The Strong CP Problem and Its Implications

PITP , 2017

Michael Dine

Department of Physics
University of California, Santa Cruz.

July, 2017

Lecture 1: The Strong CP Problem
Lecture 2: Solutions of the Strong CP Problem: An Assessment
Lecture 3: Axion Cosmology and Axion Searches: Old and New Ideas

The lectures will be self contained. But some suggested reading:

Suggested Reading: Supersymmetry and String Theory: Beyond the Standard Model by M. Dine, sections 5.4, 5.5, 19.2

There are many recent reviews. Some examples:

J. Kim:
https://arxiv.org/abs/1703.03114

D. Marsh:
https://arxiv.org/abs/1510.07633

Kawasaki et al:
https://arxiv.org/abs/1301.1123

P. Sikivie:
https://arxiv.org/abs/astro-ph/0610440

Helpful to do some review of anomalies. Choose your favorite textbook (Peskin and Schroder, Schwarz, Dine)

Lecture 1: The Strong CP Problem

1. Our understanding of QCD: Lagrangian parameters, symmetries
2. Anomalies; CP violating parameters, U(1) problem and their connections
3. Instantons as an indicator of θ dependence and consequences of the anomaly
4. Consequences of θ-dependence: the Neutron electric Dipole moment.
5. Possible Solutions

Our present understanding of the Strong Interactions

QCD: very simple lagrangian, governed by symmetries and particle content:

$$\mathcal{L} = -\frac{1}{4g^2}(F^a_{\mu\nu})^2 + \sum_f \bar{q}_f(i\not{D} - m_f)q_f \quad (1)$$

Here $F_{\mu\nu}$ is the field strength of QCD,q_f the various quark flavors.

Extremely successful.

- At very high energies, detailed, precise predictions, well verified at Tevatron and LHC (dramatically in the context of Higgs discovery.
- At low energies, precision studies with lattice gauge theory.

QCD Parameters from Lattice Gauge Theory

The parameters of the theory: α_s at some scale; quark masses. Current results from lattice simulations (summarized by the FLAG working group)

$$m_u = 2.16\ (9)(7)\mathrm{MeV} \quad m_d = 4.68\ (14)(7)\mathrm{MeV}\, m_s = 93.5(2.5)\mathrm{MeV} \tag{2}$$

(Numbers are in $\overline{MS}$ scheme at 2 GeV.)

$\alpha_s(M_Z) = 0.1192 \pm 0.0011$ from lattice gauge theory; compatible with other measurements.

Would seem we know all we could want to know

Symmetries of QCD

The u, d, and to a lesser extent the s quarks are light. Working with left-handed fields,

$$q = \begin{pmatrix} u \\ d \\ s \end{pmatrix} \quad \bar{q} = \begin{pmatrix} \bar{u} \\ \bar{d} \\ \bar{s} \end{pmatrix} \tag{3}$$

For zero quark mass, the theory has symmetry $U(3) \times U(3)$:

$$q \to Uq; \quad \bar{q} \to V\bar{q} \tag{4}$$

The symmetry is spontaneously broken to a vector subgroup, $SU(3) \times U(1)_B$:

$$\langle \bar{q}_{\bar{f}} q_f \rangle = a\Lambda^3 \delta_{\bar{f}f} \tag{5}$$

Goldstone bosons: π, K, η (eight particles). These particles behave as expected for GB's ("current algebra").

Where's the ninth? The η' is much heavier than the others. E.g.

$$m'_\eta = 958 \text{ MeV}; \; m_\eta = 548 \text{ MeV} \tag{6}$$

One feature of the axial $U(1)$ current which might account for this: it is not strictly conserved. $q \to e^{i\alpha q} \quad \bar{q} \to e^{i\alpha q}$ In four component language,

$$j_5^\mu = \bar{q}\gamma^\mu\gamma^5 q; \quad \partial_\mu j_5^\mu = \frac{N_f}{16\pi^2} F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{7}$$

Here $\tilde{F}$ (“F-dual") is

$$\tilde{F}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}; \quad F\tilde{F} = 2\vec{E}\cdot\vec{B}. \tag{8}$$

The right hand side is a total derivative, $\frac{1}{16\pi^2}\partial_\mu K^\mu$

$$K_\mu = \epsilon_{\mu\nu\rho\sigma}\left(A^a_\nu F^a_{\rho\sigma} - \frac{2}{3} f^{abc} A^a_\nu A^b_\rho A^c_\sigma\right) \tag{9}$$

So one can define a new conserved current,

$$\tilde{j}^\mu = j^\mu_5 - \frac{N_f}{16\pi^2} K^\mu \tag{10}$$

But this current is not gauge invariant, so its status requires investigation. We'll say much more about it shortly, but the charge associated with this current is not conserved. Configurations with $A^\mu \sim \frac{1}{r}$ are already important in a semi-classical treatment. So there is no mystery in the absence of a ninth Goldstone boson.

Strong CP Problem

But the assertion that $F\tilde{F}$ solves the $U(1)$ problem raises a puzzle: one can now add to the QCD lagrangian

$$\mathcal{L}_\theta = \frac{\theta}{16\pi^2} F\tilde{F}. \tag{11}$$

This term violates *parity*, and thus CP.

Again, while

$$F\tilde{F} = \partial_\mu K^\mu; \tag{12}$$

K^μ is not gauge invariant. If there are important configurations, say, in the path integral, for which the A^3 term falls off as $1/r^3$, then we can't drop the surface term. Examples of such configurations will be discussed shortly (instantons).

The possibility of a non-zero θ is related to another possible source of *CP* violation. In writing the QCD Lagrangian, we did not commit to two or four component fermions. In the language of two component fermions, we might have taken complex masses. With four component fermions, we might have included $\bar{q}\gamma_5 q$ terms in the quark lagrangian.

One might have said: what's the big deal? If we had written (two component form)

$$\bar{q}_f m_{\bar{f}f} q_f + \mathrm{c.c.} \tag{13}$$

we could, by separate $U(3)$ transformations of the q and $\bar{q}$ fields, have rendered m (the quark mass matrix) diagonal and real.

But we have just asserted that this redefinition has an anomaly. Indeed, if we think about the effect of such a transformation on the lagrangian,

$$\delta \mathcal{L} = \alpha \partial_\mu j_5^\mu = \alpha \frac{1}{16\pi^2} F\tilde{F}. \quad (14)$$

So we can trade a phase in the quark mass matrix for θ, or vice versa. The invariant quantity is often called $\bar{\theta}$,

$$\bar{\theta} = \theta - \arg\, \det m_q. \quad (15)$$

Anomaly in the current $j^\mu = \bar{q}\gamma^\mu\gamma^5 q$: A calculation

We are interested in a mass term,$me^{i\alpha}\bar{q}q$ There are many ways to think about the anomaly and I cannot do justice to the subject in ten minutes. But I can at least indicate how this effect comes about. In four component language, we have a coupling, for small α:

$$m\,\alpha\bar{q}\gamma^5 q \tag{16}$$

Let's replace m by a field (this might be one of the pseudoscalar mesons, or, as we will see later, the axion). This allows some momentum flow through the diagram.

It is convenient to work with the graph in four component notation since the fermion is massive. The basic expression has the form, after introducing Feynman parameters:

$$N_f m(q)\alpha \int\int dx_1 dx_2 f(x_1, x_2) \frac{d^4k}{(2\pi)^4[k^2+m^2]^3} \mathrm{Tr}\gamma_5 \tag{17}$$

$$(\not p_1 \not A(p_1) ptwoslash \not A(p_2))$$

$$= \frac{1}{16\pi^2} \alpha \epsilon_{\mu\nu\rho\sigma} p_1^\mu p_2^\nu A(p_1)^\rho A(p_2)^\sigma.$$

This is the anticipated result.

Exercise: Verify the expression above

Some things to note:

1. If m is a constant, the result vanishes ($p_1 = -p_2$). This is related to the total derivative, and the rather trivial (plane wave) nature of the external fields.
2. If we did the computation using the background field formalism, we would obtain, also, $\partial_\mu j^\mu = \frac{N_f}{16\pi^2} F\tilde{F}$. This result would hold for non-trivial backgrounds (magnetic monopoles in a $U(1)$ theory, instantons in the non-abelian theory) for which $\int d^4x F\tilde{F}$ is non-zero.
3. It is interesting that $\alpha F\tilde{F}$ term appears in the effective action. Surprising since would seem one could just rotate away.

Instantons

QCD is a strongly coupled theory, and providing a reliable answer to this question requires strong coupling methods (lattice gauge theory). But a semiclassical analysis, while not reliable, indicates that one cannot neglect $\int d^4x \partial_\mu K^\mu$. As a result, the axial current is not conserved and θ is physical.

In the Euclidean functional integral

$$Z = \int [dA][dq][d\bar{q}]e^{-S} \tag{18}$$

it is natural to look for stationary points of the effective action, i.e. finite action, classical solutions of the theory in imaginary time. These *instanton* solutions can be found rather easily. The following tricks simplify the construction, and turn out to yield the general solution. First, note that the Yang–Mills action satisfies an inequality, the Bogomolnyi bound:

$$\int (F \pm \tilde{F})^2 = \int (F^2 + \tilde{F}^2 \pm 2F\tilde{F}) = \int (2F^2 \pm 2F\tilde{F}) \geq 0. \tag{19}$$

So the action is bounded by $|\int F\tilde{F}|$, with the bound being saturated when

$$F = \pm \tilde{F} \tag{20}$$

i.e. if the gauge field is (anti-) self-dual.

This is a first order differential equation. Comparatively easy to solve.

The Instanton Solution

’t Hooft presented the instanton in a fashion which is useful for actual computations. Defining the symbol η:

$$\eta_{aij} = \epsilon_{aij};\ \eta_{a4i} = -\eta_{ai4} = -\delta_{ai};\quad \bar{\eta}_{a\mu\nu} = (-1)^{\delta_{a\mu}+\delta_{a\nu}}\eta_{a\mu\nu} \tag{21}$$

the instanton takes the simple form:

$$A^a_\mu = \frac{2\eta_{a\mu\nu}x^\nu}{x^2+\rho^2} \tag{22}$$

while the field strength is given by:

$$F^a_{\mu\nu} = \frac{4\eta_{a\mu\nu}\rho^2}{(x^2+\rho^2)^2} \tag{23}$$

That this configuration solves the equations of motion follows from:

$$\eta_{a\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}\eta_{a\alpha\beta}. \tag{24}$$

so $F = \tilde{F}$.
The η symbols are connected to the embedding of $SU(2)$ of the gauge group in an SU(2) subgroup of $O(4) = SU(2) \times SU(2)$. This can be understood by noting:

$$\eta_{a\mu\nu} = \frac{1}{2}\mathrm{Tr}(\sigma^a\sigma_{\mu\nu}) \quad \bar{\eta} = \mathrm{Tr}(\sigma^a\bar{\sigma}_{\mu\nu}). \tag{25}$$

Since $F = \tilde{F}$, the equations of motion are satisfied. Note the $1/r$ falloff of A^μ, as opposed to the $1/r^4$ falloff of $F_{\mu\nu}$.

Topological Charge

Asymptotically, A^μ is a pure gauge. The gauge transformation maps the three sphere onto the gauge group. $\frac{1}{16\pi^2}\int d^4x F\tilde{F}$ measures the number of times that the sphere is mapped into the gauge group.

$$\frac{1}{16\pi^2}\int d^4x F\tilde{F} = \frac{1}{16\pi^2}\int d^4x \partial_\mu K^\mu = 1 \quad (26)$$

At large x,

$$A_\mu = \frac{\eta_{a\mu\nu} x^\nu \tau^a}{x^2} \tag{27}$$

$$= i g^{-1} \partial_\mu g$$

[**Exercise:** show what $g = \frac{x_4 + i\vec{x}\cdot\vec{\tau}}{r}$].

Consequences of the Instanton

Having found a classical solution, we want to integrate about small fluctuations about it. Including the θ term, these have the form

$$\langle \mathcal{O} \rangle = e^{-\frac{8\pi^2}{g^2}} e^{i\theta} \int [d\delta A][dq][d\bar{q}] \exp\left(-\frac{\delta^2 S}{\delta A^2}\delta A^2 - S_{q,\bar{q}}\right)\mathcal{O}. \quad (28)$$

Now S contains an explicit factor of $1/g^2$. As a result, the fluctuations are formally suppressed by g^2 relative to the leading contribution. The one-loop functional integral yields a product of determinants for the fermions, and of inverse square root determinants for the bosons.

Both the bosonic and fermionic quadratic fluctuation operators have zero eigenvalues. For the bosons, these potentially give infinite contributions to the functional integral, and they must be treated separately. The difficulty is that among the variations of the fields are symmetry transformations: changes in the location of the instanton (translations), rotations of the instanton, and scale transformations.

More explicitly

$$A_\mu^a(x) = \Omega \frac{2\eta_{a\mu\nu}(x - x_0)^\nu}{x^2 + \rho^2} \tag{29}$$

where Ω denotes a global gauge transformation (or rotation).

Consider translations. For every solution, there is an infinite set of solutions obtained by shifting the origin (varying x_0). Instead of integrating over a coefficient, c_0, we integrate over the *collective coordinate* x_0 (one must also include a suitable Jacobian factor). The effect of this is to restore translational invariance in Green's functions. Similarly, the instanton breaks the rotational invariance of the theory. Correspondingly, we can find a three-parameter set of solutions and zero modes. Integrating over these rotational collective coordinates restores rotational invariance. (The instanton also breaks a global gauge symmetry, but a combination of rotations and gauge transformations is preserved.)

Finally, the classical theory is scale invariant; this is the origin of the parameter ρ in the solution. Again, one must treat ρ as a collective coordinate, and integrate over ρ. There is a power of ρ arising from the Jacobian, which can be determined on dimensional grounds. If d_O is the dimension of the operator $\mathcal{O}$, then, on dimensional grounds, one expects for the ρ dependence:

$$\int d\rho \rho^{-d_0 - 1}. \tag{30}$$

However, there is additional ρ-dependence because the quantum theory violates the scale symmetry. This can be understood by replacing $g^2 \rightarrow g^2(\rho)$ in the functional integral, and using

$$e^{-\frac{8\pi^2}{g^2(\rho)}} \approx (\rho M)^{b_0} \tag{31}$$

for small ρ. For three-flavor QCD, for example, $b_0 = 9$, the leading operator has dimension 9, and the ρ integral diverges logarithmically for large ρ. This is just the statement that the integral is dominated by the infrared, where the QCD coupling becomes strong.

Fermion functional integrals introduce a new feature. In four-component language, it is necessary to treat q and $\bar{q}$ as independent fields. (In two-component language, this corresponds to treating q and q^* as independent fields.) So at one-loop order, we need to study:

$$\not{D} q_n = \lambda_n q_n \quad \not{D} \bar{q}_n = \lambda_n \bar{q}_n \tag{32}$$

$$q(x) = \sum a_n q_n(x), \quad (33)$$

$$S = \sum \lambda_n a_n^* a_n. \quad (34)$$

Then

$$\int [dq][d\bar{q}] e^{-S} = \prod_{n=0}^{\infty} da_n da_n^* e^{-\sum_{n \neq 0} \lambda_n a_n^* a_n}. \quad (35)$$

Zero eigenvalues of the Dirac operator are special. Because the zero modes do not contribute to the action, many Green functions vanish. For example, $\langle 1 \rangle = 0$. In order to obtain a non-vanishing result, we need enough insertions of q to “soak up" all of the zero modes.

The explicit form of the zero modes is not complicated. For $SU(2)$, for simplicity:

$$\not{D}q = 0 \quad \not{D}\bar{q} = 0 \tag{36}$$

and

$$q_0 = \frac{\rho}{(\rho^2 + (x - x_0)^2)^{3/2}}\zeta, \tag{37}$$

where ζ is a constant spinor.

q here might be $u, \bar{u}, d, \bar{d}$, etc.

We can put all of this together to evaluate a Green function which violates the classical $U(1)$ symmetry of the massless theory, $\langle \bar{u}(x)u(x)\bar{d}(x)d(x)\bar{s}(x)s(x)\rangle$. There is one zero mode for each of $u, d, s, \bar{u}, \bar{d}, \bar{s}$. The fields in this Green's function can soak up all of these zero modes. The effect of the integration over x_0 is to give a result independent of x, since the zero modes are functions only of $x - x_0$. The integration over the rotational zero modes gives a non-zero result only if the Lorentz indices are contracted in a rotationally invariant manner (the same applies to the gauge indices). The integration over the instanton scale size – the conformal collective coordinate – is more problematic, exhibiting precisely the infrared divergence we discussed earlier.

So we have provided some evidence that the $U(1)$ problem is solved in QCD, but no reliable calculation. What about θ-dependence? Let us ask first about θ-dependence of the vacuum energy. In order to get a non-zero result, we need to allow that the quarks are massive. Treating the mass as a perturbation, noting the path integral contains a term $e^{i\theta}\frac{1}{16\pi^2}\int d^4xF\tilde{F} = e^{i\theta}$ we obtain a result of the form:

$$E(\theta) = C\Lambda_{\rm QCD}^9 m_u m_d m_s \cos(\theta) \int d\rho \rho^7. \quad (38)$$

So we have evidence for θ-dependence, but again cannot do a reliable calculation. That we cannot do a calculation should not be a surprise. There is no small parameter in QCD to use as an expansion parameter. Fortunately, we can use other facts which we know about the strong interactions to get a better handle on both the $U(1)$ problem and the question of θ-dependence.

Neutron Electric Dipole Moment

A particularly sensitive test of CP conservation in the strong interactions is provided by the neutron electric dipole moment. This corresponds to an operator $\bar{n}\gamma_5\sigma_{\mu\nu}nF^{\mu\nu}$ (here F is the gauge field of electrodynamics). We might first guess that d_n is of order $e\theta \,\mathrm{Fm} \approx \theta 10^{-13} cm$. Then we might expect suppression by powers of quark mass. Our instanton analysis would suggest three factors of quark mass (m_q/m_n, say). We'll see in a moment that there is only one such factor; $m_u/m_n \sim 0.002$, so from the limit on $d_n < 10^{-26}$ e cm, we would have $\theta < 10^{-10}$. We'll give a sharper estimate now.

Consider, first, the coupling of pions to nucleons (for simplicity we'll consider a limit of approximate $SU(2) \times SU(2)$ symmetry, i.e. we'll just treat the u and d quarks as light. Also for simplicity I'll take the u and d quark masses identical. I'll quote general formulas in the end, and leave the derivations to you (or to an examination of the literature).

$$\mathcal{L}_{\pi NN} = \vec{\pi} \cdot \bar{N}(\vec{\tau} i \gamma_5 g_{\pi NN} + \bar{g}_{\pi NN})N. \tag{39}$$

The second term is CP violating. Its effects are directly measurable, in principle, but we will take this to a far more sensitive test in a moment.

As an aside, we can ask what this lagrangian means. We are used to the notion of an effective action for Goldstone bosons. This makes sense; the Goldstone bosons are the light fields, and we can obtain their lagrangian by integrating out heavy fields, such as nucleons and vector mesons. But in QCD, because baryon number is conserved, we can consider a sector with a fixed, non-zero baryon number. Baryon number one is the simplest. For low momenta, we can treat the nucleons as non-relativistic and ignore nucleon-anti-nucleon pairs, and ask about pion scattering amplitudes. This subject was developed extensively some time ago by Weinberg.

For example, in the baryon number one sector, we can study low momentum pion-nucleon processes by thinking of matrix elements such as $\langle \pi\pi \ldots | \bar{q}q | \pi\pi \ldots \rangle$ in terms of a background pion field, and obtain the amplitude from the non-linear chiral lagrangian.

The CP violating term in the underlying lagrangian is obtained from the quark mass term, $m\bar{q}q + \mathrm{c.c.}$, and performing the transformation $q \to e^{i\frac{\vec{\pi}\cdot\vec{\tau}}{2f_\pi}} q$:

$$\delta\mathcal{L} = \frac{m_q\theta}{2}\bar{q}\frac{\vec{\pi}}{f_\pi}\cdot\vec{\tau}q \tag{40}$$

(compare the chiral lagrangian, $\bar{q}mq \to \mathrm{Tr}mU$, $U = e^{i\frac{\vec{\pi}}{2f_\pi}}$, now $\bar{N}UN$. We need the matrix element between an initial and final nucleon state:

$$\frac{\theta m_q}{2f_\pi}\langle N_f|\bar{q}\tau^a q|N_i\rangle \tag{41}$$

The matrix element can be obtained from standard $SU(3)$ global symmetry (Gell-Mann) arguments.

Now this coupling induces a neutron electric dipole moment. The diagram is infrared divergent as $m_\pi \to 0$, and this term is readily extracted.

$$d_n = g_{\pi NN} \bar{g}_{\pi NN} \frac{\log(M_N/m_\pi)}{4\pi^2 M_N} \tag{42}$$

Working through the details:

$$d_n = g_{\pi NN} \frac{\theta m_u m_d}{f_\pi (m_u + m_d)} \langle N_f | \bar{q} \tau^a q | N_f \rangle \ln(m_p/m_\pi) \frac{1}{4\pi^2 m_p} \tag{43}$$

$$= 5.2 \times 10^{-16} \theta cm$$

(this is calculated in an approximation which becomes more and more reliable as the masses of the light quarks become smaller).

From the experimental limit, $d_n < 3 \times 10^{-26}$ e cm, one has $\theta < 10^{-10}$.

The Strong CP Problem

This is a puzzle. Why such a small dimensionless number?

$\theta \to 0$: strong interactions preserve CP. If not for the fact that the rest of the SM violates CP, would be *natural*.

Among naturalness problems, the strong CP problem is special in that it is of almost no consequence. We don't have to invoke anthropic selection to realize that if the cosmological constant was a few orders of magnitude larger than observed, the universe would be dramatically different. The same is true for the value of the weak scale and of the light quark and lepton masses. But if θ were, say, 10^{-3}, nuclear physics would hardly be different than we observe, since effects of θ are shielded by small quark masses.

Possible Resolutions

1. $m_u = 0$ If true, $u \to e^{-i\frac{\theta}{2}\gamma_5}u$ eliminates θ from the lagrangian. An *effective* m_u might be generated from non-perturbative effects in the theory (Georgi, McArthur; Kaplan, Manohar) Could result as an accident of discrete flavor symmetries (Banks, Nir, Seiberg), or a result of “anomalous" discrete symmetries as in string theory (M.D.)
2. CP exact microscopically, $\theta = 0$; spontaneous breaking gives the CKM phase but leads, under suitable conditions, to small effective θ (Nelson, Barr). In critical string theories, CP is an exact (gauge) symmetry, spontaneously broken at generic points in typical moduli spaces. A plausible framework.
3. A new, light particle called the axion dynamically cancels off θ.