School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand

An Introduction to the Axion Solution to the Strong CP Problem

Joseph A. Wilson

Supervisor Assoc. Prof. Jenni Adams

«Compiled on 15/09/20, 16:29»

UC
UNIVERSITY OF
CANTERBURY

"Astr430" Literature Review

Abstract

Literature review on something relating to axions.

Contents

1   QCD and the Strong CP Problem

THE STANDARD MODEL of particle physics is a quantum field theory which describes all known fundamental particles and interactions to very high precision—with notable exceptions including: the absence of gravity; dark matter; and the experimentally incorrect prediction that neutrinos are massless. Alongside these shortcomings, the standard model also exhibits issues of a more philosophical nature, such as the reasons for the values of the theory's many free parameters. (For example; why are there three generations of particles? Why do the particle masses and coupling constants have the values they do?) Among these mysteries is the *strong CP problem*, a fine-tuning problem regarding the apparent symmetry of the strong force under time reversal.[1] This chapter summarises the background theory necessary to understanding the statement and origin of the strong CP problem.

The standard model was born amid the explosion of phenomenological particle physics in the early 1960s, in an effort to explain the patterns in the rapidly increasing catalogue of known particles. The introduction of the quark model explained the properties of numerous hadrons in terms of six kinds of quark, and was given the name *quantum chromodynamics* (QCD). This, together with the well-established quantum theory of electromagnetism, *quantum electrodynamics* (QED), became the first approximation to the standard model [1].

The modern statement of the standard model is that it is a Lorentz-invariant *quantum field theory* whose *gauge group* is the compact connected Lie group

$$\left(\mathrm{SU}(3) \oplus \mathrm{SU}(2) \oplus \mathrm{U}(1)\right)\big/\mathbb{Z}_6, \tag{1.1}$$

which is equipped with a particular action on matter fields. A quantum field theory is a quantised *gauge field theory*, which is a field theory whose Lagrangian is symmetric under the action of the gauge group (elaborated upon in § 1.1). Each term in the gauge group of the standard model (1.1) corresponds to a fundamental force of nature: The factor $\mathrm{SU}(3)$ is the gauge group of quantum chromodynamics, the sector of the standard model describing the interactions of colour-charged quarks and gluons via the strong force. The factor $\mathrm{SU}(2) \oplus \mathrm{U}(1)$ corresponds to the unified electroweak force, and contains another $\mathrm{U}(1)$ sub-factor corresponding to the electromagnetic force of QED.

Quantum chromodynamics was formulated quite some time after quantum electrodynamics was understood, which is a reflection that QCD is more intricate than QED. Unlike QED, quantum chromodynamics exhibits *asymptotic freedom*, meaning that the effective strength of the strong interaction between colour-changed particles *increases* with increasing separation. Thus, QCD is severely non-perturbative and evades even modern analytical treatment, except

---

[1] Time reversal T and charge–parity CP symmetry are equivalent assuming the combined charge–parity–time symmetry; i.e., $T = CP \mod CPT$.

at low energies. Asymptotic freedom also results in quark confinement, which contributed to the confusion of particle physicists in the 1960s—quarks could never be observed in isolation. The dissimilarity of QCD to QED can be credited largely to the fact that the gauge group $SU(3)$ is non-Abelian, corresponding to the fact that the gluon force carriers of QCD are themselves colour-charged, which results in asymptotic freedom.

However, quantum chromodynamics remains a highly successful and beautiful theory. Aside from the technical problems of its non-perturbative nature, QCD also exhibits a mystery of purely theoretical concern: the *strong CP problem*. This problem is of the fine-tuning variety, but it lacks even an anthropic solution and is good reason to pursue physics beyond the standard model.

The significance and role of the gauge group is outlined in the next section where a geometrical overview of gauge field theory is presented. This overview is intended to equip the reader with a geometrical idea of the kinds of objects which comprise classical gauge theories and which underpin the standard model. Then, QCD is defined and the origin of the so-called *strong CP problem* is highlighted.

«Justify focus on classical gauge theory»

1.1   Overview of Classical and Quantum Gauge Theory

Gauge theories take place on a spacetime manifold and consist of two mathematically distinct dynamical entities: a *matter field* and a *gauge field*. A classical gauge theory is completely specified by the prescription of four ingredients: the base manifold; the vector space in which the matter field takes its values; a *gauge group* equipped with an action on the matter field; and equations of motion, usually provided by a Lagrangian density. The dynamical gauge field is not prescribed at the outset—it arises naturally as a consequence of the matter field's symmetry under the action of the gauge group.

The underlying premise of gauge theory is that there may be physical redundancy in the mathematical description of a matter field at each point in spacetime. (For example, the complex phase of a wavefunction is physically irrelevant.) This redundancy results in mathematical degrees of freedoms of the matter fields which are non-physical, called *local gauge freedoms*. (Keeping with our example: local rotations of phase are a gauge freedom.) Crucially, local gauge freedoms introduce ambiguity in the notion of the physical rate of change of matter fields about a point. This is because of the point-wise independence of the matter field's gauge freedom: such a freedom means that differences in the field's value between nearby points is gauge-dependent. In other words, there is no preferred directional derivative of a matter field if a local gauge freedom is present, until the choice of a *connection* is made. The triumph of gauge theory is that, by introducing the *gauge field* to act as a connection, this ambiguity is recast as a separate set of physical degrees of freedom. The implications of this are twofold: Firstly, the gauge field yields a choice of derivative (namely, the covariant derivative with respect to the gauge field), allowing the inclusion of well-defined derivatives of the matter field in the theory's equations of motion. Secondly, the gauge field has a dynamical role in the theory, describing a new kind of field: force fields (and, after quantisation, force carrier bosons).

In geometrical language, the matter field $\psi \in \Gamma(\mathcal{E})$ is a section of a vector bundle $V \rightarrowtail$

> **Notations used in this chapter**
>
> | | |
> |---|---|
> | $A, B, \dots$ | matrices |
> | $\mathcal{A}, \mathcal{B}, \dots$ | objects with manifold structure |
> | $\rightarrowtail;\ \twoheadrightarrow;\ \leftrightarrow$ | injection; surjection; bijection |
> | $F \rightarrowtail \mathcal{E} \overset{\pi}{\twoheadrightarrow} \mathcal{M}$ | fibre bundle $\mathcal{E}$ over base space $\mathcal{M}$ with fibre F and projection $\pi : \mathcal{E} \twoheadrightarrow \mathcal{M}$ |
> | $\mathrm{T}_p\mathcal{M};\ \mathrm{T}\mathcal{M}$ | tangent space at $p \in \mathcal{M}$; tangent bundle of $\mathcal{M}$ |
> | $\mathrm{T}^r_s\mathcal{M}$ | type $\binom{r}{s}$ tensors over $\mathcal{M}$, equal to $\left(\bigotimes^r_{i=1} \mathrm{T}\mathcal{M}\right) \otimes \left(\bigotimes^s_{i=1} \mathrm{T}^*\mathcal{M}\right)$ |
> | $\mathbb{T}\mathcal{M}$ | tensor bundle of $\mathcal{M}$, equal to $\bigoplus^\infty_{r,s=0} \mathrm{T}^r_s\mathcal{M}$ |
> | $\underset{\sim}{A}$ | spacetime tensor field; i.e., a section of $\mathbb{T}\mathcal{M}$ |
> | $\underset{\sim}{\boldsymbol{A}}$ | vector field of some abstract vector space not contained in $\mathbb{T}\mathcal{M}$ |
> | $\boldsymbol{A}$ | spacetime tensor field with values in some other abstract vector space |
> | $\wedge^p V$ | pth exterior power of the vector space V; i.e., the space of V-valued p-forms |
> | $\Gamma(\mathcal{E})$ | smooth sections $\mathcal{M} \to F$ of fibre bundle $F \rightarrowtail \mathcal{E} \twoheadrightarrow \mathcal{M}$ |
> | $\Omega^p(\mathcal{M})$ | space of p-forms on $\mathcal{M}$, equal to $\Gamma(\wedge^p \mathrm{T}^*\mathcal{M})$ |
> | $\Omega^p(\mathcal{M}, V)$ | space of V-valued p-forms, equal to $\Gamma(\mathcal{E} \otimes \wedge^p \mathrm{T}^*\mathcal{M})$ where $V \rightarrowtail \mathcal{E} \twoheadrightarrow \mathcal{M}$ |

$\mathcal{E} \twoheadrightarrow \mathcal{M}$. In other words, at any point $p \in \mathcal{M}$ in spacetime the field assigns a vector, $\boldsymbol{\psi}|_p \cong V$.[2] The gauge transformations of $\boldsymbol{\psi}|_p$ at a point p form a group G under composition—but this group does not describe local gauge transformations which act on the *entire* field $\boldsymbol{\psi}$. Rather, the total space of local gauge transformations is a principle G-bundle $G \rightarrowtail \mathcal{G} \twoheadrightarrow \mathcal{M}$, consisting of smooth maps $\mathcal{M} \to G$.[3] The action of $\mathcal{G}$ on the space of matter fields $\mathcal{E}$ may be denoted

$$\boldsymbol{\psi} \overset{g}{\mapsto} \boldsymbol{\psi}' = g \cdot \boldsymbol{\psi}$$

for $\boldsymbol{\psi} \in \Gamma(\mathcal{E})$ and $g \in \Gamma(\mathcal{G})$ (i.e., $\boldsymbol{\psi} : \mathcal{M} \to V$ and $g : \mathcal{M} \to G$). This specification of an action "$\cdot$" of $\mathcal{G}$ on the vector space $\mathcal{E}$ is equivalent to the choice of a linear representation $\rho : G \to \mathrm{End}(V)$ of G on V. Thus, we may represent the group action as matrix product, $g \cdot \boldsymbol{\psi} = \rho(g)\boldsymbol{\psi} \equiv g_\rho \boldsymbol{\psi}$, remembering that both $\boldsymbol{\psi}$ and g (but not ρ) vary across $\mathcal{M}$.

Here, it is worth noting that all gauge theories are isomorphic to a gauge theory possessing only one matter field $\boldsymbol{\psi}_{\text{total}}$ with one gauge group (G, ρ). For instance, if a theory consists of different fundamental particles and forces, represented by fields $\boldsymbol{\varphi}$ and $\boldsymbol{\phi}$ with gauge groups $(G_\varphi, \rho_\varphi)$ and (G_ϕ, ρ_ϕ), then we may take the total matter field $\boldsymbol{\psi}_{\text{total}} = \boldsymbol{\varphi} \oplus \boldsymbol{\phi}$ to be the direct sum of the fields in the theory, and similarly equip it with the gauge group $G_{\text{total}} = G_\varphi \oplus G_\phi$ with representation $\rho_{\text{total}} = \rho_\varphi \oplus \rho_\phi$. This new theory is physically identical to the original. In full generality, therefore, we treat our theory as possessing one matter and one gauge field, while freely speaking of its composite parts as separate fields where convenient.

A *connection* on $\mathcal{E}$ is a derivation[4] $\underset{\sim}{\nabla} : \Gamma(\mathcal{E}) \to \Omega^1(\mathcal{M}, \mathcal{E})$ from vector fields to vector-valued 1-forms. The intended meaning is that $(\underset{\sim}{\nabla}\boldsymbol{\psi})(\underset{\sim}{X})$—that is, the action of the V-valued 1-form $\underset{\sim}{\nabla}\boldsymbol{\psi}$ on a vector $\underset{\sim}{X} \in \mathrm{T}\mathcal{M}$—gives the directional derivative of $\boldsymbol{\psi}$ along $\underset{\sim}{X}$ (with respect to the connection $\underset{\sim}{\nabla}$). Employing a basis $\{\boldsymbol{e}_a\}$ of $\mathcal{E}$ and local coordinates $\{x^\mu\}$ of $\mathcal{M}$, *any* connection

---

[2] The reason many objects in gauge theory (such as fields $\boldsymbol{\psi}$) are defined as sections of fibre bundles (and not simply as smooth maps $\boldsymbol{\psi} : \mathcal{M} \to V$) is because fibre bundles are themselves smooth manifolds which admit the construction of connections.

[3] We sometimes loosely refer to the fibre group G as the gauge group, but strictly this means the fibre bundle $\mathcal{G}$ (following [2]).

[4] More precisely, $\underset{\sim}{\nabla}$ is a $\mathcal{C}^\infty(\mathcal{M})$-linear derivation, meaning $\underset{\sim}{\nabla}(f\boldsymbol{u} + g\boldsymbol{v}) = f\underset{\sim}{\nabla}\boldsymbol{u} + g\underset{\sim}{\nabla}\boldsymbol{v}$ for scalar fields

is of the form

$$\nabla_{\sim}\psi = \underset{\sim}{\mathrm{d}}\psi + \boldsymbol{A}\psi$$
$$= \left(\partial_\mu\psi^a{}_b + A_\mu{}^a{}_b\psi^b\right)\boldsymbol{e}_a \otimes \underset{\sim}{\mathrm{d}}x^\mu$$

for some matrix-valued 1-form $\boldsymbol{A}$, where $\underset{\sim}{\mathrm{d}}$ is the exterior derivative. (More precisely, $\boldsymbol{A} \in \Omega^1(\mathcal{M}, \mathfrak{g})$ is a $\mathfrak{g}$-valued 1-form, where $\mathfrak{g}$ is the Lie algebra of the gauge group G.) If $\nabla_{\sim}\psi$ is required to transform like the matter field ψ under local gauge transformations, that is, as

$$\nabla_{\sim}\psi \overset{g}{\mapsto} g \cdot (\nabla_{\sim}\psi) = (g \cdot \nabla_{\sim})(g_\rho\psi) = g_\rho\nabla_{\sim}\psi,$$

then $\nabla_{\sim}$ is called a *covariant derivative* and the connection 1-form $\boldsymbol{A}$ consequently obeys

$$\boldsymbol{A} \overset{g}{\mapsto} g \cdot \boldsymbol{A} = g_\rho\boldsymbol{A}g_\rho^{-1} - \left(\underset{\sim}{\mathrm{d}}g_\rho\right)g_\rho^{-1}. \tag{1.2}$$

Such a connection ∇_A is not unique; it depends on the choice of the 1-form field $\boldsymbol{A}$. It is exactly this connection 1-form which is promoted to a dynamical object in a gauge theory and given the name *the gauge field*. In order for the gauge field to be incorporated in the theory's equations of motion, we would like to have a notion of its derivative. However, the covariant derivative $\nabla_{\sim} : \Gamma(\mathcal{E}) \cong \Omega^0(\mathcal{M}, \mathcal{E}) \to \Omega^1(\mathcal{M}, \mathcal{E})$ is not readily defined on 1-forms such as $\boldsymbol{A}$ until we make the natural extension[5] to a *covariant exterior derivative* $\underset{\sim}{\mathrm{d}}_\nabla : \Omega^p(\mathcal{M}, \mathcal{E}) \to \Omega^{p+1}(\mathcal{M}, \mathcal{E})$, given by

$$\underset{\sim}{\mathrm{d}}_\nabla\boldsymbol{\varphi} := \underset{\sim}{\mathrm{d}}\boldsymbol{\varphi} + \boldsymbol{A} \wedge \boldsymbol{\varphi}.$$

This enables the construction of, among other things, the curvature 2-form or *gauge field strength*

$$\boldsymbol{F} := \underset{\sim}{\mathrm{d}}_\nabla\boldsymbol{A}$$
$$= \underset{\sim}{\mathrm{d}}\boldsymbol{A} + \boldsymbol{A} \wedge \boldsymbol{A} \equiv \left(\underset{\sim}{\mathrm{d}}A^a{}_b + A^a{}_c \wedge A^c{}_b\right)\boldsymbol{e}_a \otimes \boldsymbol{e}^b,$$

where $\{\boldsymbol{e}_a\}$ and $\{\boldsymbol{e}^a\}$ form a basis and dual basis of $\mathcal{E}$. The field strength is also equivalent to

$$= \nabla_{\sim} \wedge \nabla_{\sim} \equiv [\nabla_\mu, \nabla_\nu]\underset{\sim}{\mathrm{d}}x^\mu \wedge \underset{\sim}{\mathrm{d}}x^\nu.$$

The field strength $\boldsymbol{F}$ is useful because it is geometric, in the sense that it transforms like the matter field; $\boldsymbol{F} \mapsto g \cdot \boldsymbol{F} = g_\rho\boldsymbol{F}g_\rho^{-1}$ under gauge transformations, even though $\boldsymbol{A}$ does not.

«The gauge field strength also satisfies the Bianchi identity $\underset{\sim}{\mathrm{d}}_\nabla\boldsymbol{F} = 0$.»

We have seen how the gauge field $\boldsymbol{A}$ and its strength $\boldsymbol{F}$ arise when we require the notion of a spacetime derivative for a matter field which possesses gauge freedom, and are now acquainted with the dynamical objects of a gauge theory. The matter field ψ, its derivative $\nabla_A\psi$ and the strength of the gauge field $\boldsymbol{F}$ are constructions which all transform regularly under gauge transformations. All that remains to be specified in our theory are the equations of motion, which are to be expressed in terms of these three geometrical objects in a gauge invariant manner.

$f, g \in \mathcal{C}^\infty(\mathcal{M})$ and $\nabla_{\sim}(\boldsymbol{u} \otimes \boldsymbol{v}) = \nabla_{\sim}(\boldsymbol{u}) \otimes \boldsymbol{v} + \boldsymbol{u} \otimes \nabla_{\sim}(\boldsymbol{v})$.

[5] The extension is obtained by requiring the graded Leibniz property $\underset{\sim}{\mathrm{d}}_\nabla(\varphi \otimes \boldsymbol{\psi}) = \underset{\sim}{\mathrm{d}}\varphi \otimes \boldsymbol{\psi} + (-1)^p\varphi \wedge \nabla_{\sim}\boldsymbol{\psi}$ for $\varphi \in \Omega^p(\mathcal{M})$ and $\boldsymbol{\psi} \in \Gamma(\mathcal{E})$, analogous to the usual exterior derivative.

1.1.1 Lagrangians in Field Theories

For classical gauge theories, the equations of motion may be specified as the extremisation of an action

$$S[\boldsymbol{\psi}, \nabla_A \boldsymbol{\psi}, \underset{\sim}{\boldsymbol{F}}] = \int_{\mathcal{M}} \underset{\sim}{\mathcal{L}}[\boldsymbol{\psi}, \nabla_A \boldsymbol{\psi}, \underset{\sim}{\boldsymbol{F}}],$$

where $\underset{\sim}{\mathcal{L}}$ is a local Lagrangian density. In order that the equations of motion are physically well-defined, we require the Lagrangian possesses the relevant symmetries: gauge symmetry, so that the equations of motion are gauge invariant; and Lorentz symmetry (which is automatic if $\underset{\sim}{\mathcal{L}} = \mathcal{L} \underset{\sim}{\text{vol}}$ is expressed as a volume form) [3, §7.1]. The equations of motion are invariant under adjustments to the Lagrangian density by a total derivative $\underset{\sim}{\mathrm{d}}K$, since by Stokes' theorem these contribute only to boundary terms. Therefore, a Lagrangian which is said to possess gauge symmetry is still generally permitted to transform as

$$\underset{\sim}{\mathcal{L}} \mapsto \underset{\sim}{\mathcal{L}} + \underset{\sim}{\mathrm{d}}K \tag{1.3}$$

under gauge transformations, if the fields vanish at the boundary. If a Lagrangian transforms as (1.3) under a continuous gauge symmetry parametrised by n parameters α_i, then Noether's theorem implies the existence of n conserved current densities (viz. 3-forms[6]), one for each α_i,

$$\underset{\sim}{J}_{(i)} = \frac{\partial \underset{\sim}{\mathcal{L}}}{\partial \underset{\sim}{\nabla} \boldsymbol{\psi}} \left. \frac{\partial \boldsymbol{\psi}}{\partial \alpha_i} \right|_{\mathrm{id}} - \frac{\partial \underset{\sim}{K}}{\partial \alpha_i}, \tag{1.4}$$

whose continuity equations read $\underset{\sim}{\mathrm{d}} \underset{\sim}{J}_{(i)} = 0$.

On the other hand, the Lagrangian of a quantum field theory enjoys an enlarged criterion of gauge symmetry: it is also permitted to transform as

$$\int_{\mathcal{M}} \underset{\sim}{\mathcal{L}} \mapsto \int_{\mathcal{M}} \underset{\sim}{\mathcal{L}} + n \cdot 2\pi\hbar, \tag{1.5}$$

where $n \in \mathbb{Z}$ may vary discretely under different gauges. This is because of the origin of a QFT's equations of motion in the Feynman path integral. For instance, the quantum mechanical amplitude that the fields $\boldsymbol{\psi}$ and $\boldsymbol{F}$ satisfy prescribed boundary conditions on $\partial\Omega$ surrounding some region of spacetime $\Omega \subseteq \mathcal{M}$ is given by the path integral

$$\mathscr{A} = \int_{\partial\Omega} \mathcal{D}[\boldsymbol{\psi}, \boldsymbol{A}] \exp\left\{ \frac{i}{\hbar} S[\boldsymbol{\psi}, \underset{\sim}{\nabla}_A \boldsymbol{\psi}, \underset{\sim}{\boldsymbol{A}}] \right\}, \tag{1.6}$$

where the intended meaning of $\mathcal{D}[\cdots]$ is an integration over all field configurations on Ω. If the Lagrangian were to undergo a discrete gauge transformation (1.5), the amplitude $\mathscr{A} \mapsto \mathscr{A} \exp(n \cdot 2\pi i)$ would be left invariant. In other words, physical consistency of a QFT does not require the single-valuedness of the action S, but only of $\exp(iS/\hbar)$. This leads to an enlargement of the space of possible Lagrangian densities to include, in particular, *topological terms*. These prove to be especially relevant to QFTs and to the strong CP problem itself.

---

[6]The partial derivative $\partial\underset{\sim}{\mathcal{L}}/\partial\nabla\boldsymbol{\psi}$ of a volume form $\underset{\sim}{\mathcal{L}}$ with respect to a 1-form $\nabla\boldsymbol{\psi}$ is itself a 3-form, and is defined by naturally extending the scalar partial derivative to be an anti-derivation on differential forms. For details, see [4].

1.1.2 The Yang–Mills Lagrangian and the Topological θ-Term

An important component of a gauge theory's equations of motion are the terms in the Lagrangian which describe the dynamics of the gauge field. These specify how the gauge field behaves in the vacuum (e.g., describing the classical electromagnetic field, or the quantum theory of photons, in the absence of matter). Thus, we are interested in the possible *consistent* Lagrangians which may be constructed from the gauge field $\underset{\sim}{A}$ and not the matter field $\underset{\sim}{\psi}$.

Along with requiring Lorentz and gauge invariance, consistency also requires that the quantum field theory associated to a Lagrangian be *renormaliseble*. Loosely speaking, a classical field theory is renormalisable if it can be "quantised safely". (This restricts the form of the Lagrangian considerably, but how this happens is beyond this review's scope.) Under these constraints, the only admissible QFT Lagrangians which may be constructed from the gauge field $\underset{\sim}{A}$ alone are linear combinations of

$$\langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle \equiv \frac{1}{2} \left\langle \boldsymbol{F}_{\mu\nu}, \boldsymbol{F}_{\rho\sigma} \right\rangle_{\mathrm{Ad}} g^{\mu\rho} g^{\nu\sigma} \underset{\sim}{\mathrm{vol}}, \quad \text{and} \quad \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle \equiv \frac{1}{4} \left\langle \boldsymbol{F}_{\mu\nu}, \boldsymbol{F}_{\rho\sigma} \right\rangle_{\mathrm{Ad}} \epsilon^{\mu\nu\rho\sigma} \underset{\sim}{\mathrm{vol}},$$

where $\star$ is the Hodge dual [3, § 7.1.2]. The inner product $\langle\,,\,\rangle_{\mathrm{Ad}}$ on the Lie algebra $\mathfrak{g}$ of the gauge group G is chosen such that it is gauge-invariant.[7] The inner product $\langle\,,\,\rangle_{\mathrm{Ad}}$ on $\mathfrak{g}$ is not unique; it depends on a choice of *coupling constants*. In particular, if the gauge group G is the direct sum of n simple Lie groups,[8] then $\langle\,,\,\rangle_{\mathrm{Ad}}$ is specified by the choice of exactly n coupling constants, one corresponding to each factor of G [3, § 2.5]. Physically, the coupling constants determine the relative interaction strengths of the forces associated to each factor of G, and must enter the theory as free parameters determined experimentally. (For instance, the gauge group of the standard model (1.1) has three such coupling constants for the strong $\mathrm{SU}(3)$, weak $\mathrm{SU}(2)$, and electromagnetic $\mathrm{U}(1)$ interactions.)

The term $\langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle$ is known as the *Yang–Mills Lagrangian*, and is a major component in the standard model, describing boson force carrier propagation and self-interaction (such as gluon self-interactions). The Yang–Mills Lagrangian yields the equation of motion $\mathrm{d}_\nabla \star \underset{\sim}{\boldsymbol{F}} = 0$. For the Abelian gauge group $G = \mathrm{U}(1)$ of electromagnetism, this equation, together with the Bianchi identity $\mathrm{d}_\nabla \underset{\sim}{F} = 0$, are the source-free Maxwell equations. (Since the Lie algebra of $\mathrm{U}(1)$ is $\mathbb{R}$, the 2-form $\underset{\sim}{F}$ is scalar-valued for electromagnetism.) Expressing $\underset{\sim}{F} = \mathrm{d}c t \wedge \underset{\sim}{E} + \star(\mathrm{d}c t \wedge \underset{\sim}{B})$ in terms of the non-relativistic electric and magnetic field 1-forms by choosing a spacetime split, the Yang–Mills term is the familiar electromagnetic energy density $\underset{\sim}{F} \wedge \star \underset{\sim}{F} = \underset{\sim}{B}^2 - \underset{\sim}{E}^2/c^2$.

The other term $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ is known as the *Chern–Simons term* or the *topological θ-term*, for reasons which will become apparent after a survey of its properties.

- The Chern–Simons term is odd under both time-reversal symmetry T and parity P (notice $\epsilon^{\mu\nu\rho\sigma} \mapsto -\epsilon^{\mu\nu\rho\sigma}$ under T or P) but not under charge conjugation C. This may also be seen by introducing a spacetime split, whereby $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \mathrm{tr}(\boldsymbol{E} \cdot \boldsymbol{B}) \underset{\sim}{\mathrm{vol}}$, since $\underset{\sim}{\boldsymbol{E}}$ is a vector of odd-parity and $\underset{\sim}{\boldsymbol{B}}$ is a pseudovector of even-parity. Hence, it may give rise to CP-violating dynamics.

---

[7] Recall that $\underset{\sim}{\boldsymbol{A}}$, and hence $\underset{\sim}{\boldsymbol{F}}$, are $\mathfrak{g}$-valued forms, so an inner product on $\mathfrak{g}$ is needed to produce a scalar. For this scalar to be gauge invariant, the inner product must additionally be "Ad-invariant" [3, § 7.3].

[8] Any compact connected Lie group is either of this form, or is a finite quotient of such a group, if $\mathrm{U}(1)$ is counted as "simple". [3, § 2.4.3].

- It is *topological* because it does not depend on the geometry of spacetime via the metric $g^{\mu\nu}$ (instead, all spacetime indices are contracted with $\epsilon^{\mu\nu\rho\sigma}$). Hence, an action $\int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ depends only on the integrand's topology over $\mathcal{M}$.

- Furthermore, $\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ is a total derivative of the *Chern–Simons 3-form* $\underset{\sim}{\omega}_3$,

$$\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \underset{\sim}{\mathrm{d}} \underset{\sim}{\omega}_3 = \underset{\sim}{\mathrm{d}} \operatorname{tr} \left(\underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\mathrm{d}} \underset{\sim}{\boldsymbol{A}} + \frac{2}{3} \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{A}} \wedge \underset{\sim}{\boldsymbol{A}} \right),$$

  meaning that the action $\int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ depends only on the topology of $\underset{\sim}{\boldsymbol{A}}$ on the spacetime boundary $\partial\mathcal{M}$. As such, it does not affect the classical equations of motion. However, it has important implications in the quantum theory.

- The integral is a discrete topological invariant

$$n = \frac{1}{8\pi^2} \int_{\mathcal{M}} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle = \frac{1}{8\pi^2} \int_{\partial\mathcal{M}} \underset{\sim}{\omega}_3 \in \mathbb{Z}, \tag{1.7}$$

  known as the *Pontryagin number*, the *second Chern class* [5, §1] or simply the 'winding number' [2, §2.2] of the gauge field configuration $\underset{\sim}{\boldsymbol{A}}$.

  Importantly, this means that the Chern–Simons term is *not* gauge invariant if there are topologically distinct gauge fields $\underset{\sim}{\boldsymbol{A}}$ with varying winding number.

The choice of the symbol θ in the name "θ-term" reflects the angular nature of any coefficient θ attached to the Chern–Simons term, such as

$$\underset{\sim}{\mathcal{L}}_\theta[\boldsymbol{A}] = \frac{\theta}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle \hbar. \tag{1.8}$$

The action of this Lagrangian is $\theta n \hbar$ wherever $\underset{\sim}{\boldsymbol{A}}$ has winding number n. Since this enters the path integral as $e^{i\theta n} = e^{i(\theta+2\pi)n}$, the coefficient θ, henceforth the *θ-parameter*, is only defined modulo 2π: it is an angular quantity. As a whole, (1.8) is referred to as the θ-term.

The standard model employs the Yang–Mills Lagrangian $\langle \underset{\sim}{\boldsymbol{F}} \wedge \star \underset{\sim}{\boldsymbol{F}} \rangle$, but does not find use for the θ-term. The unwanted θ-term is therefore quickly overlooked, and was historically considered an unphysical gauge-variant anomaly. «Cite.»

1.1.3 The θ-Term is a Consequence of the Non-trivial Vacuum

The θ-term is more than just a mathematical possibility which lacks physical reason for its inclusion in the Lagrangian. It is in fact central to non-Abelian gauge theories as a direct consequence of their *non-trivial vacuum structure*, which is responsible for interesting non-perturbative dynamical effects.

For an Abelian gauge group $G = \mathrm{U}(1)$ with total gauge group bundle $\mathcal{G} = \mathcal{U}(1)$, the space of asymptotically-identity gauge transformations $\Gamma^{\mathrm{id}}(\mathcal{U}(1))$ is continuously connected to the identity.[9] This means that any two gauge-equivalent gauge field configurations can be

---

[9] We consider the space of asymptotically-identity gauge transformations $\Gamma^{\mathrm{id}}(\mathcal{G})$, whose elements are maps $g : \mathcal{M} \to G$ with $g|_{\partial\mathcal{M}} = \mathrm{id}$ (or equivalently with $g(x) \to \mathrm{id}$ as $x \to \infty$), because $\underset{\sim}{\boldsymbol{A}}$ must be fixed on the boundary in order to prescribe boundary conditions. «Who realised this? 't Hooft? Wigner?»

continuously gauge transformed into each other, meaning all gauge transformations preserve the topology of the gauge field. Hence, the winding number is zero for all $\underset{\sim}{A} \in \Omega^1(\mathcal{M}, \mathfrak{u}(1))$. Consequently, the θ-term (with constant θ) is neither relevant in classical electromagnetism nor in QED, because it vanishes identically.

However, for non-Abelian gauge groups, the space of gauge transformations $\Gamma^{\text{id}}(\mathcal{G})$ may have a non-trivial topology, containing gauge transformations which are not diffeomorphic. In the case where $\mathcal{M}$ is spacetime and $\mathcal{G} = \mathcal{SU}(3)$ is the total gauge group of QCD, the space $\Gamma^{\text{id}}(\mathcal{SU}(3))$ consists of path-disconnected regions labelled by some $\nu \in \mathbb{Z}$. (In topological language, the third homotopy group $\pi_3(\Gamma^{\text{id}}(\mathcal{SU}(3))) \cong \mathbb{Z}$ is the group of integers.) This means that there are topologically inequivalent gauge field configurations $\underset{\sim}{A}$, with the winding number labelling each topological class.

Elements $g_0 : \mathcal{M} \to G$ in the identity-connected component of $\Gamma^{\text{id}}(\mathcal{G})$ are named *small* gauge transformations, and all others *large*. A large gauge transformation $g_\nu \in \Gamma^{\text{id}}(\mathcal{G})$ shifts the winding number n of a gauge field $\underset{\sim}{A}$ to $n + \nu$, giving rise to $\mathbb{Z}$-many gauge-equivalent fields $\{g_\nu \cdot \underset{\sim}{A}\}_{\nu \in \mathbb{Z}}$ which belong to distinct homotopy classes. Such gauge fields are know as *instantons*.

Despite the name, large gauge transformations are not truly gauge symmetries, in the sense that they are not genuine automorphisms of the equations of motion. Indeed, large gauge transformations may transition between states which can be distinguished by physical measurement. For instance, if the gauge field $\underset{\sim}{A}$ has winding number n, then the action of the θ-term

$$S[\underset{\sim}{A}] = \int_{\mathcal{M}} \mathcal{L}_\theta[\underset{\sim}{A}] = \theta n \hbar$$

is proportional to n. A large gauge transformation g_ν then shifts this action $S[\underset{\sim}{A}] \to S[g_\nu \cdot \underset{\sim}{A}] = S[\underset{\sim}{A}] + \theta \nu \hbar$. From the path integral (1.6), this induces relative phases $e^{i\theta\nu}$ across the domain of integration which may interfere and alter the amplitude. In other words, instantons are measurable.

The θ-Vacuum

Of particular interest are the implications of instantons for the QCD vacuum. In a Yang–Mills theory whose Lagrangian includes the term $\underset{\sim}{\mathcal{L}}_{\text{YM}} = \langle \boldsymbol{F} \wedge \star \boldsymbol{F} \rangle$, a *vacuum state* is one in which the field strength (and all other fields) vanish; $\boldsymbol{F} = \boldsymbol{0}$. Not only is this consistent with an identically vanishing gauge field $\underset{\sim}{A}_0 = \boldsymbol{0}$, but also with gauge transformations (1.2) of $\underset{\sim}{A}_0$,

$$g \cdot \underset{\sim}{A}_0 = (\underset{\sim}{\mathrm{d}} g_\rho)(g_\rho)^{-1},$$

which are called "pure gauge" configurations. Small gauge transformations are not measurable and describe the same vacuum state, whereas large ones $g_n \cdot \underset{\sim}{A}_0$ are, hence describing *distinct vacua* $|n\rangle$ labelled by winding number. Since $g_\nu \cdot |n\rangle = |n + \nu\rangle$, these states are not gauge-invariant. The only vacuum states which are invariant under the gauge (up to an overall phase) are those of the form

$$|\theta\rangle = \sum_{n \in \mathbb{Z}} e^{i\theta n} |n\rangle,$$

for some constant parameter θ. Thus, the theory possesses a topological circle of distinct gauge-invariant vacua $|\theta\rangle$ labelled by the *vacuum angle* $\theta \in [0, 2\pi)$.

This enriched vacuum $|\theta\rangle$ is equivalent to the inclusion of the θ-term in an effective Lagrangian, in the following way. Denote by $_+\langle n|m\rangle_-$ the quantum amplitude that the vacuum state $|n\rangle$ at $t \to -\infty$ evolves to $|m\rangle$ at $t \to \infty$. The vacuum-to-vacuum amplitude is then

$$_+\langle\theta|\theta\rangle_- = \sum_{n,m} e^{i\theta(n-m)} {}_+\langle m|n\rangle_- = \sum_\nu e^{i\theta\nu} \sum_n {}_+\langle n|n+\nu\rangle_-, \tag{1.9}$$

with all summations over $\mathbb{Z}$. In the path integral formulation (1.6), the amplitude $_+\langle n|n+\nu\rangle_-$ can be expressed explicitly as

$$_+\langle n|n+\nu\rangle_- = \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu] \exp\left\{\frac{i}{\hbar}\int_\Omega \underset{\sim}{\mathcal{L}}\right\}, \tag{1.10}$$

where $\mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu]$ means that the path integral is over all instanton gauge fields with winding number ν, since only those induce a transition $|n\rangle \mapsto |n+\nu\rangle$. Combining (1.9) and (1.10), the amplitude of evolution from the gauge invariant vacuum to itself is

$$\begin{aligned}_+\langle\theta|\theta\rangle_- &= \sum_\nu e^{i\theta\nu} \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu] \exp\left\{\frac{i}{\hbar}\int_\Omega \underset{\sim}{\mathcal{L}}\right\} \\ &= \sum_\nu \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}};\nu] \exp\left\{\frac{i}{\hbar}\int_\Omega \underset{\sim}{\mathcal{L}} + i\theta\nu\right\},\end{aligned}$$

which, using (1.7) to write $i\theta\nu$ as the θ-term in the Lagrangian, is

$$= \int_{\partial\Omega} \mathcal{D}[\underset{\sim}{\boldsymbol{A}}] \exp\left\{\frac{i}{\hbar}\int_\Omega \Big[\underbrace{\underset{\sim}{\mathcal{L}} + \frac{\theta}{8\pi^2}\langle\underset{\sim}{\boldsymbol{F}}\wedge\underset{\sim}{\boldsymbol{F}}\rangle\hbar}_{\underset{\sim}{\mathcal{L}}_{\text{eff}}}\Big]\right\}.$$

The effects of the non-trivial vacuum are encapsulated in the effective Lagrangian

$$\underset{\sim}{\mathcal{L}}_{\text{eff}} = \underset{\sim}{\mathcal{L}} + \frac{\theta}{8\pi^2}\langle\underset{\sim}{\boldsymbol{F}}\wedge\underset{\sim}{\boldsymbol{F}}\rangle\hbar.$$

It is in this sense that the θ-term arises in the Lagrangian due to the non-trivial QCD vacuum.

1.1.4 Dirac Fermion Fields and the Chiral Anomaly

«Need flowing introduction.»

An important kind of matter field in the standard model is the *Dirac fermion field* φ, which transforms under the spin-$\frac{1}{2}$ representation of the Lorentz group. Dirac fermions take their values in a 4-component complex space, denoted $\mathbb{C}_{\text{D}}^4$. The entire matter content of the standard model, including quarks and leptons, is comprised purely of such fermion fields.

The Dirac matrices $\boldsymbol{\gamma}^\mu$ form a basis for the algebra of spacetime, the Clifford algebra $\mathcal{C}l_{1,3}(\mathbb{C})$, satisfying $\boldsymbol{\gamma}^{(\mu}\boldsymbol{\gamma}^{\nu)} = \eta^{\mu\nu}$, and are used to write 4-component Dirac fermions in the spin-$\frac{1}{2}$ representation. An inner product on fermions $\langle\psi,\varphi\rangle \equiv \bar{\psi}\varphi \in \mathbb{R}$ is provided by the Dirac adjoint

$\bar{\boldsymbol{\psi}} := \boldsymbol{\psi}^\dagger \boldsymbol{\gamma}^0$, so that Lorentz-invariant quantities may be naturally constructed. Finally, fermions may be separated into left-handed $\boldsymbol{\varphi}_+$ and right-handed $\boldsymbol{\varphi}_-$ components with the projection operators $\boldsymbol{\varphi}_\pm = \frac{1}{2}(1 \pm \boldsymbol{\gamma}^5)$ where $\boldsymbol{\gamma}^5 = i\boldsymbol{\gamma}^0\boldsymbol{\gamma}^1\boldsymbol{\gamma}^2\boldsymbol{\gamma}^3$. In theories which violate parity, left- and right-handed fermions may experience different interactions. (For instance, left-handed neutrinos interact in the standard model, while right-handed neutrinos are completely inert.)

The simplest fermion equation of motion (namely, the Dirac equation) derives from the *Dirac Lagrangian density*

$$\mathcal{\underset{\sim}{L}}_{\text{Dirac}} = \bar{\boldsymbol{\varphi}}\big(i\hbar c\boldsymbol{\gamma}^\mu\partial_\mu - mc^2\big)\boldsymbol{\varphi}\ \underset{\sim}{\text{vol}},$$

which describes a single non-interacting spin-$\frac{1}{2}$ fermion of mass m. The Dirac Lagrangian may be localised in the presence of a gauge symmetry, giving

$$\mathcal{\underset{\sim}{L}}_{\text{Dirac}} = \bar{\boldsymbol{\psi}}\big(i\boldsymbol{\gamma}^\mu\nabla_\mu - m\big)\boldsymbol{\psi}\ \underset{\sim}{\text{vol}}, \tag{1.11}$$

where $\underset{\sim}{\nabla}\boldsymbol{\psi} \equiv (\nabla_\mu\boldsymbol{\psi})\,\mathrm{d}x^\mu$ is the covariant derivative with respect to the gauge field, and where we have now begun to employ units in which $\hbar = c = 1$ for brevity. The matter field $\boldsymbol{\psi} = \boldsymbol{\varphi}_{(1)} \oplus \cdots \oplus \boldsymbol{\varphi}_{(n)}$ may now more generally be comprised of multiple fermion fields, which may be rotated into one another by the gauge group (an example being *flavour symmetry* in QCD). The localised Dirac Lagrangian describes multiple fermion types and their interactions with the gauge bosons. The electromagnetic interactions of charged fermionic matter are described by (1.11) in the case of a $\mathrm{U}(1)$ gauge symmetry, where the gauge field $\boldsymbol{\underset{\sim}{A}} \equiv \underset{\sim}{A}$ is the electromagnetic vector potential.

The Anomalous Axial Symmetry

In quantum theories of fermions, the θ-term makes another important appearance, arising in the context of the *chiral anomaly*. The Dirac Lagrangian classically possesses two fundamental $\mathrm{U}(1)$ symmetries which rotate fermion phases *vectorially* or *axially*. The classical axial symmetry is violated upon quantisation, an effect known as the chiral anomaly.

To illustrate, the vectorial symmetry $\mathrm{U}(1)_V$ is the invariance of the Dirac Lagrangian under global Abelian transformations known as *vector fermion rotations*. The action of $\mathrm{U}(1)_V$ is a simple phase rotation $\psi \mapsto e^{i\alpha}\psi$, where α is the parameter. The conserved Noether current density associated to this symmetry is

$$\underset{\sim}{J}_V = \hbar c\, \bar{\psi} \star \underset{\sim}{\gamma}\, \psi, \qquad\qquad \text{i.e.,} \quad j_V^\mu = \hbar c\, \bar{\psi}\, \gamma^\mu \psi,$$

where $\underset{\sim}{\gamma} = \gamma_\mu\, \mathrm{d}x^\mu$. (These 3-form and vector representations are related by $\underset{\sim}{J}_V = \star \underset{\sim}{j}_V$.) The associated continuity equation reads $\mathrm{d}\underset{\sim}{J}_V = 0$ (i.e., $\partial_\mu j_V^\mu = 0$). Classically, this corresponds to conservation of charge, and in the quantum regime, to conservation of the number of charged particles.

On the other hand, the axial symmetry $\mathrm{U}(1)_A$ is invariance under *axial fermion rotations*, which transform left- ψ_+ and right-handed ψ_- fermion components differently:

$$\psi \overset{\mathrm{U}(1)_A}{\mapsto} e^{i\gamma^5\theta}\psi, \qquad\qquad \text{i.e.,} \quad \psi_\pm \overset{\mathrm{U}(1)_A}{\mapsto} e^{\pm i\theta}\psi_\pm. \qquad (1.12)$$

«Isn't the mass term not invariant here?» The associated Noether current density,

$$\underset{\sim}{J}_A = \hbar c\, \bar{\psi} \star \underset{\sim}{\gamma}\, \gamma_5\, \psi, \qquad\qquad \text{i.e.,} \quad j_A^\mu = \hbar c\, \bar{\psi}\, \gamma^\mu \gamma_5\, \psi,$$

is conserved classically. The charge associated to $\underset{\sim}{J}_A$ is the number of left-handed particles minus the number of right-handed, named the *baryon number* in QCD. However, the $\mathrm{U}(1)_A$ symmetry is *anomalous*, meaning it does not survive the quantisation procedure, and is not an exact symmetry of the quantum theory.[10] Specifically, an anomalous symmetry does not leave invariant the integral measure $\mathcal{D}[\cdots]$ in (1.6) of the path integral in the quantum theory [2] Instead, the continuity equation $\mathrm{d}\underset{\sim}{J} = 0$ fails by the presence of none-other than the θ-term,

$$\mathrm{d}\underset{\sim}{J}_A \propto \langle \underset{\sim}{F} \wedge \underset{\sim}{F} \rangle,$$

with the constant of proportionality depending on the model. This means that the axial current $\underset{\sim}{J}_A$ is not conserved—and hence CP symmetry is violated—in the presence of instantons, where $\int \langle \underset{\sim}{F} \wedge \underset{\sim}{F} \rangle$ is nonzero.

By Noether's theorem (1.4), this is equivalent to the addition of a total derivative to the Lagrangian $\underset{\sim}{\mathcal{L}}_{\text{eff}} = \underset{\sim}{\mathcal{L}} + \mathrm{d}K$. This total derivative is precisely the θ-term, because

$$\mathrm{d}\underset{\sim}{J}_A = \langle \underset{\sim}{F} \wedge \underset{\sim}{F} \rangle = \frac{\partial\, \mathrm{d}\underset{\sim}{K}}{\partial\theta} \qquad \Longrightarrow \qquad \mathrm{d}\underset{\sim}{K} = \theta\, \langle \underset{\sim}{F} \wedge \underset{\sim}{F} \rangle,$$

where here θ is the axial rotation parameter appearing in (1.12). The axial current $\underset{\sim}{J}'_A$ of this new Lagrangian $\underset{\sim}{\mathcal{L}}_{\text{eff}}$ is indeed conserved. In other words, an axial rotation does not leave the Lagrangian invariant, but instead generates an effective θ-term [7, § 8].

«What does this mean?? Relevant because Peccei-Quinn mechanism depends on chiral symmetry breaking. Write something about the gauge invariant QCD vacuum $|\theta\rangle$. "Recognizing the complicated nature of the QCD vacuum, effectively adds an extra term to the QCD Lagrangian." - [lecture notes on axions] »

---

[10] Indeed, baryon number need not conserved in QCD (although violations have not been observed). This may provide a possible explanation of *baryogenesis*—the mechanism by which the early universe produced more matter than antimatter [6, § 24.6].

1.2 QCD and the Strong CP Problem

Quantum chromodynamics describes the strong interactions among hadronic matter. Fields which interact via the strong force are called *colour-charged*, and in QCD, the colour-charged fields are the *quarks*.

In QCD with N_c colours, a quark is a colour-charged Dirac fermion, represented by a matter field $\psi = \varphi \otimes c$ with values in $\mathbb{C}_D^4 \otimes \mathbb{C}_C^{N_c}$, where $\varphi(x) \in \mathbb{C}_D^4$ is a plain Dirac fermion and $c(x) \in \mathbb{C}_C^{N_c}$ is a N_c-component vector in *colour space*. A quark field ψ can either be viewed as a fermion with components in colour space, or equivalently as a N_c-tuple of fermions. The gauge group is $\mathrm{SU}(N_c)$, and its action on ψ is to transform colour space under the fundamental representation; i.e., $\psi \mapsto \psi' = \varphi \otimes \mathsf{U}c$ with $\mathsf{U} \in \mathrm{SU}(N_c)$. The gauge field $\underset{\sim}{A}$, named the *gluon field*, is an $\mathfrak{su}(N_c)$-valued 1-form, which can equivalently be viewed as a collection of $\dim \mathfrak{su}(N_c) = N_c^2 - 1$ independent 1-form fields, or eight distinct gluon types. QCD can be constructed with N_f quark types—or *flavours*—by taking the matter field to be a direct product of N_f quark fields, each sharing the same $\mathrm{SU}(N_c)$ gauge action.

The Lagrangian of pure QCD is a sum of the Dirac and Yang–Mills Lagrangians,

$$\underset{\sim}{\mathcal{L}}_{\mathrm{QCD}} = \bar{\psi}\big(i\gamma^\mu \nabla_\mu - \mathsf{m}\big)\psi \underset{\sim}{\mathrm{vol}} - \big\langle \underset{\sim}{F} \wedge \star \underset{\sim}{F}\big\rangle + \frac{\theta}{8\pi^2}\big\langle \underset{\sim}{F} \wedge \underset{\sim}{F}\big\rangle,$$

where $\psi \equiv \psi^{(1)} \oplus \cdots \oplus \psi^{(N_f)}$ is the matter field separated into its quark flavours, and $\mathsf{m} = m_1 \oplus \cdots \oplus m_{N_f}$ is a diagonal matrix of quark masses. With indices written explicitly, and $\hbar$ and c temporarily reinstated for completeness, the Lagrangian density may be spelled as

$$\mathcal{L}_{\mathrm{QCD}} = \sum_{q=1}^{N_f} \bar{\psi}_{a\mathfrak{c}}^{(q)}\big(i\hbar c\gamma^{\mu a}{}_b \nabla_\mu - m_q c^2 \delta^a{}_b\big)\psi_{(q)}^{b\mathfrak{c}} - \frac{\hbar}{4}F^{\mathfrak{a}}{}_{\mathfrak{b}\mu\nu}F_{\mathfrak{a}}{}^{\mathfrak{b}\mu\nu} + \frac{\theta\hbar}{8\pi^2}F^{\mathfrak{a}}{}_{\mathfrak{b}\mu\nu}F_{\mathfrak{a}}{}^{\mathfrak{b}}{}_{\rho\sigma}\epsilon^{\mu\nu\rho\sigma},$$

where Latin and Fraktur indices denote $\mathbb{C}_D^4$ fermion components and $\mathbb{C}_C^3$ colour components, respectively. «Is this all correct?»

Yukawa Couplings and the Measurable $\bar{\theta}$-parameter

The QCD sector of the standard model is an extension of pure QCD with three colours and six quarks. The quarks are partitioned into *up type* and *down type*, and again into three *generations*, each with varying masses and charges under the other components of the standard model gauge group.

generation	I	II	III
up type	u $2.2\,\mathrm{MeV}/c^2$	c $1.3\,\mathrm{GeV}/c^2$	t $170\,\mathrm{GeV}/c^2$
down type	d $4.7\,\mathrm{MeV}/c^2$	s $0.1\,\mathrm{GeV}/c^2$	b $4.2\,\mathrm{GeV}/c^2$

Pure QCD contains a mass term $\bar{\psi}\mathsf{m}\psi$, giving each quark $\psi^{(q)}$ an intrinsic mass m_q. This is *not* the mechanism by which quarks exhibit mass in the standard model—it cannot be, since is $\bar{\psi}\mathsf{m}\psi$ not invariant under axial rotations. «Why must axial rotations be a symmetry at all?»

Instead, quarks obtain mass via the *Higgs mechanism*, whereby ψ is coupled to the Higgs field h in *Yukawa interaction terms* in the Lagrangian [3, § 7.6.6].

$$\mathscr{L}_{\text{mass}} = \mathbb{R}\big(h\bar{\psi}_+ \mathsf{m}\,\psi_-\big) = \frac{1}{2}\big(h\bar{\psi}_+ \mathsf{m}\,\psi_- + h^\dagger\bar{\psi}_- \mathsf{m}^\dagger\psi_+\big)$$

The Lagrangian of the standard model QCD sector is thus

$$\mathscr{L}_{\text{QCD}}^{\text{SM}} = \big[\bar{\psi}i\boldsymbol{\gamma}^\mu\nabla_\mu\psi + h\,\mathbb{R}\big(\bar{\psi}_+ \mathsf{m}\,\psi_-\big)\big]\,\underset{\sim}{\text{vol}} - \langle \underset{\sim}{\boldsymbol{F}} \wedge \star\underset{\sim}{\boldsymbol{F}}\rangle + \frac{\theta}{8\pi^2}\,\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}}\rangle. \qquad (1.13)$$

Under independent axial rotations of each of the quark fields, the Yukawa mass term is invariant if the quark masses also shift phase. In addition, each independent $\mathrm{U}(1)_A$ rotation generates a corresponding θ-term, due to the chiral anomaly (§ 1.1.4). Thus, the full QCD Lagrangian (1.13) is invariant under transformations of the form

$$\psi_{(q)} \mapsto e^{i\boldsymbol{\gamma}^5 a_q/2}\psi_{(q)}, \qquad m_q \mapsto e^{-ia_q}m_q, \qquad \theta \mapsto \theta + \sum_{q=1}^{N_f} a_q, \qquad (1.14)$$

where α_q parametrise the N_f independent $\mathrm{U}(1)_A$ rotations. To aid physical interpretation, these $\mathrm{U}(1)_A$ freedoms are exploited in order to *normalise* the Yukawa mass terms by making the mass phases real. «This is only a guess.»

The fact that the θ-parameter may be redefined by axially rotating the quark fields means that θ is not directly observable. However, this gauge freedom can be fixed by defining

$$\bar{\theta} = \theta + \arg\det\mathsf{m} = \theta + \arg\prod_{q=1}^{N_f} m_q,$$

which is invariant under (1.14), as the two right-hand terms transform by $\pm\sum a_q$. The Lagrangian of the QCD sector of the standard model, complete with the $\bar{\theta}$-term, is thus

$$\mathscr{L}_{\text{QCD}}^{\text{SM}} = \big[\bar{\psi}i\boldsymbol{\gamma}^\mu\nabla_\mu\psi + h\,\mathbb{R}\big(\bar{\psi}_+ \mathsf{m}\,\psi_-\big)\big]\,\underset{\sim}{\text{vol}} - \langle \underset{\sim}{\boldsymbol{F}} \wedge \star\underset{\sim}{\boldsymbol{F}}\rangle + \frac{\bar{\theta}}{8\pi^2}\,\langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}}\rangle. \qquad (1.15)$$

The Statement of the Strong CP Problem

Proceeding with the assumption that $\bar{\theta} \neq 0$, one finds that the strong force now violates CP symmetry. A physical prediction of the standard model modified with a CP-violating QCD sector (1.15) is that the neutron is expected to possess an electric dipole moment of approximate magnitude $|d_n| \approx 10^{-18}\,e\,\text{cm}$. In reality, current measurements [8] of the neutron's electric dipole moment yield a tight upper bound of $|d_n| \lesssim 10^{-26}\,e\,\text{cm}$, which in turn implies a stringent constraint on the $\bar{\theta}$-parameter, $|\bar{\theta}| \lesssim 10^{-10}$ [6]. Thus, the $\bar{\theta}$-term is not considered to be part of the standard model.

However, $\bar{\theta}$ can only be zero if apparently unrelated parameters of the standard model perfectly cancel each other: the vacuum angle θ, a QCD parameter; and the quark mass phases $\arg\det\mathsf{m}$, deriving from multiple electroweak parameters. One therefore expects $\bar{\theta}$ to be $\mathcal{O}(1)$

in Nature, and its extremely small value is a problem of fine-tuning. The strong CP problem is the question, "why, then, is $\bar{\theta}$ so small?"

At first sight, the strong CP problem may not appear to be a problem at all. After all, QCD is a theory whose Lagrangian possibly—but not necessarily—admits a term $\propto \langle \boldsymbol{F} \wedge \tilde{\boldsymbol{F}} \rangle$ which gives rise to CP-violating interactions in the strong force, predicting an electric dipole moment of the neutron. Empirical data is consistent with the neutron's electric dipole moment (and hence the CP-violating term) being zero. From a phenomenological perspective, it is satisfactory to simply leave the $\bar{\theta}$-term out of the theory's Lagrangian and end the story there. Indeed, a tautological way to 'resolve' the strong CP problem is to simply require that CP be a symmetry of the strong force. However, this only begs the question of why CP symmetry appears to be preserved in some sectors of the standard model while it is broken in others.

Furthermore, there is a strong sense that the inclusion of the CP-violating term is "natural". That is, we lack reason to exclude it on a theoretical basis: it is Lorentz and gauge invariant, etc.; it is a theoretical prediction of the non-trivial vacuum structure of QCD (instantons); and arises via the chiral anomaly. From an empirical perspective, if no CP-violating interactions were observed in Nature, then $\bar{\theta}$ could be justifiably set to zero on the basis of symmetry. However, the weak interaction is explicitly parity-violating.[11] Hence, the fact that the $\bar{\theta}$-term violates CP is not theoretically satisfactory reason for its exclusion.

The strong CP problem differers from other fine-tuning problems in the standard model in the sense that it is of almost no consequence to everyday physics. Variation of the θ-parameter hardly affects nuclear physics at all because its effects are suppressed by the quark masses [9]. On the other hand, variations of the cosmological constant, for example, predict universes drastically different to our own, and similarly for the value of the weak scale, or the quark and lepton masses. Such fine-tuning problems at least have anthropic solutions—but the strong CP problem does not.[12] The strong CP problem is therefore a compelling theoretical indication that the standard model remains incomplete.

---

[11] In fact, the standard model is asymmetric under all combinations of charge conjugation, C; parity P; and time-reversal T modulo the prevailing combined CPT symmetry.

[12] Given a theory linking the presence of dark matter to the smallness of θ, an anthropic solution may exist if it turns out that dark matter is necessary for, e.g., galaxy formation (investigated in [9]).

2 Solutions to the Strong CP Problem

«Introduce and outline solutions»

2.1 The Massless Quark Solution

The simplest resolution to the strong CP problem is to stipulate that at least one quark was in fact massless. If this were true, then $\det \mathsf{m}$ would vanish, and the parameter $\bar{\theta} = \theta + \arg \det \mathsf{m}$ would be rendered unphysical. The *massless quark solution* is the claim that $\bar{\theta} \approx 0$ because the up quark is massless, $m_u = 0$.

At first sight, this economical resolution to the strong CP problem appears to be in contradiction with the experimentally determined masses of the quarks, all of which are non-zero (including the up quark, with mass $\sim 2.2\,\mathrm{MeV}$). However, it was realised in the mid-1980s that the mass of the up quark has *two* contributions in the standard model Lagrangian: not only the Yukawa mass m_u (the 'bare mass') as introduced above, but also a non-perturbative contribution m_{eff} from topological effects (i.e., instantons) [10]. Only the bare quark masses contribute to the value of $\bar{\theta}$ via the quark mass matrix m. Importantly, it was plausible that this secondary source of the up quark's mass could be of order $m_{\mathrm{eff}} \approx 2.2\,\mathrm{MeV}$, allowing m_u to vanish while still preserving the up quark's overall mass.

The massless up quark hypothesis remained controversial until *lattice gauge theory* had advanced sufficiently to made numerical simulations of non-perturbative effects in QCD possible. Consensus was reached by 2019 that the instanton contribution m_{eff} is not sufficiently large, and hence that the massless up quark hypothesis is false [10–12]. Instead, another mechanism is required to explain the smallness of the $\bar{\theta}$-parameter.

2.2 The Peccei–Quinn Mechanism

Perhaps the most famous resolution to the strong CP problem is the Peccei–Quinn theory of the *axion*, first proposed in 1977 [13]. In essence, the axion solution involves extending the standard model in order to promote the original $\bar{\theta}$-parameter to a field $\bar{\theta}(x) = \bar{\theta}_{\mathrm{SM}} + a(x)$ in such a way that it is dynamically relaxed to zero. In doing so, a new massive boson described by the scalar axion field $a(x)$ is necessarily introduced. There are different inequivalent ways to extend the standard model to realise the axion solution, but all Peccei–Quinn axion models share the same necessary features:

- The extended Lagrangian $\underset{\sim}{\mathcal{L}}_{+\mathrm{PQ}}^{\mathrm{SM}}$ possesses an additional global chiral symmetry $\mathrm{U}(1)_{\mathrm{PQ}}$.

The exact action of this *Peccei–Quinn symmetry* depends on the particular axion model, and is not of central importance. The defining feature of $\mathrm{U}(1)_{\mathrm{PQ}}$ is that it is chiral, so that a $\mathrm{U}(1)_{\mathrm{PQ}}$ transformation by α radians anomalously induces a θ-term $\alpha \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle$ in the effective Lagrangian (via the chiral anomaly).

- The Peccei–Quinn symmetry $\mathrm{U}(1)_{\mathrm{PQ}}$ is *spontaneously broken*, and the single resulting Nambu–Goldstone boson is named the axion field, $a(x)$. Being a Nambu–Goldstone boson, the axion transforms as $a(x) \mapsto a(x) + \alpha$ under a $\mathrm{U}(1)_{\mathrm{PQ}}$ rotation of α radians. The chiral anomaly results in a potential for the axion $a(x)$ (giving rise to axion mass) with a potential minimum occurring where $a(x) = -\bar{\theta}_{\mathrm{SM}}$.

If the extra $\mathrm{U}(1)_{\mathrm{PQ}}$ symmetry was indeed an exact gauge symmetry, then the strong CP problem is trivially solved, because a $\mathrm{U}(1)_{\mathrm{PQ}}$ rotation by $\bar{\theta}$ radians cancels the $\bar{\theta}$-term in the effective Lagrangian, meaning the dynamics of the theory are equivalent to one in which $\bar{\theta} = 0$. However, the main result of Peccei and Quinn [13] is that $\mathrm{U}(1)_{\mathrm{PQ}}$ need not be an exact symmetry of the theory: if $\mathrm{U}(1)_{\mathrm{PQ}}$ is spontaneously broken, then $\bar{\theta}$ is still driven to zero because it obtains a potential from the chiral anomaly.

A Toy Axion Model

To illustrate the Peccei–Quinn mechanism explicitly, consider a minimal toy axion model, which involves the addition of two fields to the standard model: a complex scalar field Φ known as the *parent field* (since it begets the axion field); and an additional fermion $\boldsymbol{q}$. The extended Lagrangian takes the form

$$\underset{\sim}{\mathcal{L}}_{\mathrm{PQ}} = \underset{\sim}{\mathcal{L}}_{\mathrm{SM}}^{\bar{\theta}} + \langle \underset{\sim}{\mathrm{d}\Phi}, \underset{\sim}{\mathrm{d}\Phi} \rangle + \mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-) \underset{\sim}{\mathrm{vol}} + \underset{\sim}{\mathcal{L}}_q, \tag{2.1}$$

where $\underset{\sim}{\mathcal{L}}_{\mathrm{SM}}^{\bar{\theta}}$ is the standard model Lagrangian (including the $\bar{\theta}$-term); $\langle \underset{\sim}{\mathrm{d}\Phi}, \underset{\sim}{\mathrm{d}\Phi} \rangle$ is a kinetic term for the parent field; $\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-)$ is a Yukawa coupling term; and $\underset{\sim}{\mathcal{L}}_q$ stands for any other terms involving the new fermion $\boldsymbol{q}$. The action of $\mathrm{U}(1)_{\mathrm{PQ}}$ on these fields is

$$\Phi \mapsto e^{i2\alpha}\Phi, \qquad\qquad \boldsymbol{q} \mapsto e^{i\gamma^5 \alpha}\boldsymbol{q} \quad \text{or} \quad \begin{cases} \bar{\boldsymbol{q}}_+ \mapsto e^{i\alpha}\bar{\boldsymbol{q}}_+, \\ \boldsymbol{q}_- \mapsto e^{i\alpha}\boldsymbol{q}_-, \end{cases} \tag{2.2}$$

which indeed leaves $\langle \underset{\sim}{\mathrm{d}\Phi}, \underset{\sim}{\mathrm{d}\Phi} \rangle$ and $\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-)$ invariant. However, since $\boldsymbol{q}$ undergoes an axial rotation, the entire (effective) Lagrangian $\underset{\sim}{\mathcal{L}}_{\mathrm{PQ}}$ is only invariant with the simultaneous subtraction of $\alpha \langle \boldsymbol{F} \wedge \boldsymbol{F} \rangle$, which arises via the chiral anomaly. Therefore, the entire action of $\mathrm{U}(1)_{\mathrm{PQ}}$ is to transform $\bar{\theta} \mapsto \bar{\theta} - \alpha$, as well as the fields as in (2.2). At this stage, the theory predicts CP violation in the strong sector in areas where the axions and instantons interact such that $\bar{\theta}(x) \neq 0$.

The final component of the axion solution is to make the parent field Φ spontaneously break. This may be done by adding a "Mexican hat" potential to $\underset{\sim}{\mathcal{L}}_{\mathrm{PQ}}$ of the form

$$V(|\Phi|) = \lambda(|\Phi|^2 - f_a^2)^2,$$

where the parameter f_a is interpreted as the *axion scale*, the energy below which axion dynamics are relevant.«???» Below this energy scale, the parent field Φ relaxes to some minimum of the form

$$\Phi = |\Phi|e^{i\arg\Phi} = f_a e^{ia/f_a},$$

where a varies across space. This field $a(x)$ is the Nambu–Goldstone boson associated with the spontaneous breaking of $\mathrm{U}(1)_{\mathrm{PQ}}$ by the parent field Φ, and is identified as the axion field. After spontaneous breaking, the Yukawa term involves a complex mass, which can be normalised by a $\mathrm{U}(1)_{\mathrm{PQ}}$ rotation by $-a/f_a$

$$\mathbb{R}(\bar{\boldsymbol{q}}_+ \Phi \boldsymbol{q}_-) = f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ e^{ia/f_a} \boldsymbol{q}_-) \mapsto f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ \boldsymbol{q}_-).$$

This axial rotation of $\boldsymbol{q}$ induces a corresponding rotation of $\theta \mapsto \theta - a/f_a$, so that Lagrangian (2.1) is effectively

$$\underset{\sim}{\mathcal{L}}_{\mathrm{PQ}} = \underset{\sim}{\mathcal{L}}_{\mathrm{SM}} + f_a^2 \langle \underset{\sim}{\mathrm{d}}a, \underset{\sim}{\mathrm{d}}a \rangle + f_a \mathbb{R}(\bar{\boldsymbol{q}}_+ \boldsymbol{q}_-) + \underset{\sim}{\mathcal{L}}_q + \left(\bar{\theta} - \frac{a}{f_a}\right) \frac{1}{8\pi^2} \langle \underset{\sim}{\boldsymbol{F}} \wedge \underset{\sim}{\boldsymbol{F}} \rangle, \qquad (2.3)$$

after spontaneous symmetry breaking of Φ.

Peccei and Quinn showed that the last term in (2.3) provides an effective potential $V(a)$ whose minimum occurs at $V(\bar{\theta} f_a) = 0$, giving the axion a vacuum expectation value $\langle a \rangle = \bar{\theta} f_a$ and a mass $m_a = \partial V/\partial a$.

3   Axions in Cosmology

Bibliography

[1] S. Weinberg, "The Making of the standard model", The European Physical Journal C-Particles and Fields **34**, 5–13 (2004).

[2] D. Tong, *Cambridge, Department of Applied Mathematics and Theoretical Physics, Lecture Notes: Gauge Theory*, URL: `http://www.damtp.cam.ac.uk/user/tong/gaugetheory.html`. Last visited on 2020/05/24, 2018.

[3] M. Hamilton, *Mathematical Gauge Theory: With Applications to the Standard Model of Particle Physics*, Universitext (Springer International Publishing, 2017).

[4] E. Mitsou, *Differential form description of the Noether-Lagrange machinery, vielbein/gauge-field analogies and energy-momentum complexes*, 2013.

[5] E. Witten, "Quantum field theory and the Jones polynomial", Communications in Mathematical Physics **121**, 351–399 (1989).

[6] P. D. Group et al., "Review of Particle Physics", Progress of Theoretical and Experimental Physics **2020**, 083C01 (2020) `10.1093/ptep/ptaa104`.

[7] B. Gripaios, *Lectures on Physics Beyond the Standard Model*, 2015.

[8] C. Abel et al., "Measurement of the Permanent Electric Dipole Moment of the Neutron", Phys. Rev. Lett. **124**, 081803 (2020).

[9] M. Dine, L. S. Haskins, L. Ubaldi, and D. Xu, "Some remarks on anthropic approaches to the strong CP problem", Journal of High Energy Physics **2018** (2018) `10.1007/jhep05(2018)171`.

[10] C. Alexandrou et al., *Ruling out the massless up-quark solution to the strong CP problem by computing the topological mass contribution with lattice QCD*, 2020.

[11] M. Dine, P. Draper, and G. Festuccia, "Instanton effects in three flavor QCD", Phys. Rev. D **92**, 054004 (2015).

[12] S. Aoki et al., *Review of lattice results concerning low-energy particle physics*, 2016.

[13] R. D. Peccei and H. R. Quinn, "CP Conservation in the Presence of Pseudoparticles", Phys. Rev. Lett. **38**, 1440–1443 (1977).