Axions and Topology in QCD

Rasmus Nielsen

Last edited: June 13, 2022

These are a complimentary set of notes for the *Lattice Journal Club* presentation, held on *June 14th 2022*. The notes largely cover the same material presented during the journal club meeting, but also discuss further details on certain topics. Additional information on derivations and related topics can be found in the references. Enjoy!

Contents

1 Topological configurations and the θ-vacuum

In perturbative *Quantum field theory*, we need to expand around some *classical solution* of the field equations. Typically, we further restrict our attention to static (*non time-dependent*) classical solutions, and recover the time-dependent solutions via Lorentz transforms. We will from here on out refer to the static classical solutions as *vacuum configurations*, or sometimes *vacuum sates*. Given the current context, this begs the question: *what are the vacuum configurations of QCD?*. In fact, let us start even simpler. What are the vacuum configurations of pure $SU(3)$ Yang Mills theory?

I order identify the vacuum configurations of pure $SU(3)$ YM, we first write down the familiar action in Minkowski space:

$$S[A] = -\frac{1}{2g^2}\int_{\mathbb{R}^4} d^4x \;\operatorname{tr} F_{\mu\nu}F^{\mu\nu} \quad , \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \tag{1}$$

Where we use the signature choice: $(-,+,+,+)$. If need be, we can also expand the $\mathfrak{su}(3)$ Lie-algebra valued gauge field A, in terms of a set of 8 generators T_a:

$$A_\mu = A_\mu^a T_a \quad , \quad \operatorname{tr} T_a T_b = \frac{1}{2}\delta_{ab} \tag{2}$$

The above action S is, in particular, invariant under all $SU(3)$ gauge transformations:

$$A_\mu \to U\left(A_\mu + i\partial_\mu\right)U^{-1} \quad , \quad U \in SU(3) \tag{3}$$

If we now insist on looking only at static gauge configurations, while also working in *static gauge* where $A_0 = 0$, we find that the pure $SU(3)$ YM action takes the following form:

$$S = \frac{1}{g^2}\int_{\mathbb{R}^3} d^3x \;\operatorname{tr} \boldsymbol{B}^2 \geq 0 \tag{4}$$

Where $B^i = \frac{1}{2}\varepsilon^{ijk}F_{jk}$, is the chromo-magnetic field. Note that the above expression is simply the energy $\mathcal{E}$ of pure YM in the absence of a chromo-electric field $E^i = F^{0i}$:

$$\mathcal{E} = \int_{\mathbb{R}^3} d^3x\, T^{00} = \frac{1}{g^2}\int_{\mathbb{R}^3} d^3x \;\operatorname{tr}\big(\boldsymbol{E}^2 + \boldsymbol{B}^2\big) \geq 0 \tag{5}$$

Where $T^{\mu\nu}$ is the *energy-momentum tensor*. It is clear that the gauge field configurations which minimize the energy, are those for which $F = 0$. We know that if A is given as a *pure gauge* configuration, the field strength F will vanish. It turns out that the converse is also true, by virtue of the *non-abelian Stoke's theorem*:

$$F_{\mu\nu} = 0 \quad \Leftrightarrow \quad A_\mu = iU\partial_\mu U^{-1} \tag{6}$$

Thus, the vacuum configurations of pure $SU(3)$ YM, takes on the following simple form:

$$A_i^{\text{vac}} = iU\partial_i U^{-1} \quad , \quad A_0^{\text{vac}} = 0 \tag{7}$$

Where, in order to stay in the static $A_0 = 0$ gauge, we must insist that the $SU(3)$ transformations U be *time-independent*.

1.1 Winding of non-abelian gauge fields

How can we classify all the possible vacuum configurations of pure $SU(3)$ Yang Mills? We start by noting that every vacuum configuration A^{vac}, is specified by some *time-independent* group element $U(\boldsymbol{x}) \in SU(3)$. Next, recall that we can generate any $SU(3)$ group element via the exponential map:

$$U(\boldsymbol{x}) = \exp\{i\omega(\boldsymbol{x})\} \tag{8}$$

Where $\omega(\boldsymbol{x})$ is some element of the *Lie algebra* $\mathfrak{su}(3)$. For our current purposes, it is convenient to distinguish between 2 important types of $U(\boldsymbol{x})$ transformations, depending on their behaviour at spatial infinity:

1. Gauge transformations for which: $U(\boldsymbol{x}) \to$ constant as $|\boldsymbol{x}| \to \infty$. These are the gauge transformations which relate **physically equivalent states**.

2. Gauge transformations for which: $U(\boldsymbol{x}) \not\to$ constant as $|\boldsymbol{x}| \to \infty$. These are to be thought of as **genuine symmetries of the system** and so **do not** transform between physically equivalent states.

From here on out, we will not concern ourselves with type 2. gauge transformations, as they do not play a critical part in the following discussion. To ease the following analysis slightly, notice that we can restrict our attention to gauge transformations for which $U(\infty) = 1$. Gauge transforms which approach a different constant group element V at infinity can then be obtained by multiplying $U(\boldsymbol{x})$ by V.

Because we now restrict our attention to transformation for which $U(\boldsymbol{x}) \to 1$, no matter the direction from which we approach spatial infinity, we make the following small but important observation: since all group transformations approach the same value at spatial infinity, we can treat spatial infinity as a *single point*. This changes the *topology* of space from the standard $\mathbb{R}^3$ topology to S^3; the topology of the 3-dimensional sphere. Thus, we can think of space dependent $SU(3)$ transformations as maps from S^3 into $SU(3)$, which maps the point at infinity to 1:

$$U : S^3 \longmapsto SU(3) \quad , \quad U(\infty) = 1 \tag{9}$$

The set of all such maps has a name in mathematics: *the 3rd homotopy group of* $SU(3)$, and is often denoted $\pi_3(SU(3))$. As the name indicates, this set has a group structure under composition of maps, and it can be shown that for the case of $SU(3)$, we have:

$$\pi_3(SU(3)) = \mathbb{Z} \tag{10}$$

The above result tells us that the gauge transformations: $U : S^3 \longmapsto SU(3)$, fall into distinct classes, which are labelled by an integer refereed to as the *winding number*. An explicit example of a gauge transformation with winding number n is given by the following $SU(2) \subset SU(3)$ transformation:

$$U(\boldsymbol{x}) = \exp\{i\phi(r)\,\hat{n}\cdot\boldsymbol{\sigma}\} = \cos\phi(r) + i\hat{n}\cdot\boldsymbol{\sigma}\,\sin\phi(r) \tag{11}$$

Here $r = |\boldsymbol{x}|$, $\hat{n}$ is the radial unit vector-field in $\mathbb{R}^3$, and $\boldsymbol{\sigma}$ are the Pauli-matrices. If the function ϕ satisfies $\phi(0) = \pi$ and $\phi(r) \to 2\pi n$ as $r \to \infty$, the winding number of the above gauge transformation will be exactly n.

It turns out that we explicitly compute the winding number n for any given map U, using the following integral expression:

$$n(U) = \frac{1}{24\pi^2}\int_{S^3} d^3x\, \varepsilon_{ijk}\,\mathrm{tr}\big(U\partial_i U^{-1}\big)(U\partial_j U^{-1})(U\partial_k U^{-1}) \in \mathbb{Z} \tag{12}$$

The fact that the above integral always yields an integer is somewhat non-trivial, and we will not make any attempts to proof this claim here. Additionally, it can be shown that the winding number is additive under group multiplication. That is:

$$n(U_1U_2) = n(U_1) + n(U_2) \tag{13}$$

This implies that we can transform a vacuum configuration with winding number n_1, to one with winding number $n_1 + n_2$, by use of an $SU(3)$ element U_2 with winding number n_2:

$$iU_1\partial_i U_1^{-1} \to U_2 \left(iU_1\partial_i U_1^{-1} + i\partial_i\right) U_2^{-1} = i(U_2U_1)\partial_i(U_2U_1)^{-1} \tag{14}$$

The fact that these $SU(3)$ transformations, and by extension the vacua A^{vac}, can be classified by an integer n, implies that these elements are not continuously connected. Thus, in order to go from one vacuum to another, one necessarily has to go through non-vacuum configurations. In other words, vacua with different winding numbers are separated by a potential barrier. Classically, this means that the system cannot evolve from one distinct vacuum state into another. Quantum mechanically however, it is possible for *tunnelling* to occur, which we will now discuss.

1.2 Instantons and tunnelling amplitudes

We want to understand the possibility of tunnelling between vacua with different winding numbers in pure $SU(3)$ YM. To be more precise, we want to compute the probability amplitude for transitioning between a vacuum state $|n\rangle_-$ with winding number n at $t = -\infty$, to another vacuum state $|m\rangle_+$ with winding number m at $t = +\infty$. This amplitude can be expressed as a *path integral* in the following way:

$$_+\langle m|n\rangle_- = \int_{A_{(n)}}^{A_{(m)}} \mathcal{D}A \ \exp\{iS[A]\} \tag{15}$$

In order to better understand the dominant contribution to the above amplitude, we perform a *Wick rotation*: $t \to -it$ and $A_0 \to iA_0$. One can think of this as a type of coordinate transformation, and so should not change the value of the path integral (*up to a phase*). After performing the Wick rotation, we end up with the following expression:

$$_+\langle m|n\rangle_- = \int_{A_{(n)}}^{A_{(m)}} \mathcal{D}A \ \exp\{-S_E[A]\} \tag{16}$$

The quantity $S_E[A]$ is known as the *Euclidean action*, and is simply given by the following:

$$S_E[A] = \frac{1}{2g^2}\int_{\mathbb{R}^4} d^4x \ \mathrm{tr}\, F_{\mu\nu}F_{\mu\nu} \quad , \quad F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu] \tag{17}$$

Where we now work with Euclidean signature: $(+,+,+,+)$, and keeping track of upper and lower indices is therefore irrelevant.

In the path integral formalism, the dominant field configurations are those which *minimize* the Euclidean action subject to the boundary conditions; e.i. the *classical solutions*. In Minkowski space we already know that no classical solutions exist, which connect vacua with different winding numbers. There are two reasons for this:

1. Vacua with differing values of winding number n are not continuously connected.

2. The vacua are all field configurations which **minimize** the energy $\mathcal{E}$.

Thus, in order to go between vacua with different winding numbers, one must go through configurations with higher energy, and so energy conservation would be violated! In Euclidean space however, the energy function $\mathcal{E}_E$ picks up a minus sign after Wick rotation:

$$\mathcal{E}_E = -\frac{1}{g^2}\int_{\mathbb{R}^3} d^3x \ \operatorname{tr}\big(\boldsymbol{E}^2 + \boldsymbol{B}^2\big) \leq 0 \tag{18}$$

This means that configurations which were formerly minima of the energy in Minkowski space, now become **maxima** of the energy! It is no longer a problem finding solutions which connect distinct vacua, and simultaneously conserve energy. We shall now investigate which of these configurations also *minimize* the Euclidean action $S_E[A]$. At this point, it would be an appropriate time to interject that classical solutions which connect different vacua in Euclidean space, when no such solutions exist in Minkowski space, are often referred to as *instanton solutions*, or simply *instantons*.

The standard approach to finding minimal solutions w.r.t any action is of course to employ the famous *Euler-Lagrange equations*, which in our case yields:

$$D_\nu F_{\mu\nu} = \partial_\mu F_{\mu\nu} - i[A_\nu, F_{\mu\nu}] = 0 \tag{19}$$

The above equations are horribly complicated *second order non-linear* equations in A, and attempting to find a solution would most certainly be a rather painful process. Fortunately, there is in our case a better approach. The trick is to rewrite the Euclidean action $S_E[A]$ as follow:

$$S_E[A] = \frac{1}{4g^2}\int_{\mathbb{R}^4} d^4x \ \operatorname{tr}\left(F_{\mu\nu} \mp \tilde{F}_{\mu\nu}\right)^2 \pm \frac{1}{2g^2}\int_{\mathbb{R}^4} d^4x \ \operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu} \tag{20}$$

Where the quantity $\tilde{F}_{\mu\nu}$ is know as the *dual field strength*. It is given in terms of $F_{\mu\nu}$ as so:

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\rho\sigma}F_{\rho\sigma} \tag{21}$$

The first term in the above rewriting of $S_E[A]$ is clearly bound from below by zero, and thus we find that the Euclidean action is bound from below by the second term:

$$S_E[A] \geq \pm\frac{1}{2g^2}\int_{\mathbb{R}^4} d^4x \ \operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu} \tag{22}$$

By straight forward, albeit somewhat tedious, algebraic manipulation, one finds that the above integrand is in fact a total derivative:

$$\int_{\mathbb{R}^4} d^4x \ \operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu} = \int_{\mathbb{R}^4} d^4x \ \partial_\mu K_\mu \quad , \quad K_\mu = 2\varepsilon_{\mu\nu\rho\sigma}\operatorname{tr}\left(A_\nu\partial_\rho A_\sigma - \frac{2i}{3}A_\nu A_\rho A_\sigma\right) \tag{23}$$

In order for the Euclidean action to be finite, we must have that $F_{\mu\nu} \to 0$ as $|x| \to \infty$. As we have already argued in the previous section, this means that the gauge field must approach a *pure gauge* configuration:

$$F_{\mu\nu} = 0 \quad \Leftrightarrow \quad A_\mu = iU\partial_\mu U^{-1} \tag{24}$$

We can now make use of the above condition to simplify the form of K_μ at infinity. First, we note the fact that $F_{\mu\nu} = 0$ at infinity gives us the following relation:

$$F_{\rho\sigma} = \partial_\rho A_\sigma - \partial_\sigma A_\rho - i[A_\rho, A_\sigma] = 0 \quad \Rightarrow \quad \varepsilon_{\mu\nu\rho\sigma}\partial_\rho A_\sigma = i\varepsilon_{\mu\nu\rho\sigma}A_\rho A_\sigma \tag{25}$$

The above relation allows us to simplify the expression for K_μ in the limit as $|x| \to \infty$:

$$K_\mu \to \frac{2i}{3}\varepsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\,A_\nu A_\rho A_\sigma \quad \text{as} \quad |x| \to \infty \tag{26}$$

We now make use of the fact that A_μ must approach a pure gauge configuration as we approach infinity, and find that K_μ must then approach:

$$K_\mu \to \frac{2}{3}\varepsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big(U\partial_\nu U^{-1}\big)(U\partial_\rho U^{-1})(U\partial_\sigma U^{-1}) \quad \text{as} \quad |x| \to \infty \tag{27}$$

Given the asymptotic form of K_μ above, we can now rewrite the lower bound integral as a surface integral over the 3-sphere S^3_∞ at infinity:

$$\int_{\mathbb{R}^4} d^4x \;\mathrm{tr}\,F_{\mu\nu}\tilde{F}_{\mu\nu} = 16\pi^2\left[\frac{1}{24\pi^2}\int_{S^3_\infty} d^3x\;\hat{n}_\mu\varepsilon_{\mu\nu\rho\sigma}\,\mathrm{tr}\big(U\partial_\nu U^{-1}\big)(U\partial_\rho U^{-1})(U\partial_\sigma U^{-1})\right] \tag{28}$$

If we choose to work in radial coordinates: $x = (r, \theta, \phi, \psi)$, the normal vector will be of the form: $\hat{n} = (1, 0, 0, 0)$. Thus, we can finally write the lower bound on the Euclidean action $S_E[A]$, as follow:

$$S_E[A] \geq \pm\frac{8\pi^2}{g^2}\nu(U) \tag{29}$$

Where the quantity $\nu(U)$ is given by the following integral over the 3-sphere S^3_∞ at infinity:

$$\nu(U) = \frac{1}{24\pi^2}\int_{S^3_\infty} d^3x\;\varepsilon_{ijk}\,\mathrm{tr}\big(U\partial_i U^{-1}\big)(U\partial_j U^{-1})(U\partial_k U^{-1}) \in \mathbb{Z} \tag{30}$$

The quantity $\nu(U)$ is often refereed to as the *instanton number* for a given field configuration. The astute reader will probably have noticed, that the expression for the instanton number $\nu(U)$ above look suspiciously similar to the expression for the winding number $n(U)$ presented earlier. Formally, the expressions are completely identical, and so it should be no surprise that also $\nu(U)$ is integer-valued. The only difference between the two quantities is that the integrations are carried out over different, but related, 3-spheres. We will come back to the connection between $\nu(U)$ and $n(U)$ shortly, when we discuss how to enforce the boundary conditions on the path integral.

So far, we have managed to show that the Euclidean action for pure $SU(3)$ Yang Mills theory, can be rewritten in the following way:

$$S_E[A] = \frac{1}{4g^2}\int_{\mathbb{R}^4} d^4x \ \mathrm{tr}\left(F_{\mu\nu} \mp \tilde{F}_{\mu\nu}\right)^2 \pm \frac{8\pi^2}{g^2}\nu(U) \tag{31}$$

Since it is not possible to continuously deform between gauge configuration A with differing instanton numbers ν, we conclude that the minima of the Euclidean action are the field configurations which satisfy:

$$\tilde{F}_{\mu\nu} = \pm F_{\mu\nu} \tag{32}$$

The above equations are *first order non-linear* in A; a considerable improvement from the second order non-linear equation one gets from the standard Euler-Lagrange approach. A field strength which satisfy $\tilde{F}_{\mu\nu} = +F_{\mu\nu}$ is said to be *self-dual*, and one which satisfy $\tilde{F}_{\mu\nu} = -F_{\mu\nu}$ said to be *anti self-dual*. A simple example of a $\nu = 1$ instanton solution can be written as follow:

$$A_\mu(x) = \frac{x^2}{x^2+\rho^2} iU\partial_\mu U^{-1} \quad , \quad U(x) = \frac{x^\mu \sigma_\mu}{\sqrt{x^2}} \tag{33}$$

Where $\sigma_\mu = (1, -i\vec{\sigma})$. A similar instanton solution for the case of $\nu = -1$ can also readily be written down. One simply needs to replace σ_μ with $\bar{\sigma}_\mu = (1, i\vec{\sigma})$. Solutions with other values of ν can be found via the *Atiyah Drinfeld Hitchin Manin* (*ADHM*) construction. We shall not go into the details on this construction here. Instead, we now turn to investigate which, if any, of the instanton solutions respect the boundary conditions of the probability amplitude path integral; A must have winding number n at $t = -\infty$ and winding number m at $t = +\infty$.

1.2.1 Back to the boundary conditions

In order to determine which instanton solutions satisfy the right boundary conditions, it is crucial to understand the connection between the winding number $n(U)$ and the instanton number $\nu(U)$. To investigate this connection, we return for a moment back to Minkowski space. Also in this setting, $\operatorname{tr} F_{\mu\nu}\tilde{F}^{\mu\nu}$ can be written as a total derivative, and we can express its integral over all of $\mathbb{R}^4$ as a boundary integral:

$$\int_{\mathbb{R}^4} d^4x \ \operatorname{tr} F_{\mu\nu}\tilde{F}^{\mu\nu} = \int_{\mathbb{R}^4} d^4x \ \partial_\mu K^\mu = \int_{\mathbb{R}^3_+} d^3x \ K^0 - \int_{\mathbb{R}^3_-} d^3x \ K^0 \tag{34}$$

Where in the above, we have taken advantage of the fact that K^μ runs orthogonal to the time-like boundary at infinity: This implies that no current '*leaks out*' of the system. It is easy to prove this statement by imposing the gauge choice: $A_0 = 0$.

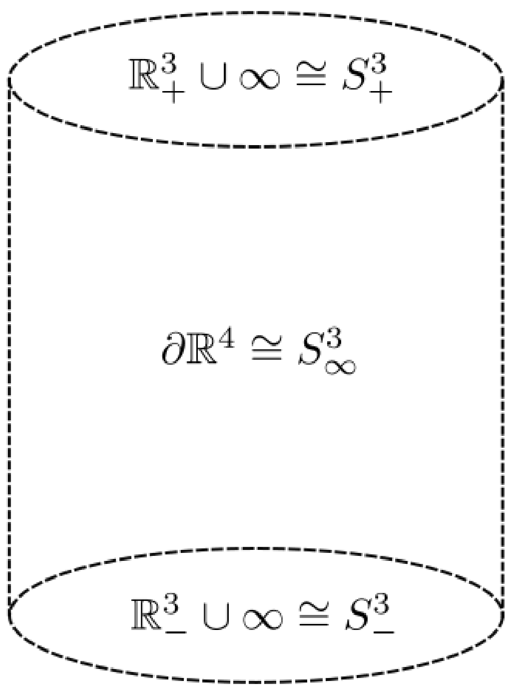

Figure 1: Diagram of $\mathbb{R}^4$. One can either think of the boundary as one big 3-sphere S^3_∞, or as two distinct spatial boundaries, $\mathbb{R}^3_+$ and $\mathbb{R}^3_-$, together with the curved time-like boundary. The spatial boundaries can subsequently be compactified to separate 3-spheres, S^3_+ and S^3_-.

If we now assume that the system starts out in a vacuum configuration with winding number n on the spatial boundary at $t = -\infty$, and ends up in another vacuum configuration with winding number m on the spatial boundary at $t = +\infty$, the integrals over K^0 exactly yield these winding numbers:

$$\frac{1}{16\pi^2}\int_{\mathbb{R}^3_-} d^3x\; K^0 = \frac{1}{24\pi^2}\int_{S^3_-} d^3x\; \varepsilon_{ijk}\,\mathrm{tr}\big(U\partial_i U^{-1}\big)\big(U\partial_j U^{-1}\big)\big(U\partial_k U^{-1}\big) = n(U) \tag{35}$$

$$\frac{1}{16\pi^2}\int_{\mathbb{R}^3_+} d^3x\; K^0 = \frac{1}{24\pi^2}\int_{S^3_+} d^3x\; \varepsilon_{ijk}\,\mathrm{tr}\big(U\partial_i U^{-1}\big)\big(U\partial_j U^{-1}\big)\big(U\partial_k U^{-1}\big) = m(U) \tag{36}$$

Thus, we find that the integral of $\mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}$ over all $\mathbb{R}^4$ can be related to the difference of winding numbers between the two spatial boundaries:

$$\int_{\mathbb{R}^4} d^4x\;\; \mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} = 16\pi^2\,[m(U) - n(U)] \tag{37}$$

Now recall that we derived a very similar looking relation for the instanton number $\nu(U)$ earlier in this section, by way of similar arguments:

$$\int_{\mathbb{R}^4} d^4x\;\; \mathrm{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} = \int_{\mathbb{R}^4} d^4x\; \partial_\mu K^\mu = \int_{S^3_\infty} d^3x\; K^0 = 16\pi^2\,\nu(U) \tag{38}$$

Thus, we find the following simple relation between the winding numbers $n(U)$ and $m(U)$, and the instanton number $\nu(U)$:

$$\nu(U) = m(U) - n(U) \tag{39}$$

The above relation tells us, that if we want to consider the amplitude of staring in a configuration with winding number n, and ending in one with winding number m, we must allow only configurations with $\nu = m - n$ in the path integral. The dominant contributions will then come from instanton solutions which have exactly this instanton number.

To summarize, we have found that the leading contribution to the tunnelling amplitude ${}_+\langle m|n\rangle_-$, is given in terms of instanton solutions with $\nu = m - n$:

$${}_+\langle m|n\rangle_- = \int_{A_{(n)}}^{A_{(m)}} \mathcal{D}\delta A\;\; \exp\big\{-S_E[A_{(\mathrm{inst})} + \delta A]\big\} \sim \sum_{\mathrm{inst}} \exp\bigg\{-\frac{8\pi^2}{g^2}\nu\bigg\} \tag{40}$$

Where δA denote fluctuations around the instanton solution $A^{(\text{inst})}$, and the quantity in the exponential furthest to the right is simply $S_E[A^{(\text{inst})}]$. Furthermore, the sum runs over odd sequences of ν instantons and $-\nu$ anti-instantons. The possibility for distinct vacua to tunnel between one another has some very interesting implications for the quantum theory of $SU(3)$ YM, as we shall now discuss.

1.3 The emergence of the θ-term

The fact that tunnelling can occur between vacua with different winding numbers implies that these states are not *stationary*; e.i they evolve in time. Thus, they can not by *energy eigenstate* of the quantum system. In fact, the *true vacuum state*, to a good approximation, can be taken to be a linear super-position of the $|n\rangle$ vacua. But which super-position? Well, we know that the vacuum states $|n\rangle$ are not invariant under gauge transformations, since any U with non-zero winding number maps between these states. Thus, the correct super-position must be an eigenstate of the unitary operators $T^{n\to m}$, which takes $|n\rangle \to |m\rangle$. It is straight forward to verify that states of the following form satisfy this condition:

$$|\theta\rangle = \sum_{n=-\infty}^{\infty} e^{-in\theta}\,|n\rangle \tag{41}$$

These states are parametrized by a continuous variable θ, which we can take to lie on the interval $[0, 2\pi]$. Note that since $|\theta\rangle$ is an eigenstate of $T^{n\to m}$ with eigenvalue $e^{-i\theta}$, all physical states in the *Hilbert space* must also have this eigenvalue under gauge transformations. Otherwise, a general super-position would pick up relative phases between elements.

Now that we have a good approximation to the true $SU(3)$ YM vacuum, let us attempt to compute one of the most central objects in any field theory; the vacuum to vacuum transition amplitude, also known as the *partition function*. To be more precise, we want to compute the transition amplitude between the asymptotic state $|\theta\rangle_-$ at $t = -\infty$, and the asymptotic state $|\theta\rangle_+$ at $t = +\infty$:

$${}_+\langle\theta|\theta\rangle_- = \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} e^{im\theta}e^{-in\theta}\;{}_+\langle m|n\rangle_- = \sum_{n=-\infty}^{\infty}\sum_{\nu=-\infty}^{\infty} e^{i\nu\theta}\;{}_+\langle n+\nu|n\rangle_- \tag{42}$$

Note that ${}_+\langle n+\nu|n\rangle_-$ can be expressed as a path integral over configurations with instanton number ν, as explained earlier. Furthermore, when restricting to these configurations, we can write ν as an integral of $\operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu}$ over all $\mathbb{R}^4$. Thus, we find that:

$$e^{i\nu\theta}\,{}_+\langle n+\nu|n\rangle_- = \int_{A_{(n)}}^{A_{(n+\nu)}} \mathcal{D}A\ \exp\left\{-S_E[A] + \frac{i\theta}{16\pi^2}\int_{\mathbb{R}^4} d^4x \operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu}\right\} \tag{43}$$

Summing the above result over all winding numbers n and all instanton numbers ν, we find that the proper partition function: $Z[\theta] \equiv {}_+\langle\theta|\theta\rangle_-$, is given by the following expression:

$$Z[\theta] = \int \mathcal{D}A\ \exp\{-S_E[A] - S_\theta[A]\} \tag{44}$$

Where the boundary conditions are now: **any** vacuum to **any** vacuum. We see that the partition function look almost exactly like we would have expected if the theory only contained the trivial vacuum $A = 0$. The only difference is the emergence of the θ-term:

$$S_\theta[A] = -\frac{i\theta}{16\pi^2}\int_{\mathbb{R}^4} d^4x \operatorname{tr} F_{\mu\nu}\tilde{F}_{\mu\nu} \tag{45}$$

Without the knowledge of all the arguments we have now gone through, one might have though that inclusion of a θ-term in the YM action would be allowed, but not necessary. We now know this not to be true, and it is in fact necessary to include in order for the partition function to be properly gauge invariant.

2 The strong CP-problem and Axions

There are two sources of CP violation in the Standard Model. The first source comes entirely from the *electro-weak sector*, and can be described by the so-called *Cabibbo Kobayashi Maskawa* (*CKM*) matrix, which appears in the interaction terms between quarks and the weak $W_\pm$ bosons. We shall not discuss this source of CP violation any further here.

The second source of CP violation comes from an interesting interplay between the *strong* and *electro-weak* sectors. This source of CP violation was a bit of a mystery for a long time, since it was unmistakeably predicted by the SM, but did not appear to show up in experiments! This discrepancy came to be known as the *strong CP problem*. Before we further discuss this apparent paradox, and its eventual resolution, we need to first discuss the most crucial aspect of the underlying theory; namely *the chiral anomaly.*

2.1 The chiral anomaly

It is an interesting fact of nature, that not all symmetries which holds classically survives the transition to quantum mechanics. Symmetries which do not survive the transition are said to be *anomalous.* Consider as an example the following fermionic action, which couple a set of quarks q, $\bar{q}$ to the $SU(3)$ gauge field A:

$$S_{\text{kinetic}}[A, q, \bar{q}] = \int_{\mathbb{R}^4} d^4x \ \left[\bar{q}_L i\not{D} q_L + \bar{q}_R i\not{D} q_R\right] \quad , \quad D_\mu = \partial_\mu - iA_\mu \tag{46}$$

Where $\not{D} = \gamma^\mu D_\mu$, and the gamma matrices satisfy the Clifford algebra: $\{\gamma^\mu, \gamma^\nu\} = -2\eta^{\mu\nu}$, in signature $(-, +, +, +)$. The above action turns out to be invariant under *chiral rotations* of the spinor components q_L, q_R:

$$q_L \to e^{i\epsilon/2}\, q_L \quad , \quad q_R \to e^{-i\epsilon/2}\, q_R \tag{47}$$

Using the above transformation properties for the chiral components of q, it can easily be checked that the action is invariant under chiral rotations. This is all well and good. Classically our theory is invariant, but quantum mechanically we are not quite home free yet. This is most intuitively seen in the path integral formalism, where the spinor fields also appear in the *measure* of the partition function:

$$Z = \int \mathcal{D}q\mathcal{D}\bar{q} \ \exp\{iS_{\text{kinetic}}[A, q, \bar{q}]\} \tag{48}$$

This means that we also need to check whether or not the measure stays invariant under chiral rotation. It turns out that it **does not!** In fact, after some fairly heavy analysis, one can show that the measure transform in the following way under chiral rotation:

$$\mathcal{D}q\mathcal{D}\bar{q} \longrightarrow \mathcal{D}q\mathcal{D}\bar{q} \ \exp\left\{\frac{i\epsilon}{16\pi^2}\int_{\mathbb{R}^4} d^4x \ \operatorname{tr} F_{\mu\nu}\tilde{F}^{\mu\nu}\right\} \tag{49}$$

Thus, the action effectively picks up a θ-like term under chiral rotations. This implies that we can actually transform the θ-term away via a chiral rotation! So much for all that work we did in the last section... Fortunately, we have no good reason to believe that any of the quarks in nature are completely massless. Given that this is the case, we shall now investigate what changes when we add a quark mass-term to the action above.

2.2 CP-violation in theory

After spontaneous breaking of the electro-weak $SU(2)\times U(1)_Y$ gauge symmetry, the coupling between the Higgs doublet H and quarks q, generates mass terms of the following form:

$$S_{\text{mass}}[A,q,\bar{q}] = \int_{\mathbb{R}^4} d^4x\ \left[-\frac{v}{\sqrt{2}}y_q\bar{q}_R q_L - \frac{v}{\sqrt{2}}y_q^*\bar{q}_L q_R\right] \tag{50}$$

Where v is related to the expectation value of H. The Yukawa-coupling parameter y_q will in general be a complex number, but since the two mass terms in the action are complex conjugates of each other, the action remains real-valued. However, if y_q is complex-valued, the action will **violate CP symmetry!** To be precise, for a complex-valued y_q, the action will be invariant under C but not P:

$$\text{C}: \quad y_q\bar{q}_R q_L + y_q^*\bar{q}_L q_R \to y_q\bar{q}_R q_L + y_q^*\bar{q}_L q_R \tag{51}$$

$$\text{P}: \quad y_q\bar{q}_R q_L + y_q^*\bar{q}_L q_R \to y_q\bar{q}_L q_R + y_q^*\bar{q}_R q_L \tag{52}$$

It turns out that the θ-term also violates CP symmetry. As was the case for the quark mass terms, the θ-term is also invariant under C but not P:

$$\text{C}: \quad \text{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} \to \text{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{53}$$

$$\text{P}: \quad \text{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} \to -\,\text{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{54}$$

On first sight, it appears that we have two independently CP violating part of the total action. However, this is not quite the case. By performing a chiral rotation with parameter $\epsilon = -\arg(y_q)$, we can '*move*' all the CP violation to the θ-term.

$$S_{\text{mass}}[A,q,\bar{q}] + S_\theta[A] \to \int_{\mathbb{R}^4} d^4x\ \left[-\frac{v}{\sqrt{2}}|y_q|\bar{q}_R q_L - \frac{v}{\sqrt{2}}|y_q|\bar{q}_L q_R + \frac{\bar{\theta}}{16\pi^2}\,\text{tr}\, F_{\mu\nu}\tilde{F}^{\mu\nu}\right] \tag{55}$$

Where $\bar{\theta} = \theta - \arg(y_q)$. In the electro-weak sector of the Standard Model, the interaction terms between the $SU(2)$ Higss doublet and the quarks also generate quadratic couplings between different flavours of quarks. All the quadratic quark terms are collectively described by the two Yukawa-matrices: Y_d and Y_u. Also in this more complicated setup, it is possible to obtain mass terms with real-valued Yukawa-mass parameters, this time at the expense of introducing the following modified θ-angle:

$$\bar{\theta} = \theta - \arg\det(Y_d Y_u) \tag{56}$$

It turns out that the existence of $\bar{\theta}$ has physically measurable implications. In particular, a non-zero value of $\bar{\theta}$ induces an electric dipole moment d_N for the Neutron. However, very precise experiments have been able to obtain a strict bound on the value of d_N, and in turn on the value of $\bar{\theta}$. The experimental results look as follow:

$$|d_N| < 2.9 \times 10^{-26} e \cdot \mathrm{cm} \quad \Rightarrow \quad |\bar{\theta}| < 10^{-10} \tag{57}$$

This absurdly precise, but nevertheless required, fine-tuning of the $\bar{\theta}$ angle is known as the *strong CP problem*. There is currently no known solution to this problem within the framework of the Standard Model. As far as we can tell, an extension of the SM is required to find a resolution to this apparent paradox. We shall now describe one such extension, curtsey of *Roberto Peccei* and *Helen Quinn*.

2.3 The Peccei-Quinn mechanism

The original idea of Peccei and Quinn was to postulate a new symmetry: $U(1)_{PQ}$, with an accompanying complex scalar field φ. This scalar field then couples to the d-type quarks (*down, strange bottom*), replacing the Higgs coupling terms for these quarks. Unfortunately, it turns out that this original model has been ruled out by experiments. Nevertheless, the general idea lives on, and several similar models now exist which are still compatible with observations. We shall now discuss the, arguable, most conceptually simple of these models; the *Kim Shifman Vainshtein Zakharov* (*KSVZ*) model.

2.3.1 The KSVZ model

Like the original model by Peccei and Quinn, the KSVZ model introduces the $U(1)_{PQ}$ symmetry, alongside the complex scalar field φ. Additionally, this model also introduces a new chiral quark pair Q_L, Q_R, which transforms trivially under $SU(2) \times U(1)_Y$. Moreover, the left-chiral component Q_L have charge 1 under $U(1)_{PQ}$, the right chiral-component have charge 0, and the complex scalar φ has charge -1:

$$\text{PQ}: \quad Q_L \to e^{-i\epsilon} Q_L \quad , \quad Q_R \to Q_R \quad , \quad \varphi \to e^{+i\epsilon} \varphi \tag{58}$$

Given the above transformation properties, we see that it is possible to construct Yukawa-type coupling terms between Q_L, Q_R and φ:

$$S_{\text{Yukawa}}[\varphi, Q, \bar{Q}] = \int_{\mathbb{R}^4} d^4x \; \left[-y_Q \, \varphi \, \bar{Q}_R Q_L - y_Q \varphi^* \bar{Q}_L Q_R \right] \tag{59}$$

In this case, we can take the Yukawa-coupling parameter y_Q to be real-valued, since we can always remove any non-zero phase by absorbing it into φ and φ^*. Apart from the Yukawa terms, it is also possible to add a standard kinetic and potential term for the scalar field φ:

$$S_{\text{scalar}}[\varphi] = \int_{\mathbb{R}^4} d^4x \; \left[-|\partial_\mu \varphi|^2 - V(|\varphi|) \right] \tag{60}$$

In order to be consistent with observation, the new quarks Q must be heavily massive, and so Yukawa-mass terms need to be generated via *spontaneous symmetry breaking* of the $U(1)_{PQ}$ symmetry. This is done by assuming a standard *Mexican hat potential* for φ:

$$V(\varphi) = \lambda \left(\frac{f_a^2}{2} - |\varphi|^2 \right)^2 \tag{61}$$

Where λ controls the strength of the potential, and f_a controls the energy scale of symmetry breaking. The above potential is clearly minimized when $|\varphi|$ takes the following form:

$$|\varphi_{\text{min}}| = \frac{f_a}{\sqrt{2}} \quad \Rightarrow \quad \varphi_{\text{min}} = \frac{f_a}{\sqrt{2}} e^{-ia/f_a} \tag{62}$$

Where the real-valued scalar field a is known as the **Axion**. In perturbation theory, we seek to understand fluctuations around the minima of the fields in our theory. Thus, with the above form of φ_{min} in mind, it is convenient to decompose φ in the following way:

$$\varphi = \frac{1}{\sqrt{2}} \left(f_a + \rho_a \right) e^{-ia/f_a} \tag{63}$$

Where ρ_a are the perturbations in the radial direction around the circle of minima. Given this decomposition, the Yukawa-coupling action S_{Yukawa}, can be rewritten as follow:

$$S_{\text{Yukawa}}[\rho_a, a, Q, \bar{Q}] = \int_{\mathbb{R}^4} d^4x \; \left[-\frac{f_a}{\sqrt{2}} y_Q \, e^{-ia/f_a} \, \bar{Q}_R Q_L - \frac{f_a}{\sqrt{2}} y_Q \, e^{ia/f_a} \, \bar{Q}_L Q_R + \dots \right] \tag{64}$$

We now see that this procedure has effectively generated a complex-valued Yukawa-mass parameter, with argument a/f_a. Just as we did for the Yukawa-mass parameter y_q, we can now remove this argument a/f_a at the cost of reintroducing it as an effective θ-term:

$$S_\theta[A,a] \to \int_{\mathbb{R}^4} d^4x \, \frac{1}{16\pi^2} \left(\bar{\theta} + \frac{a}{f_a}\right) \operatorname{tr} F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{65}$$

By redefining the Axion field by a constant shift: $a \to a - \bar{\theta} f_a$, we obtain the following simple coupling between the Axion field a and the gauge field A:

$$S_\theta[A,a] \to \int_{\mathbb{R}^4} d^4x \, \frac{1}{16\pi^2} \frac{a}{f_a} \operatorname{tr} F_{\mu\nu}\tilde{F}^{\mu\nu} \tag{66}$$

It is exactly this coupling which solves the *strong CP problem*! To see how, it is most convenient to Wick rotate to Euclidean space, where the θ-term naturally splits from the rest of the action in the partition function:

$$Z[a] = \int \mathcal{D}\Phi \exp\{-S_E[\Phi] - S_\theta[A,a]\} = \int \mathcal{D}\Phi \exp\left\{-S_E[\Phi] + i\frac{a}{f_a}\nu[A]\right\} \tag{67}$$

Where $\Phi = (A, \rho_a, Q, \bar{Q})$. We have also made use of the fact that $S_\theta[A,a]$ can be written in terms of the gauge field dependent winding number $\nu[A]$. Note that all factors of the integrand, apart from $\exp\left\{i\frac{a}{f_a}\nu[A]\right\}$, are real-valued:

$$|\mathcal{D}\Phi \exp\{-S_E[\Phi]\}| = \mathcal{D}\Phi \exp\{-S_E[\Phi]\} \tag{68}$$

Using this observation, it is now straight forward to prove the following property of $Z[a]$:

$$\begin{aligned} Z[a] &= \int \mathcal{D}\Phi \exp\left\{-S_E[\Phi] + i\frac{a}{f_a}\nu[A]\right\} \leq \int \left|\mathcal{D}\Phi \exp\left\{-S_E[\Phi] + i\frac{a}{f_a}\nu[A]\right\}\right| \\ &= \int |\mathcal{D}\Phi \exp\{-S_E[\Phi]\}| = \int \mathcal{D}\Phi \exp\{-S_E[\Phi]\} = Z[0] \\ &\Rightarrow \quad Z[a] \leq Z[0] \end{aligned} \tag{69}$$

Lastly, we make use of the connection between the Axion dependent partition function $Z[a]$, and the *effective potential* of the Axion $V_{\text{eff}}(a)$, in order to arrive at the following result:

$$Z[a] = Z[0] \exp\{-\mathcal{V}\, V_{\text{eff}}(a)\} \quad \text{and} \quad Z[a] \leq Z[0] \quad \Rightarrow \quad V_{\text{eff}}(0) \leq V_{\text{eff}}(a) \tag{70}$$

Where $\mathcal{V}$ is simply the volume of Euclidean space. Thus, we see that the effective potential for the Axion is minimized at $a = 0$, and so all the strong CP violation has effectively been removed! Note also that $Z[a] = Z[a + 2\pi n f_a]$, and so the Axion potential in fact has infinitely many minima: $a = 2\pi n f_a$, for $n \in \mathbb{Z}$.

The arguments which led us to concluded that $V_{\text{eff}}(a)$ has minima at $a = 2\pi n$ were somewhat indirect, in the sense that we never derived an explicit expression for $V_{\text{eff}}(a)$. It would be reassuring if we could find a limit in which it is possible to derive such an expression. It turns out that several such limits exists, but we will discuss only one of these limit here: the *Dilute Instanton Gas Approximation (DIGA)*. In this limit, the complete partition function is approximated by taking into account only fluctuations around widely separated single instanton solutions; that is, instantons with $\nu = \pm 1$:

$$Z[a] = \sum_{n_I=\infty}^{\infty} \sum_{n_{\bar{I}}=\infty}^{\infty} \frac{1}{n_I!\, n_{\bar{I}}!} Z_I^{n_I} \, Z_{\bar{I}}^{n_{\bar{I}}} \, e^{i(n_I - n_{\bar{I}})\, a/f_a} \tag{71}$$

Where n_I is the number of single instantons ($\nu = +1$), $n_{\bar{I}}$ is the number of single anti-instantons ($\nu = -1$), and Z_I, $Z_{\bar{I}}$ are the single instanton and single anti-instanton partition functions respectively. The above sums can be carried out explicitly. Exploiting the fact that $Z_I = Z_{\bar{I}}$, we find that:

$$Z[a] = \exp\left\{ Z_I \, e^{ia/f_a} + Z_{\bar{I}} \, e^{-ia/f_a} \right\} = \exp\{(Z_I + Z_{\bar{I}}) \cos(a/f_a)\} \tag{72}$$

Using again the relationship between $Z[a]$ and $V_{\text{eff}}(a)$, we arrive at the following result:

$$V_{\text{eff}}(a) = \frac{Z_I + Z_{\bar{I}}}{\mathcal{V}} \left[1 - \cos(a/f_a)\right] \tag{73}$$

We see that this potential exactly has minima at $a = 2\pi n f_a$, just as we derived from the more indirect arguments above. It should be noted that the DIGA approximation is only valid at high energies, since QCD is only perturbative in this regime. It is however possible to also obtain a result for the potential using *chiral perturbation theory* (χPT) at low energies. Although the explicit form of the potential is very different is this regime, it too has minima located at $a = 2\pi n f_a$. For the sake of completeness, we include the form of the χPT potential below:

$$V_{\text{eff}}(a) = -m_\pi^2 f_\pi^2 \sqrt{1 - \frac{4 m_u m_d}{(m_u + m_d)^2} \sin\left(\frac{a}{2 f_a}\right)} \tag{74}$$

References

[1] David Tong (2018), *Gauge Theory*,
damtp.cam.ac.uk/user/tong/gaugetheory/gt.pdf.

[2] Sidney Coleman (1977), *The Uses of Instantons*,
physics.mcgill.ca/ jcline/742/Coleman-Instantons.pdf

[3] L. Di Luzio, M. Giannotti, E. Nardi, L. Visinelli (2020), *The Landscape of QCD Axion Models*,
arxiv.org/abs/2003.01100.

[4] Guy D. Moore (2017), *Axion Dark Matter on the Lattice*,
arxiv.org/abs/1709.09466

[5] S. Borsanyi, M. Dierigl, Z. Fodor, S.D. Katz, S.W. Mages, D. Nogradi, J. Redondo, A. Ringwald, K.K. Szabo (2015), *Axion Cosmology, Lattice QCD and the Dilute Instanton Gas Approximation*,
arxiv.org/abs/1508.06917

[6] Stefan Vandoren and Peter van Nieuwenhuizen (2008), *Lectures on Instantons*,
arxiv.org/abs/0802.1862

[7] David J. E. Marsh (2015), *Axion Cosmology*,
arxiv.org/abs/1510.07633

[8] Flip Tanedo (2010), *Instantons and their Applications*,
classe.cornell.edu/ pt267/files/documents/A_instanton.pdf

[9] Edward Shuryak (2018), *Nonperturbative Topological Phenomena in QCD and Related Theories*,
arxiv.org/abs/1812.01509

[10] T. Schaefer, E. Shuryak (1996), *Instantons in QCD*,
arxiv.org/abs/hep-ph/9610451