\mathcal{L}_{3A} = -\frac{1}{2}gf_{bc}^a (\partial_\mu A_\nu^a - \partial_\nu A_\mu^a) A_b^\mu A_c^\nu \quad (13.52)$$

In momentum space, the fields and the field derivatives are replaced with

$$A_\mu^a(x) = \int \frac{d^4 p}{(2\pi)^4} e^{-ip \cdot x} A_\mu^a(p) \quad \partial_\mu A_\nu^a(x) = -ip_\mu A_\nu^a(p) e^{-ip \cdot x} \quad (13.53)$$

Hence, the momentum space (13.52) has the form

$$\mathcal{L}_{3A} = -\frac{ig}{2}f_{bc}^a \int \frac{d^4 p_1 d^4 p_2 d^4 p_3}{(2\pi)^{12}} e^{-i(p_1+p_2+p_3) \cdot x} [p_{1\mu} g_{\rho\nu} - p_{1\nu} g_{\rho\mu}] A_a^\rho(p_1) A_b^\mu(p_2) A_c^\nu(p_3) \quad (13.54)$$

The action is then

$$S_{3A} = -\frac{ig}{2}f_{bc}^a \int \frac{d^4 p_1 d^4 p_2 d^4 p_3}{(2\pi)^{12}} (2\pi)^4 \delta^{(4)}(p_1 + p_2 + p_3) [p_{1\mu} g_{\rho\nu} - p_{1\nu} g_{\rho\mu}] A_a^\rho(p_1) A_b^\mu(p_2) A_c^\nu(p_3) \quad (13.55)$$

We now note that there is antisymmetry in the indices μ , ν and ρ . This can be made explicit:

$$p_{1\mu} g_{\rho\nu} - p_{1\nu} g_{\rho\mu} = g_{\mu\nu} (p_1 - p_2)_\rho + g_{\nu\rho} (p_2 - p_3)_\mu + g_{\rho\mu} (p_3 - p_1)_\nu \quad (13.56)$$

Hence

$$\mathcal{L}_{3A} = -\frac{ig}{2}f_{bc}^a \int \frac{d^4 p_1 d^4 p_2 d^4 p_3}{(2\pi)^{12}} A_a^\mu(p_1) A_b^\nu(p_2) A_c^\rho(p_3) [g_{\mu\nu} (p_1 - p_2)_\rho + g_{\nu\rho} (p_2 - p_3)_\mu + g_{\rho\mu} (p_3 - p_1)_\nu] \quad (13.57)$$

Here, we might be tempted to simply take functional derivatives. However, we recall from Part II that the path integral formalism allows us to read off the contribution of a Lagrangian to \mathcal{M}_{fi} by stripping off the external fields^a and adding i . So, going back to position space, the three-gauge boson vertex is

Definition 13.15 (Three-gauge boson vertex)

$$gf_{bc}^a [g^{\mu\nu} (p_1 - p_2)^\rho + g^{\nu\rho} (p_2 - p_3)^\mu + g^{\rho\mu} (p_3 - p_1)^\nu] \quad (13.58)$$

where the previous factor of $\frac{1}{2}$ disappeared during summation.

^aHere we are talking about momentum space fields, so the momentum space 4-integrals and the exponential term (which is unity anyway) are removed too.

Derivation 13.6 (Four-gauge boson vertex) The part of the Lagrangian that gives rise to the four-gauge boson vertex is

$$\mathcal{L}_{4A} = -\frac{1}{4}g^2 f_{bc}^a f_{b'c'}^a A_\mu^b A_\nu^c A^{b'\mu} A^{c'\nu} \quad (13.59)$$

There are no field derivatives, so the only item we have to Fourier-transform is the fields themselves. The Lagrangian is

$$\mathcal{L}_{4A} = -\frac{g^2}{4} f_{bc}^a f_{b'c'}^a \int \prod_{i=1}^4 \frac{d^4 p_i}{(2\pi)^4} e^{-i(p_1+p_2+p_3+p_4)\cdot x} g^{\mu\nu} g^{\rho\sigma} A_\mu^b(p_1) A_\nu^c(p_2) A_\rho^{b'}(p_3) A_\sigma^{c'}(p_4) \quad (13.60)$$

Integrating over the 4-coordinates x gives us the action

$$S_{4A} = -\frac{g^2}{4} f^{abc} f_a^{de} \int \prod_{i=1}^4 \frac{d^4 p_i}{(2\pi)^4} (2\pi)^4 \delta^{(4)}(p_1 + p_2 + p_3 + p_4) g^{\mu\nu} g^{\rho\sigma} A_\mu^b(p_1) A_\nu^c(p_2) A_\rho^d(p_3) A_\sigma^e(p_4) \quad (13.61)$$

Again, due to momentum conservation, we can remove the term $e^{-i(p_1+p_2+p_3+p_4)\cdot x}$ which is unity. The structural constants remain antisymmetric. We can exploit this to make the substitution

$$f^{abc} f_a^{de} g_{\sigma\nu} g_{\mu\rho} = f^{abc} f_a^{de} (g_{\sigma\nu} g_{\mu\rho} - g_{\mu\nu} g_{\sigma\rho}) + f^{adc} f_a^{be} (g_{\nu\sigma} g_{\mu\rho} - g_{\mu\sigma} g_{\nu\rho}) + f^{abd} f_a^{ce} (g_{\sigma\mu} g_{\nu\rho} - g_{\mu\nu} g_{\sigma\rho}) \quad (13.62)$$

Finally, we make this substitution, strip off the fields and introduce a factor of i . This gives, up to a relabelling of constants

Definition 13.16 (Four-gauge boson vertex)

$$-ig^2 [f_e^{ab} f^{cde} (g^{\mu\rho} g^{\nu\sigma} - g^{\mu\sigma} g^{\nu\rho}) + f_e^{ac} f^{bde} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\sigma} g^{\nu\rho}) + f_e^{ad} f^{bce} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\rho} g^{\nu\sigma})] \quad (13.63)$$

where the previous factor of $\frac{1}{4}$ disappeared during summation.

Before proceeding, one might want to revise the derivation of the Feynman rule of a generic vertex in Note 6.3.

Derivation 13.7 (Ghost field propagator) The free part of the ghost Lagrangian is

$$\mathcal{L} = -\partial^\mu \bar{c}^a \partial_\mu c^a \quad (13.64)$$

In momentum space, the Lagrangian becomes:

$$\mathcal{L} = -\bar{c}^a(p) p^2 c^a(p) \quad (13.65)$$

Stripping the Faddeev-Popov ghost and antighost fields, we find that the kernel is

$$D_{ab}^{-1}(p) = p^2 \delta^{ab} \quad (13.66)$$

Finally, we invert the kernel and eliminate the pole singularity. This gives the propagator as

Definition 13.17 (Ghost field propagator)

$$D_{ab}(p) = \frac{i\delta^{ab}}{p^2 + i\varepsilon} \quad (13.67)$$

Derivation 13.8 (Ghost field vertex) Starting from the interaction term

$$\mathcal{L} = -g f_{ab}^c (\partial^\mu \bar{c}^a) A_\mu^c b^b \quad (13.68)$$

We Fourier-transform all fields into momentum space. The ghost field transforms identically as our Fermion scalar field:

$$c^b(x) = \int \frac{d^4 r}{(2\pi)^4} e^{-ir\cdot x} c^b(r) \quad (13.69)$$

The momentum space Lagrangian is then

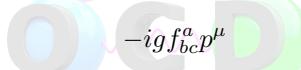
$$\mathcal{L} = -g f_{ab}^c \int \frac{d^4 p}{(2\pi)^4} \frac{d^4 q}{(2\pi)^4} \frac{d^4 r}{(2\pi)^4} (ip^\mu) e^{-i(p+q+r)\cdot x} \bar{c}^a(p) A_\mu^c(q) c^b(r) \quad (13.70)$$

Integrate over the 4-coordinates x give us the action:

$$S = -g f_{ab}^c \int \frac{d^4 p}{(2\pi)^4} \frac{d^4 q}{(2\pi)^4} \frac{d^4 r}{(2\pi)^4} (ip^\mu)(2\pi)^4 \delta^{(4)}(p+q+r) \bar{c}^a(p) A_\mu^c(q) c^b(r) \quad (13.71)$$

Stripping off the fields, relabelling colour indices and adding the factor i gives

Definition 13.18 (Ghost-gauge boson vertex)



$$-igf_{bc}^a p^\mu \quad (13.72)$$

where p^μ is the incoming momentum carried by the ghost field c^b .

Finally, we are in a position to write down the Feynman rules for non-Abelian gauge theories.

Quote 13.4 There's some complicated stuff which makes it a bit more complicated.

Felix Halbwedl, 6 July 2025

Theorem 13.1 (Non-Abelian gauge theory Feynman rules) For a given Feynman diagram in an arbitrary non-Abelian gauge theory, the scattering amplitude matrix elements with a complex factor $i\mathcal{M}_{fi}$ are constructed as follows:

NON-ABELIAN GAUGE THEORY FEYNMAN RULES (PARTIAL)	
For each	Add to expression
Incoming and outgoing fermion	$\bar{u}_\alpha(s, p)$ and $u_\alpha(s, p)$
Incoming and outgoing antifermion	$v_\alpha(s, p)$ and $\bar{v}_\alpha(s, p)$
Incoming and outgoing gauge boson	$\epsilon^{*\mu}(\lambda, p)$ and $\epsilon^\mu(\lambda, p)$
Internal gauge boson line	$\frac{-i\delta^{ab} g^{\mu\nu}}{p^2 + i\varepsilon}$
Internal fermion line	$\frac{i(p+m)}{p^2 - m^2 + i\epsilon}$
Internal ghost line	$\frac{i\delta^{ab}}{p^2 + i\varepsilon}$
Internal loop	$\int d^4 l_n \frac{1}{(2\pi)^4}$ over corresponding line
Fermion-gauge boson vertex	$-igT^a \gamma^\mu$
Three-gauge boson vertex	$gf_{bc}^a [g^{\mu\nu}(p_1 - p_2)^\rho + g^{\nu\rho}(p_2 - p_3)^\mu + g^{\rho\mu}(p_3 - p_1)^\nu]$
Four-gauge boson vertex	$-ig^2 [f_e^{ab} f^{cde} (g^{\mu\rho} g^{\nu\sigma} - g^{\mu\sigma} g^{\nu\rho}) + f_e^{ac} f^{bde} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\sigma} g^{\nu\rho}) + f_e^{ad} f^{bce} (g^{\mu\nu} g^{\rho\sigma} - g^{\mu\rho} g^{\nu\sigma})]$
Ghost-gauge boson vertex ^a	$-gf_{bc}^a p^\mu$
Any vertex ^b	$(2\pi)^4 \delta^3(k_i - k_f)$

Finally, before taking a well-deserved break, remove a factor of $(2\pi)^4 \delta^3(p - q)$, where p and q are the total initial and detected momenta.

^aHere, a is the generator index while b and c are ghost generator indices.

^bThis term enforces 4-momentum conservation.

13.4 BRST symmetry

Quote 13.5 The BRST symmetry is in some sense a remnant of the gauge symmetry that persists even after the gauge has been fixed.

Matthias Gaberdiel and Aude Gehrmann-De Ridder, 2011

In non-Abelian gauge theories, an additional symmetry known as the so-called BRST (Carlo Becchi, Alain Rouet, Raymond Stora and Igor Tyutin) symmetry supercedes gauge symmetries. Let us begin by

recognising that the previous gauge-invariance of the Lagrangian no longer exists in (13.46)

$$\mathcal{L} = \underbrace{-\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \bar{\psi}\gamma^\mu\psi A_\mu^a T_a}_{\text{gauge-invariant}} - \underbrace{\frac{1}{2\xi}(\partial_\mu A^\mu)^2 + \partial_\mu \bar{c}^a D_\mu^{ac} c^c}_{\text{not gauge-invariant}}$$

due to a failure of the gauge-fixing and ghost terms to be gauge invariant.

For reasons that will become apparent later, we introduce an auxiliary field B^a , so-called due to its lack of kinetic terms giving rise to algebraic equations of motion:

Definition 13.19 (Auxiliary field)

$$B^a = \frac{1}{\xi}\partial^\mu A_\mu^a \quad (13.73)$$

The Lagrangian is rewritten as

$$\mathcal{L} = -\frac{1}{4}F_a^{\mu\nu}F_{\mu\nu}^a - \frac{\xi}{2}B^a B_a + (\partial_\mu \bar{c}^a)D_\mu^{ac} c^c + \bar{\psi}(i\gamma^\mu\partial_\mu - m)\psi - \bar{\psi}\gamma^\mu\psi A_\mu^a T_a \quad (13.74)$$

We are now in a position to define a so-called *BRST operator*:

Definition 13.20 (BRST operator) The BRST operator Q represents a BRST transformation on our good friend, the field ψ . In this sense, an infinitesimal transformation can be written as Q coupled to a *smol* parameter ϵ :

$$\delta\phi = \epsilon Q\phi \quad (13.75)$$

The central property that defines the BRST operator is its *nilpotency*, which we know is just a fancy way of saying that the square of Q is zero:

$$QQ\phi = 0 \quad (13.76)$$

Theorem 13.2 (BRST transformations) The infinitesimal changes brought about by the BRST operator is thus, for the quantities we are interested in

$$\delta A_\mu^a = \epsilon D_\mu^{ab} c^b \quad (13.77)$$

$$\delta\psi = ig\epsilon c^a T^a \psi \quad (13.78)$$

$$\delta c^a = -\frac{1}{2}g\epsilon f^{abc} c^b c^c \quad (13.79)$$

$$\delta\bar{c}^a = \epsilon B^a \quad (13.80)$$

$$\delta B^a = 0 \quad (13.81)$$

Exercise 13.1 Show explicitly that all five transformations satisfy nilpotency by imposing them twice.

Note 13.3 (BRST transformations as a generalisation of gauge transformations) Notably, if we define a ‘rescaled’ ghost parameter

$$\alpha^a(x) = g\epsilon c^a(x) \quad (13.82)$$

We will find that, amazingly, the transformations (13.77) and (13.78) are nothing but gauge transformations:

$$\delta A_\mu^a = D_\mu \alpha^a \quad \delta\psi = i\alpha_a T^a \psi \quad (13.83)$$

Despite this equivalence, we have clearly defined, in the last three equations, distinctly new ghost field, antighost field and auxiliary field transformations. So what does this imply?

- Our BRST transformations are nothing but a generalisation of gauge transformations.
- By *promoting* the gauge parameters to anticommuting ghost fields, the symmetry structure is extended to the gauge-fixing and Faddeev-Popov ghost terms in the Lagrangian.

Definition 13.21 (BRST-closed and BRST-exact) A field is called *BRST-closed* if

$$Q\phi = 0 \quad (13.84)$$

and *BRST-exact* if

$$\phi = Q\chi \quad (13.85)$$

for some field χ .

All BRST-exact fields are BRST-closed. This is because

$$Q\phi = Q(Q\chi) = Q \times 0 = 0 \quad (13.86)$$

Theorem 13.3 (BRST-exactness of field products) If ϕ_1 is a BRST-closed field while $\phi_2 = Q\chi_2$ is BRST-exact, then the product $\phi_1\phi_2$ is BRST-exact:

$$Q(\phi_1\phi_2) = Q\phi_1\chi_2 \pm \phi_1Q\chi_2 = \pm\phi_1\phi_2 \quad (13.87)$$

where the sign depends on whether ϕ_1 is a bosonic field or a fermion field.

An important implication of BRST symmetry is the so-called *Slavnov-Taylor identity*, which, as we previously mentioned, is a generalisation of the Ward-Takahashi identity. A good description of its significance comes from the great man himself:

Quote 13.6 Slavnov-Taylor identities: relations between Green functions, which provide the gauge invariance of the quantum theory including non-Abelian gauge fields.

Andrei A. Slavnov, 2010

Now let us relate this to BRST symmetry. We now know that the gauge invariance of the Lagrangian is superseded by BRST invariance in non-Abelian gauge theories. Whereas the gauge symmetry in QED reflects itself in the Ward-Takahashi identity, BRST symmetry reflects itself in the Slavnov-Taylor identity.

Theorem 13.4 (Slavnov-Taylor identity) For an *effective Lagrangian* defined as

$$\mathcal{L}_{\text{eff}} = -\frac{1}{4}F_{\mu\nu}^a F_{\mu\nu}^a + \frac{1}{2\xi}(\partial_\mu A_\mu)^2 - \partial_\mu \bar{c}^a (D_\mu c)^a \quad (13.88)$$

The following path integral is always zero:

$$\int e^{i \int (\mathcal{L}_{\text{eff}} + J_\mu^a(x) A_\mu^a(x)) dx} \left(-\frac{1}{\xi} \partial_\mu A_\mu^a(y) + \int \bar{c}^a(y) J_\mu^b(z) [D_\mu c(z)]^b dz \right) \prod_x dA_\mu(x) d\bar{c}(x) dc(x) = 0 \quad (13.89)$$

The LHS is often called the *Slavnov-Taylor operator*.

Essentially, through the Slavnov-Taylor identity:

Quote 13.7 It is shown that all the divergences of the Yang-Mills theory can be removed by means of a renormalization of the charge and the wave function.

Andrei A. Slavnov, 23 June 1971

However, this is out of the scope of this book for the time being.

13.5 Quantum chromodynamics

Quote 13.8 Whole QCD in one week? Respect.

Felix Halbwedl, on a poorly designed standard model course in a certain

university, 12 October 2024

Now that we have formulated a generic non-Abelian gauge theory, it would seem as if QCD is nothing but a formality of setting $N = 3$ and inserting the corresponding generators. However, now everything carries a physical meaning:

When in QCD...	...becomes...
Fermion	Quark
Gauge boson	Gluon
Gauge boson field A_μ^a	Gluon field G_μ^a
$SU(N)$ generators	$SU(3)$ generators (Gell-Mann matrices)
Generator index	Colour index
(Arbitrary) coupling constant g	Strong coupling g_s

We first calculate the generators. In $SU(3)$, the transformations are

$$U = e^{i\alpha_a \frac{1}{2} \lambda^a} \quad (13.90)$$

In our adjoint representation, we have $N^2 - 1 = 3^2 - 1 = 8$ generators. Conventionally, we relabel our generators $T^a = \frac{1}{2}\lambda^a$ to be twice a series of matrices λ^a . These are the so-called *Gell-Mann matrices*⁹:

Definition 13.22 (Gell-Mann matrices)

$$\begin{aligned} \lambda_1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \lambda_2 &= \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} & \lambda_3 &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ \lambda_4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} & \lambda_5 &= \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \\ \lambda_6 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} & \lambda_7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} & \lambda_8 &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \end{aligned} \quad (13.91)$$

Yet this is not the end of the story. One might remember from undergraduate high energy physics that spin is governed by $SU(2)$. Its generators are Pauli matrices, whose eigenvectors are then spin eigenstates. In QCD, which is governed by $SU(3)$, many analogies to spin can be produced.

Firstly, $SU(3)$ Gell-Mann matrices are the equivalent of $SU(2)$ Pauli matrices. Just like how Pauli matrices give rise to spin, Gell-Mann matrices, as we now need to discuss the so-called *colours* which, like spin, are an intrinsic property of particles. The eigenstates of Gell-Mann matrices are different ‘basis states’ in the *colour space*. Linear combinations of these states then describe physical colour configurations of quarks and gluons.

Definition 13.23 (Colour basis) The eigenvectors of the Gell-Mann matrices are known as the *colour basis*:

$$c_1 = |\mathbf{r}\rangle = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad c_2 = |\mathbf{g}\rangle = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad c_3 = |\mathbf{b}\rangle = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (13.92)$$

Like $U(1)$ QED, where the single generator’s ‘eigenscalar’ gives rise to one type of electrical charge (and by that, negative charge), from the 3 colour eigenstates, we have 3 colours and 3 anticolours:

- *Colour charge*, which includes **red**, **green** and **blue**.
- *Colour anticharge*, which includes **antired**, **antigreen** and **antiblue**.

Remark 13.2 Due to Noether’s theorem, colour charge is conserved.

A general *colour state* can therefore be represented as

⁹Sometimes, we define a ‘zeroth’ Gell-Mann matrix $\lambda_0 = \mathbb{I}_3$. This is not a ‘real’ Gell-Mann matrix because $2\lambda_0$ does not satisfy $\det U = 1$ and is not a generator.

Definition 13.24 (Colour state)

$$c = \alpha|\mathbf{r}\rangle + \beta|\mathbf{g}\rangle + \gamma|\mathbf{b}\rangle \quad (13.93)$$

where α , β and γ are complex numbers that represent probabilities and phase relationships in quantum states, same as complex spin states.

Just like how Pauli matrices¹⁰ can be combined to form the so-called *ladder operators* to move between spin states, we can also combine the Gell-Mann matrices to form ladder operators to move between colour states:

Definition 13.25 (QCD ladder operators)

$$\underbrace{T_{\pm} = \frac{1}{2}(\lambda_1 \pm i\lambda_2)}_{\text{red} \rightarrow \text{green}} \quad \underbrace{V_{\pm} = \frac{1}{2}(\lambda_4 \pm i\lambda_5)}_{\text{red} \rightarrow \text{blue}} \quad \underbrace{U_{\pm} = \frac{1}{2}(\lambda_6 \pm i\lambda_7)}_{\text{green} \rightarrow \text{blue}} \quad (13.94)$$

There exist two more quantities of interest which arise from the SU(3) generators. If one inspects the whole range of Gell-Mann matrices, they will find that only two are diagonal: λ_3 and λ_8 . As it turns out, two new quantities are hidden in plain sight: the *colour isospin* I_3^c and the *colour hypercharge* Y^c ¹¹.

Definition 13.26 (Colour isospin) The matrix form of the colour isospin is

$$I_3^c = \frac{1}{2}\lambda_3 = \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (13.95)$$

from which one can derive the colour isospin for each colour along the basis $(\mathbf{r}, \mathbf{g}, \mathbf{b})$ as

$$I_3^c|\mathbf{r}\rangle = +\frac{1}{2}|\mathbf{r}\rangle \quad I_3^c|\mathbf{g}\rangle = -\frac{1}{2}|\mathbf{g}\rangle \quad I_3^c|\mathbf{b}\rangle = 0 \quad (13.96)$$

Definition 13.27 (Colour hypercharge) The matrix form of the colour hypercharge is

$$Y^c = \frac{1}{\sqrt{3}}\lambda_8 = \frac{1}{3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix} \quad (13.97)$$

from which one can derive the colour hypercharge for each colour along the basis $(\mathbf{r}, \mathbf{g}, \mathbf{b})$ as

$$Y^c|\mathbf{r}\rangle = \frac{1}{3}|\mathbf{r}\rangle \quad Y^c|\mathbf{g}\rangle = \frac{1}{3}|\mathbf{g}\rangle \quad Y^c|\mathbf{b}\rangle = -\frac{2}{3}|\mathbf{b}\rangle \quad (13.98)$$

In summary, the colour isospin and colour hypercharge each unit colour charge holds are as follows:

Charge	COLOUR CHARGE IN EACH UNIT COLOUR					
	red	green	blue	antired	antigreen	antiblue
I_3^c	1/2	-1/2	0	-1/2	1/2	0
Y^c	1/3	1/3	-2/3	-1/3	-1/3	2/3

Still, some puzzles remain when we consider colour neutrality. Naively, one might think that:

- Particles that have zero red, green and blue colour charge are colour-neutral. However, as we will see later, a baryon, which is colour-neutral, can have the structure $\mathbf{r} + \mathbf{g} + \mathbf{b}$.
- Particles that have zero I_3^c and Y^c are colour-neutral. However, as we will see later, gluons that correspond to λ_3 and λ_8 in (13.101) have zero I_3^c and Y^c merely because they are related to λ^3 and λ_8 , yet no gluons are colour-neutral.

This is strange.

¹⁰Pauli matrices

¹¹The naming corresponds to the isospin I_3 and the hypercharge Y we will see later in electroweak theory.

Quote 13.9 Some Puzzles.*Niklas Beisert, 2017*

In fact, both are very common perceptions that are often not clarified in literature. The second misunderstanding is particularly common because one often draws a 2D coordinate system, where I_3^c is the x -axis, and Y^c is the y -axis, and particles with coordinates $(0, 0)$ are likened, in layman's terms, to equal units of **red**, **green** and **blue** light mixing together to form white light in the RGB colour system. Here is where we reveal a plot twist: whether a particle is colour-neutral is related to neither having zero **red**, **green** and **blue** colour charges nor having zero I_3^c and Y^c , although a colour-neutral particle must satisfy the latter.

Theorem 13.5 (Colour neutrality) A particle is colour-neutral if its colour state c defined in (13.93) is invariant when transformed by all 8 Gell-Mann matrices.

One implication of this is the so-called *colour confinement* which, although verified experimentally, has not been proved analytically.

Theorem 13.6 (Colour confinement) A particle that is not colour-neutral cannot be observed on its own. Only colour-neutral particles can exist as standalone particles. Colour-neutral particles are known as *colour singlets* or *singlets*.

This has some implications:

- Gluons are not colour-neutral, and hence not free particles. This means that the strong coupling decreases with distance in what is known as *asymptotic freedom*, and that the strong force exists as a short-range force.
- Quarks, which carry colour charge, cannot be isolated, although (colour-neutral) quark-antiquark jets can. The formation of these jets, which are hadrons, is called *hadronisation*.
- Hypothetical *glueballs*, which are singlets made up of multiple gluons, are free particles.

Quote 13.10 Chromodynamics is weird, with couplings that grow with distances and...

Alessio Serafini, 27 March 2025

Note 13.4 (Conserved charge) It is worth noting that there is a charge associated with each Gell-Mann matrix (and with that, each generator), among them I_3^c and Y^c . These are the conserved charges in QCD, not the colours **red**, **green** and **blue** we often call 'colour charge'.

Now, we are in a position to briefly discuss the particles of interest in QCD: quarks and gluons. On their own, quarks are not very interesting, as colour confinement forbids their standalone appearance. Rather, our particles of interest are *hadrons* which, as our sneak peek in the last section suggested, are the natural conclusion of hadronisation.

A particle is a hadron as long as it is made of quarks. This means that a hadron could be either a fermion (pion, kaon, etc.) or a boson (proton, neutron, baryon, etc.). There are two kinds of hadrons:

- *Baryons*, which have an odd number of quarks (usually 3).
- *Mesons*, which have an odd number of quarks (usually 2: a quark and its antiquark).

The existence of gluons is also intuitive, perhaps more so if we work them out conceptually 'backwards'. It is now known that there are a total of 8 gluons. Interestingly, we also happen to have 8 Gell-Mann matrices. This is more than just a coincidence. In fact, every gluon is exactly a Gell-Mann matrix:

Definition 13.28 (Gluons) The collection of all 8 gluons is known as a so-called *octet*. We distinguish between two types of gluons:

- 6 gluons are not colour-neutral and mediate changes between different colour states^a:

$$g_1 = \frac{1}{\sqrt{2}}(|\mathbf{r}\rangle\langle\bar{\mathbf{g}}| + |\mathbf{g}\rangle\langle\bar{\mathbf{r}}|) \quad g_2 = \frac{i}{\sqrt{2}}(|\mathbf{r}\rangle\langle\bar{\mathbf{g}}| - |\mathbf{g}\rangle\langle\bar{\mathbf{r}}|) \quad g_4 = \frac{1}{\sqrt{2}}(|\mathbf{r}\rangle\langle\bar{\mathbf{b}}| + |\mathbf{b}\rangle\langle\bar{\mathbf{r}}|) \quad (13.99)$$

$$g_5 = \frac{i}{\sqrt{2}}(|\mathbf{r}\rangle\langle\bar{\mathbf{b}}| - |\mathbf{b}\rangle\langle\bar{\mathbf{r}}|) \quad g_6 = \frac{1}{\sqrt{2}}(|\mathbf{g}\rangle\langle\bar{\mathbf{b}}| + |\mathbf{b}\rangle\langle\bar{\mathbf{g}}|) \quad g_7 = \frac{i}{\sqrt{2}}(|\mathbf{g}\rangle\langle\bar{\mathbf{b}}| - |\mathbf{b}\rangle\langle\bar{\mathbf{g}}|) \quad (13.100)$$

These gluons serve to change the colour of a quark, so they must carry both colour charges and anticolour charges. For example, a gluon changing a **red** quark into a **blue** quark must itself carry **red** and **antiblue** colour charges.

- 2 gluons are colour-neutral. They do not change colours but still participate in the strong interaction:

$$g_3 = \frac{1}{\sqrt{2}}(|\mathbf{r}\rangle\langle\bar{\mathbf{r}}| - |\mathbf{g}\rangle\langle\bar{\mathbf{g}}|) \quad g_8 = \frac{1}{\sqrt{6}}(|\mathbf{r}\rangle\langle\bar{\mathbf{r}}| + |\mathbf{g}\rangle\langle\bar{\mathbf{g}}| - 2|\mathbf{b}\rangle\langle\bar{\mathbf{b}}|) \quad (13.101)$$

^aThis makes sense, as $3! = 6$.

Exercise 13.2 Insert the colour state vectors into the expressions and verify that each gluon is indeed a Gell-Mann matrix.

For completeness, we also have a *singlet* which is utterly unphysical:

$$g_9 = \frac{1}{\sqrt{3}}(|\mathbf{r}\rangle\langle\bar{\mathbf{r}}| + |\mathbf{g}\rangle\langle\bar{\mathbf{g}}| + |\mathbf{b}\rangle\langle\bar{\mathbf{b}}|) \quad (13.102)$$

By inserting the colour state vector, one can see that this is exactly \mathbb{I}_3 , which is *not* a Gell-Mann matrix (which we know to be traceless) but rather the extra traceful generator when one moves from SU(3) to U(3). Had this singlet indeed existed in real life, it would have implied a boson that would not interact with colour charges and imply a long-range force, which is not observed.

At this point, one might be tempted to ask why we mention this non-existent g_9 at all. This is because of the well-known expression

$$3 \otimes \bar{3} = 8 \oplus 1 \quad (13.103)$$

Our ‘retroactive’ verification of the Gell-Mann matrices from the gluon expressions tells us that Gell-Mann matrices can be essentially thought of as a combination of colour eigenvector inner products. However, if we take all possible inner product permutations of our colour eigenvectors, we get 9, not 8, matrices. This is the significance of (13.103): the 9 matrices we acquired from colour eigenvectors are the physical octet and the unphysical singlet.

Finally, let us compare SU(3) with some other Lie groups that encode physical concepts.

- SU(3) vs. U(1): U(1) electrical charge is something we can easily observe. However, colour is not physically observable: neither quarks nor gluons can be observed on their own. Despite this, QCD clearly cannot function without formulating this colour structure. In other words, colours do not directly correspond to physical quantities, but we know that *it's there*.
- SU(3) & U(1) vs. SU(2) in spin: While SU(2) spin does not correspond to any single gauge theory¹², U(1) electrical charge and SU(3) colours correspond to the gauge theories of QED and QCD respectively. Spin eigenstates correspond to global symmetries, while colour eigenstates and electrical charge correspond to local gauge symmetries¹³.

The second point offers a good opportunity to review the difference between global symmetries and local gauge symmetries. While the Pauli matrices (in the context of spin) represent actual physical transformations, the sole generator \mathbb{I}_1 in U(1) symmetry does not. Instead, it allows for arbitrary local changes in the phase of the electron field. As we have just seen, the Gell-Mann matrices do not represent physical transformations either.

¹²While the weak part of electroweak theory is governed by SU(2), the two merely use the same Lie group and do not relate to each other conceptually.

¹³Colour charges correspond to gauge symmetries, but not gauge fixing, as we can clearly see that colour charges have nothing to do with the gauge-fixing term in the Lagrangian.

Chapter 14

Electroweak theory

Quote 14.1 If you're not regularly questioning your own sanity, your life choices and the sanity of everybody else in the field, are you really specialising in HEP?

Dominic Hirtler, 8 August 2025

The weak force on its own is governed by quantum flavourdynamics (QFD), a non-Abelian gauge theory corresponding to the $SU(2)$ group. However, we rarely discuss QFD on its own. This is because, in pursuit of our final objective to construct a theory of everything, we want to unify the 3 out of 4 forces that we have managed to quantise. Through the so-called electroweak unification in the 1960s, we know that the electromagnetic and weak forces are collectively governed by electroweak theory, which is defined in relation to $SU(2) \times U(1)$ groups, where symmetry breaking gives mass to weak bosons. In this regard, QED and QFD are nothing but EFTs at specific energy levels.

14.1 Quantum flavourdynamics

We can introduce the weak force by a motivating example, which happens to be the way it was historically discovered. Consider the the so-called *beta decay*:



To conserve energy and momentum, we must have an additional particle $\bar{\nu}_e$ which we now call the *neutrino*.

Definition 14.1 (Neutrinos) There are a total of 6 neutrinos:

- The electron neutrino ν_e and antineutrino $\bar{\nu}_e$, which are associated with the electron and the positron.
- The muon neutrino ν_μ and antineutrino $\bar{\nu}_\mu$, which are associated with the muon and the antimuon.
- The tau neutrino ν_τ and antineutrino $\bar{\nu}_\tau$, which are associated with the tau and the antitau.

However, (14.1) is not the whole picture. To see why, it is necessary to introduce the concept of quark *flavours*.

Definition 14.2 (Flavours) We primarily categorise quarks by flavours, of which there are six types:

- **Up-type quarks:** up (u), charm (c) and top (t) have charge $\frac{2}{3}$.
- **Down-type quarks:** down (d), strange (s) and bottom (b) have charge $-\frac{1}{3}$.

Note 14.1 It does not take a genius to see that flavours with positive-sounding names have positive charge, while those with negative-sounding names have negative charge.

Definition 14.3 (Generation) There are three generations of quarks, charged leptons and neutrinos:

- First generation: up and down quarks, electron, electron neutrino
- Second generation: charm and strange quarks, muon, muon neutrino
- Third generation: top and bottom quarks, tau, tau neutrino

Conceptually, the flavours are highly similar to the colours we saw previously:

- To change colours, we need virtual particles called gluons, from which the strong force arises.
- Likewise, to change flavours, we need additional virtual particles from which the weak force arises.

Definition 14.4 (W and Z bosons) These new flavour-changing virtual particles are the W and Z bosons. We have:

- The W^+ boson, which has charge 1.
- The W^- boson, which has charge -1 .
- The Z boson, which has charge 0^a

^aFor this reason, it is sometimes labelled Z^0 , and its current is known as a *neutral current*.

Now let us relate this entire blurb to (14.1). It is well-known that the quark compositions of a neutron and a proton are udd and uud respectively. If we remove ud on each side, we have

$$d \rightarrow u + e^- + \bar{\nu}_e \quad (14.2)$$

Finally, we acknowledge that a flavour-changing virtual particle exists. As we have changed the $-\frac{1}{3}$ down quark to the $\frac{2}{3}$ up quark, a total charge of -1 must leave the down quark. Hence, our virtual particle must be W^- :

$$d \rightarrow u + W^- \rightarrow u + e^- + \bar{\nu}_e \quad (14.3)$$

Now that we have introduced the conceptual lead-up to the weak force, we can discuss its Lie theory and its unification with QED. In doing so, we will introduce two important quantities: the *isospin* or *weak isospin* I_3 and the *hypercharge* Y .

14.2 Lie theory

The Lie group that governs QFD is the so-called *isospin group* $SU(2)_L$. This group acts only on left-handed fermions, but is otherwise identical to the normal $SU(2)$ group. The choice of $SU(2)_L$ was motivated by experiments in the 1950s, which showed that parity is violated in weak interactions. In other words, only left-handed chiral particle spinors and right-handed chiral anti-particle spinors participate in the weak interaction.

Under $SU(2)_L$, the left-handed and right-handed 4-spinors transform as

$$\psi_L \rightarrow \psi'_L = e^{i\theta_a T^a} \psi_L \quad \psi_R \rightarrow \psi'_R = \psi_R \quad (14.4)$$

where T^a are the generators and θ_a are the parameters.

Now we discuss the generators themselves. We have labelled generators of $SU(2)_L$ representations as T^a , which are related to Pauli matrices by

$$T^a = \frac{\sigma^a}{2} \quad (14.5)$$

By solving for the eigenvectors of the Pauli matrices, we see that the flavour eigenstates are

$$|\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad (14.6)$$

which are the basis vectors of the isospin doublet.

From them, we can recover the Pauli matrices as well as the identity \mathbb{I}_2 :

$$\sigma^1 = |\downarrow\rangle\langle\uparrow| + |\uparrow\rangle\langle\downarrow| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma^2 = i|\downarrow\rangle\langle\uparrow| - i|\uparrow\rangle\langle\downarrow| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad (14.7)$$

$$\sigma^3 = |\uparrow\rangle\langle\uparrow| - |\downarrow\rangle\langle\downarrow| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad \mathbb{I}_2 = |\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow| = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (14.8)$$

The recovery of the Pauli matrices finally puts us in a position to introduce the isospin.

Definition 14.5 (Isospin) The matrix form of the isospin is

$$I_3 = \frac{1}{2}\sigma^3 \quad (14.9)$$

from which one can derive the isospin^a along the basis (\uparrow, \downarrow) as

$$I_3|\uparrow\rangle = +\frac{1}{2}|\uparrow\rangle \quad I_3|\downarrow\rangle = -\frac{1}{2}|\downarrow\rangle \quad (14.10)$$

^aSome texts treat this as the *3rd component of the isospin*, with the other two components arising from σ^1 and σ^2 .

In QCD, the 8 Gell-Mann matrices led to 8 gluons. Here, as we have 3 Pauli matrices, one can reasonably say that we have 3 gauge bosons, each corresponding to a gauge field. We then relabel A_μ^a as W_μ^a where $a = (1, 2, 3)$.

In real physics, we are not interested in these fields, which take real values and are not eigenstates of electric charge. The actually physical fields are instead linear combinations of W_μ^1 and W_μ^2 which are the gauge fields for the W^+ and W^- bosons we observe in real life:

Definition 14.6 (W bosons)

$$W^+ = \frac{1}{\sqrt{2}}(W_\mu^1 - iW_\mu^2) \quad W^- = \frac{1}{\sqrt{2}}(W_\mu^1 + iW_\mu^2) \quad (14.11)$$

These are complex-valued fields which, as we will see later, are eigenstates of the charge operator Q with eigenvalues $+1$ and -1 , which are the charges they carry.

However, we note two funny points:

- The W_μ^3 field is yet to correspond to anything physical. There is then an apparent redundancy of it.
- The Z boson, which we know from modern-day hindsight to exist, is missing.

Indeed, it soon became clear that $SU(2)_L$ alone, which does not predict the existence of the Z boson, was deficient in describing even the weak force itself. This problem was solved by *electroweak unification*, which remarkably demonstrates that 2 out of 4 fundamental forces are actually effective field theories of electroweak theory.

The question is then how we can do this. We cannot simply add the QED and QFD Lagrangians together because while the QED Lagrangian conserves parity, the QFD Lagrangian does not. We can see this by investigating the interaction term of the QED Lagrangian after the explicit insertion of (6.104)

$$J_{\text{QED}}^\mu = Q_e \bar{\psi}_e \gamma^\mu \psi_e = Q_e (\bar{e}_R \gamma^\mu e_R + \bar{e}_L \gamma^\mu e_L) \quad (14.12)$$

This term clearly observes $U(1)$ invariance. However, when transformed under $SU(2)_L$, the existence of the term $\bar{e}_L \gamma^\mu e_L$ prevents the invariance of the current.

This can be resolved by realising that we have been wrong to assume that QED currents, etc., can be used as-is. In fact, our original $U(1)$ group is a manifestation of the effective field theory nature of QED. The actual¹ Lie group is $U(1)_Y$. It is *mathematically identical* to $U(1)$, but its gauge field and charge are considered distinct from the photon field and electrical charge Q brought about by the original $U(1)$:

- The gauge field of $U(1)_Y$ is labelled B^μ .

¹Read: less obviously EFT-looking, if further unification is to be believed.

- The charge of $U(1)_Y$ is the *hypercharge* Y .

Remark 14.1 The colour isospin I_3^c and colour hypercharge Y^c are analogous to I_3 and Y . Both isospins correspond to the third generator of their respective Lie groups.

As it turns out, the Z boson and the photon are nothing but orthogonal linear combinations of the fields W_μ^3 and B_μ .

Definition 14.7 (Z boson and photon)

$$Z_\mu = W_\mu^3 \cos \theta_W - B_\mu \sin \theta_W \quad A_\mu = W_\mu^3 \sin \theta_W + B_\mu \cos \theta_W \quad (14.13)$$

where θ_W is the so-called *Weinberg angle* related to the $SU(2)_L$ coupling g and the $U(1)_Y$ coupling g' :

$$\tan \theta_W = \frac{g'}{g} \quad (14.14)$$

Quote 14.2 Where is the Weinberg?

Felix Halbwedl, 2 April 2025

Remark 14.2 One funny implication exists here: The photon is a neutral boson like Z , putting them in the same category.

As we will see later, the construction of the electroweak Lagrangian is extremely tedious as a result of symmetry breakings. For now, we are content to first study the covariant derivative, which will put us closer to a position to construct (or rather guess) the Lagrangian. As we already have all the gauge fields, this is unsurprisingly simple:

Definition 14.8 (Electroweak covariant derivative)

$$D_\mu = \partial_\mu - ig \frac{\sigma^a}{2} W_\mu^a - ig' \frac{Y}{2} B_\mu \quad (14.15)$$

where, again, g and g' are the $SU(2)_L$ and $U(1)_Y$ couplings.

We can prove that a portion of this is the QED covariant derivative contribution $-ieQA_\mu$.

Derivation 14.1 (Electrical charge) We first rewrite (14.13) as

$$W_\mu^3 = Z_\mu \cos \theta_W + A_\mu \sin \theta_W \quad B_\mu = -Z_\mu \sin \theta_W + A_\mu \cos \theta_W \quad (14.16)$$

By inserting this into (14.15), we see that

$$D_\mu = \partial_\mu - i \left[\left(\frac{g}{2} \sigma^3 \cos \theta_W - \frac{g'}{2} Y \sin \theta_W \right) Z_\mu + \left(\frac{g}{2} \sigma^3 \sin \theta_W + \frac{g'}{2} Y \cos \theta_W \right) A_\mu \right] \quad (14.17)$$

Recognising the part that corresponds to $-ieQA_\mu$, we write

$$eQ = g \frac{\sigma^3}{2} \sin \theta_W + g' \frac{Y}{2} \cos \theta_W \quad (14.18)$$

We know that electric charge is a linear combination of I_3 and Y :

$$Q = aI_3 + bY = a \frac{\sigma^3}{2} + bY \rightarrow eQ = ea \frac{\sigma^3}{2} + ebY \quad (14.19)$$

Comparing this to the previous expression gives

$$ea = g \sin \theta_W \quad eb = \frac{g'}{2} \cos \theta_W \quad (14.20)$$

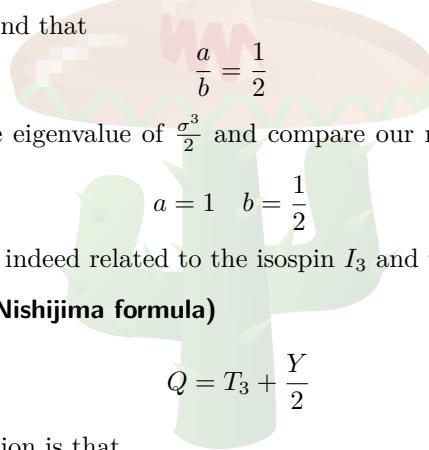
a and b can then be easily solved as

$$a = \frac{g \sin \theta_W}{e} \quad b = \frac{g' \cos \theta_W}{2e} \quad (14.21)$$

Dividing one formula by another yields

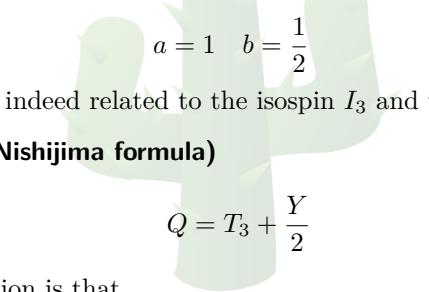
$$\frac{a}{b} = \frac{g \sin \theta_W}{g' \cos \theta_W} \frac{g' \cos \theta_W e}{2e} \quad (14.22)$$

If we consult (14.14), we will find that



$$\frac{a}{b} = \frac{1}{2} \quad (14.23)$$

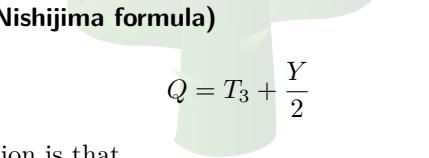
We acknowledge that I_3 is the eigenvalue of $\frac{\sigma^3}{2}$ and compare our results against (14.18). It is then clear that



$$a = 1 \quad b = \frac{1}{2} \quad (14.24)$$

Hence, the electric charge Q is indeed related to the isospin I_3 and the hypercharge Y .

Theorem 14.1 (Gell-Mann-Nishijima formula)



$$Q = T_3 + \frac{Y}{2} \quad (14.25)$$

From (14.21), another implication is that

$$e = g \sin \theta_W = g' \cos \theta_W \quad (14.26)$$

14.3 Spinors

Now we want to construct the fermion 4-spinors (Dirac spinors) of this theory. Similar to the Weyl basis in QED (6.104), we can split a 4-spinor into two parts, an upper 2-spinor (Weyl spinor) and a lower 2-spinor. We call 4-spinors made up of two non-zero 2-spinors *doublets*, and 2-spinors as well as 2-spinors made up of 1 non-zero 2-spinor and 1 zero 2-spinor *singlets*. There are two types of doublets:

- Doublets representing neutrino-lepton pairs.
- Doublets representing up-type quark-down-type quark pairs.

In either case, the two particles in the doublet *must* belong to the same generation.

Definition 14.9 (Neutrinos and leptons)

- Isospin doublets:

$$\psi_L^1 = \begin{pmatrix} \nu_e \\ e_L \end{pmatrix} \quad \psi_L^2 = \begin{pmatrix} \nu_\mu \\ \mu_L \end{pmatrix} \quad \psi_L^3 = \begin{pmatrix} \nu_\tau \\ \tau_L \end{pmatrix} \quad (14.27)$$

- Isospin singlets:

$$\psi_R^1 = \begin{pmatrix} 0 \\ e_R \end{pmatrix} \quad \psi_R^2 = \begin{pmatrix} 0 \\ \mu_R \end{pmatrix} \quad \psi_R^3 = \begin{pmatrix} 0 \\ \tau_R \end{pmatrix} \quad (14.28)$$

The superscript denotes the generation of the singlet/doublet. For a lepton of interest $\ell = (e, \mu, \tau)$, ν_ℓ and ℓ are not particles themselves (as this would make no sense), but rather their corresponding 2-spinors.

Note 14.2 (Sterile neutrinos) The implication should be clear: while a lepton has left-handed and right-handed components, there are no right-handed neutrinos (better known as *sterile neutrinos*) in the standard model. However, they exist in many BSM theories.

Definition 14.10 (Quarks)

- Quark doublets:

$$Q_L^1 = \begin{pmatrix} u_L \\ d_L \end{pmatrix} \quad Q_L^2 = \begin{pmatrix} c_L \\ s_L \end{pmatrix} \quad Q_L^3 = \begin{pmatrix} t_L \\ b_L \end{pmatrix} \quad (14.29)$$

- Quark up-type singlets:

$$Q_{R\uparrow}^1 = \begin{pmatrix} u_R \\ 0 \end{pmatrix} \quad Q_{R\uparrow}^2 = \begin{pmatrix} c_R \\ 0 \end{pmatrix} \quad Q_{R\uparrow}^3 = \begin{pmatrix} t_R \\ 0 \end{pmatrix} \quad (14.30)$$

- Quark down-type singlets:

$$Q_{R\downarrow}^1 = \begin{pmatrix} 0 \\ d_R \end{pmatrix} \quad Q_{R\downarrow}^2 = \begin{pmatrix} 0 \\ s_R \end{pmatrix} \quad Q_{R\downarrow}^3 = \begin{pmatrix} 0 \\ b_R \end{pmatrix} \quad (14.31)$$

We summarise the structures of these doublets and singlets:

- Doublets are always left-handed.
- Singlets are always right-handed.
- Up-type quarks structurally correspond to neutrinos.
- Down-type quarks structurally correspond to leptons.

This is due to the weak part of the electroweak Lie group being $SU(2)_L$.

Using Noether's theorem, each gauge symmetry corresponds to a conserved current. In electroweak theory, the hypercharge current J_Y^μ supersedes the QED 4-current J_{QED}^μ . From the Gell-Mann-Nishijima formula, an analogous relation can be acquired for the QED, isospin and hypercharge currents:

$$\begin{aligned} J_Y^\mu &= 2 \left(J_{\text{QED}}^\mu - J_3^\mu \right) \\ &= \bar{e}_R \gamma^\mu Y_{R,e} e_R + \bar{\nu}_R \gamma^\mu Y_{R,\nu} \nu_R + \bar{\chi}_L \gamma^\mu Y_{L,\chi} \chi_L \\ &= \bar{e}_R \gamma^\mu Y_{R,e} e_R + \bar{\nu}_R \gamma^\mu Y_{R,\nu} \nu_R + \bar{e}_L \gamma^\mu Y_{L,\chi} e_L + \bar{\nu}_L \gamma^\mu Y_{L,\chi} \nu_L \end{aligned} \quad (14.32)$$

where χ_L is the doublet $(\nu_L, e_L)^T$, and $Y_{R,e}$, $Y_{R,\nu}$ and $Y_{L,\chi}$ are the hypercharge of e_R , ν_R and χ_L . Let us compare this with the standalone isospin and QED currents:

$$J_\mu^3 = \bar{\chi}_L \gamma_\mu \frac{\sigma_3}{2} \chi_L = \frac{1}{2} \bar{\nu}_L \gamma_\mu \nu_L - \frac{1}{2} \bar{e}_L \gamma_\mu e_L \quad J_\mu^{\text{QED}} = Q_e (\bar{e}_R \gamma_\mu e_R + \bar{e}_L \gamma_\mu e_L) = J_\mu^3 + \frac{1}{2} J_\mu^Y \quad (14.33)$$

By matching the coefficients, we find that

$$Y_{R,e} = 2Q_e \quad Y_{R,\nu} = 0 \quad Y_{L,e} = 2Q_e + 1 \quad Y_{L,\nu} = -1 \quad (14.34)$$

In summary, electroweak theory is not governed by one Lie group, but rather two groups which combine to form the $SU(2)_L \times U(1)_Y$ group. There exists a total of 4 gauge bosons, emerging from the 3 and 1 generators of the two groups put together, which are not physically observable. Instead, they combine to form our good friends, the W and Z bosons and the photons. This is also known as the *Glashow-Weinberg-Salam (GWS) theory*.

Quote 14.3 In some sense, you already worked together with the Great Weiny.

Felix Halbwedel, 2 April 2025

14.4 Local gauge symmetry breaking

In Part III, we made a superficial discussion on global spontaneous symmetry breaking. Here, we see how local gauge symmetries are also spontaneously broken. In electroweak theory, this leads to the so-called *Higgs boson*. Before deriving the actual Higgs boson, we will demonstrate how local gauge symmetry takes place via a toy model based on QED².

Derivation 14.2 (Abelian toy model) We restrict ourselves to purely the free fermion sector. The Lagrangian is identical to its QED counterpart except for the addition of one extra self-interacting term

²This is similar to how we introduced Feynman diagrams via the toy model of ϕ^4 theory.

$-\lambda(\phi^*\phi)^2$. This term is equivalent to the ϕ^4 theory self-interacting term, but with the Grassmann-oddness of the fermion field taken into account. With covariant derivatives written explicitly, this self-interacting free fermion Lagrangian is

$$\mathcal{L} = (\partial_\mu - ieA_\mu)\phi^*(\partial^\mu + ieA^\mu)\phi - \mu^2\phi^*\phi - \lambda(\phi^*\phi)^2 \quad (14.35)$$

where μ and λ are mass- and ϕ^4 -like couplings.

Like (8.60), we perturb the field around the VEV, which in this case is $v^2 = -\mu^2/\lambda$, where the implication is that we must have either $\mu^2 < 0$ and $\lambda > 0$ or $\mu^2 > 0$ and $\lambda < 0$. However, as the fermion field is complex, we introduce a real component $\eta(x)$ and a complex component $i\xi(x)$

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \eta(x) + i\xi(x)] \quad (14.36)$$

where the significance of the factor $\frac{1}{\sqrt{2}}$ will be revealed later. The toy model Lagrangian then reads

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\xi)^2 + \frac{1}{2}(\partial_\mu\eta)^2 - \frac{1}{2}(2v^2\lambda^2)\eta^2 + \frac{1}{2}e^2v^2A_\mu A^\mu - evA_\mu\partial^\mu\xi \quad (14.37)$$

Here:

- ξ has only a kinetic term, making it a massless Goldstone boson.
- η has both a kinetic term and a mass term, making it a massive scalar particle with mass $\sqrt{2}\lambda v$.
- A_μ has both a kinetic term^a and a mass term, making it a massive vector particle with mass ev .
- η and A_μ interact by dint of the term $-evA_\mu\partial^\mu\xi$.

The massive nature of A^μ is the outstanding problem, as it raises the degrees of freedom from 2 to 3. There are two implications:

- We need a gauge transformation to kill 1 of the 3 degrees of freedom.
- This extra degree of freedom (which might not be the massive A_μ) represents an unphysical field.

Indeed, this is where we see the significance of the factor $\frac{1}{\sqrt{2}}$ in (14.35), which can be written to the first order expansion as

$$\phi(x) = \frac{1}{\sqrt{2}}[v + \eta(x)]e^{i\xi(x)/v} \quad (14.38)$$

So in some sense, the Goldstone boson field $\xi(x)$ is not a field but rather a parameter like $\Lambda(x)$ in (6.131), but even this statement is incomplete.

Quote 14.4 So you are no booker, but booked, in some sense.

Felix Halbwedl, on the author, 22 May 2025

Let us remind ourselves that we are still concerned with the same gauge transformation (6.131), only using a slightly different toy model Lagrangian. It then becomes clear that $\xi(x)$ is in fact a rescaled version $\Lambda(x)$:

$$\frac{\xi(x)}{v} = \Lambda(x) \quad (14.39)$$

Ultimately, by rewriting the (unphysical) Goldstone boson field $\xi(x)$ as a gauge transformation, it is absorbed into instances of the gauge field A_μ , which we now relabel as \tilde{A}_μ . In literature, this is known as the eating of the Goldstone boson by the gauge field.

Quote 14.5 Yum. Yum. Yum.

Gauge field \tilde{A}_μ , as it eats the Goldstone boson $\xi(x)$

We relabel the *real* scalar field η as h . The Lagrangian then reads

$$\mathcal{L} = \frac{1}{2}(\partial_\mu h)^2 - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 \tilde{A}_\mu \tilde{A}^\mu - \lambda v h^3 - \frac{1}{4}\lambda h^4 + \frac{1}{2}e^2 \tilde{A}_\mu^2 h^2 + v e^2 \tilde{A}_\mu^2 h \quad (14.40)$$

where h is the Higgs field.

^aThis is the field strength term we omitted earlier in (14.35).

^bThis may seem off, but it is actually a taste of things to come in (14.40).

So what have we learned? Essentially, the Goldstone bosons can be treated as gauge transformations and hence be eaten by (i.e. absorbed into) the gauge field. This is a gauge choice known as the *unitary gauge*.

14.5 Higgs mechanism

We have made clear that what we did in the last section was nothing but a toy model. Now we will demonstrate that this eating of the Goldstone bosons also manifests in non-Abelian gauge theories and is exactly what endows mass to the W and Z bosons.

Derivation 14.3 (W boson, Z boson and photon masses) First, we see that spontaneous symmetry breaking allows us to predict the mass of the W and Z bosons and the photon. Again, we start with a similar-looking Lagrangian

$$\mathcal{L} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2 \phi^\dagger \phi - \lambda (\phi^\dagger \phi)^2 \quad (14.41)$$

where D_μ is the covariant derivative in (14.15), and ϕ is a so-called doublet of scalar fields given by

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \equiv \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} \quad (14.42)$$

The vacuum expectation value lies at

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad v = \sqrt{\frac{\mu^2}{\lambda}} \quad (14.43)$$

Again we perturb this vacuum expectation value by both a real and a complex component. This is

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) + i\xi^a(x) \end{pmatrix} \quad (14.44)$$

The unitary gauge is implemented by recognising the complex part as yet another complex exponential to the first-order expansion:

$$\phi(x) = \frac{1}{\sqrt{2}} e^{i \frac{\xi^a(x) \sigma^a}{v}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (14.45)$$

Now we compute the Lagrangian. In doing so, it is often convenient to combine the gauge fields into a single 2×2 matrix \mathcal{W}_μ :

$$\mathcal{W}_\mu = \frac{g}{2} \sigma^a W_\mu^a + \frac{g'}{2} B_\mu \mathbb{I} \quad (14.46)$$

This is possible as the $SU(2)_L$ part is written in terms of the 2×2 Pauli matrices σ^a , while the hypercharge term from $U(1)_Y$ is proportional to the identity matrix \mathbb{I} . The covariant derivative reads

$$D_\mu \phi = (\partial_\mu - i\mathcal{W}_\mu) \phi \quad (14.47)$$

Let us write \mathcal{W}_μ explicitly:

$$\mathcal{W}_\mu = \frac{g}{2} \begin{pmatrix} W_\mu^3 & W_\mu^1 - iW_\mu^2 \\ W_\mu^1 + iW_\mu^2 & -W_\mu^3 \end{pmatrix} + \frac{g'}{2} \begin{pmatrix} B_\mu & 0 \\ 0 & B_\mu \end{pmatrix} \quad (14.48)$$

This simplifies to

$$\mathcal{W}_\mu = \begin{pmatrix} \frac{g}{2} W_\mu^3 + \frac{g'}{2} B_\mu & \frac{g}{\sqrt{2}} W_\mu^+ \\ \frac{g}{\sqrt{2}} W_\mu^- & -\frac{g}{2} W_\mu^3 + \frac{g'}{2} B_\mu \end{pmatrix} \quad (14.49)$$

After the eating of the Goldstone bosons, the gauge boson terms of $(D_\mu \phi)^\dagger (D^\mu \phi)$ thus reads

$$\begin{aligned} (D_\mu \phi)^\dagger (D^\mu \phi)_{\text{gauge bosons}} &= \frac{1}{8} v^2 g^2 \left(|W_\mu^1|^2 + |W_\mu^2|^2 \right) + \frac{1}{8} v^2 (g' B_\mu - g W_\mu^3) (g' B^\mu - g W^{3\mu}) \\ &= \left(\frac{1}{2} g v \right)^2 W_\mu^+ W^{-\mu} + \frac{1}{8} v^2 (g' B_\mu - g W_\mu^3)^2 + M_A^2 (g' W_\mu^3 + g B_\mu)^2 \\ &= M_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} M_Z Z_\mu Z^\mu \end{aligned} \quad (14.50)$$

where we have recognised that

Definition 14.11 (W bosons, Z boson and photon masses)

$$M_W = \frac{1}{2}gv \quad M_Z = \frac{1}{2}v\sqrt{g^2 + g'^2} \quad M_A = 0 \quad (14.51)$$

What is the takeaway from this? The photon mass is zero, while the masses of the W bosons, the Z boson and the photon only depend on the VEV field and the two couplings, which are nothing but physical constants that we can measure experimentally. Hence, our theory has predicted the mass (or lack thereof) of the four particles. This endowment of mass by the perturbation (which we know from earlier to be the Higgs boson) is the so-called *Higgs mechanism*.

^aThis is because the Higgs doublet has $Y = 1$.

Derivation 14.4 (Higgs boson mass) The only field that remains to be derived is the Higgs field itself. We again recognise the perturbation $\eta(x)$ as the Higgs field h and evaluate the Lagrangian. Using the electroweak covariant derivative (14.15), the kinetic term becomes

$$(D_\mu \phi)^\dagger (D^\mu \phi) = \frac{1}{2} \partial_\mu h \partial^\mu h + \text{etc.} \quad (14.52)$$

where the unwritten terms are the gauge boson mass terms we discussed earlier, and the scalar field kinetic term is

$$\frac{1}{2}(\partial_\mu h)^2 \quad (14.53)$$

Now we investigate the scalar potential or the so-called *Higgs potential*, which reads

Definition 14.12 (Higgs potential)

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2 \quad (14.54)$$

From (14.45), we can write

$$\phi^\dagger \phi = \frac{1}{2}(v + h)^2 \quad (14.55)$$

Inserting this into the potential gives it the sombrero-looking structure

$$\begin{aligned} V(h) &= \mu^2 \frac{1}{2}(v + h)^2 + \lambda \left(\frac{1}{2}(v + h)^2 \right)^2 \\ &= \frac{\mu^2}{2}(v^2 + 2vh + h^2) + \frac{\lambda}{4}(v^2 + 2vh + h^2)^2 \end{aligned} \quad (14.56)$$

We expand this up to second order in h^a :

$$V(h) = \text{const.} + (\mu^2 v + \lambda v^3)h + \left(\frac{1}{2}\mu^2 + \frac{3}{2}\lambda v^2 \right) h^2 + \dots \quad (14.57)$$

But at the minimum of the potential, v , we must have

$$\left. \frac{\partial V}{\partial h} \right|_{h=0} = 0 \rightarrow \mu^2 v + \lambda v^3 = 0 \quad (14.58)$$

So the linear term cancels, and the quadratic term becomes:

$$\frac{1}{2}(2\lambda v^2)h^2 \quad (14.59)$$

Hence, the mass of the Higgs boson is

Definition 14.13 (Higgs boson mass)

$$M_H = \sqrt{2\lambda}v \quad (14.60)$$

This expression hides a sinister implication. Unlike the other couplings g and g' , which are constrained by the need of gauge invariance in the $SU(2)_L$ and $U(1)$ groups, λ arises from the Yukawa interaction and is a free parameter that can assume any value in the standard model. That is to say, we have not predicted the mass of the Higgs boson.

Quote 14.6 The mass of the Higgs boson cannot be determined from theory (...) This might be the clue that the standard model is no fundamental theory, but rather an effective one.

Felix Halbwedl, 31 January 2024

^aThe higher-order terms correspond to interactions which do not concern the mass of the Higgs boson, which is what we are first interested in.

The Higgs boson endows masses to not only bosons, but also fermions. The latter is accomplished using the so-called *Yukawa interaction* or *Yukawa coupling*. Here, we need to make a spontaneous symmetry breaking of *the Higgs field itself*.

Derivation 14.5 (Fermion mass) If one considers only one flavour, the Yukawa interaction Lagrangian between a fermion field and the Higgs doublet is:

Definition 14.14 (One-flavour Yukawa Lagrangian)

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L H \psi_R + \text{h.c.} \quad (14.61)$$

where y_f is the Yukawa coupling constant for fermion f , H is the Higgs doublet, and h.c. denotes the hermitian conjugate version of the previous terms^a.

^aThat is to say, we repeat the same terms twice, once without and once with hermitian conjugation. This is to make sure that the Lagrangian is Hermitian.

The Higgs doublet looks very much like the fermion scalar field doublet:

$$H = \begin{pmatrix} H^+ \\ H^0 \end{pmatrix} \quad (14.62)$$

After spontaneous symmetry breaking, the neutral component H^0 acquires a VEV:

$$H_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \quad (14.63)$$

In the unitary gauge, the Higgs doublet becomes

$$H(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \quad (14.64)$$

where $h(x)$ is the *physical* Higgs boson field.

Substituting this into the Yukawa interaction yields

$$\mathcal{L}_{\text{Yukawa}} = -y_f \bar{\psi}_L \left(\frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \right) \psi_R + \text{h.c.} \quad (14.65)$$

Only the lower component of the Higgs doublet contributes:

$$\mathcal{L}_{\text{Yukawa}} = -\frac{y_f}{\sqrt{2}} (v + h(x)) \bar{\psi}_L \psi_R + \text{h.c.} = -\frac{y_f v}{\sqrt{2}} \bar{\psi} \psi - \frac{y_f}{\sqrt{2}} h(x) \bar{\psi} \psi + \text{h.c.} \quad (14.66)$$

We now identify the mass term and the Higgs-fermion interaction term as

$$-m_f \bar{\psi} \psi - \frac{y_f}{\sqrt{2}} h(x) \bar{\psi} \psi \quad (14.67)$$

where we have recovered the fermion mass as

Definition 14.15 (Fermion mass)

$$m_f = \frac{y_f v}{\sqrt{2}} \quad (14.68)$$

That is to say, the mass term actually originates from spontaneous symmetry breaking and the Higgs mechanism emerging from it. The extra term represents the Higgs-fermion vertex and is proportional to the fermion mass. Interestingly, we see that the fermion mass cannot be predicted from theory either due to its dependence on the arbitrary coupling y_f that emerges from the Yukawa interaction and is not constrained by gauge invariance.

This is where we reveal the ‘white lie’ that we ran with up to this point in the book. In all previous chapters, we made the assumption that fermions have massive scalar fields on their own. Now that we have introduced the Higgs mechanism, we know that this is not true. Rather, the fermions start out massless and only acquire mass when one considers the Higgs mechanism in the Yukawa interaction term. The newfound mass term emerging from the Higgs mechanism is then actually the mass term we have unconditionally assumed to exist in previous chapters all along.

The Higgs mechanism gives rise to one more funny implication. From the structure of the Yukawa Lagrangian, it is clear that the Higgs field gives mass to fermions by coupling left- and right-chiral fields. Once the Higgs gets a vacuum expectation value, this becomes the Yukawa coupling term $m\bar{\psi}_L\psi_R$ which ties chirality and mass together.

14.6 Quark and neutrino mixing

There is one loose end to tie up before putting everything together. We discuss the vertices that result in the mixing of particles. Even though we will not discuss the electroweak Feynman rules for the time being, we acknowledge that:

- A W boson can decay into a quark pair that has the same amount of charge (i.e. +1 and -1). Or rather, a up/down-type quark becomes a down/up-type quark by emitting a $W^+/'W^-$ boson.
- A W^- boson can decay into a lepton and its corresponding *antineutrino*. Or rather, a lepton can become its corresponding *neutrino* by emitting a W^- boson. The same applies to all charges reversed (i.e. W^+ boson).

The coupling itself is the same in all possible vertices. This is known as *lepton universality*. However, the mass eigenstates of the quarks and the neutrinos, which are 4-spinors, are actually rescaled, unlike in QCD. We show this for the first case, which rescales the quark mass eigenstates.

Derivation 14.6 (Quark mixing) In (14.61), we only considered the Yukawa Lagrangian with one generation. This can be generalised into three generations:

- The fermion mass coupling m_f becomes a 3×3 matrices Y_u and Y_d , which are the *Yukawa coupling matrices* for up- and down-type quarks across all generations.
- The left doublet $\bar{\psi}_L$ becomes \bar{Q}_L , an *array* of 3 left doublets across all generations.
- ψ_R become u_R and d_R , 3-vectors of right-handed up- and down-type singlets across all generations.

So the Yukawa Lagrangian becomes

Definition 14.16 (Yukawa Lagrangian)

$$\mathcal{L}_Y = -\bar{Q}_L Y_d H d_R - \bar{Q}_L Y_u \tilde{H} u_R + \text{h.c.} \quad (14.69)$$

where y_f is the Yukawa coupling constant for fermion f and H is the Higgs doublet.

Like in (14.68), the up- and down-type quark masses are generated via symmetry breaking and the Higgs mechanism. This time, however, the quark masses are matrices directly proportional to the Yukawa coupling matrices:

$$M_u = \frac{v}{\sqrt{2}} Y_u \quad M_d = \frac{v}{\sqrt{2}} Y_d \quad (14.70)$$

The mass matrices M_u and M_d are generally not diagonal, but we can perform biunitary transformations to make them so. The new basis they live in is the so-called *mass eigenbasis*.

$$M_u^{\text{diag}} = V_{uL}^\dagger M_u V_{uR} \quad M_d^{\text{diag}} = V_{dL}^\dagger M_d V_{dR} \quad (14.71)$$

These unitary matrices unitarily diagonalise the up- and down-type quark mass matrices by rotating the quark fields as

$$u_L \rightarrow V_{uL} u_L \quad u_R \rightarrow V_{uR} u_R \quad d_L \rightarrow V_{dL} d_L \quad d_R \rightarrow V_{dR} d_R \quad (14.72)$$

Let us now look at the weak charged current interaction, which couples only the left-handed quark doublets:

$$\mathcal{L}_{\text{CC}} = \frac{g}{\sqrt{2}} \bar{u}_L^\alpha \gamma^\mu W_\mu^+ d_L^\alpha + \text{h.c.} \quad (14.73)$$

After rotating to the mass basis, this becomes

$$\mathcal{L}_{\text{CC}} = -\frac{g}{\sqrt{2}} \bar{u}_L^{\text{mass}} \gamma^\mu W_\mu^+ (V_{uL}^\dagger V_{dL}) d_L^{\text{mass}} + \text{h.c.} \quad (14.74)$$

The term $V_{uL}^\dagger V_{dL}$ essentially encodes the mismatch between the rotations that are needed to diagonalise the up- and down-type quark mass matrices. It is known as the *CKM* (Nicola Cabibbo, Makoto Kobayashi and Toshihide Maskawa) *matrix*:

$$V_{\text{CKM}} = V_{uL}^\dagger V_{dL} \quad (14.75)$$

There are 3 generations of up and down quarks, so V_{CKM} has 9 components. These are experimentally shown as

Definition 14.17 (CKM matrix)

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \approx \begin{pmatrix} 0.974 & 0.225 & 0.004 \\ 0.225 & 0.973 & 0.041 \\ 0.009 & 0.040 & 0.999 \end{pmatrix} \quad (14.76)$$

The specific component V_{ij} is then inserted into the specific vertex involving the i and j quarks and the W boson as a scaling factor.

Now let us see the CKM matrix in action, with it, the coupling reads

$$\mathcal{L}_{\text{CC}} = -\frac{g}{\sqrt{2}} \bar{u}'_i \gamma^\mu V_{ij} d'_{jL} W_\mu^+ + \text{h.c.} \quad (14.77)$$

where W^+ denotes the W^+ boson propagator, and \bar{u}' and d' are the original \bar{u} and d quark (mass) eigenstates, which are 4-spinors, that absorbed the CKM matrix. For our convenience, and by convention, we absorb the CKM matrix components entirely into the down-type spinors instead of the up-type spinors. Hence, for given quark mass eigenstates d , s and b , the CKM acts as follows:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix} \quad (14.78)$$

Technically, it is possible to make the conventions such that the CKM components are absorbed by up-type quark mass eigenstates instead:

$$\begin{pmatrix} u' \\ c' \\ t' \end{pmatrix} = V_{\text{CKM}}^\dagger \begin{pmatrix} u \\ c \\ t \end{pmatrix}$$

However, we did not do so historically.

Finally, some miscellaneous comments on the CKM matrix can be made:

- It is a unitary matrix.

- It arises from misalignment between the flavour bases of the up- and down-type left-handed quarks.
- Each element $(V_{CKM})_{ij}$ gives the strength of the coupling between up-type quark u_i and down-type quark d_j via the charged weak interaction mediated by the W (but not Z) bosons.
- The diagonal entries (e.g., V_{ud}, V_{cs}, V_{tb}) are close to 1. This shows that generation-changing transitions are suppressed^a, and most weak transitions happen within the same generation.
- In QCD, the analogous matrix is simply \mathbb{I}_3 , meaning that the strong interaction forbids generation changing.

^aThe term ‘suppressed’ means that they are insignificant.

Remark 14.3 An interesting case is Kaon mixing, where 4 CKM components appear in the transition amplitude.

Now we consider the second case, where the neutrino mass eigenstates are rescaled³. The Lagrangian that represents the lepton-neutrino charge current is

$$\mathcal{L}_{CC} = -\frac{g}{\sqrt{2}} \bar{\ell}_{\alpha L} \gamma^\mu U_{\alpha i} \nu_{i L} W_\mu^- + \text{h.c.} \quad (14.79)$$

Where $U_{\alpha i}$ is a neutrino-mixing matrix called the *PMNS* (Bruno Pontecorvo, Ziro Maki, Masami Nakagawa, and Shoichi Sakata) *matrix*. The derivation is identical, but with a few relabelings:

- The left 3-quark doublet array \bar{Q}_L becomes its lepton counterpart \bar{L}_L .
- The 3-quark singlet arrays u_R and d_R become their lepton and neutrino counterparts ℓ_R and ν_R respectively⁴.

Remark 14.4 Both the CKM and PMNS matrices have 3 mixing angles and 1 physical CP-violating phase. For Majorana neutrinos, there are 2 additional CP-violating Majorana phases.

14.7 Feynman rules

Due to the large number of Feynman diagram components, we will not go through with the quantisation. In any case, most of the relevant techniques have been covered up to this point. We will simply state the Lagrangian and list the Feynman rules.

Definition 14.18 (Electroweak theory Lagrangian) The total Lagrangian is

$$\mathcal{L} = \mathcal{L}_{\text{gauge field}} + \mathcal{L}_{\text{matter field}} - V(\phi) + \mathcal{L}_{\text{Higgs, gauge}} + \mathcal{L}_{\text{Higgs, matter}} \quad (14.80)$$

We list all its parts:

- The gauge field Lagrangian is

$$\mathcal{L}_{\text{gauge field}} = -\frac{1}{4} W_{\mu\nu}^i W_i^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \quad (14.81)$$

where the field strength tensors read

$$W_{\mu\nu}^i = \partial_\mu W_\nu^i - \partial_\nu W_\mu^i - g \varepsilon^{ijk} W_\mu^j W_\nu^k \quad B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \quad (14.82)$$

- The matter field Lagrangian is

$$\mathcal{L}_{\text{matter field}} = \sum_L \bar{L} \gamma^\mu \left(i \partial_\mu + g \frac{\sigma_i}{2} W_\mu^i + g' \frac{Y}{2} B_\mu \right) L + \sum_R \bar{R} \gamma^\mu \left(i \partial_\mu + g' \frac{Y}{2} B_\mu \right) R \quad (14.83)$$

- The part of the Higgs field Lagrangian that concerns the gauge field is

$$\mathcal{L}_{\text{Higgs, gauge}} = \left| \left(i \partial_\mu + g \frac{\sigma_i}{2} W_\mu^i + g' \frac{Y}{2} B_\mu \right) \phi \right|^2 \quad (14.84)$$

³Again, we can technically rescale the lepton mass eigenstates instead, but as a result of convention, we do not do this.

⁴In the minimal standard model, the neutrino is massless, and hence there is no neutrino mixing. However, we now know that the neutrinos have very *smol* masses, so we ultimately still have to worry about the PMNS matrix.

- $V(\phi)$ is the ‘sombrero’ Higgs potential (14.54).
- The part of the Higgs field Lagrangian that concerns the matter field is

$$\mathcal{L}_{\text{Higgs, matter}} = -G_1(\bar{L}\phi R + \bar{R}\phi L) - G_2(\bar{L}\phi_c R + \bar{R}\phi_c L) + \text{h.c.} \quad (14.85)$$

where $\phi_c = i\sigma^2\phi^*$, and G_1 and G_2 are the Yukawa coupling s for $I_3 = -\frac{1}{2}$ and $I_3 = \frac{1}{2}$ respectively.

Theorem 14.2 (Electroweak theory Feynman rules) For a given Feynman diagram in electroweak theory, the scattering amplitude matrix elements with a complex factor $i\mathcal{M}_{fi}$ are constructed as follows:

ELECTROWEAK THEORY FEYNMAN RULES (PARTIAL)	
For each	Add to expression
Incoming and outgoing fermion	$\bar{u}_\alpha(s, p)$ and $u_\alpha(s, p)$
Incoming and outgoing antifermion	$v_\alpha(s, p)$ and $\bar{v}_\alpha(s, p)$
Incoming and outgoing gauge boson	$\epsilon^{*\mu}(\lambda, p)$ and $\epsilon^\mu(\lambda, p)$
Internal photon line	$\frac{-ig^{\mu\nu}}{p^2+i\varepsilon}$
Internal W boson line	$\frac{-i\left(g^{\mu\nu} - \frac{p_\mu p_\nu}{m_W^2}\right)}{p^2 - m_W^2 + i\varepsilon}$
Internal Z boson line	$\frac{-i\left(g^{\mu\nu} - \frac{p_\mu p_\nu}{m_Z^2}\right)}{p^2 - m_Z^2 + i\varepsilon}$
Internal fermion line	$\frac{i(p+m_f)}{p^2 - m_f^2 + i\varepsilon}$
Internal Higgs line	$\frac{i}{p^2 - m_f^2 + i\varepsilon}$
Internal ghost line	$\frac{i\delta^{ab}}{p^2 + i\varepsilon}$
Internal loop	$\int d^4l_n \frac{1}{(2\pi)^4}$ over corresponding line
Any vertex ^a	$(2\pi)^4 \delta^3(k_i - k_f)$

GAUGE BOSON VERTICES

Triple gauge boson vertices:

- $W_\mu^+ W_\nu^- \gamma_\rho$:

$$-ie [(p_+ - p_-)_\rho g_{\mu\nu} + (p_- - p_\gamma)_\mu g_{\nu\rho} + (p_\gamma - p_+)_\nu g_{\rho\mu}] \quad (14.86)$$

- $W_\mu^+ W_\nu^- Z_\rho$:

$$-ie \cot \theta_W [(p_+ - p_-)_\rho g_{\mu\nu} + (p_- - p_Z)_\mu g_{\nu\rho} + (p_Z - p_+)_\nu g_{\rho\mu}] \quad (14.87)$$

Quartic gauge boson vertices:

- $W_\mu^+ W_\nu^- \gamma_\rho \gamma_\sigma$:

$$-ie^2 (2g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) \quad (14.88)$$

- $W_\mu^+ W_\nu^- Z_\rho Z_\sigma$:

$$-ie^2 \cot^2 \theta_W (2g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) \quad (14.89)$$

- $W_\mu^+ W_\nu^- Z_\rho \gamma_\sigma$:

$$-ie^2 \cot \theta_W (2g_{\mu\nu}g_{\rho\sigma} - g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}) \quad (14.90)$$

FERMION-GAUGE BOSON VERTICES

Photon-fermion-fermion:

- $\gamma \bar{f} f$:

$$-ie Q_f \gamma^\mu \quad (14.91)$$

where Q_f is the fermion’s electric charge.

Z boson-fermion-fermion:

- $Z\bar{f}f$:

$$-i \frac{g}{\cos \theta_W} \gamma^\mu \left(I_3^f P_L - Q_f \sin^2 \theta_W \right) \quad (14.92)$$

where I_3^f is the weak isospin, Q_f is the electric charge, $P_L = \frac{1}{2}(1 - \gamma^5)$.

Charged current:

- $W^\pm \bar{f}_u f_d$:

$$-i \frac{g}{\sqrt{2}} \gamma^\mu P_L V_{ud} \quad (14.93)$$

where V_{ud} is the CKM matrix element.

- $W^\pm \bar{\nu}_\ell \ell$

$$-i \frac{g}{\sqrt{2}} \gamma^\mu P_L \quad (14.94)$$

VERTICES INVOLVING THE HIGGS

Gauge-Higgs interactions:

- $HW_\mu^+ W_\nu^-$:

$$ig M_W g_{\mu\nu} \quad (14.95)$$

- $HZ_\mu Z_\nu$:

$$i \frac{g M_Z}{\cos \theta_W} g_{\mu\nu} \quad (14.96)$$

- $H\gamma\gamma$ and $HZ\gamma$: These two vertices only appear at loop level. They are loop-generated via triangles and have no simple Feynman rule.

Higgs self-interactions:

- HHH :

$$-i\lambda v = -i \frac{3M_H^2}{v} \quad (14.97)$$

- $HHHH$:

$$-i\lambda = -i \frac{3M_H^2}{v^2} \quad (14.98)$$

where λ is the Higgs self-coupling.

Higgs-fermion-fermion:

- $H\bar{f}f$

$$-i \frac{m_f}{v} \quad (14.99)$$

Finally, before taking a well-deserved break, remove a factor of $(2\pi)^4 \delta^3(p - q)$, where p and q are the *total* initial and detected momenta.

^aThis term enforces 4-momentum conservation.

Standard Model of Elementary Particles and Gravity

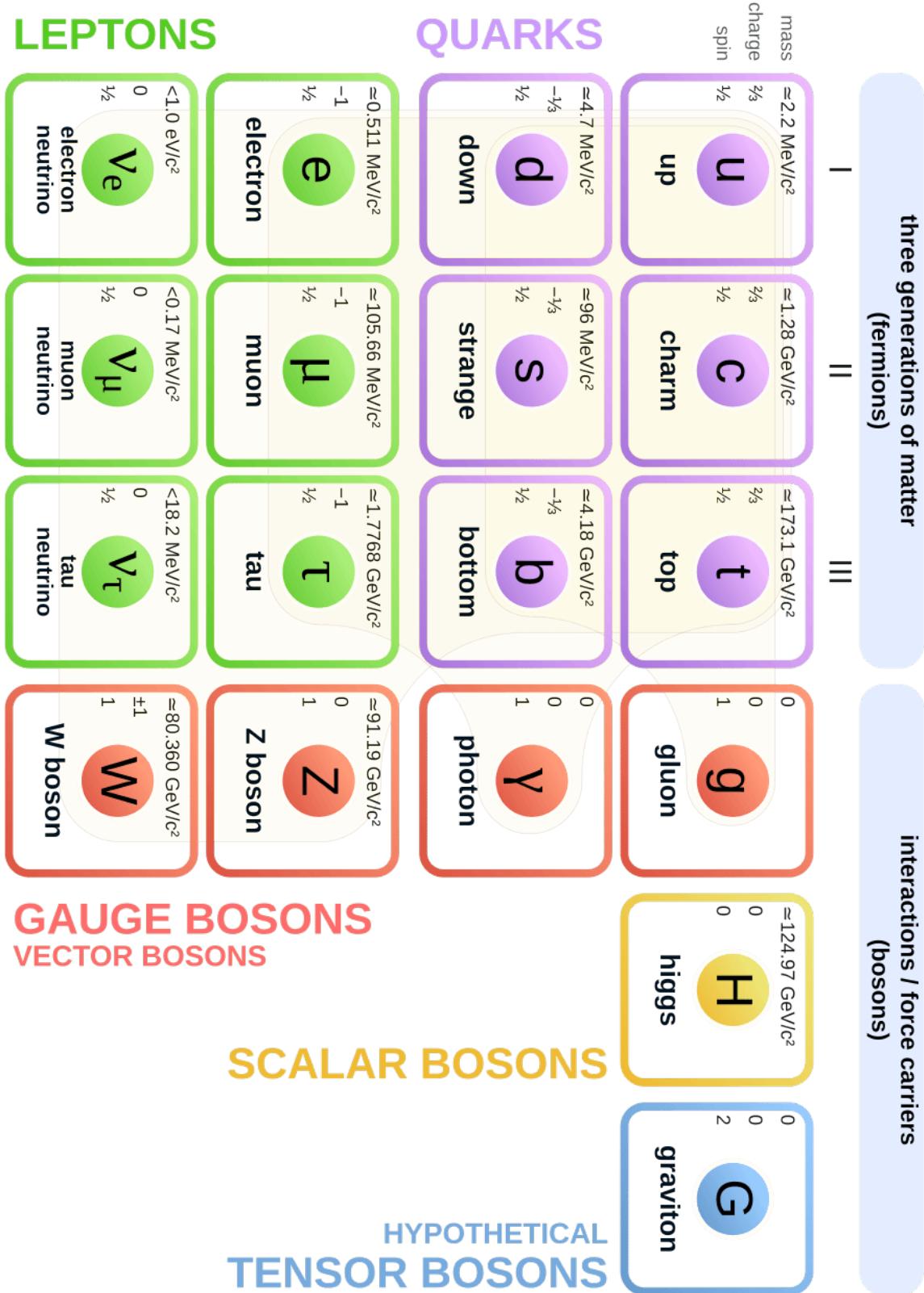


Figure 14.1: The standard model and the graviton.

Chapter 15

Experimental high energy physics

Quote 15.1 Without experimentalists, we still would be a bunch of philosophers, scratching their long white beards, while talking about the same stuff over and over again, (and again).

Felix Halbwedel, 8 August 2025

Everything we have learned up to this point then gives rise to experimental high energy physics. Without experimentalists, theoretical high energy physics would not have existed. The running of the coupling constants in renormalisation and the failure of HEP theory to predict the Higgs mass are good examples of this. Indeed, even as theoreticians, it helps to know a silver of experimental high energy physics for the sake of context.

15.1 Probability amplitude

Now that we have learned the Feynman rules of the standard model, we are in a position to calculate M_{fi} and with it, S_{fi} . However, it should be apparent at this point that we did not actually find an expression for *the probability itself*. This will be the first goal of this chapter. From (5.19), (5.22) and (5.26), it would seem as if S_{fi} is a good candidate for the probability density. However, this is actually not the case. Let us use the Heisenberg picture S -matrix (5.26) as an example. Squaring it gives

$$|S_{qp,H}|^2 = |\langle 0 | a_{q_1,H} \cdots a_{q_m,H} a_{p_1,H}^\dagger \cdot a_{p_n,H}^\dagger | 0 \rangle|^2 \quad (15.1)$$

This looks innocuous enough, but note that the probability density should be dimensionless, while each $a_{q_m,H}$ and $a_{p_n,H}^\dagger$ have mass dimension $-3/2$, hence giving $|S_{qp,H}|^2$ a dimension of $-3(m+n)$. Because of this, $|S_{qp,H}|^2$ *cannot* be the probability amplitude itself.

With the benefit of hindsight, we can already expect this somewhat. If $|S_{qp,H}|^2$ is the end point, why even bother with the LSZ formula (5.28) and M_{fi} ? As it turns out, we can prove that \mathcal{M}_{fi} has a mass dimension of $[\mathcal{M}_{fi}] = 4 - (m+n)$. For a $2 \rightarrow 2$ process, which has dominated our examples so far, this gives $[\mathcal{M}_{fi}] = 0$. Hence, we see that while \mathcal{M}_{fi} is not quite it either.

However, there is a very nice general relationship between \mathcal{M}_{fi} and the probability. A general formula reads

$$P = \text{init.} \times |\mathcal{M}_{fi}|^2 \text{LIPS}(m) \quad (15.2)$$

where init. is some initial factor and LIPS(m) is the Lorentz-invariant phase space we will see later. Let us calculate this explicitly:

Derivation 15.1 (Probability density) We first adopt the shorthand notation

$$A_p = \prod_{j=1}^n a_{p_j} \quad B_q = \prod_{j=1}^m a_{q_j,H} \quad (15.3)$$

From the Born rule (3.17), the probability density of scattering involving m final particles is

$$dP(q_1, \dots, q_m) = \text{Tr}[d\Pi_{q_1} \cdots d\Pi_{q_m} \rho] \quad (15.4)$$

In momentum space, we have

$$d\Pi_{q_m} = \frac{d^3 q_m}{(2\pi)^3} a^\dagger |0\rangle \langle 0| a_{q_m} \quad (15.5)$$

where $1/(2\pi)^3$ is a normalisation factor resulting from normalising the previously unnormalised plane wave solution ϕ

$$\langle 0| a_p a_p^\dagger |0\rangle = (2\pi)^3 \delta(0) \quad (15.6)$$

In any case, by inserting (15.5), (15.4) becomes

$$\begin{aligned} dP(q_1, \dots, q_m) &= V^{-n} \int \frac{d^{3n} k}{(2\pi)^{3n}} \frac{d^{3m} q}{(2\pi)^{3m}} \langle 0| A_k B_q^\dagger |0\rangle \langle 0| B_q A_p^\dagger |0\rangle \langle 0| A_p A_k^\dagger |0\rangle \\ &= V^{-n} \frac{d^{3m} q}{(2\pi)^{3m}} \langle 0| A_k B_q^\dagger |0\rangle \langle 0| B_q A_p^\dagger |0\rangle \\ &= \frac{1}{V^n} \frac{d^{3m} q}{(2\pi)^{3m}} |S_{qp}|^2 \end{aligned} \quad (15.7)$$

which is the simplified form of the probability density.

S_{fi} is related to \mathcal{M}_{fi} via (5.44). However, we acknowledge the relativistic normalisation of single-particle states:

$$\langle p' | p \rangle = (2E_p V) \delta_{p,p'} \quad (15.8)$$

Hence, the relationship actually reads

$$S_{fi} = \delta_{fi} + i(2\pi)^4 \delta^{(4)}(p_f - p_i) \mathcal{M}_{fi} \prod_{j=1}^{n+m} \frac{1}{\sqrt{2E_j V}} \quad (15.9)$$

For a transition from n initial particles to m final particles, the relevant part of S_{fi} is

$$S_{fi} = i(2\pi)^4 \delta^{(4)}(p_f - p_i) \frac{\mathcal{M}_{fi}}{\sqrt{\prod_{j=1}^n 2E_{p_j} V \prod_{k=1}^m 2E_{q_k} V}} \quad (15.10)$$

Here, there is a peculiarity we can exploit. Theoretically, the momenta in the momentum conservation-enforcing term $\delta^4(p_i - p_f)$ exist throughout the entirety of the spacetime. In a real experiment, however, incoming and outgoing states are localised. To deal with this, we assume the interaction happens essentially in a 4D box. That is, over a time interval T and localised in some volume V . First, we can perform the rewrite

$$(2\pi)^4 \delta^4(p_i - p_f) = \int_{VT} d^4 x e^{i(p_f - p_i)x} \quad (15.11)$$

Considering the conservation of momentum ($p_f = p_i$), we have

$$(2\pi)^4 \delta^4(0) = VT \rightarrow \delta^4(0) = \frac{VT}{(2\pi)^4} \quad (15.12)$$

Now we can insert this into (15.7), which we then insert into (15.7). The differential of the *transition rate* W_{fi} , which is the probability per unit time, is then solved by differentiating both sides w.r.t. time T :

Definition 15.1 (Transition rate differential) The differential of the transition rate is given by

$$dW_{fi} = |\mathcal{M}_{fi}^2| V \prod_{j=1}^n \frac{1}{2E_{p_j} V} \text{LIPS}(m) \quad (15.13)$$

In this expression, we have crammed all Lorentz-invariant terms together as the so-called *Lorentz-invariant phase space*, which is defined with respect to m particles in the final state:

Definition 15.2 (Lorentz-invariant phase space)

$$\text{LIPS}(m) \equiv (2\pi)^4 \delta^4(p_i - p_f) \prod_{k=1}^m \frac{d^3 q_k}{(2\pi)^3 2E_k} \quad (15.14)$$

Remark 15.1 If the final particles are identical, $\text{LIPS}(m)$ is divided by $m!$.

Very nicely, T has disappeared, and the expression now has no explicit dependence on the time interval.

Now we consider specific cases of the transition rate. When there is only one initial particle ($n = 1$), the process is then called a *decay*. Using $n = 1$, we find that the transition rate for decays, or rather the *decay rate*, is

Definition 15.3 (Decay rate)

$$\Gamma_{fi} = \frac{1}{2m} \int |\mathcal{M}_{fi}|^2 \text{LIPS}(m) \quad (15.15)$$

In experimental high energy physics, this quantity is often called the *decay width* and written as Γ without the indices.

Again, very nicely, the two V s have cancelled themselves out.

In the case where the end product consists of 2 particles, the decay rate reduces to

$$\Gamma_{fi} = \frac{1}{2m} \int |\mathcal{M}_{fi}|^2 q_f d\Omega \quad (15.16)$$

where Ω is the so-called *solid angle*. It is related with the polar angle θ and the azimuthal angle ϕ by the following:

$$d\Omega = d(\cos \theta) d\phi \quad (15.17)$$

Remark 15.2 In most cases, there is no dependence on the azimuthal angle ϕ , so integrating it simply yields 2π .// A quantity ultimately related to the transition rate is the *cross-section*, which is the transition rate for a single particle *per unit beam flux*. We first define the *particle flux*:

Definition 15.4 (Particle flux) The particle flux for a beam with velocity v_1 and a density of particles of $1/V^a$ and a target with velocity v_2 is

$$F = \frac{|v_1 - v_2|}{V} \quad (15.18)$$

This is simply the number of particles per unit area which run past each other per unit time.

^ai.e. 1 particle in a volume of V .

We can find the cross-section differential by dividing the transition rate differential by the flux:

$$d\sigma = \frac{d\Gamma_{fi}}{F} = \frac{1}{|v_1 - v_2|} \frac{1}{4E_1 E_2} |\mathcal{M}_{fi}|^2 \text{LIPS}(m) \quad (15.19)$$

This expression is Lorentz-invariant. We then integrate and find

Definition 15.5 (Cross-section)

$$\sigma = \frac{1}{4\sqrt{(p_1 \cdot p_2)^2 - m_1^2 m_2^2}} \int |\mathcal{M}_{fi}|^2 \text{LIPS}(m) \quad (15.20)$$

There is a very nice standard result for $2 \rightarrow 2$ processes. In the *CM* (centre-of-mass) *frame*:

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2 s} \frac{|p_f|}{|p_i|} |\mathcal{M}|^2 = \frac{1}{64\pi^2 s} |\mathcal{M}|^2 \quad (15.21)$$

There is also the *lab frame* (which is centred) around instruments, like the detector. In practice, this means that the target of the scattering is initially at rest. By convention, we then have:

- Particle 1 has energy E_1 and is the incoming projectile.
- Particle 2 is the target at rest and has mass M .
- Particle 3 has energy E_2 is the outgoing projectile.
- Particle 4 is the recoiled target and (expectedly) has mass M .

The cross-section per solid angle is then

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2} \left(\frac{E_3}{ME_1} \right)^2 |\mathcal{M}_{fi}|^2 \quad (15.22)$$

Finally, two useful experimental quantities can be derived from what we have so far.

Derivation 15.2 (Number of events) By integrating (15.19), we find

$$W_{fi} = \sigma F \quad (15.23)$$

That is to say, the rate of a process is the product of the cross-section and the flux. If the experiment is taking place in a collider, the flux is also known as the *instantaneous luminosity* with the symbol \mathcal{L} .

Definition 15.6 (Integrated luminosity) The aptly-named *integrated luminosity* is the time integral of the instantaneous luminosity:

$$L = \int \mathcal{L} dt \quad (15.24)$$

Remark 15.3 Note the similarity with the Lagrangian density and the Lagrangian? We don't know why either.

The integrated luminosity becomes useful if we integrate Γ_{fi} and F by time. Then, we get

Theorem 15.1 (Number of events)

$$N = \sigma F \quad (15.25)$$

where N is the *number of events*.

Derivation 15.3 (Flight distance) As mentioned, decay concerns a single unstable particle. Here, the quantities of our concern are the average distance it travels before decaying, or *flight distance*, and the average time before it decays, or *lifetime*. The latter is calculated via

Theorem 15.2 (Lifetime)

$$\Gamma_{fi} = \frac{\hbar}{\tau} \quad (15.26)$$

Here τ lies in the CM frame.

The flight distance, in terms of the lab and CM frames lifetime respectively, is then

Theorem 15.3 (Number of events)

$$L = vt_{\text{lab}} - v\gamma\tau_{CM} = \gamma\beta c\tau_{CM} \quad (15.27)$$

where, we have encountered our very good friends, the SR factors β and γ :

$$\beta = \frac{v}{c} \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}} \quad (15.28)$$

15.2 $2 \rightarrow 2$ processes

We will now put everything we have learned so far to the test by considering $2 \rightarrow 2$ processes. The simplest $2 \rightarrow 2$ processes are all QED processes. The easiest examples is the so-called *Bhabha scattering*:

$$e^+ e^- \rightarrow e^+ e^- \quad (15.29)$$

This can take two forms:

- The *s-channel*, where an electron and a positron annihilate into a photon, which then produces an electron-positron pair. The photon line is horizontal in the diagram, and the photon momentum is the difference between the momenta of the incoming and outgoing electrons/positrons.

- The *t-channel*, where an electron emits a photon that is absorbed by a positron (or vice versa). The photon line is vertical in the diagram, and the photon momentum is the sum of the incoming/outgoing electron-positron pair.

But something is fishy here. If one rotates the *s*-channel diagram by 90 degrees. The real question is, why do we do this? As we know, in a Feynman diagram, time increases (abstractly, and with no scale) from left to right. Suppose the top-left particle in the *s*-channel diagram is a positron. The arrow then points to the top-left. However, rotating it clockwise¹ by 90 degrees will cause the arrow to point towards the top-right. As it points rightwards, it is no longer a positron, but an electron. This then gives us the same scattering, but in the *t*-channel.

This is known as *crossing symmetry*, which means that antiparticles can be seen as nothing but particles going backwards in time. Critically, this does *not* mean that their scattering amplitudes are the same. In the most general case, where the four particles have different momenta, the *s*- and *t*-channel momentum transfers (which, in our case, is the photon momentum) will be different, and as \mathcal{M}_{fi} explicitly depends on the momentum transfer, we cannot assume its two incarnations to be the same.

We can then ‘relate’ the two scattering amplitudes via the fact that the only different quantity are their respective internal lines, whose denominator can be conveniently defined via the so-called *Mandelstam variables*:

Definition 15.7 (Mandelstam variables) The Mandelstam variables s , t and u correspond to the *s*-, *t*- and *u*-channels^a respectively.

$$s = (p_1 + p_2)^2 \quad t = (p_1 - p_3)^2 \quad u = (p_1 - p_4)^2 \quad (15.30)$$

s , t and u are equal to the four-momentum exchange q^2 in their own channels. They are Lorentz-invariant and satisfy

$$s > 0 \quad t < 0 \quad u < 0 \quad s + t + u = m_1^2 + m_2^2 + m_3^2 + m_4^2 \quad (15.31)$$

^aWe will discuss the *u*-channel later.

Let us go through some more examples. The first archetypal process we consider is *Møller scattering*, which reads

$$e^- e^- \rightarrow e^- e^- \quad (15.32)$$

Here we also have two channels:

- The *t*-channel, where an electron emits a photon that is absorbed by another electron. The photon line is vertical in the diagram, and the photon momentum is the difference between the momenta of the incoming and outgoing electrons.
- The *u*-channel, which modifies the *t*-channel so that the two outgoing electrons are swapped.

Again we consider crossing symmetry. The *t*- and *u*-channel scattering amplitudes may seem identical, but this is false due to a sinister caveat. The swapping of the outgoing electrons in the *u*-channel means that the *momenta* of the two outgoing electrons are *also* swapped. Thus, the momentum transfers of the two channels are different, resulting in different \mathcal{M}_{fi} .

The next process of interest is *Compton scattering*, which reads

$$e^- \gamma \rightarrow e^- \gamma \quad (15.33)$$

Here we also have two channels:

- The *s*-channel, where an electron absorbs a photon and emits another photon later on. The internal line is that of the electron itself.
- The *u*-channel, which modifies the *s*-channel so that the incoming and outgoing photons are swapped. Essentially, the electron first emits a photon and then absorbs another.

The final process to consider is the category of electro-muon interactions. Yet again, there are two channels:

¹If anticlockwise, then the same argument applies for the bottom-left electron, which will point towards the bottom-right.

- The *s*-channel, where an electron and a positron annihilate to form a photon, which then decays into a muon and an antimuon. This is known as *electron-muon pair production* and is given by

$$e^+ e^- \rightarrow \mu^+ \mu^- \quad (15.34)$$

By rotating this, the positively-charged leptons become negatively-charged (or vice versa), in which case we have *electron-muon scattering*, which is basically the *t*-channel of *electron-muon pair production*.

- The *t*-channel, where an electron emits a photon, which is absorbed by a muon (or vice versa). This is known as *electron-muon scattering* and is given by

$$e^- \mu^- \rightarrow e^- \mu^- \quad (15.35)$$

The cross-section can be calculated as

$$\frac{d\sigma}{dq^2} = \frac{e^4}{8\pi q^4} \left(1 + \left(1 + \frac{q^2}{s} \right)^2 \right) \quad (15.36)$$

Now there remains the question of why all processes lack one of the 3 possible channels:

- There is no *u*-channel in Bhabha scattering. This is because the final particles are different (electron and positron), and the *u*-channel is only possible *when the final particles are identical*.
- There is no *s*-channel in Møller scattering. This is because charge must be conserved at all vertices, and two electrons cannot annihilate into an (electrically-neutral) photon.
- There is no *t*-channel in Compton scattering. This is also because we cannot have two electrons annihilating into one photon, which is charge-neutral.
- There is no *u*-channel in electron-muon scattering either as the outgoing particles are not identical.

Remark 15.4 The *s*-channel is so-called because the relevant variable is the centre-of-mass energy $s = (p_1 + p_2)^2 = m^2$. The *t*- and *u*-channels are so-called simply because the letters *t* and *u* follow *s*. Finally, we consider how to actually do $2 \rightarrow 2$ calculations. The first step is to calculate $|\mathcal{M}_{fi}|^2$:

Derivation 15.4 (Calculating $|\mathcal{M}_{fi}|^2$)

- Draw the (lowest-order) Feynman diagram. It also helps to draw a diagram of the kinematics.
- Choose a reference frame. Whenever possible, use the CM frame as it is preferred for its simplicity.
- Choose the axes. By convention:
 - The direction of particle motion of the z -axis.
 - In $2 \rightarrow 2$ processes, one can always choose $\phi = 0$ or $\phi = \pi$.
- Construct the scattering amplitude \mathcal{M}_{fi} by identifying the currents^a and internal lines (propagators).
- Calculate $|\mathcal{M}_{fi}|^2$. As \mathcal{M}_{fi} is nothing but a number, and the vertical bars denote the squared complex modulus, $|\mathcal{M}_{fi}|^2$ is simply \mathcal{M}_{fi} times its complex conjugate:

$$|\mathcal{M}_{fi}|^2 = \mathcal{M}_{fi} \times \mathcal{M}_{fi}^* \quad (15.37)$$

^aThis consists of external legs and vertices. In a $2 \rightarrow 2$ interaction, it has the form external leg spinor-vertex-external leg spinor. Note also that the adjoint spinor is on the left while the normal spinor is on the right.

Derivation 15.5 (Final steps)

Spins/polarisations and indistinguishable paths also affect the calculation of $|\mathcal{M}_{fi}|^2$:

- If there are spins or colours in the initial and final states, we first square the probability amplitude, and then average over the initial spin/polarisation states and sum over the final spin/colour states.

This is possible as the probability amplitudes of the states cannot interfere with each other.

$$|\mathcal{M}|^2 = \frac{1}{N_{\text{spin}}^{\text{init}} N_{\text{colour}}^{\text{init}}} \sum_{\text{spins, colours}} |\mathcal{M}|^2 \quad (15.38)$$

where $N_{\text{spin}}^{\text{init}}$ and $N_{\text{colour}}^{\text{init}}$ are spin and colour factors. Assuming that the beam is unpolarised^a, they are

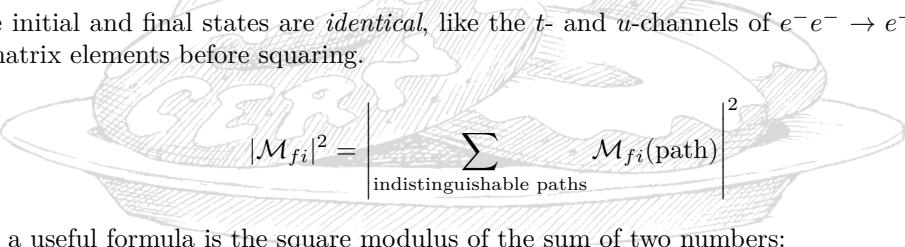
$$N_{\text{spin}}^{\text{init}} = \prod_{i \in \text{initial}} n_{\text{spin}}(i) \quad n_{\text{spin}}(i) = \begin{cases} 2s_i + 1 & \text{massive particle of spins}_i, \\ 2 & \text{massless vector (photon, gluon): two helicities} \\ 2 & \text{massless fermion if both helicities populated} \\ 1 & \text{chiral-only case (e.g. SM LH neutrino in weak interactions)} \end{cases} \quad (15.39)$$

$$N_{\text{colour}}^{\text{init}} = \prod_{i \in \text{initial}} \dim R_i \quad \dim R_i = \begin{cases} N, & \text{quark or antiquark (fundamental/anti-fundamental)} \\ N^2 - 1 & \text{gluon (adjoint)} \\ 1 & \text{colour singlet (e.g. photon, } W/Z, \text{ lepton)} \end{cases} \quad (15.40)$$

where $N = 3$ in the real world as we have the SU(3) group.

There is no weak factor $N_{\text{weak}}^{\text{init}}$ or electroweak factor $N_{\text{EW}}^{\text{init}}$ because weak/electroweak charges label physical particle species (electrons, neutrinos, W^\pm , etc.) which we can and do prepare in definite states, whereas QCD colours are unobservable (due to colour confinement) and must be averaged over.

- If the initial and final states are *identical*, like the t - and u -channels of $e^- e^- \rightarrow e^- e^-$, we sum the matrix elements before squaring.



$$|\mathcal{M}_{fi}|^2 = \left| \sum_{\text{indistinguishable paths}} \mathcal{M}_{fi}(\text{path}) \right|^2 \quad (15.41)$$

Here, a useful formula is the square modulus of the sum of two numbers:

$$\begin{aligned} |\mathcal{M}_{fi}|^2 &= |\mathcal{M}_1 + \mathcal{M}_2|^2 \\ &= |\mathcal{M}_1|^2 + |\mathcal{M}_2|^2 + 2 \operatorname{Re} \mathcal{M}_1^* \mathcal{M}_2 \end{aligned} \quad (15.42)$$

Here we see the motivation for summing before squaring: $2 \operatorname{Re} \mathcal{M}_1^* \mathcal{M}_2$ is the superposition resulting from the two identical states, which physically represents the interference (constructive or destructive) between them. We then need to account for this by summing before squaring.

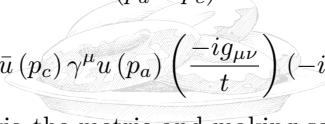
- In real life, many processes have both characteristics (e.g. $e^- e^- \rightarrow e^- e^-$). For these cases, we *first* sum the states that have indistinguishable paths, and *then* square everything and sum the spins.

This then puts us in a position to calculate the decay rate via (15.15).

^aWe will not consider the polarised case, where a normalised density matrix is imposed after squaring, but before summing.

At this point, two examples are in order:

Derivation 15.6 ($e^- \mu^- \rightarrow e^- \mu^-$ scattering) The process lives in the t -channel. if we label the incoming particles with indices a and b and outgoing particles c and d , the propagator denominator can then be written with the Mandelstam variable $t = (p_a - p_c)^2$. The transition amplitude (with a factor of i) then reads



$$i\mathcal{M}_{fi} = (-ie)\bar{u}(p_c)\gamma^\mu u(p_a) \left(\frac{-ig_{\mu\nu}}{t} \right) (-ie)\bar{u}(p_d)\gamma^\nu u(p_b) \quad (15.43)$$

Shifting the index of one γ^μ down via the metric and making some superficial simplifications, we find

that

$$\mathcal{M}_{fi} = \frac{e^2}{t} \bar{u}(p_c) \gamma^\mu u(p_a) \bar{u}(p_d) \gamma_\mu u(p_b) \quad (15.44)$$

Now we essentially have two vertices, each attached to two external lines. Let us simplify the first one:

$$(\bar{u}(p_c) \gamma^\mu u(p_a))^* = \left(u(p_a)^\dagger \gamma^{\mu\dagger} \gamma^{0\dagger} u(p_c) \right) = \left(u(p_a)^\dagger \gamma^{\mu\dagger} \gamma^0 u(p_c) \right) \quad (15.45)$$

Using the last identity and then the second identity in (6.12), we derive

$$\left(u(p_a)^\dagger \gamma^{\mu\dagger} \gamma^0 u(p_c) \right) = \left(u(p_a)^\dagger \gamma^0 \gamma^{\mu\dagger} \gamma^0 u(p_c) \right) = \left(u(p_a)^\dagger \gamma^0 \gamma^\mu u(p_c) \right) = (\bar{u}(p_a) \gamma^\mu u(p_c)) \quad (15.46)$$

The same follows for the second vertex. Substituting and squaring gives

$$|\mathcal{M}_{fi}|^2 = \frac{e^4}{t^2} \bar{u}(p_c) \gamma^\mu u(p_a) \bar{u}(p_d) \gamma_\mu u(p_b) \bar{u}(p_a) \gamma^\nu u(p_c) \bar{u}(p_b) \gamma_\nu u(p_d) \quad (15.47)$$

As discussed earlier, we sum over final-state spins and average over initial-state spins. Here, $N_{\text{spin}}^{\text{init}}$ is $2 \times 2 = 4$. This gives

$$\begin{aligned} \frac{1}{4} \sum_{\text{spins}} |\mathcal{M}_{fi}|^2 &= \frac{e^4}{4t^2} \left[\gamma_{\alpha\beta}^\mu (\not{p}_a + m_e)_{\beta\zeta} \gamma_{\zeta\eta}^\nu (\not{p}_c + m_e)_{\eta\alpha} \right] \left[\gamma_{\mu,\alpha'\beta'} (\not{p}_b + m_\mu)_{\beta'\zeta'} \gamma_{\nu,\zeta'\eta'} (\not{p}_d + m_\mu)_{\eta'\alpha'} \right] \\ &= \frac{e^4}{4t^2} \text{Tr} \left(\gamma^\mu (\not{p}_a + m_e) \gamma^\nu (\not{p}_c + m_e) \right) \text{Tr} \left(\gamma_\mu (\not{p}_b + m_\mu) \gamma_\nu (\not{p}_d + m_\mu) \right) \end{aligned} \quad (15.48)$$

where we have contracted the Dirac spinor indices. Note that while four pairs of indices are contracted, there is only one trace. This is because, no matter the order you do the calculation, three of them (say, β , ζ and η) are used up in the multiplication, leaving only one index (in our example α) which then requires tracing^a. Using Theorem 6.4 and the Mandelstam variables, we find that

$$\frac{1}{4} \sum_{\text{spins}} |\mathcal{M}_{fi}|^2 = \frac{2e^4}{t^2} \left(s^2 + u^2 - 4(m_e^2 + m_\mu^2)(s+u) + 6(m_e^2 + m_\mu^2)^2 \right) \quad (15.49)$$

In colliders, the masses are *smol* compared to the energy. This allows us to assume the ultra-relativistic limit and set the masses to zero. We then get

$$\frac{1}{4} \sum_{\text{spins}} |\mathcal{M}_{fi}|^2 = \frac{2e^4}{t^2} (s^2 + u^2) \quad (15.50)$$

^aNote also that very lazily, we are not following the Einstein summation convention.

Derivation 15.7 ($e^-e^- \rightarrow e^-e^-$ scattering) We briefly mention this process because of crossing symmetry. As a result of it, we have two channels, so we need to compute the t - and u -channel transition amplitudes. Respectively, we have

$$i\mathcal{M}_1 = \frac{ie^2}{t} \bar{u}(p_c) \gamma^\mu u(p_a) \bar{u}(p_d) \gamma_\mu u(p_b) \quad i\mathcal{M}_2 = -\frac{ie^2}{u} \bar{u}(p_d) \gamma^\mu u(p_a) \bar{u}(p_c) \gamma_\mu u(p_b) \quad (15.51)$$

The negative sign may seem arbitrary, but it carries a deep meaning. This arises from the Grassmann-odd nature of Fermion fields, which dictates that they anticommute. Thus, the sign changes when the fields for electrons c and d are switched.

But why do we *switch* the fields in the first place, instead of just *relabeling* them as one another, which is what we thought we were doing up to now? This arises from the fact that they represent two distinct Wick contractions of the fields and thus, two distinct processes, albeit with identical particles^a.

The final noteworthy point is that we need to sum the contributions before squaring them, like in (15.42).

^aRemember that if the end particles are different, there is no t -channel whatsoever.

15.3 Rise of quarks

Experimentally, the quark model was discovered by studying *electron-proton scattering*. The cross-section of electron-proton scattering is more complex than our previous generic cross-section because the proton cannot be considered massless. A naive calculation of the electron-proton scattering cross-section is the *Mott cross-section*:

Definition 15.8 (Mott cross section)

$$\left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} = \frac{\alpha^2}{4E_1^2 \sin^4(\theta/2)} \frac{E_3}{E_1} \cos^2(\theta/2) \quad (15.52)$$

where α is the fine structure constant defined in (12.98).

However, this fails to consider that protons are not point-like. Thus, it is *not* the correct cross-section. Instead, the *real* cross-section is derived via the Rosenbluth formula:

Definition 15.9 (Rosenbluth formula)

$$\frac{d\sigma}{d\Omega} = \left(\frac{d\sigma}{d\Omega}\right)_{\text{Mott}} \left[\frac{G_E^2 + \tau G_M^2}{1 + \tau} + 2\tau G_M^2 \tan^2(\theta/2) \right] \quad (15.53)$$

where G_E and G_M are so-called *form factors*, and τ is a factor given by

$$\tau = \frac{Q^2}{4m_p^2} > 0 \quad (15.54)$$

Q is defined in relation to the internal photon line momentum q :

$$Q^2 = -q^2 \quad (15.55)$$

But we are going about this topic in the way everything was discovered historically. At this point, we did not yet know that protons have internal structures, not to mention the fact that we do not know how the structure constants G_E and G_M are defined in the first place.

Here, it is then necessary to introduce the so-called concept of *deep inelastic scattering* (DIS), which collides leptons with hadrons. To understand this, consider our innocent electron-proton scattering:

- If the scattering is *elastic*, then the electron and the proton simply collide and bounce off each other, which we expect.
- If the scattering is *inelastic*, however, the proton breaks apart.

The reason why DIS is so-called, or rather why it is ‘deep’, lies in the fact that the lepton is set to a very high energy. This corresponds to a very short wavelength and the ability to probe ‘deep inside’ the hadron of interest. First, we introduce two variables. The first is the so-called *Bjorken x*, also written as *Bjorken-x*:

Definition 15.10 (Bjorken x)

$$x \equiv \frac{Q^2}{2p_2 \cdot q} \quad (15.56)$$

Its physical significance is not immediately obvious, and will be discussed later on.

Quote 15.2 Uh, Bjork?

Teknolust, 2002

The second is the *inelasticity y*:

Definition 15.11 (Inelasticity)

$$y \equiv \frac{p_2 \cdot q}{p_2 \cdot p_1} \quad (15.57)$$

The two can be seen to be related by

$$Q^2 = (s - m_p^2) xy \quad (15.58)$$

The DIS cross-section is then

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[\left(1 - y - \frac{m_p^2 y^2}{Q^2} \right) \frac{F_2(x, Q^2)}{x} + y^2 F_1(x, Q^2) \right] \quad (15.59)$$

In DIS, we have $Q^2 \ll m_p^2 y^2$, so we have

$$\frac{d^2\sigma}{dxdQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1 - y) \frac{F_2(x, Q^2)}{x} + y^2 F_1(x, Q^2) \right] \quad (15.60)$$

where $F_1(x, Q^2)$ and $F_2(x, Q^2)$ are *structure functions*.

Remark 15.5 We can return to elastic scattering by setting $x = 1$. This returns us to (15.53). From observation, three points are historically noted.

1. The following relations hold:

$$Q^2 = 4E_1 E_3 \sin^2 \frac{\theta}{2} \quad x = \frac{Q^2}{2m_p(E_1 - E_3)} \quad y = 1 - \frac{E_3}{E_1} \quad (15.61)$$

2. The dependence of $F_1(x, Q^2)$ and $F_2(x, Q^2)$ on Q^2 are *smol*, so we can practically write $F_1(x)$ and $F_2(x)$. This is known as *Bjorken scaling*.
3. The structure functions are related via the Bjorken x :

$$F_2 = 2xF_1 \quad (15.62)$$

The factor 2 implies that the structure functions describe spin- $\frac{1}{2}$ objects.

It was then theorised that the proton, and indeed, all hadrons, have substructures and are made up of spin- $\frac{1}{2}$ *partons*. This is known as the *parton model*. However, both are historical terms. Today, we know them as *quarks* and the *quark model*.

15.4 Electron-quark scattering

It is then apparent that when we study electron-proton scattering, what we are really studying is electron-quark scattering. The first immediately interesting thing is the role of the Bjorken x . Let us consider the absorption of a photon by a quark within a proton. The photon carries momentum x while the proton carries momentum p_2 , of which ξp_2 goes to the quark that participates in the collision. For our convenience, let us define the so-called *infinite momentum frame*:

Definition 15.12 (Infinite momentum frame) In the infinite momentum frame, the momenta of the particles go up to infinity. That is to say, the frame is moving infinitely fast towards or away from the particles.

As the momenta are infinitely large, we are allowed to set the proton mass to effectively zero. Hence, we have

$$(\xi p_2 + q)^2 = m_{\text{quark}}^2 = 0 \quad (15.63)$$

which gives us

$$\xi = \frac{Q^2}{2p_2 \cdot q} = x \quad (15.64)$$

which shows us that the Bjorken x is the fraction of the proton's momentum that went to the quark that collided with the photon.

Aphorism 15.1 (Felix Halbwedl, 15 August 2025) If your nucleon is a *bad tetrad* and you hit it really hard, it will break apart into *s'mol* bits whose rest masses will sum up to a mass exceeding the bad tetrad's mass. The ratio between your poor bad tetrad's mass and the *s'mol* bit's mass sum is crudely your x . But you already knew that, because you're a good tetrad.

Let us then calculate the electron-quark scattering cross section. It is almost identical to the electron-muon scattering cross section, except the muon charge is replaced by the quark charge C , in units of e :

$$\frac{d\sigma}{dq^2} = \frac{2\pi\alpha^2 C^2}{q^4} \left(1 + \left(1 + \frac{q^2}{s} \right)^2 \right) \quad (15.65)$$

Recalling the definitions of y and Q , we can rewrite this as

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1-y) + \frac{y^2}{2} \right] \quad (15.66)$$

However, this merely describes the collision between our photon and 1 of the (depending on the hadron) at least 2 quarks. To get the total cross-section, we then need to sum over all quarks:

$$\frac{d^2\sigma}{dx dQ^2} = \frac{4\pi\alpha^2}{Q^4} \left[(1-y) + \frac{y^2}{2} \right] \sum_i C_i^2 \mathcal{P}^i(x) dx \quad (15.67)$$

where i spans over all possible quarks of the hadron. Compare this with (15.60), and we will find

Theorem 15.4 (Structure functions)

$$F_2(x) = x \sum_i C_i^2 \mathcal{P}^i(x) dx \quad F_1(x) = \frac{1}{2} \sum_i C_i^2 \mathcal{P}^i(x) dx \quad (15.68)$$

This shows that we can indeed write the two structure functions as functions of the Bjorken x only. The only thing new to us is then $\mathcal{P}^i(x)$, which is the *parton distribution functions* (PDFs) of a hadron. For example, for a proton and a neutron respectively, we have

Definition 15.13 (Parton distribution functions for protons and neutrons)

$$\mathcal{P}^i(x) = u^p(x), u^p(x), d^p(x) \quad \mathcal{P}^i(x) = u^n(x), d^n(x), d^n(x) \quad (15.69)$$

where $u(x)$, $d(x)$, etc. are the individual PDFs for up quarks, down quarks, etc. The upper indices p and n denote that they are the quarks of a proton and a neutron respectively. However, we will soon see that they are not merely eye candy either.

Yet there is one more loose end emerging from QCD. We know that a proton is made up of uud , a neutron is made up of udd , etc. But recall also that quarks are bound together through the strong force, which are mediated through gluons. Both quarks and these existing gluons can radiate more gluons, much like the emission and absorption of photons in QED. The result is ‘free’ gluons that eventually split into quark-antiquark pairs we call *sea quarks*:

$$g \rightarrow q\bar{q} \quad (15.70)$$

The sea quarks do not contribute to the quark structure due to their nature as quark-antiquark pairs. The ‘existing’ quarks that actually contribute to the quark structure are then known as *valence quarks*. Expectedly, the sea quarks do not live forever. They are virtual excitations in the hadrons and are thus unstable. Eventually, they annihilate and recombine themselves into a gluon:

$$q\bar{q} \rightarrow g \quad (15.71)$$

Both the splitting of gluons into sea quarks and the combining of sea quarks into gluons happen continually inside the proton.

Derivation 15.8 (Structure functions) So what is the implication of our new friends, the sea quarks?

Firstly, they exist as bookkeeping devices for the energy inside a hadron. Behind the gluons and the valence quarks, the sea quarks have the third largest (or rather the smallest) contribution to energy. The fact that gluons hold so much momentum is partly because many sea quark pairs annihilate back into gluons, which replenishes the gluon population.

One more loose end exists for the sea quarks, which can be seen when we consider the structure

function for the proton

$$F_2^p \left(\frac{4}{9} u^p(x) + \frac{1}{9} d^p(x) + \frac{4}{9} \bar{u}^p(x) + \frac{1}{9} \bar{d}^p(x) \right) \quad (15.72)$$

Naively, we know from *uud* that a proton is made up of two up quarks and a down quark (or so we think up to this point), so why do we have antiquark contributions?

The answer is that our previous picture of a proton being made up of only *uud* is a simplified one. Rather than saying that we only have two up quarks and one down quark, what *uud* actually means is that:

- We have one more up quark than down quarks.
- There is at least one down quark, and that one (minimum) down quark is the valence quark *d* we see in *uud*.

In other words, *uud* does not tell us about the total number of quarks. This number could be much larger, due to the existence of sea quarks. But due to the nature of sea quarks, extra quarks always come with an equal number of corresponding antiquarks, which then explain the existence of the antiquark PDFs.

For the neutron, we then have

$$F_2^n \left(\frac{4}{9} u^n(x) + \frac{1}{9} d^n(x) + \frac{4}{9} \bar{u}^n(x) + \frac{1}{9} \bar{d}^n(x) \right) \quad (15.73)$$

As the sea quarks are meaningless in the big picture, it makes sense to decompose the PDFs into valence and sea quarks. For the proton, we have

$$u^p(x) = u_v(x) + s(x) \quad d^p(x) = d_v(x) + s(x) \quad (15.74)$$

where u_v and d_v are the PDFs for valence up and down quarks, and $s(x)$ is the *common sea contribution* from $u\bar{u}$ and $d\bar{d}$ pairs.

Definition 15.14 (Isospin symmetry) Here a simplification can be made. To begin with, compared to the QCD energy scale ($\Lambda_{\text{QCD}} \sim 200\text{MeV}$), the difference between the up and down quarks' masses is not that apparent:

$$m_u \approx 2.2\text{MeV} \quad m_d \approx 4.7\text{MeV} \quad (15.75)$$

While their charges are different, this is also *smol* compared to the QCD scale. Meanwhile, the strong coupling is identical between the two, so the two can be more or less treated the same. That is to say, if we swap all *us* and *ds*, QCD physics stays the same:

$$u \rightarrow d \quad d \rightarrow u \quad (15.76)$$

This is known as *isospin symmetry*.

Isospin symmetry applies to both valence and sea quarks. Firstly, thanks to isospin symmetry in the sea, we are allowed to write

$$\bar{u}^p = \bar{d}^p = s(x) \quad (15.77)$$

For the neutron, we simply swap *u* and *d* PDFs of the proton, again thanks to isospin symmetry:

$$u^n(x) = d_v(x) + s(x) \quad d^n(x) = u_v(x) + s(x) \quad (15.78)$$

By substituting (15.74) and (15.78) into (15.72) and (15.73) respectively, we find the 'real' structure functions for the proton and the neutron:

Theorem 15.5 (Proton and neutron structure functions)

$$F_2^p(x) = x \left[\frac{4}{9} u_v(x) + \frac{1}{9} d_v(x) + \frac{5}{9} s(x) \right] \quad (15.79)$$

$$F_2^n(x) = x \left[\frac{4}{9} d_v(x) + \frac{1}{9} u_v(x) + \frac{5}{9} s(x) \right] \quad (15.80)$$

15.5 Feynman diagrams

Now that we have made an overview of the entire standard model, the process of drawing a Feynman diagram is nothing but a summary of what we have learned:

- List the particle contents of both sides, with any particles consisting of quarks broken down to their quark contents.
- Link all particles that appear on both sides. These particles then do not participate in the interaction.

Now consider the virtual particles:

- A photon can only couple to particles that carry EM charge. i.e. it could produce a quark-antiquark/ W^+W^- /charged lepton-antilepton pair, and would find its other end at another particle-antiparticle pair. This makes sense, as photons are produced in annihilation. However, due to its nature, a photon **must** have a charged particle involved, so neutrinos, etc. are disallowed.
- A gluon can only couple to particles that carry colour charge. i.e. it can only produce gluon pairs and quark-antiquark pairs. It would find itself emitted by a particle that appears on both sides. The selection is completely arbitrary.
- W bosons can produce:
 - Lepton-antineutrino or antilepton-neutrino pairs.
 - A quark and an antiquark that whose charges add up to $\pm e$.

The type of W boson is determined by the charge of the end product:

- An end-product of $+e$ charge (e.g. u and \bar{d}) indicates a W^+ boson.
- An end-product of $-e$ charge (e.g. \bar{c} and s) indicates a W^- boson.

If both ends of a W boson have one particle on each side, we do not denote the sign of the W boson as it can go either way. Sometimes a W is called a ‘charged current’.

- The Z boson, capable of producing neutrino-antineutrino pairs and quark-antiquark pairs, is largely uninteresting. However, if there is a neutrino-antineutrino pair, it **must** come from a Z boson.

Finally, as a sanity check:

- If you have a vertex connecting a lepton to a quark, [then there is definitely something wrong!](#)

[And that's it!](#)

15.6 Detectors

Quote 15.3 If you carry anything at Geneva Airport, the customs officials will not even blink an eye. If you do that in any other airport, you're in big trouble.

Andreas Korn, on transporting particle detectors, 1 February 2024

inally, we veer dangerously close to both engineering and just plain popular science by introducing the archetypal detector used in modern experimental high energy physics:

- **Tracking detector²:** All charged particles leave tracks and have their momenta measured in trackers. If particles of opposite charges went through a magnet, the opposite tracks are recorded in the tracker, allowing particles and antiparticles to be distinguished. Particles of opposite charge leave tracks in opposite directions.

²Also known as a *tracker*.

- **Electromagnetic calorimeter:** γ , e^+ and e^- produce electromagnetic showers that deposit energies. Muons deposit *smol* energies.
- **Hadronic calorimeter:** Quarks and gluons are not observable due to colour confinement but produce hadron jets via hadronisation. Each jet then initiates a *hadronic shower* which then deposits energy in hadronic calorimeters. Muons deposit *smol* energies.
- **Muon spectrometer:** The higher mass of muons means that they do not produce showers, and the aforementioned *smol* energies are not useful experimentally. Rather, muons are measured in the outermost muon spectrometer and deposit momenta.
- **Magnet:** Often this takes the form of a solenoid magnet. It does not detect anything on itself and is not necessary in all particle detectors. However, it is always present in *collider* detectors. This is because the magnet enables the direct measurement of charged particle momenta from track curvature. Without it, one is forced to infer momenta from energy deposits in calorimeters, which is less precise. Also, due to the magnet, particles of opposite charge make curved paths in opposite directions, allowing them to leave tracks in opposite directions in the tracker.

There are, however, a few special cases:

- While *neutrinos* elude all detectors, they can be constructed via the missing momentum of a process.
- *b*-hadrons, *c*-hadrons and τ leptons can be identified (or rather *tagged*) by reconstructing the displaced secondary vertex. This is because they decay slowly enough to be displaced from the primary vertex.
- Short-lived particles can be statistically identified by reconstructing the invariant mass of their decay products.

To differentiate particles that go through the same detectors and have identical charge but different masses, two possibilities exist:

- Run ‘em through a tracker and a hadronic calorimeter to determine their momenta and energies.
- Run ‘em through a tracker and measure the time of flight.

In both cases, we are able to compare the particles’ masses in the end.

Now we briefly discuss the hardware. In collider detectors like ATLAS, the configuration, in order of increasing distance from the beam pipe, is as follows:

- The tracking detector.
- The magnet. This may seem confusing because the magnet has to first curve the charged particles before they go through the tracker. However, the ‘magnet’ here is nothing but the hardware (e.g. solenoid coil for ATLAS). The field itself is inside the tracker, and the particles start curving immediately when they enter the tracker. This is intuitive when one realises that *had* the hardware been placed inside the tracking volume, the particles would have *scattered* because of this, degrading the position resolution.
- The EM and hadronic calorimeters. Here only energies are deposited, so the magnet can be placed before it without complications.
- The muon spectrometer. This is because the muons, uniquely, have very high penetration. They do not interact much with the calorimeters and thus (as mentioned before) lose little energy from them, whereas almost all non-muons do not survive the calorimeters. It is then reasonable to place the muon spectrometer last to produce a clean sample.

One last comment concerns what we mean when we say ‘tracking detector’, ‘magnet’ or ‘calorimeter’. When we speak of an instrument here, we actually speak of a *category of instruments* instead of any specific design, which varies by detector. In CERN alone, the tracking detector varies by experiment. We again proceed in order of increasing distance from the beam pipe:

- In ATLAS, the tracking detector consists of a pixel detector, a silicon strip detector (also known as a semiconductor tracker) and a transition radiation tracker. The first two are known as the *inner tracking* whereas the last is known as the *outer tracking*.
- In ALICE, the inner tracking also consists of a pixel detector and a silicon strip detector. The outer tracking is a time projection chamber.
- In CMS, there is only an inner tracking: a pixel detector and a silicon strip detector.

Quote 15.4 With the pixel detector, regions of interest can be set according to needs and a range of scattering angles in horizontal and vertical dimension is enabled.

Christina Bömer, 2020

Chapter 16

Instead of a postscript

Quote 16.1 There is no end, there will be just a point where the base case of the base case finally works.

Felix Halbwedl, 24 July 2025

16.1 Beyond the standard model

And this is where we are so far: the so-called *standard model* which describes the strong, weak and electromagnetic forces. However, many problems remain. Among them:

- Even though there is a unification of electromagnetic and weak forces, we still have two distinct couplings g and g' instead of one.
- There is no unification between electroweak theory and QCD¹.
- Quote 14.6.

The standard model is well-understood to be insufficient for describing physics. Hence, there is a need to go *beyond the standard model* (BSM). The next immediate task is to construct a *grand unified theory* (GUT), which unify electroweak theory and QCD.

Quote 16.2 The aim of grand unified theory is to rectify this by unifying strong, weak and electromagnetic interactions in a grand unified gauge theory with a single coupling constant.

David Bailin and Alexander Love, 1986

From the 1970s on, there have been various attempts to do this, including the SU(5) (which has since somewhat fallen out of favour) and SO(10) grand unified theories. However, problems remain. For example, all GUTs predict proton decay, which is yet to be observed. There is also the problem with supersymmetry which has likewise fallen out of favour due to strong evidence against it.

Should this be accomplished, we then have the problem of gravity. As shown in Part IV, gravity is infamously non-renormalisable in four dimensions, which has so far prevented us from constructing a well-accepted theory of quantum gravity. Again, this has not prevented us from trying.

Quote 16.3 Imagine a modern version of Einstein who would have learned all the standard techniques of field theory description. We would have eventually constructed GR from the field theory perspective, just years later.

Lavinia Heisenberg, March 2019

The possibility of developing a theory of quantum gravity was first looked into in the 1960s by Richard Arnowitt, Stanley Deser and Charles W. Misner, who developed the so-called *ADM formalism* which will

¹Despite this, we sometimes write the standard model as $SU(3) \times SU(2)_L \times U(1)_Y$. This, however, is a *direct product* between the groups instead of a simple group, and does not imply electroweak-QCD unification.

be covered in the (hopefully) upcoming appendix of this book.

The ADM formalism itself does not quantise gravity but rather provides a framework for it that resolves some prominent problems. For example, the $3 + 1$ decomposition of spacetime into a series of constant time hypersurfaces by applying a pair of tetrad fields on the metric allows us to write out the canonical momentum, which is not well-defined in curved space. In the end, this gives us the Wheeler-DeWitt equation, which is the equation of motion of gravity written in the field theory formalism. Ironically, however, the ADM formalism is primarily known to today's relativists for its strong applications in the field of numerical relativity.

Many (and perhaps too many) theories of quantum gravity exist, and it is impossible to list all of them. And even if we somehow quantise gravity, there is *still* the problem of unifying it with other forces. From the 1970s on, the so-called *string theory* emerged as a strong candidate for a fundamental theory that unifies all four forces. String theory reached its heyday in the 1990s and still inhabits the public consciousness today as the foremost theory of everything. However, it too declined in prestige in the recent decades, not so much because of any strong evidence against it but rather because of there being no strong evidence for it, which has raised some concerns that string theory is a philosophical theory rather than a physical one. Today, string theory only retains a significant following in its birthplace of North America.

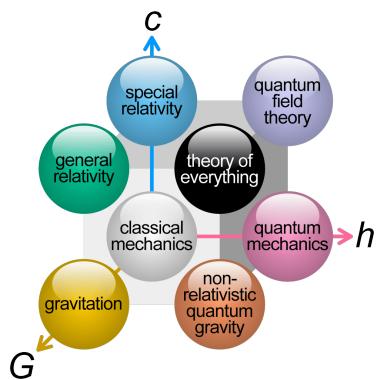


Figure 16.1: The so-called cGh (speed of light, gravitational constant and the Planck constant) *cube* illustrates attempts at unifying physics.

16.2 Where do we go from here?

To keep this book as short as possible, many important topics are regrettably left unaddressed:

- IR and UV singularities
- Instantons
- Anomalies
- Conformal field theory
- Gravitation
- ...and so on.

Some of these topics will likely be added to this book in future editions.

Quote 16.4 Our two statements actually describe the same, but are written down on two different sides of the same medal.

As they said, it's not the difficulty of the territory. You can avoid the most dangerous cliffs if you take the right route. It's more the sheer size which bends your knee, as you have to cross long distances in the realm of QFT.

Felix Halbwedel, on Quote 1.1 and Quote 1.2, 17 November 2024



Figure 16.2: ‘There is nothing we can do...’