On the Yang–Mills Existence and Mass Gap Problem

The essential mathematical background and why we care about the mass gap

Fotis I. Giasemis

Dissertation Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Mathematical and Theoretical Physics

University of Oxford
Trinity 2019

Abstract

The Yang–Mills existence and mass gap problem is a famous unsolved problem and one of the seven Millennium Prize problems defined by the Clay Mathematics Institute (CMI). The problem can be split into two distinct parts: First, it should be proven that, given a compact simple gauge group G, there exists a non-trivial quantum Yang–Mills gauge theory on $\mathbb{R}^4$. Then, the second part involves proving that this theory has a mass gap.

In other words, for the mass gap part, it should be proven that the mass of the least massive particle predicted by the theory is strictly positive. For example, for the strong nuclear interaction, it should be proven that glueballs have a lower mass bound, i.e. their mass satisfies: mass $\geq \Delta > 0$. It is important to confirm that there exists a mass gap, because a mass gap might imply that some forces in nature, like Yang–Mills fields, have a finite range, as opposed to the infinite range of the electromagnetic force. Generally, in the physics community, the existence of the mass gap is considered as proven but it does not satisfy the standards of mathematical rigour.

The aim of the dissertation is to study this problem. First, the essential mathematical background will be explored and then the problem will be revisited, with focus on the mass gap part. Then, we will look at the evidence for the existence of the mass gap that is convincing enough for physicists, and ultimately, we will try to understand where it falls short of the essential mathematical rigour. Finally, some of the different methods for solving the problem, and where these methods fail, will be described.

Contents

Chapter 1

Introduction

1.1 A brief history

In the beginning of the 20th century it was realised that classical physics was inadequate to describe the physical world on subatomic scales. This led to the development of quantum mechanics which, at the time, was a groundbreaking theory. One of its most puzzling aspects is its interpretation: how does the mathematical theory of quantum mechanics correspond to reality? This new theory managed to explain a lot of physical phenomena that were not previously understood.

However, quantum mechanics of particles was not the complete picture. The concept of a field made its first appearance in Maxwell's theory of electromagnetism in the 19th century. By the early 1930s it was clear that a theory that incorporates both concepts of a field and a particle was necessary. This was the beginning of quantum field theory (QFT), were classical fields were promoted to quantum fields, in a way analogous to the first quantization. The surprising result of this theory is that the distinction between particles and fields breaks down and particles are realised as different excitations of the underlying field.

Gauge theories are important as the successful QFTs describing elementary particle physics. Maxwell's theory is the classical example of a gauge theory, where its gauge symmetry group is the abelian group $U(1)$. The concept of a gauge was introduced by Hermann Weyl in 1918, in his attempt to unify general relativity with electromagnetism, and four years later Schrödinger proposed that Weyl's gauge theory could be used in the quantum mechanical description of the electron. Similar proposals were made independently by Vladimir Fock and Fritz London some years later. The first widely recognized gauge theory, the $U(1)$ symmetry of electromagnetism, was popularized by Pauli in his 1941 paper "Relativistic Field Theories of Elementary Particles" [1].

In 1954 [2], in attempt to construct a theory that would describe the strong interaction, Chen Ning Yang and Robert Mills generalized the gauge invariance of electromagnetism. They tried to build a theory with symmetry group $SU(2)$ instead of $U(1)$. This idea, together with the mechanism of *spontaneous symmetry breaking* through which massless gauge bosons acquire mass, later led to the unification of electromagnetism with the weak interaction, the electroweak theory with gauge group $SU(2) \times U(1)$. On the other hand, the theory of the strong interaction, quantum chromodynamics (QCD), was completely described, in the 1970s, by a non-abelian gauge theory in which the gauge group is $SU(3)$. The full theory, including the strong and the electroweak interaction, is now known as the *standard model*, with symmetry group $G = SU(3) \times SU(2) \times U(1)$. The tremendous success of the standard model in unifying three of the four known fundamental forces as well as its accurate experimental predictions exemplify the importance of gauge theories in physics.

1.2 Why quantum field theory?

Classical properties of gauge theory are within the reach of established mathematical methods and, indeed classical non-abelian gauge theory played an important role in the study of three- and four-dimensional manifolds. However, one does not yet have a mathematically complete example of a quantum gauge theory in four dimensional space-time. Quantum field theory started to have a central role in physics in the 20th century and it is likely that it will be important for 21st century mathematics as well.

Resulting from the study of quantum field theory, new mathematical ideas have been produced. For example, from the analysis side, new measures have been constructed to suit the needs of the particular theory such as Euclidean-invariant measures on spaces of generalized functions. Renormalization theory provides a mathematical framework for the study of singularities in QFTs, and results from this theory also apply to other areas of mathematics. On the algebraic side, it has led to new discoveries involving topics such as quantum groups.

Most importantly, geometry has plenty of examples of new mathematical structures that have their roots in the study of quantum field theory. One famous example is *mirror symmetry*, a relationship between Calabi–Yau manifolds. Initially discovered by physicists, mirror symmetry became interesting to mathematicians around 1990 when it was proven that this relationship could be used as a tool in enumerative geometry, effectively a theory of counting in geometry. Today, mirror symmetry is a major research topic in pure mathematics. Other examples include Donaldson theory of 4-manifolds and the Jones polynomial of knot theory.

For these reasons the scientific advisory board of the CMI has chosen a millennium problem about quantum gauge theory. A solution of the problem would require understanding of one deep unsolved mystery, the existence of a mass gap, and also producing a mathematically complete description of a quantum gauge theory in four dimensions.

Chapter 2

Gauge fields and gauge theory

2.1 Gauge transformations

We begin by discussing invariance in field theories as presented in [3]. Very often, for a specific theory, there exists a freedom in the choice of the configuration of the fields. A transformation between two distinct descriptions is called a *gauge transformation*. After a gauge transformation, all the observables are identical to the initial ones and the underlying invariance is called a *gauge invariance*. As an example we look at the complex scalar field theory

$$\mathcal{L} = (\partial_\mu \phi)^\dagger (\partial^\mu \phi) - m^2 \phi^\dagger \phi \, .$$

This theory has a $U(1)$ symmetry, i.e. we can do the transformation

$$\phi(x) \mapsto \phi(x) e^{i\alpha} \, ,$$

and the Lagrangian is invariant. Because α does not depend on x—it is the same at every point in spacetime—this is called a *global transformation*. This is in contrast to *local transformations*, where $\alpha = \alpha(x)$. Now, this theory is not invariant under local $U(1)$ transformations, but can be made invariant by introducing a new field $A_\mu(x)$ via the *covariant derivative* D_μ defined as

$$D_\mu = \partial_\mu + iqA_\mu(x)\,, \tag{2.1}$$

where q is known as the coupling strength. Then, insisting that $A_\mu(x)$ transforms as $A_\mu(x) \mapsto A_\mu(x) - \frac{1}{q}\partial_\mu\alpha(x)$ the new Lagrangian

$$\mathcal{L} = (D_\mu\phi)^\dagger(D^\mu\phi) - m^2\phi^\dagger\phi \tag{2.2}$$

is invariant. In the Lagrangian above we have omitted the gauge kinetic term but we will see how it enters into the discussion in the next section. Summarising, doing the two transformations simultaneously

$$\phi(x) \mapsto \phi(x)e^{i\alpha(x)}$$

$$A_\mu(x) \mapsto A_\mu(x) - \frac{1}{q}\partial_\mu\alpha(x)$$

our theory with Lagrangian 2.2 is unchanged. Finally, such a theory is called a *gauge theory* and the relevant $A_\mu(x)$ is known as a *gauge field.*

2.2 The QED Lagrangian

Let us now review quantum electrodynamics (QED) from the modern viewpoint, as in [4]. We begin with the complex-valued Dirac field $\psi(x)$ and insist that our theory should be invariant under the local transformation

$$\psi(x) \mapsto e^{i\alpha(x)}\psi(x)\,. \tag{2.3}$$

The question that we need to answer is the following: how do we construct the Lagrangian that is invariant under this transformation? Of course, when no derivatives are involved, this is easy to answer: add terms that are invariant to global phase rotations. For example, $m\bar{\psi}\psi$, where $\bar{\psi} = \psi^\dagger\gamma^0$.

As we saw in the previous section, when derivatives are involved things are more difficult. The directional derivative is defined as usual

$$n^\mu \partial_\mu \psi = \lim_{\epsilon \to 0} \frac{1}{\epsilon}[\psi(x + \epsilon n) - \psi(x)] \,.$$

But, in order for this definition to make sense, the two fields that are being subtracted need to have the same transformation laws. The simplest way to do this is to introduce the scalar quantity $U(y, x)$ that obeys

$$U(y, x) \mapsto e^{i\alpha(y)} U(y, x) e^{-i\alpha(x)} \tag{2.4}$$

simultaneously with 2.3. Then, set $U(x, x) = 1$ and then work infinitesimally

$$U(x + \epsilon n, x) = 1 - ie\epsilon n^\mu A_\mu(x) + \mathcal{O}(\epsilon^2) \,. \tag{2.5}$$

In this way a new vector field $A_\mu(x)$ has appeared as the coefficient of the expansion. Then, the covariant derivative takes the form 2.1, with $q = e$. By inserting the above in 2.4 we find the transformation law of $A_\mu(x)$, as we saw it in the previous section.

Now to find a kinetic term for the Lagrangian we need to use the explicit expression of the comparator $U(y, x)$. Extending equation 2.5 to the next term in the expansion of ϵ we get

$$U(x + \epsilon n, x) = \exp\left[-ie\epsilon n^\mu A_\mu(x + \frac{\epsilon}{2} n) + \mathcal{O}(\epsilon^3)\right] \,.$$

Define the product of comparators around the four corners of a square in the $(1, 2)$-plane as in

$$\mathbf{U}(x) \equiv U(x, x + \epsilon\hat{2}) U(x + \epsilon\hat{2}, x + \epsilon\hat{1} + \epsilon\hat{2}) U(x + \epsilon\hat{1} + \epsilon\hat{2}, x + \epsilon\hat{1}) U(x + \epsilon\hat{1}, x) \,. \tag{2.6}$$

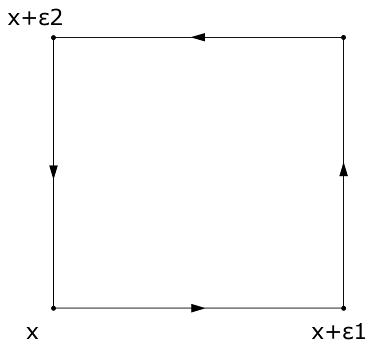


Figure 2.1: Construction of the field strength by comparisons around a small square in the $(1, 2)$-plane.

The transformation law 2.4 of U implies that $\mathbf{U}$ is locally invariant. Now take the limit $\epsilon \to 0$ to get

$$\mathbf{U}(x) = 1 - i\epsilon^2 e[\partial_1 A_2(x) - \partial_2 A_1(x)] + \mathcal{O}(\epsilon^3)\,.$$

Hence, the structure $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is locally invariant. We omit the rest of the argument, which can be found in 15.1 in [4]

The important conclusion is the following. We started by stipulating that the electron field obeys the local symmetry 2.3. From this, we showed the existence of an electromagnetic vector potential. Furthermore, the symmetry principle implies that the most general renormalizable Lagrangian in four dimensions, invariant under parity or time reversal, has the form

$$\mathcal{L} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{2.7}$$

where $F_{\mu\nu}$ is the electromagnetic tensor. Note that we have used Feynman slash notation and Einstein summation convention. This is the Maxwell-Dirac Lagrangian used in QED. Since, the group $U(1)$ is abelian, this theory is an abelian gauge theory.

2.3 The Yang–Mills Lagrangian

Just as the simple geometrical argument of the previous section starting from $U(1)$ symmetry gave us Maxwell's equations, in a much similar way, Yang and Mills generalised the argument for an arbitrary continuous and compact symmetry group G, called the *gauge group*. In this case, for an irreducible representation r of G, the Lagrangian is [4]

$$\mathcal{L}_{\text{pure YM}} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} \,, \tag{2.8}$$

where now the index a runs over all the independent generators of the local symmetry. The index r, in the generators t^a_r, refers to the representation r of G. The infinitesimal transformation laws are

$$\psi \mapsto (1 + i\alpha^a t^a)\psi$$

$$A^a_\mu \mapsto A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c \,,$$

where f^{abc} are the *structure constants* of the Lie algebra obeying

$$[t^a, t^b] = i f^{abc} t^c.$$

The field tensor $F^a_{\mu\nu}$ is defined by

$$[D_\mu, D_\nu] = -igF^a_{\mu\nu} t^a \,,$$

or more explicitly

$$F^a_{\mu\nu} = \partial_\mu A^a_\mu - \partial_\nu A^a_\mu + g f^{abc} A^b_\mu A^c_\nu \,,$$

using the expression for the covariant derivative

$$D_\mu = \partial_\mu - igA^a_\mu t^a_r \,.$$

Any such gauge theory, where the local symmetry group G is the $SU(N)$ group is called a *Yang–Mills gauge theory*. In general, the theories with a non-abelian symmetry group G are called non-abelian gauge theories.

To construct a theory of $SU(N)$ Yang–Mills fields interacting with fermions, we add the Dirac Lagrangian $\bar{\psi}(i\partial\!\!\!/ - m)\psi$ to the pure Yang–Mills Lagrangian 2.8 and perform the usual substitution

$$\partial_\mu \to D_\mu \,.$$

The new Lagrangian is then

$$\mathcal{L} = \bar{\psi}(iD\!\!\!\!/ - m)\psi - \frac{1}{4}F^a_{\mu\nu}F^{a\mu\nu}\,, \tag{2.9}$$

which looks almost identical to the QED Lagrangian 2.7. However, now ψ is an N-multiplet

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_N \end{pmatrix}\,,$$

transforming as

$$\psi_i \mapsto (\delta_{ij} + i\alpha^a(t^a)_{ij})\psi_j \,.$$

In section 4.1 we will see the definition of QCD according to the above.

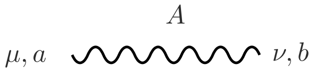
The Yang–Mills Lagrangian and all the above can be described in a more elegant way in the coordinate free approach of *differential forms*, see, for example, [5].

2.4 Quantization

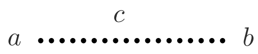
The next step is to quantize the theory. From this procedure we obtain the propagators of the theory and the Feynman rules for fermions and gauge bosons. Also, it should be noted that in this procedure new particles called *Fadeev–Popov ghosts*, which have spin 0 but fermionic statistics, are generated. These fermionic scalars must be included in internal lines for consistency even though they never appear in external states. They compensate the unphysical degrees of freedom still contained in the gauge fields A_μ using a covariant gauge-fixing condition as $\partial_\mu A^\mu = 0$. It is speculated that in order to have Lorentz invariance in a perturbative gauge theory, ghosts are unavoidable [6]. For completeness, we give the Feynman rules for a Yang–Mills theory (that has not undergone spontaneous symmetry breaking) here.[1] For a more detailed discussion and derivation of the rules, see any textbook such as [7], [4] or [8].

[1]Figures from [8].

Propagators

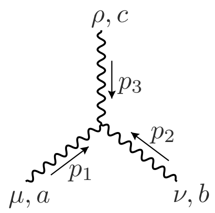


$$-\mathrm{i}\delta_{ab}\left[\frac{\eta_{\mu\nu}}{k^2+\mathrm{i}\epsilon}-(1-\xi)\frac{k_\mu k_\nu}{(k^2)^2}\right]$$



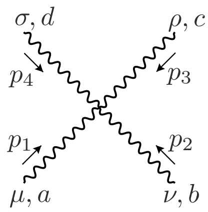
$$\delta_{ab}\ \frac{\mathrm{i}}{k^2+\mathrm{i}\epsilon}$$

Triple Gauge Interactions



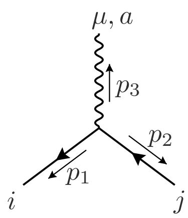
$$-g_s f^{abc}[\ \eta^{\mu\nu}(p_1-p_2)^\rho+\eta^{\nu\rho}(p_2-p_3)^\mu$$
$$+\eta^{\rho\mu}(p_3-p_1)^\nu]$$
$$p_1+p_2+p_3=0$$

Quartic Gauge Interactions



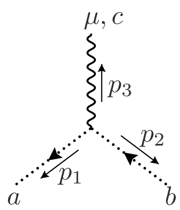
$$-\mathrm{i}g_s^2\Big[\ f_{eab}f_{ecd}(\eta_{\mu\rho}\eta_{\nu\sigma}-\eta_{\mu\sigma}\eta_{\nu\rho})$$
$$+f_{eac}f_{edb}(\eta_{\mu\sigma}\eta_{\rho\nu}-\eta_{\mu\nu}\eta_{\rho\sigma})$$
$$+f_{ead}f_{ebc}(\eta_{\mu\nu}\eta_{\rho\sigma}-\eta_{\mu\rho}\eta_{\nu\sigma})\Big]$$
$$p_1+p_2+p_3+p_4=0$$

Fermion Gauge Interactions



$$-\mathrm{i}\,g_s\gamma^\mu T^a_{ij}$$

Ghost Interactions



$$g_s\ f^{abc}p_1^\mu$$
$$p_1+p_2+p_3=0$$

Chapter 3

The problem

3.1 Quantum fields

A *quantum field* (local quantum field operator) is an operator-valued generalized function on space-time obeying certain axioms. We first have a brief look through the required properties as described at a physical level of precision, and then turn to a more rigorous definition of them.

3.1.1 Canonical quantization

As described in many textbooks, see for example [8], we start with a classical field ϕ. A *field* is map which associates to each point in space-time x a k-tuple of values $\phi_a(x)$, $a = 1, ..., k$, which transform under some representation of the Poincaré group. Then, with the Lagrangian

$$\mathcal{L} = \mathcal{L}(\phi_a, \partial_\mu \phi_a)\,,$$

we can construct the action

$$S[\phi_a] = \int_\Omega d^4x \mathcal{L}(\phi_a, \partial_\mu \phi_a)$$

and from the varying the fields ϕ_a we obtain the Euler-Lagrange equations

$$\frac{\partial \mathcal{L}}{\partial \phi_a} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi_a)} = 0\,.$$

Also, we define the momentum π_a conjugate to ϕ_a as

$$\pi_a = \frac{\partial \mathcal{L}}{\partial \dot{\phi}_a}\,.$$

The next step is to promote each field ϕ_a to an operator $\hat{\phi}_a$. Now $\hat{\phi}$ associates to each point in space-time x a k-tuple of operators $\hat{\phi}_a(x)$, $a = 1, ..., k$, and therefore is an operator-valued function. Consequently, we have promoted the classical field ϕ to a *quantum field* $\hat{\phi}$. Similarly, we promote π_a to $\hat{\pi}_a$. Finally, we impose the equal-time canonical commutation/anti-commutation relations

$$[\phi_a(\vec{x}, t), \pi^b(\vec{y}, t)] = i\delta^{(3)}(\vec{x} - \vec{y})\delta_a^b \tag{3.1}$$

$$[\phi_a(\vec{x}, t), \phi_b(\vec{y}, t)] = [\pi_a(\vec{x}, t), \pi_b(\vec{y}, t)] = 0\,, \tag{3.2}$$

where $[.,.]$ is a commutator or an anti-commutator depending on our theory. Now, since the fields are operators we also have to define the space of states upon which the fields act on. For free theories, this is the Fock space defined as

$$\text{vacuum:} \quad |0\rangle\,, \text{ such that } \hat{a}(\vec{p})\,|0\rangle = 0\,,$$

$$\text{general state:} \quad |\Psi\rangle = \hat{a}^\dagger(\vec{p_1}) \cdots \hat{a}^\dagger(\vec{p_n})\,|0\rangle\;,$$

where $\hat{a}^\dagger(\vec{p})$ and $\hat{a}(\vec{p})$ are the creation and annihilation operators in terms of which the fields can be expanded.

This is the general outline for the procedure of quantizing a field.

3.1.2 Mathematical formulation

Turning to the more mathematical treatment, we describe the axioms, known as Gårding–Wightman axioms, for the notions of field and field theory in Streater & Wightman (1964) [9].

General comments

Quantum field theory, even for the case of a free theory, suffers from ultraviolet (UV) divergences that arise from the unconstrained internal momenta in loops. This leads to the collection of techniques known as renormalization. Consequently, defining the field at a point is problematic. The Wightman formulation overcomes this problem by introducing the idea of a smeared field. For example, in the case of the electric field, $\mathcal{E}(\mathbf{x}, t)$ is not a well-defined operator, while the *smeared field* $\mathcal{E}(f) = \int d\mathbf{x}dt f(x)\mathcal{E}(\mathbf{x}, t)$ is. Here, the *test function* f is C^∞ and of compact support.

Now, we discuss the Wightman axioms.

W0 — assumptions of relativistic quantum theory

The basic idea of the zeroth axiom is that there is a Hilbert space $\mathcal{H}$, and the Poincaré group acts unitarily on that Hilbert space. In this way, we can define the energy-momentum operator P^μ, for which $P^\mu P_\mu = m^2$ is interpreted as the square of the mass. Also, the eigenvalues of P^μ lie in the forward cone. The last part of the axiom is that there exists a state $\Omega \in \mathcal{H}$, known as the vacuum, that is invariant under the action of the Poincaré group and it is unique, up to a phase.

W1 — assumptions about the domain and continuity of the field

The first axiom has to do with the domain of definition of the fields. For each test function f there exists a set of operators $\phi_1(f), ..., \phi_n(f)$, that together with their adjoints are defined on a domain D that is dense in $\mathcal{H}$ and contains the vacuum. The fields ϕ are operator-valued *tempered distributions*, i.e. generalized functions that admit a Fourier transform.

W2 — transformation law of the field

The second axiom describes the transformation law that fields $\phi_i(f)$ obey. They transform under some representation of the Lorentz group.

W3 — local commutativity, sometimes called microscopic causality

An important part of a quantum field theory is its causal structure, which is addressed by the third axiom. If the support of f and the support of g are space-like separated (in analogy to two fields being space-like separated) then

$$[\phi_i(f), \phi_j(g)]_\pm = 0\,,$$

where $[.,.]_\pm$ denotes the commutator $(-)$ or the anticommutator $(+)$.

Consequences

From these axioms several theorems follow. For example, the CPT theorem, i.e. that any Lorentz invariant local quantum field theory with a Hermitian Hamiltonian must have CPT symmetry, and the spin-statistics theorem. [10]

Existence of theories which satisfy the Wightman axioms

These axioms were defined for $\mathbb{R}^4$ but one can generalize the Wightman axioms to other dimensions other than $D = 4$. In fact, interacting theories that satisfy the axioms have been constructed in $D = 2$ and $D = 3$. However, in $D = 4$ there is no proof that the Wightman axioms can be satisfied by interacting theories. In particular, the existence part of the problem we have been examining is exactly that: to prove that the axioms can be satisfied by gauge theories in $\mathbb{R}^4$.

3.2 Statement of the problem

Now we turn to the official statement of the problem in [11]. To establish existence of four-dimensional quantum gauge theory with gauge group G, one should define a quantum field theory in the above sense. Since the vacuum vector Ω is Poincaré invariant, it is an eigenstate of the Hamiltonian with zero energy, namely $H\Omega = 0$. We know that the spectrum of H is within $[0, \infty)$, and we say that a quantum field theory has a *mass gap* Δ if H has no spectrum in the interval $(0, \Delta)$ for some $\Delta > 0$. The supremum of such Δ is the mass m and we require $m < \infty$. The statement of the problem is as follows.

Yang–Mills Existence and Mass Gap

Prove that for any compact simple gauge group G, a non-trivial quantum Yang–Mills theory exists on $\mathbb{R}^4$ and has a mass gap $\Delta > 0$. Existence includes establishing axiomatic properties at least as strong as those cited in [9], [12].

3.3 Existence

To illustrate the difficulty of the problem, we might compare it with classical Yang–Mills theory. By minimizing the Yang–Mills action, we obtain a system of non-linear partial differential equations. The most basic mathematical question here is to specify a class of initial conditions for which we can guarantee existence and uniqueness of solutions. Such questions have seen great development in recent years, using techniques from the theory of linear PDE's.

By contrast, there is no rigorous definition of quantum Yang–Mills theory, because of the difficulties of renormalization. The simplest starting point is to consider Yang–Mills theory on a lattice, or in other words, a graph Γ. That is, we can define the partition function

$$Z[\gamma, G] = \int \prod dU_i \, e^{-S} \, ,$$

over a finite subgraph $\gamma \in \Gamma$. The integral is over all holonomies in γ, S is the Yang–Mills action and the measure is the product of Haar measure for the holonomy on each edge in γ. Then, the question is whether there is a sensible way to define the limit of $Z[\gamma]$ for increasingly larger subgraphs $\gamma \in \Gamma$. This limiting process is the equivalent of taking $a \to 0$. Taking this limit will clearly involve renormalization.

In this limit, the axioms satisfied by the correlation functions in a continuum quantum field theory, and other axioms regarding the construction of the Hilbert space and the operator interpretation of the theory, have to be established. This is the existence part of the problem. [13]

Chapter 4

The mass gap

Let us now turn to the mass gap part of the problem.

4.1 QCD and its properties

We consider QCD, a well-studied example of non-abelian gauge theory. QCD is $SU(3)$ gauge theory interacting with fermions (quarks of the various flavours) each assigned to the fundamental representation of the local gauge group $SU(3)$. In other words, the Lagrangian is of the form 2.9

$$\mathcal{L}_{\mathrm{QCD}} = \sum_f \bar{\psi}_f (i\not{D} - m)\psi_f - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} \,,$$

where f is the flavour index, $G = SU(3)$ and

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} ,$$

with $\{\psi_i\}_{i=1,2,3}$ usually identified with the three different *colours* of the quarks (red, green and blue).

We saw in the previous chapter the definition of the mass gap: some constant $\Delta > 0$ such that every excitation of the vacuum has energy at least Δ. The existence of a mass gap is very important for the theory described because the absence of it might imply that the Yang-Mills fields are long-ranged, like, for example, QED. On the other hand, QCD describes the strong interaction, which is considered to be short-ranged. Therefore a mass gap is one of the required properties of QCD. Also, solving the question of the mass gap will force mathematical physicists to understand exactly what the observables of QCD are.

In order for QCD to describe the strong interaction, it must have three properties [11]:

(1) It must have a mass gap.

(2) It must have *quark confinement*.

(3) It must have *chiral symmetry breaking*.

Although, both experimental evidence and computer simulations [1] carried out since the 1970s suggest that QCD fulfils the above requirements, these properties are still not completely understood theoretically. Only in simplified models of the theory, i.e. severely truncated ones, these properties can be seen in theoretical calculations. In particular, standard perturbation theory fails in the infrared regime, i.e. low energies, where QCD is strongly coupled. This property of QCD at low energies is known as *infrared slavery*. On the other hand, in the ultraviolet regime, QCD is *asymptotically free*, that is, at high energies the interactions between quarks become weaker and weaker, and hence perturbation theory becomes reliable. This feature of QCD was discovered in 1973 [14], and we will have a brief look at it in the next section.

[1] For example, see [15].

4.1.1 Asymptotic freedom

The beta function, $\beta(g)$, of a theory, describes the dependence of the theory's coupling constant, g, on the energy scale, μ, of a given process. It is defined as

$$\beta(g) = \frac{\partial g}{\partial \log(\mu)} \, .$$

When $\beta(g)$ vanishes, the coupling parameter is independent of the scale and the theory is said to be scale invariant. On the other hand, if the beta function is non-zero, the theory has *scale dependence.* Interestingly, the coupling parameters of a quantum field theory can be scale dependent even though the classical field theory is scale invariant. This is known as a scale or conformal anomaly. There are various ways of explicitly calculating the beta function, and one example is perturbation theory.

For QCD, the one-loop beta function with n_f flavours and n_s scalar coloured bosons is [4]

$$\beta(g) = -\frac{g^3}{(4\pi)^2}\left(11 - \frac{n_s}{3} - \frac{2n_f}{3}\right) \, .$$

For appropriate values of n_f, n_s, the expression above is negative, indicating that QCD is asymptotically free. At high energies, the coupling becomes weak, and at low energies, the coupling becomes increasingly strong, which presumably leads to confinement. We will discuss confinement in the next section. This behaviour is experimentally confirmed, as shown in the figure. The coupling α is defined as usual

$$\alpha := \frac{g^2}{4\pi} \, .$$

It is also worth pointing out that even though the classical theory has no mass scale, the quantum theory, after including quantum corrections, has developed a scale dependence. This introduces a scale, which is the scale at which confinement occurs.

[1] Figure from [16].

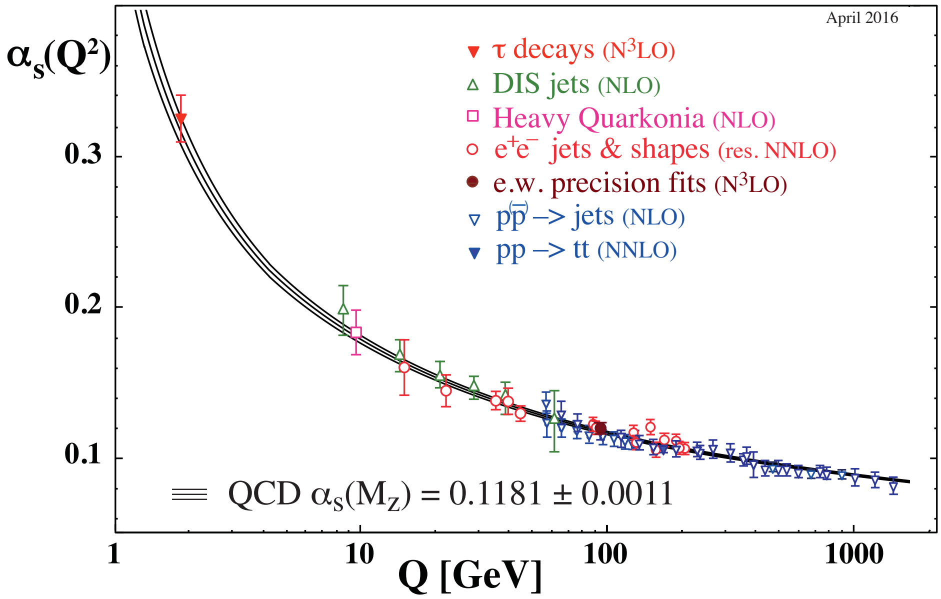


Figure 4.1: Measurements of the strong coupling α_s as a function of the energy scale Q.[1]

4.1.2 Colour confinement

The mass gap problem is closely related to colour confinement: the expected phenomenon in Yang–Mills theory/QCD that colour charged particles (e.g. quarks or gluons) cannot be isolated. Instead, quarks and gluons combine to form colourless combinations known as hadrons or mesons. There are also bound states composed only from gluons and they are called glueballs. When one tries to separate two quarks, the potential energy increases linearly until a quark anti-quark pair is created and the result is two different bound states of quarks. This process is known as string breaking. In accelerator experiments, string breaking occurs multiple times and it is observed in the form of hadronization jets [17].

It is important to note here that even though infrared slavery, which we saw in the previous section, naïvely looks like a promising explanation of confinement, in [18] it is shown that this is not the full explanation.

Confinement is related to the mass gap in the following simple way. Even if the constituent parts of a bound (due to confinement) system may be massless, the system will have a strictly positive mass. Therefore, if colour confinement is proven to be a property of QCD, the mass gap will be understood. As we mentioned earlier, there is no analytic proof of colour confinement for any non-abelian gauge theory.

As of now, the best we can do is to study computer simulations of lattice gauge theory.

4.1.3 Lattice QCD

Lattice QCD is a non-perturbative approach to the solution of QCD. It is a lattice gauge theory formulated on a grid or lattice of points in space-time. Then, one recovers continuum QCD by taking the limit of the spacing of the lattice to zero. Formulating the theory on lattice of spacing a naturally introduces a momentum cutoff of the order $1/a$, and hence the theory is regularized. As a result, lattice QCD is mathematically well-defined. This is the reason why lattice QCD is used for the investigation of non-perturbative phenomena in QCD such as colour confinement. Indeed, our best evidence that colour confinement is really a property of QCD comes from Monte Carlo simulations of lattice gauge theory. In order to explain what a lattice field theory is and the important related concepts, we will look at the Ising model of ferromagnetism.

The Ising model

We will mostly follow [19]. Consider a solid with a cubic structure. Each atom on the D-dimensional lattice can be on one of two states, "spin up" or "spin down", with the magnetic moment oriented along the direction of spin. The Hamiltonian is

$$H = -J \sum_{x} \sum_{\mu=1}^{D} s(x) s(x+\hat{\mu}) \,, \tag{4.1}$$

where $s(x) = 1$ represents an atom at x with spin up, $s(x) = -1$ represents spin down and J is a positive constant. At low temperatures in any dimension $D > 1$, most spins tend to point in the same direction and this is due to the interaction between neighbouring spins. This is known as an *ordered state*. On the other hand, at high temperatures, the system is in a *disordered state*, where the average spin is zero. Now, according to the usual principles of statistical mechanics, the probability of a spin configuration $\{s(x)\}$ at a temperature T is given by

$$\text{Prob}[\{s(x)\}] = \frac{1}{Z} \exp\left[-\frac{H}{kT}\right] ,$$

and

$$Z = \sum_{\{s(x)\}} \exp\left[\beta \sum_x \sum_{\mu=1}^{D} s(x)s(x+\hat{\mu})\right] ,$$

where $\beta = J/kT$. Now observe that the Hamiltonian $H[\{s(x)\}]$, and the probability distribution $\text{Prob}[\{s(x)\}]$, are invariant under the transformation of each spin by

$$s(x) \to s'(x) = zs(x) \quad \text{where } z = \pm 1 . \tag{4.2}$$

The two transformations for $z = 1$ and for $z = -1$ form a group, known as $\mathbb{Z}_2$. The operation 4.2 is a global transformation. Considering the average spin

$$\langle s \rangle = \sum_{\{s(x)\}} \frac{1}{N_{\text{spins}}} \left(\sum_{x'} s(x')\right) \text{Prob}[\{s(x)\}] ,$$

it is apparent now that it should be zero. From this argument it would appear that permanent magnets are impossible. This is correct in the sense that it is impossible to have permanent magnets at a finite temperature for infinitely long time. So the above is formally true, but for "practical" purposes completely wrong. It is therefore useful to introduce an external magnetic field h

$$H_h = -J \sum_x \sum_{\mu=1}^{D} s(x)s(x+\hat{\mu}) - h \sum_x s(x) ,$$

$$Z_h = \sum_{\{s(x)\}} \exp\left[-H/kT\right] ,$$

so that $\langle s \rangle \neq 0$ at any temperature, and then consider the limit

$$m = \lim_{h \to 0} \lim_{N_{\text{spins}} \to \infty} \frac{1}{Z_h} \sum_{\{s(x)\}} \frac{1}{N_{\text{spins}}} \left(\sum_{x'} s(x') \right)$$

$$\times \exp \left[\beta \sum_{x} \sum_{\mu=1}^{D} s(x) s(x+\hat{\mu}) + \frac{h}{kT} \sum_{x} s(x) \right] .$$

In this case, we can have $m = \langle s(x) \rangle \neq 0$, and we say that the $\mathbb{Z}_2$ global symmetry is spontaneously broken. That is, despite the invariance of the Hamiltonian, an observable (such as the magnetization m) which is not invariant under the $\mathbb{Z}_2$ symmetry can nevertheless give a non-zero expectation value. At high temperatures the symmetry is unbroken. In general, in the unbroken symmetry phase, the symmetry of the Hamiltonian implies the vanishing of the expectation values of observables that are not invariant under the symmetry group. In the broken phase, non-invariant observables can have non-zero expectation values. It turns [19] out that the symmetry-breaking phase transition (e.g. what happens when heating an Ising ferromagnet beyond its *Curie temperature*) appears only for $D \geq 2$.

Gauge invariance: the unbreakable symmetry

The spin system described is an example of a lattice field theory. The points of the lattice are known as *sites* and the lines joining neighbouring sites are *links*. Consider a local gauge transformation. The trick now is to associate the dynamical degrees of freedom with the links of the lattice. Denote by $s_\mu(x)$ the spin associated to the link between sites x and $x + \hat{\mu}$. The Hamiltonian for the gauge-invariant Ising model is

$$H = -J \sum_{x} \sum_{\mu=1}^{D-1} \sum_{\nu > \mu}^{D} s_\mu(x) s_\nu(x+\hat{\mu}) s_\mu(x+\hat{\nu}) s_\nu(x) ,$$

and it is invariant under

$$s_\mu(x) \to z(x) s_\mu(x) z(x+\hat{\mu}) ,$$

where the $z(x) = \pm 1$ can be chosen to be different at each site. This is the local $\mathbb{Z}_2$ gauge symmetry of the Ising model. As a consequence, we have [19] [20] Elitzur's theorem: *A local gauge symmetry cannot break spontaneously. The expectation value of any gauge non-invariant local observable must vanish.* Therefore, the average spin on a link will be zero, even if we introduce an external field h. We have to look, instead, to gauge-invariant observables. These can be constructed by taking the product of spins on links around a closed loop C

$$W(C) = \left\langle \prod_{(x,\mu)\in C} s_\mu(x) \right\rangle .$$

This is a particular example of a *Wilson loop*, and it is analogous to the construction of $\mathbf{U}$ in equation 2.6. These loops were in introduced by Kenneth G. Wilson in 1974 [21] in an attempt to formulate QCD non-perturbatively.

One can generalize the construction for $\mathbb{Z}_2$ to any symmetry group G, just by choosing link variables which are elements of G. Then, the action, written in terms of Wilson loops, has to be evaluated for different configurations and summed over to make the total sum. This is done numerically, after performing a Wick rotation on the action to get the Euclidean action. After Wick rotation, the field theory can be regarded as a statistical (rather than quantum) system and here the powerful Monte Carlo method comes in. In a lattice Monte Carlo, the idea is to replace the integral over all configurations, weighted by the distribution $\frac{1}{Z}\, e^{-S[U]}$, by an average over a finite set of sample lattice configurations $\{U^{(n)}\}_{n=1,\ldots,N_{\mathrm{conf}}}$

$$\begin{aligned} \langle Q \rangle &= \int \mathcal{D}U Q[U] \frac{1}{Z}\, e^{-S[U]} \\ &\approx \frac{1}{N_{\mathrm{conf}}} \sum_{n=1}^{N_{\mathrm{conf}}} Q[U^{(n)}] \, , \end{aligned}$$

where the sample configurations are generated stochastically. One has finally to perform an extrapolation to the continuum by approaching the limiting case in which the lattice spacing $a \to 0$ [22].

4.2 Clustering

Now, we look at another important implication of the existence of a mass gap. Locality in QFT is implemented by using local interactions in the Lagrangian density, i.e. only products of fields and their derivatives at a single point [23]. It turns out that the structure of correlators in a QFT is central to understanding characteristics of the theory, one of them being locality. In particular, one is interested in the spacelike asymptotic behaviour of truncated correlators comprised of field clusters. This behaviour is characterised by the *cluster decomposition theorem.* The proof of this theorem relies on using an axiomatic QFT approach, similar to what we saw in section 3.1.2. [24]

Having the *cluster decomposition property*, the vacuum expectation value of a product of operators in some region A and some other operators in some region B, far away from A, asymptotically equals the product of the expectation value of the product of operators in A times the expectation value of operators in B. Consequently, sufficiently separated regions behave independently.

Going back to the mass gap, not only does it have a physical significance as we saw in previous sections, but also its existence implies clustering [11]. Let $\vec{x} \in \mathbb{R}^3$ be a point in space. Also, let H and $\vec{P}$ denote the energy and momentum, generators of time and space translation. For any positive constant C less than the mass gap Δ and for any local quantum field operator $\mathcal{O}(\vec{x}) = e^{-i\vec{P}\cdot\vec{x}}\mathcal{O}e^{i\vec{P}\cdot\vec{x}}$ such that $\langle\Omega|\mathcal{O}|\Omega\rangle$=0, one has

$$|\langle\Omega|\mathcal{O}(\vec{x})\mathcal{O}(\vec{y})|\Omega\rangle| \leq \exp\left(-C|\vec{x}-\vec{y}|\right) , \tag{4.3}$$

for sufficiently large $|\vec{x}-\vec{y}|$. As before, $\Omega \in \mathcal{H}$ is the vacuum vector. Equation 4.3 implies $\lim_{|\vec{x}-\vec{y}|\to\infty} |\langle\Omega|\mathcal{O}(\vec{x})\mathcal{O}(\vec{y})|\Omega\rangle| = 0$ which translates into the fact that sufficiently separated operators are independent. This clustering property, may make it possible to apply mathematical results established on $\mathbb{R}^4$ to any 4-manifold [11] and hence the mass gap may also be important for mathematical applications of four-dimensional quantum gauge theory to geometry.

Chapter 5

Concluding remarks

In this work, we discussed the mathematical concepts that are essential in order to understand the Yang–Mills problem. Starting with gauge transformations, we motivated the QED Lagrangian and consequently generalized the procedure to write down the Yang–Mills Lagrangian. Then, we described quantum fields at the physical level of precision and compared this description to the corresponding mathematical formulation in the context of the Wightman axioms.

The official statement of the problem was then quoted from [11] and the mass gap was defined. We then briefly looked at the challenges of solving the existence part of the problem. A more extensive discussion of the mass gap was given in the next chapter. Using QCD as an example of non-abelian gauge theory, we examined asymptotic freedom and colour confinement, and how these concepts are related to the QCD mass gap. Interestingly, these concepts are not mathematically understood but have only been shown to be true in lattice QCD computer simulations, where the theory has been severely simplified.

Apart from the physical implications of the mass gap, we also saw that there are important mathematical implications. One particularly important implication stems from the clustering property. The clustering property may make it possible to apply mathematical results established on $\mathbb{R}^4$ to any four-dimensional manifold. Therefore, the mass gap may be important for applications of quantum gauge theory to geometry. This is one of the reasons that the CMI scientific advisory board has chosen a problem about quantum gauge theory in four dimensions to be part of the millennium prize problem list.

Acknowledgements

I would like to thank my supervisor Prof Philip Candelas for stimulating and enlightening discussions as well as my academic advisor Prof Xenia de la Ossa for her useful guidance throughout the MMathPhys course. Also, I would like to express my sincere appreciation to all my lecturers and in particular Dr Lucian Harland-Lang for his insightful lectures on quantum field theory. In the late stages of this work I received helpful input by Prof Christopher Beem, whom I would also like to thank.

Bibliography

[1] W Pauli. *Relativistic Field Theories of Elementary Particles*. Rev. Mod. Phys. 13 (1941) 203-232. doi: 10.1103/RevModPhys.13.203.

[2] C N Yang, R L Mills. *Conservation of Isotopic Spin and Isotopic Gauge Invariance*. Phys. Rev. 96 (1954) 191-195. doi: 10.1103/PhysRev.96.191.

[3] T Lancaster, S J Blundell. *Quantum Field Theory for the Gifted Amateur*. New York, NY: Oxford University Press, 2014.

[4] M E Peskin, D V Schroeder. *An Introduction to Quantum Field Theory*. Boca Raton, FL: CRC Press, 2018.

[5] T Eguchi, P B Gilkey, A J Hanson. *Gravitation, Gauge Theories and Differential Geometry*. Phys. Rept. 66 (1980) 213. doi: 10.1016/0370-1573(80)90130-1.

[6] M D Schwartz. *Quantum Field Theory and the Standard Model*. New York, NY: Cambridge University Press, 2014.

[7] S Weinberg. *The Quantum Theory of Fields, Volumes I and II*. Cambridge, UK: Cambridge University Press, 1995.

[8] M Kachelriess. *Quantum Fields, From the Hubble to the Planck Scale*. New York, NY: Oxford University Press, 2018.

[9] R F Streater, A S Wightman. *PCT, Spin and Statistics, and All That*. Princeton, NJ: Princeton University Press, 2000.

[10] R F Streater. *Wightman Quantum Field Theory*. Scholarpedia, 4(5):7123, (2009). doi: 10.4249/scholarpedia.7123.

[11] A Jaffe, E Witten. *Quantum Yang–Mills Theory, Official Problem Description*. 2000. http://www.claymath.org/sites/default/files/yangmills.pdf.

[12] K Osterwalder, R Schrader. *Axioms for Euclidean Green's Functions*. Commun. Math. Phys. 31 (1973) 83-112. doi: 10.1007/BF01645738.

[13] M R Douglas. *Report on the Status of the Yang–Mills Millenium Prize Problem*. 2004. http://www.claymath.org/sites/default/files/ym2.pdf.

[14] D J Gross, F Wilczek. *Ultraviolet Behavior of Nonabelian Gauge Theories*. Phys. Rev. Lett. 30 (1973) 1343-1346. doi: 10.1103/PhysRevLett.30.1343.

[15] M Creutz. *Monte Carlo Study of Quantized SU(2) Gauge Theory*. Phys. Rev. D21 (1980) 2308-2315. `doi: 10.1103/PhysRevD.21.2308`.

[16] M Tanabashi et al. *Review of Particle Physics*. Phys. Rev. D98 (2018) no.3, 030001. `doi: 10.1103/PhysRevD.98.030001`.

[17] A Ali, G Kramer. *Jets and QCD: A Historical Review of the Discovery of the Quark and Gluon Jets and its Impact on QCD*. Eur. Phys. J. H36 (2011) 245-326. `doi: 10.1140/epjh/e2011-10047-1`.

[18] C Alabiso, G Schierholz. *Infrared Slavery and Quark Confinement*. Nucl. Phys. B110 (1976) 81. `doi: 10.1016/0550-3213(76)90421-1`.

[19] J Greensite. *An Introduction to the Confinement Problem*. Lect. Notes Phys. 821 (2011) 1-211. `doi: 10.1007/978-3-642-14382-3`.

[20] S Elitzur. *Impossibility of Spontaneously Breaking Local Symmetries*. Phys. Rev. D12 (1975) 3978-3982. `doi: 10.1103/PhysRevD.12.3978`.

[21] K G Wilson. *Confinement of Quarks*. Phys. Rev. D10 (1974) 2445-2459. `doi: 10.1103/PhysRevD.10.2445`.

[22] R Iengo. *Quantum Field Theory, An Arcane Setting for Explaining the World*. San Rafael, CA: Morgan & Claypool Publishers, 2018. `doi: 10.1088/978-1-6432-7053-1`.

[23] D Tong. *Lectures on String Theory*. `arXiv:0908.0333v3 [hep-th] 23 Feb 2012`.

[24] P Lowdon. *Confinement and the Cluster Decomposition Property in QCD*. Nucl. Part. Phys. Proc. 282-284 (2017) 168-172. `doi: 10.1016/j.nuclphysbps.2016.12.032`.