Instantons in Particle Physics

Yuhsin Tsai

Institute for High Energy Phenomenology,
Newman Laboratory of Elementary Particle Physics,
Cornell University, Ithaca, NY 14853, USA

E-mail: yt237@cornell.edu

Abstract

In this report we review the basic idea of instanton, especially its connection to the vacuum structure of the non-abelian gauge theory. The solution of the QCD $U(1)$-problem is described, and the baryon and lepton number violation using instanton effect is discussed.

1 Introduction

After Yang and Mills's first paper of the isotopic gauge invariance [1], people's understanding of the non-abelian gauge theory has changed dramatically in this fifty five years. Not only in discovering the gauge interactions and field representations in nature, but also have some idea about how the vacuum looks like. As we will review in this paper, instantons play a central rule in generating the vacuum structure of the nonabelian gauge theory. This is why instantons are widely used in different aspects of particle physics.

This report focus on the basic property of instantons and their importance in the SM physics, such as solving the $U(1)$-problem and giving the baryon and lepton number violation. We will very closely follow the discussion of Coleman [2]. In order to use the newer notations and some topological properties of the gauge theory, we also consult Ryder [3] and Terning [4]. To give a big picture without focusing too much on the algebra, we put detailed derivations in the Appendix. As you can easily see, this report is 'about' fifteen pages besides the Appendices.

The structure of this review is as follows: In Sec. 2, we define instantons in a $1+1$D quantum mechanical system as the barrier penetration between two potential wells. After learning how to use instantons to calculate the tunneling amplitude, in Sec 3, we move to the non-abelian gauge theory and looking for its vacuum structure. As we will see, instantons in this case also generate barrier penetration between different vacua. Using the idea, we identify and solve the QCD $U(1)$-problem in Sec. 4. In Sec. 5, we use the tool developed in the $U(1)$ problem to generate the baryon and lepton number violation. A short conclusion is given in the end.

2 Instanton and the tunneling amplitude

Instanton describes the tunneling between different vacua. The vacua can be as trivial as the minimum of scalar potential or can be as subtle as the gauge field configuration on the boundary. As a warm up, let us begin from the tunneling between potential wells in 1+1D quantum mechanics in the semiclassical limit (small $\hbar$).

Different from the standard quantum mechanical way of solving the differential equation, here we calculate the penetration amplitude using the path integral with an Euclidean action. Why Euclidean action? This is because when using the Euclidean action S_E, e^{-S_E} gives the tunneling amplitude. To see this, let us write the tunneling amplitude between points a and b with energy E and potential barrier V in the WKB approximation

$$\exp\left(-\frac{1}{\hbar}\int_a^b [2m(V-E)]^{\frac{1}{2}} dx\right). \tag{2.1}$$

When $E > V$, the integral becomes

$$\int p dx = \int p\dot{x} dt = \int (H+L) dt = \int (E+L) dt = \int L\, dt. \tag{2.2}$$

In the last equality we shift the energy to zero. As we can see, the WKB factor in this case gives the usual action. When $E < V$, the way to get the same action result is to change $V \to -V$. For the equation of motion $m\ddot{x} = -\partial V/\partial x$, this is identical to changing the time $t \to it$. That is, when using the Euclidean action, we recover the penetration amplitude directly.

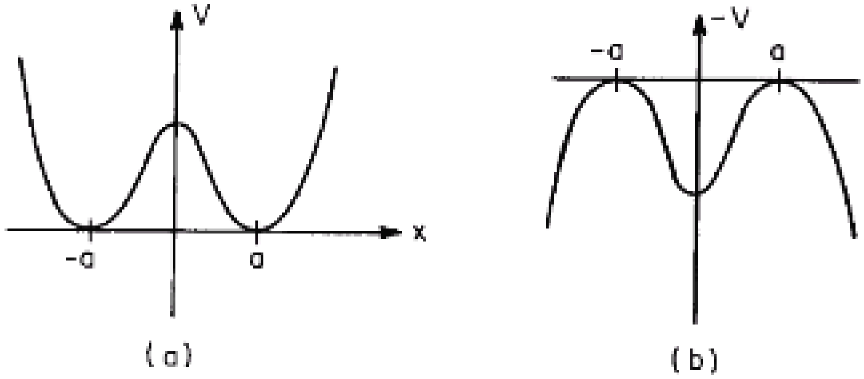


Figure 1: The 1D double wells. The graph is copied from [2].

The goal of this section is to calculate the correlation function of a particle with unit mass moving between two vacua in Fig. 1(a),

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle = N\int [dx]e^{-S/\hbar}, \quad x_i, x_f = \pm a, \tag{2.3}$$

where S is the Euclidean action and N is a normalization factor. The action in this case is

$$S = \int_{-T/2}^{T/2} dt \left[\frac{1}{2}\left(\frac{dx}{dt}\right)^2 + V\right]. \tag{2.4}$$

Since the potential becomes negative in the Euclidean action, the tunneling between $x = \pm a$ becomes the oscillation between the two points in Fig. 1(b). Let us focus on an interesting case first, with the particle being static at $x = -a$ when $t = -T/2$ and then moves to $x = a$ at $t = T/2$. The equation of motion in this case is the one with vanishing E,

$$dx/dt = (2V)^{\frac{1}{2}}. \tag{2.5}$$

Equivalently,

$$t = t_1 + \int_0^x dx'(2V)^{-\frac{1}{2}}, \tag{2.6}$$

where t_1 is the time when $x = 0$. The solution is sketched in Fig. 2. For large t, x approaches a, and eq. (2.5) can approximated by

$$dx/dt = \omega(a - x). \tag{2.7}$$

Thus we have

$$(a - x) \propto e^{-\omega t}. \tag{2.8}$$

This means the solution is a well-localized objects, having a size on the order of $1/\omega$. This object is called instanton. When having a solution going from a to $-a$, we call it anti-instanton. The instanton action can be written as

$$S_0 = \int dt(dx/dt)^2 = \int_{-a}^{a} dx(2V)^{\frac{1}{2}}. \tag{2.9}$$

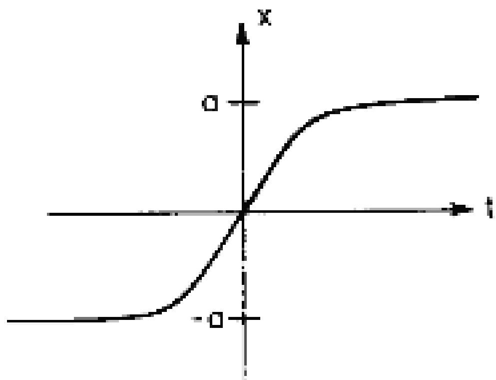


Figure 2: The one instanton solution for barrier penetration. [2]

This goes back to the WKB factor in eq. (2.1) as we expect when using the Eucliden action.

Go back to the corelation function in eq. (2.3). If the particle starts from $-a$ and goes back to $-a$, we have

$$\langle -a|e^{-H(-a)\frac{\Delta t_n}{\hbar}}|-a\rangle\langle -a|Ins(t_n)|a\rangle\langle a|e^{-H(a)\frac{\Delta t_{n-1}}{\hbar}}|a\rangle\langle a|Ins(t_{n-1})|-a\rangle...\langle a|Ins(t_1)|-a\rangle\langle -a|e^{-H(-a)\frac{\Delta t_1}{\hbar}}|-a\rangle. \tag{2.10}$$

Here $\Delta t_n \equiv t_n - t_{n-1}$. $Ins(t)$ means the one-instanton-effect centered at t that changes the particle position between $\pm a$. There are even numbers of instantons in this case. If the process is $-a \to a$, there will be odd number of instantons. As we discuss before, instantons are well time-localized. This means even if it changes the position, the particle stays at $\pm a$ in most of the time. Since the potential at $\pm a$ are symmetric, the Lagrangian is almost a constant through the time evolution (besides the short 'instanton appearace'). We then be able to use this 'dilute-gas approximation' to get

$$\langle -a|e^{-H(-a)\frac{\Delta t_n}{\hbar}}|-a\rangle\langle a|e^{-H(a)\frac{\Delta t_{n-1}}{\hbar}}|a\rangle...\langle -a|e^{-H(-a)\frac{\Delta t_1}{\hbar}}|-a\rangle K^n = \langle -a|e^{-HT/\hbar}|-a\rangle \times K^n. \tag{2.11}$$

where K denotes the correction from the short intanton appearence. Define $\omega^2 \equiv V''(\pm a)$. In Appendix. A, we calculate

$$\langle -a|e^{-HT/\hbar}|-a\rangle = \left(\frac{\omega}{\pi\hbar}\right)^{\frac{1}{2}} e^{-\omega\Delta t}. \tag{2.12}$$

One can use functional integral to show [2]

$$K = (S_0/2\pi\hbar)^{\frac{1}{2}} \left|\frac{\det(-\partial_t^2+\omega^2)}{\det'(-\partial_t^2+V'')}\right|^{\frac{1}{2}}, \tag{2.13}$$

where det′ indicates that the zero eigenvalue is removed.

For the instanton part of the action, the total action of these widely separated objects is nS_0. Also, we can parameterize the time integral as

$$\int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{t_1} dt_2 ... \int_{-T/2}^{t_{n-1}} dt_n = T^n/n!. \tag{2.14}$$

Combining all the ingredients together, we have

$$\langle -a|e^{-HT/\hbar}|-a\rangle = \left(\frac{\omega}{\pi\hbar}\right)^{\frac{1}{2}} e^{-\omega T/2} \sum_{even\ n} \frac{(KTe^{-S_0/\hbar})^n}{n!}, \tag{2.15}$$

while for the $-a \to a$ case, we sum over odd ns. This gives the correlation function between two vacua defined in eq. (2.3)

$$\langle \pm a|e^{-HT/\hbar}|-a\rangle = \left(\frac{\omega}{\pi\hbar}\right)^{\frac{1}{2}} e^{-\omega T/2}\frac{1}{2}\left[\exp(Ke^{-S_0/\hbar}T) \mp \exp(-Ke^{S_0/\hbar}T)\right]. \tag{2.16}$$

For the ground state energy, let us define the energy eigenstates $|n\rangle$ as

$$\langle \pm a|e^{-HT/\hbar}|-a\rangle = \sum_n e^{-E_nT/\hbar}\langle \pm a|n\rangle\langle n|-a\rangle. \tag{2.17}$$

In large T limit, the correlation function only relates to the ground state energy . Comparing to eq. (2.16), this gives

$$E_\pm = \frac{1}{2}\hbar\omega \pm \hbar K e^{-S_0/\hbar}. \tag{2.18}$$

That is, the instanton that corresponds to the barrier penetration breaks the degeneracy of the two ground states. If we call these eigenstates $|+\rangle$ and $|-\rangle$, we have

$$\langle a|-\rangle\langle -|-a\rangle = -\langle a|+\rangle\langle +|-a\rangle = \frac{1}{2}\left(\frac{\omega}{\pi\hbar}\right)^{\frac{1}{2}}. \tag{2.19}$$

To sum up, instantons are well time-localized objects describing the barrier penetration between two vacua. By using the Euclidean action and the dilute-gas approximation, we can calculate the correlation function and the ground state energy.

3 The vacuum of the gauge theory

In this section, we review the role of instantons in describing the tunneling between vacua that relate to different gauge field configurations. Limited by the pages, I will skip some detailed proof and focus on the big picture of the relation between winding-number, $|n\rangle$-vacuum, $|\theta\rangle$-vacuum and instantons.

3.1 The winding number

The winding number is a topological quantity that denotes different mapping classes between a group configuration space and a coordinate space. The simplest example is the mapping between $U(1)$ and a circle. Topologically, both of them are S^1. The $U(1)$ generator of the trivial mapping is

$$h^{(0)}(\theta) = 1. \tag{3.1}$$

We can have the identity mapping

$$h^{(1)}(\theta) = e^{i\theta}, \tag{3.2}$$

or even more complicated mappings

$$h^{(\nu)}(\theta) = e^{i\nu\theta}. \tag{3.3}$$

These mappings can not be connected by a continuous deformation. They belong to different 'homotopy classes' of the mappings. We call the integer number ν as the winding number[1] of the mapping between S^1 and S^1. The fancy convention is

$$\pi_1(S^1) = \mathcal{Z}. \tag{3.4}$$

Not only for abeliean groups, in the non-abelian case, we can also have homotopy classes betwen specific mappings. For example, the mapping between $SU(2)$ and the boundary of 4D Euclidean spacetime (S^3) gives

$$\pi_3(S^3) = \mathcal{Z}. \tag{3.5}$$

One way of seeing this is that we can parametrize the $SU(2)$ operator into (here $x \cdot \sigma \equiv x_i\sigma_i$)

$$h^\nu(x) = \left(\frac{x_4 + ix\cdot\sigma}{\sqrt{\tau^2}}\right)^\nu, \quad \tau^2 = x_4^2 + x^2. \tag{3.6}$$

i.e. with certain winding number ν, each point on the boundary corresponds to a group element. An important formula of calculating the winding number from the operators is

$$\nu = \frac{1}{24\pi^2}\int d^3S\hat{n}_\mu\epsilon_{\mu\nu\lambda\sigma}(\partial_\nu hh^{-1}, \partial_\lambda hh^{-1}\partial_\sigma hh^{-1}). \tag{3.7}$$

One can show that this integral is topologically invariant.

3.2 The gauge field on the boundary

The question we are interested in is the non-perturbative gauge theory with the field strength part of the action looks like $e^{-\frac{8\pi^2}{g^2}}$. This means, we want to study the semi-classical theories with finite actions (large g but not too large).

For an $SU(2)$ gauge theory with finite action (for conventions, see Appendix. B)

$$S = \int d^4x\left[\frac{1}{4g^2}(F,F) - i\bar{\psi}D_\mu\gamma^\mu\psi\right], \tag{3.8}$$

the gauge field on the boundary can only be in a pure gauge form

$$A_\mu = -\frac{i}{g}(\partial_\mu h)h^{-1} + \mathcal{O}(\tau^{-2}). \tag{3.9}$$

Ignoring the $\mathcal{O}(\tau^{-2})$ term that is suppressed at large radius, the gauge field on the boundary can be written into the following form using eq. (3.6) (with $\nu = 1$)

$$A_i = \frac{i}{g\tau^2}(x_i - \sigma_i(\sigma\cdot x + ix_4)), \qquad A_4 = \frac{-1}{g\tau^2}\sigma\cdot x. \tag{3.10}$$

[1]For negative ν, we mean the mapping is 'wided' in the opposite direction.

When doing this, we map the gauge field configuration to the Euclidean boundary S^3.

Besides the usual field strength term (F, F) in the Lagrangian, there is another gauge invariant term that should exist

$$\int d^4x \frac{1}{4}(F, \tilde{F}), \qquad \tilde{F}_{\mu\nu} \equiv \epsilon_{\mu\nu\lambda\sigma} F^{\lambda\sigma}. \tag{3.11}$$

Doing some algebra, we can show that $(F, \tilde{F})$ can be written into a totally derivative term

$$(F, \tilde{F}) = \partial_\mu K_\mu, \qquad K_\mu = 2\epsilon_{\mu\nu\lambda\delta}\left(A_\nu, \partial_\lambda A_\sigma + \frac{2}{3}A_\lambda A_\sigma\right). \tag{3.12}$$

Using eq. (3.10), we have

$$K_\mu = \frac{2x_\mu}{g^2\tau^4}, \tag{3.13}$$

and a nonvanishing volume integral

$$\frac{1}{4}\int d^4x(F, \tilde{F}) = \int K_\perp d^3S = \frac{8\pi^2}{g^2}. \tag{3.14}$$

One interesting thing is, the same integral can also be connected to eq. (3.7)

$$\frac{1}{4}\int d^4x(F, \tilde{F}) = \int K_\perp d^3S = \frac{1}{2}\int d^3S \hat{n}_\mu \epsilon_{\mu\nu\lambda\sigma}\left(A_\nu, \partial_\lambda A_\sigma + \frac{2}{3}A_\lambda A_\sigma\right), \tag{3.15}$$

$$= \frac{1}{3g^2}\int d^3S \hat{n}_\mu \epsilon_{\mu\nu\lambda\sigma}(\partial_\nu h h^{-1}, \partial_\lambda h h^{-1} \partial_\sigma h h^{-1}) = \frac{8\pi^2}{g^2}\nu. \tag{3.16}$$

This means, the integral in eq. (3.14) is given by the nontrivial mapping between $SU(2)$ and the Euclidean S^3 with winding number $\nu = 1$.

3.3 The n-vacuum and the θ vacuum

In fact, the nontrivial gauge field A_μ we use in the previous section is the instanton in 4D. As we just see, one instanton effect changes the winding number of the system by one. To see this more clearly, let us do the winding number integral again by using the boundaries as shown in Fig. 3.

$$\nu = \frac{1}{24\pi^2}\left[\int_{I-II} d^3S\ \epsilon_{4ijk}(\bar{A}_i, \bar{A}_j \bar{A}_k) + \int_{-\infty}^{\infty} dx_4 \int_{III} d^2S\ \hat{n}_i \epsilon_{i\mu\nu\lambda}(\bar{A}_\mu, \bar{A}_\nu \bar{A}_\lambda)\right]. \tag{3.17}$$

Here $\bar{A}_\mu \equiv (\partial_\mu h)h^{-1}$. To do the integral, we need to know the instanton solution inside the boundary, which is[2]

$$A_\mu = \frac{\tau^2}{\tau^2 + \rho^2}\left(\frac{-i}{e}\right)(\partial_\mu h)h^{-1}. \tag{3.18}$$

[2]We can get this by solving $F_{\mu\nu} = \tilde{F}_{\mu\nu}$ and match the boundary condition. The equation comes from getting the equality of the Schwartz inequality $(F, F) \geq (F, \tilde{F})$, which gives the lower bound of the action.

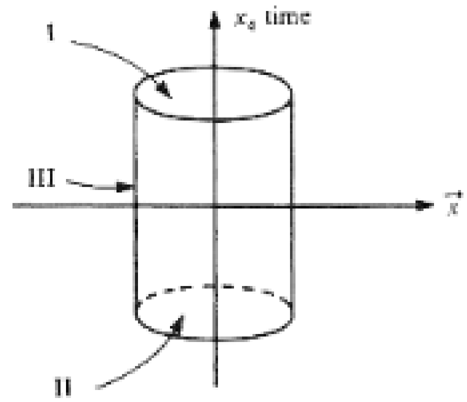


Figure 3: I and II are the hypersurfaces $x_4 \to \infty$ and $x_4 \to -\infty$ and III is the hypersurface joining them. The diagram is copied from [3]

The ρ denotes the 'size' of the instanton. As we can see, A_μ goes back to eq. (3.9) when $\tau \to \infty$. Since the winding number is gauge invariant, it is convenient to choose a gauge such that $A'_4 = 0$ so that the integral over the 'cylinder III' vanishes[3]. The integral then becomes

$$\nu = \frac{1}{24\pi^2}\left[\int_I d^3S\ \epsilon_{4ijk}(\bar{A}_i, \bar{A}_j\bar{A}_k) - \int_{II} d^3S\ \epsilon_{4ijk}(\bar{A}_i, \bar{A}_j\bar{A}_k)\right] = \nu_I - \nu_{II}. \tag{3.19}$$

i.e. the instanton really changes the winding number of the system.

Another way of interpreting the result is that as time evolves from $-\infty$ to ∞, a vacuum (with homotopy class ν_{II}) evolves into another vacuum (with homotopy class ν_I). The instanton solution then represents the transition between one vacuum class to another. The Yang-Mills vacuum is therefore infinitely degenerate, consisting of an infinite number of homotopically non-equivalent vacua. Since they are denoted by different winding numbers, we call these vacua the $|n\rangle$-vucua. We can also do the Fourier transform of it

$$|\theta\rangle = \sum_n e^{in\theta}|n\rangle. \tag{3.20}$$

This gives the so called $|\theta\rangle$-vacua.

3.4 The energy of a θ vacuum

Let us calculate some quantities of a θ vacuum that we can use for the $U(1)$ problem. First, using the dilute-gas approximation, the correlation function $\langle\theta|e^{-HT}|\theta\rangle$ is contributed by n instantons and $\bar{n}$ anti-instantons

$$\langle\theta|e^{-HT}|\theta\rangle = \sum_{n,\bar{n}} \left(Ke^{-S_0}\right)^{n+\bar{n}} \frac{(VT)^{n+\bar{n}}}{n!\bar{n}!} e^{i(n-\bar{n})\theta}. \tag{3.21}$$

[3]Such a gauge transform is $A'_\mu = UA_\mu A^{-1} - i(\partial_\mu U)U^{-1}$ with $U = \exp[\frac{ix\cdot\sigma}{(\tau^2+\rho^2)^{1/2}}\theta]$, $\theta = \tan^{-2}[\frac{x_4}{(\tau^2+\rho^2)^{1/2}} - \frac{\pi}{2}]$.

Similar to the way we derive eq. (2.16), here e^{-S_0} is the one instanton action, K denotes the 'instanton appearance' effect (similar to eq. (2.13)), $[(VT)^{n+\bar{n}}/(n!\bar{n}!)]$ comes from the 4-volume integral as in eq. (2.14), and $e^{i(n-\bar{n})\theta}$ is the Fourier transform factor as in eq. (3.20). The sum can be written as

$$\langle\theta|e^{-HT}|\theta\rangle = \exp(KVTe^{-S_0}e^{i\theta})\exp(KVTe^{-S_0}e^{-i\theta}) = \exp(2KVTe^{-S_0}\cos\theta), \tag{3.22}$$

which gives the energy density

$$E(\theta)/V = -2Ke^{-S_0}\cos\theta. \tag{3.23}$$

After putting in the instanton action $S_0 = 8\pi^2/g^2$ and the correct K value (coming from the calculation of the instanton degree of freedom, dimensional analysis and the RG running of the coupling), the result becomes

$$E(\theta)/V = -A\cos\theta\ e^{-8\pi^2/g^2}g^{-8}\int_0^\infty \frac{d\rho}{\rho^5}(\rho M)^{8\pi^2\beta_1}, \tag{3.24}$$

where g is the gauge coupling, M is the renormalization scale, β_1 is a number that can be calculated from RG, and A is a constant independent of ρ, g and M.

The next thing we want to calculate is the expectation value of $(F, \tilde{F})$. By the translational invariance,

$$\langle\theta|(F(x), \tilde{F}(x))|\theta\rangle = \frac{1}{VT}\int d^4x\langle\theta|(F,\tilde{F})|\theta\rangle = \frac{32\pi^2}{g^2VT}\langle\theta|\nu|\theta\rangle, \tag{3.25}$$

here we have used eq. (3.14). This gives

$$\langle\theta|(F(x), \tilde{F}(x))|\theta\rangle = \frac{32\pi^2\int[dA]\nu e^{-S}e^{i\nu\theta}}{g^2VT\int[dA]e^{-S}e^{i\nu\theta}} = -\frac{32\pi^2 i}{g^2VT}\frac{d}{d\theta}\ln\left(\int[dA]e^{-S}e^{i\nu\theta}\right). \tag{3.26}$$

Since

$$\int[dA]e^{-S}e^{i\nu\theta} = \langle\theta|e^{-HT}|\theta\rangle, \tag{3.27}$$

we have

$$\langle\theta|(F(x), \tilde{F}(x))|\theta\rangle = -\frac{64\pi^2 i}{g^2}Ke^{-S_0}\sin\theta. \tag{3.28}$$

4 The U(1) problem

The most successful use of instantons in the SM is to solve the U(1) problem. In this section, we define what is the U(1) problem first and then give a solution using the concept of instantons.

4.1 The Goldstone-boson of the $\mathrm{U}(1)_A$

In SM, QCD is the formal description of the strong force. The quarks including up and down types in each of the three families interact with each other through changing the SU(3) gauge fields. This gives successful prediction in experiments and has been well accepted. However, one interesting feature of the strong interaction is that besides the gauge symmetry, we also have some approximated global symmetry that gives corresponding selection rules to a very good level. For example, when ignoring the u and d quark masses in the high energy scale (like GeV scale), the $U(2)_L \times U(2)_R$ transform acting on the chiral spinors

$$\psi_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \qquad \psi_R = \begin{pmatrix} u_R \\ d_R \end{pmatrix}, \tag{4.1}$$

is invariant in the kinetic term

$$\mathcal{L} \ni \bar{\psi}_L i D_\mu \gamma^\mu \psi_L + \bar{\psi}_L i D_\mu \gamma^\mu \psi_L. \tag{4.2}$$

Because of the quark condensation, this global symmetry is spontaneously broken into

$$U(2)_L \times U(2)_L \to SU(2)_D \times U(1)_B, \tag{4.3}$$

where the $\mathrm{SU}(2)_D$ gives the isospin symmetry and the $\mathrm{U}(1)_B$ gives the baryon number conservation in the strong dynamics. The three Goldstone bosons from $SU(2)_L \times SU(2)_R \to SU(2)_D$ becomes the three light pions, which is a strong evidence that the spontaneous symmetry breaking really happens.

Extending the same idea to the broken $U(1)_A$, which has the transform

$$\psi_f \to e^{-i\alpha\gamma^5}\psi_f, \tag{4.4}$$

(Now ψ_f is written into the four component Dirac spinor with flavor $f = 1, 2$) and the associated current

$$j^5_\mu = \sum_{f=1}^{2} \bar{\psi}_f \gamma_\mu \gamma_5 \psi_f, \tag{4.5}$$

the SSB (spontaneous symmetry breaking) of it should also give a Goldstone boson with the same mass scale as pions. However, we do not see it. This gives the most naive version of the $U(1)$ problem, i.e., where is the Goldstone of the broken $U(1)_A$?

The reason why we say "naive" is that the SSB argument we just give is not quite correct. In fact, the $U(1)_A$ is broken by anomaly in the perturbation theory. In the limit of N massless quarks,

$$\partial^\mu j^5_\mu = \frac{N}{32\pi^2}(F_{\mu\nu}, \tilde{F}_{\lambda\sigma}). \tag{4.6}$$

This means j^5_μ is not a conserved current. It looks like we just solve the problem since $U(1)_A$ is not a symmetry in the beginning. However, if we redifine the current into

$$J^5_\mu \equiv j^5_\mu - \frac{N}{32\pi^2}\epsilon_{\mu\nu\lambda\sigma}(A^\nu, F^{\lambda\sigma} - \frac{2}{3}A^\lambda A^\sigma) \equiv j^5_\mu - \frac{N}{32\pi^2}G_\mu, \tag{4.7}$$

since

$$\partial_\mu G_\mu = (F_{\mu\nu}, \tilde{F}_{\lambda\sigma}), \tag{4.8}$$

J^5_μ becomes a gauge-variant but conserved current[4]. Gauge-variant means it is not observable. Nevertheless, because its charge

$$Q_5 = \int d^3x J^5_0 \tag{4.9}$$

is conserved, this theory when realized in the Goldstone mode would demand the existence of a pseudoscalar meson with mass m_0 [6][5]

$$m_0 \simeq m_\pi. \tag{4.10}$$

To see this we can use the standard current-algebra technique to obtain a Ward identity [7] [6]

$$\begin{aligned} m_0^2 f_0^2 &= i\frac{m_0^2 - k^2}{m_0^2}\left[ik^\nu \int d^4x e^{-ik\cdot x}\left\langle 0\left|T(\partial^\mu J^5_\mu(0)J^5_\nu(x))\right|0\right\rangle + \int d^4x e^{-ik\cdot x}\left\langle 0\left|\delta(x_0)\left[\partial^\mu J^5_\mu(0), J^5_0(x)\right]\right|0\right\rangle\right], \\ &= i\frac{m_0^2 - k^2}{m_0^2}\left[ik^\nu \int d^4x e^{-ik\cdot x}\left\langle 0\left|T(\partial^\mu J^5_\mu(0)J^5_\nu(x))\right|0\right\rangle + m_\pi^2 f_\pi^2\right], \end{aligned} \tag{4.11}$$

where f_0 is the isoscalar meson decay constant. In low energy limit $k^\nu \to 0$, if there is no zero mass pole in the first term on the right-hand side, we relate $m_0^2 f_0^2$ to a the second integral term which is $m_\pi^2 f_\pi^2$,

$$m_0^2 f_0^2 = m_\pi^2 f_\pi^2 \tag{4.12}$$

To go further, one can write J^5_μ into the sum of an $SU(3)$ octet and a singlet. Since all the pseudoscalar octet decay constants are equal, $f_0 \simeq f_\pi$, we have the mass relation in eq. (4.10). This brings the trouble back; we have to explain where this pseudo scalar goes.

It is pointed out by Kogut and Susskind [8] that one way to avoid the problem is for the gauge-variant J^5_μ to be coupled to a massless 'particle', then the first integral in the right side gets a pole and does not drop out when $k^\nu \to 0$. There then be no constraint on m_0. Since J^5_μ is gauge-variant, this gauge-dependent massless particle does not couple to the physical quantity and brings no further problems.

To identify this massless field, Kogut and Susskind study the Schwinger model - massless spinor electrodynamics, in 1 + 1D in a covariant gauge. This model is solvable and has the ingredients we want:

- a gauge-invariant axial current with an anomalous divergence.
- a gauge-variant but conserved axial current.
- a $U(1)_A$ breaking without Goldstone poles in the gauge-invariant Green's functions.

---

[4]The new added gauge-variant term can be seen as spurions [5]. That is the way we change an explicit SSB (of $U(1)_A$ from anomaly) into a spontaneous one. The price of doing this is that we should have a Goldstone-like field couples to the this gauge-variant term. This is what we want to show here.

[5]A more precise result should be $m_0 \leq \sqrt{3}m_\pi$.

[6]I do not know how to derive this, just to describe the problem is in a more quantitative way.

What Kogut and Susskind find is that after using boson fields to describe the 1D system (this is the feature of the Schwinger model), two free massless fields ϕ_+ and ϕ_-, create quanta of positive and negative norm. This gives the propagators carrying opposite signs. All gauge-invariant quantities couple to the sum of these fields $(\phi_+ + \phi_-)$. This has zero propagator and is free of the new poles. On the other hand, the gauge-variant quantities couples to $\partial_\mu(\phi_+ - \phi_-)$. The coupling to ϕ_- then carries an additional sign which compensates the sign in the propagator and thus gives the pole we want. This setup is called the *Goldstone dipole*.

According to this, the missing degree of freedom coming from the SSB does not exist in the gauge-invariant greens functions but are shown in the gauge-variant ones. We then be able to formulate the $U(1)$-problem in a more precise way: *Is the $U(1)_A$ in QCD spontaneously broken via a Goldstone dipole?*

4.2 QCD (baby version)

't Hooft gives a brilliant solution of the $U(1)$ problem when connecting this chiral symmetry to instanton [9]. To get an idea about how this is done, let us solve the $U(1)$-problem in a baby version QCD first. The basic steps are as follows

- Show the relation between the $U(1)_A$ transform and the shift of the θ-vacua.
- Show that the $U(1)_A$ is SSB when staying in a θ-vacuum.
- Look for the Goldstone-dipole in the gauge-variant Green's functions.

4.2.1 $U(1)_A$ and the θ-vacua

In this baby theory the gauge symmetry is $SU(2)$, and there exists only a single isodoublet quark with zero mass. The action looks like

$$S = \int d^4x \left[\frac{1}{4g^2}(F,F) - i\bar{\psi} D_\mu \gamma_\mu \psi \right]. \tag{4.13}$$

There are θ-vacua in this case, with the same quantities derived in Sec. 3.4,

$$E(\theta)/V = -2K\cos\theta e^{-S_0}, \qquad \langle\theta|(F,\tilde{F})|\theta\rangle = -\frac{64\pi^2 i}{g^2} K e^{-S_0}\sin\theta. \tag{4.14}$$

The only difference for quarks is now the K-factor (see eq. (2.13)) contains

$$\det\left[\frac{i\not{D}}{i\not{\partial}}\right] = \det\left[\frac{i(\partial_\mu + A_\mu)\gamma_\mu}{i\not{\partial}}\right], \tag{4.15}$$

where A_μ is the field of an instanton. To calculate this, we need to find the eigenfields of ψ under the $i\not{D}$ operator. As we will show now, there exists zero eigenvalue modes when there is a nontrivial winding number. This makes the determinant vanish, as does $E(\theta)/V$ and $\langle\theta|(F,\tilde{F})|\theta\rangle$.

Decomposing the fermion into different eigenfunctions of $i\not{D}$,

$$i\not{D}\psi_r = \lambda_r \psi_r. \tag{4.16}$$

Since $iD\!\!\!/$ is Hermitian, all λ_r's are real. Using γ^5, we can get the other set of the eigenfields

$$iD\!\!\!/\gamma^5\psi_r = -\lambda_r\gamma^5\psi_r. \tag{4.17}$$

Thus non-vanishing eigenvalues always occur in pairs of opposite sign. The eigenfunctions of γ^5 are the zero eigenfunctions of $iD\!\!\!/$ (since $\gamma_5^2 = 1$, we have $\chi_r = \pm 1$)[7]

$$\gamma_5\psi_r = \chi_r\psi_r, \;\; (\lambda_r = 0). \tag{4.18}$$

Let us denote the number of zero eigenfunctions by $n_\pm$. As proved in Appendix. C.2, we have

$$n_- - n_+ = \nu. \tag{4.19}$$

That is, there is a zero eigenvalue in any gauge field of non-zero winding number. When there is an instanton background centering at X and having size ρ, the zero mode eigenfunction is

$$\psi_0(x - X, \rho) = \rho\left[\rho^2 + (x - X)^2\right]^{-\frac{3}{2}} u, \tag{4.20}$$

where u is a constant spinor. When having n widely separated instantons and anti-instantons, there are n such eigenfunctions centered about each object.

The result of having zero modes is that, the $E(\theta)$ and the expectation value of $(F, \tilde{F})$ with different θ are always zero. All the θ-vacua are degenerate. Where does this degeneracy come from? As we show in Appendix.X, when applying a $U(1)_A$ transform ($\psi \to \exp(-i\alpha\gamma^5)\psi$) to the θ-vacuum expectation value of quark mutilinears $\phi^{(i)}$'s, we have

$$\left[\frac{\partial}{\partial\alpha} + 2\frac{\partial}{\partial\theta}\right]\left\langle\theta\left|\phi^{(1)}(x_1)...\right|\theta\right\rangle = 0. \tag{4.21}$$

This means the $U(1)_A$ transform rotates one θ-vacuum into another. When staying in one $|\theta\rangle$, $U(1)_A$ is spontaneously broken, and the θ-vacua are the many vacua that appear when a symmetry suffers SSB. This explains where the degeneracy comes from - it comes from the $U(1)_A$ invariance.

4.2.2 The SSB of $U(1)_A$

In fact, it might be too early to say that the SSB really occurs. If the ∂_α and ∂_θ for every Green's function just vanish independently, the θ-vacua would have nothing to do with the $U(1)_A$ transform. To show that the vacuum really 'feels' the $U(1)_A$ transform, let us calculate the Green's function of a fermion condensate $\bar{\psi}\psi$ that breaks $U(1)_A$.

$$\langle\theta\left|\sigma_\pm(x)\right|\theta\rangle = \frac{\displaystyle\int[dA][d\psi][d\bar{\psi}]e^{-S}e^{i\nu\theta}\sigma_\pm(x)}{\displaystyle\int[dA][d\psi][d\bar{\psi}]e^{-S}e^{i\nu\theta}}, \tag{4.22}$$

where

$$\sigma_\pm = \frac{1}{2}\bar{\psi}(1 \pm \gamma^5)\psi. \tag{4.23}$$

[7]To see this, when having $\chi_r = 1$, $iD\!\!\!/\gamma^5\psi_r = iD\!\!\!/\psi_r = \lambda_r\psi_r = \lambda_r\gamma^5\psi_r$, which contradicts eq. (4.17) if λ_r is non-zero.

Since under $U(1)_A$

$$\sigma_\pm \to \sigma_\pm + \delta\sigma_\pm = \frac{1}{2}\left(\bar{\psi} - i\bar{\psi}\gamma^5\delta\alpha\right)\left(1 \pm \gamma^5\right)\left(\psi - i\gamma^5\psi\delta\alpha\right), \tag{4.24}$$

The $\sigma_\pm$ are eigenfunctions of the $U(1)_A$ transform

$$\partial\sigma_\pm/\partial\alpha = \delta\sigma_\pm/\delta\alpha = \mp 2i\sigma_\pm. \tag{4.25}$$

If eq. (4.22) is non-zero, the vacuum expectation value feels the $U(1)_A$ transform, and the SSB really happens.

To calculate the Green's function, we use the dilute-gas approximation to multiply all the instanton and anti-instons together like what we do in Sec. 3.4. The similar factors such as

$$K = 2g^{-8}\int_0^\infty \frac{d\rho}{\rho^5} f(\rho M) \tag{4.26}$$

still exists. For the functional integral of fermions, if there exists one instanton, the winding number $\nu = 1$ gives one zero mode and the integral containing $\det(i\not{D})$ vanishes. The only non-zero case is when having $\bar{\psi}_0\psi_0$ in the integral, such that the zero mode fermions in the action contract with it and the determinant of the rest of the fermions exists. This case is exactly the same as the one in eq. (4.22). For the σ_- in the numerator, the Fermi integral under one instanton background is[8]

$$\frac{1}{2}\psi_0^\dagger(x-X,\rho)(1-\gamma^5)\psi_0(x-X,\rho)\prod_{\lambda_r\neq 0}\lambda_r = \psi_0^\dagger(x-X,\rho)\psi_0(x-X,\rho)\,{\det}'(i\not{D}). \tag{4.27}$$

Here $\det'$ denotes a determinant with vanishing eigenvalues removed. Since the determinant term does not depend on the instanton location X, the integral over X is trivial

$$\int d^4X\psi_0^\dagger\psi_0 = 1. \tag{4.28}$$

For $[dA]$, the instanton action gives a factor $e^{-\frac{8\pi^2}{g^2}}$, and there is also a factor $e^{i\theta}$ for $\nu = 1$. In the denomenator, since there is no fermions in the integral that contract the zero modes, there can be no instantons and anti-instantons. Thus we only have $\det(i\not{\partial})$.

Putting all the ingredient together, the Green's function becomes

$$\langle\theta\,|\sigma_\pm(x)|\,\theta\rangle = e^{-\frac{8\pi^2}{g^2}}e^{\mp i\theta}g^{-8}2\int_0^\infty\frac{d\rho}{\rho^5}f(\rho M)\frac{\det'(i\not{D})}{\det(i\not{\partial})}. \tag{4.29}$$

Using dimensional analysis, since the eigenvalues of $i\not{D}$ should have 1/length, the det' which has one less eigenvalue gives

$$\frac{\det'(i\not{D})}{\det(i\not{\partial})} = \rho \times h(\rho M), \tag{4.30}$$

where h is an function of a quantity. The expectation value carries dimension $1/(\text{length})^3$ as expected. We can go further by using the RG running of the coupling to figure out f and h, but the more important result for us now is that eq. (4.22) is non-zero and the SSB of $U(1)_A$ do occur.

[8] Here we use the fact that one instanton has one zero eigenfunction with $\gamma^5\psi_0 = -\psi_0$ and none with $\gamma^5\psi_0 = \psi_0$. The anti-instanton is reversed.

4.2.3 The Goldstone dipole

Now we know spontaneous symmetry breaking occurs. Are there Goldstone bosons? Let us look for them in

$$\langle\theta\,|\sigma_+(x)\sigma_-(0)|\,\theta\rangle\,, \tag{4.31}$$

which is the propagator of the fermion-condensate that breaks the symmetry. If there is a Goldstone pole couple in the gauge invariant quantity, it must be in this propagator[9].

The calculation of this is similar to the one we just did. The only difference now is since there are two σ's, we can have either no instantons or one instanton and one anti-instanton that get rid of the zero modes. The first case gives the usual one-loop perturbation theory expression, having a two-quark cut, but no Goldstone pole. The second case just gives the product $\langle\theta|\sigma_+|\theta\rangle\langle\theta|\sigma_-|\theta\rangle$. This also has no Goldstone pole. We can also check other gauge-invariant quantities, and the Goldstone never shows up.

Following Kogut and Susskind's idea, let us check if the gauge-variant current J^5_μ couples to the Goldstone dipole. For example,

$$\langle\theta|J^5_\mu(x)\sigma_-(0)|\theta\rangle = \langle\theta|j^5_\mu(x)\sigma_-(0)|\theta\rangle + \frac{1}{32\pi^2}\langle\theta|G_\mu(x)\sigma_-(0)|\theta\rangle. \tag{4.32}$$

The first term on the right side is gauge-invariant and non-conserved by the explicit breaking, which has nothing to do with the Goldstone poles. However, the second term does. The way to see this is that the integration of the totally derivative term

$$\int d^4x\partial_\mu\langle\theta|G_\mu\sigma_-(0)|\theta\rangle = \int d^4x\langle\theta|(F,\tilde{F})\sigma_-(0)|\theta\rangle = 32\pi^2\langle\theta|\sigma_-(0)|\theta\rangle \neq 0 \tag{4.33}$$

(here we have used eqs. (3.14) and (4.29)) is non-vanishing. This means the Green's function connecting to $x\to\infty$ is non-zero. Since only a massless field can propagate like this, there must be a Goldstone dipole couple to this.

As a reminder, the nonvanishing integral requires a nontrivial winding number in $\langle\theta|\sigma_-(0)|\theta\rangle$, i.e. the instanton background is vital to the argument. On the other hand, since there is no instanton configuration for $\langle\theta|J^5_\mu J^5_\lambda|\theta\rangle$, we have $\nu=0$, and there is no poles in it. To summarize, in the dilute-gas approximation, the $SU(2)$ gauge theory with one massless fermion-doublet contains:

- SSB of $U(1)_A$.
- no Goldstone poles in gauge-invariant Green's functions.
- no Goldston dipoles in the propagator of gauge-variant conserved current.
- a Goldstone dipole in the Green's function of a gauge-variant conserved current and a gauge-invariant operator.

This gives the Goldstone dipole of Kogut and Susskind.

[9] I think one way to see this is, the fermion-condensate is the 'higgs' in the dual picture that generates SSB. The Green's function should include the propagator of Goldstone-bosons if there exists any.

4.3 QCD (the real version)

Real QCD in the chiral $U(2) \times U(2)$ limit differs from the baby model in two respects. Firstly, we have triplet quarks with gauge group $SU(3)$ but not doublet quarks with $SU(2)$. Second, we have three massless quarks (u and d) rather than one.

For the symmetry part, there is a remarkable theorem due to Raoul Bott that states that any continuous mapping of S^3 into G can be continuously deformed into a mapping into an $SU(2)$ subgroup of G. Thus, everything we used about the winding number is the same. The only thing we have to change is the g^{-8} from the integration of the instanton phase space now becomes g^{-12}.

Changing the number of massless fermions effects more. As we show in the Appendix. C.2, when having two massless fermions, the sum rule in eq. (4.19) changes to

$$n_- - n_+ = 2\nu. \tag{4.34}$$

This means $\nu = 1$ gives two vanishing eigenvalues instead of one. We then have $\langle\theta|\sigma_-|\theta\rangle$ vanishing in the one instanton background. The way to show the SSB in this case is to consider the expectation value of two σ's[10]

$$\frac{1}{2}\epsilon_{ij}\epsilon_{kl}\bar{\psi}_i(1-\gamma^5)\psi_k\bar{\psi}_j(1-\gamma^5)\psi_l = \det[(\bar{\psi}_R\psi_L)_{ff'}], \tag{4.35}$$

where ff' are the flavor (u and d) indices that $SU(2)_L \times SU(2)_R$ acts on.

All in all, the result we find in the previous section still applies, $U(1)_A$ is broken spontaneously in the instanton background, and the missing degree of freedom becomes the Goldstone dipole couple to the gauge-variant but conserved axial current J^5_μ.

5 The baryon and the lepton number violation

One very important concept we developed when solving the $U(1)$-problem is that the instanton changes the number of zero modes in eq. (4.34). This special sum rule is called the 't Hooft term [9]:[11]

$$n_{\rm L} - n_{\rm R} = \sum_r n_r 2C(r), \tag{5.1}$$

This corresponds to the one-instant case, where n_r is the number of fermions in the representation r and $C(r)$ is defined in eq. (B.4). When having an instanton effect, it creates the LH fermions, while for an anti-instanton effect it creates the RH fermions.

Since the instanton effect also exists in the electroweak symmetry $SU(2)\times U(1)$, we can change the number of fermions charged under $SU(2)$. The fermions that have non-trivial $C(r)$ in $SU(2)$ are the three (generation f) lepton doublets (ν^f_L, ℓ^f_L) and three quark doublets $(u^{f,r}_L, d^{f,r}_L)$. For each flavor of the leptons, $C(r) = 1/2$ and $n_r = 1$. For each flavor of the quarks, we have $C(r) = 1/2$

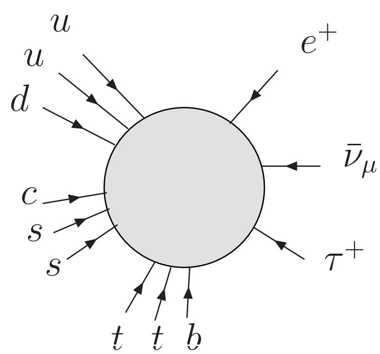
[10] Instanton does not break chiral $SU(2)_L \times SU(2)_R$, and the θ-vacuum is invariant under this. This means the non-zero vacuum configuration has to be a chiral $SU(2)_L \times SU(2)_R$ singlet.

[11] The origin of this comes from the non-conservation of the chiral current $\partial_\mu j^5_\mu \propto (F, \tilde{F})$. When doing the volume integral, the nonconservation of the LH and RH fermions is proportional to the winding number.

and $n_r = 3$ for three colors. When having one anti-instanton, since only the LH fermions are effected by the $SU(2)$, the LH anti-fermions are generated and satisfy [9]

$$\Delta e_L + \Delta\nu_{Le} = \Delta\mu_L + \Delta\nu_{L\mu} = \Delta\tau_L + \Delta\nu_{L\tau} = -1, \qquad \Delta u_L + \Delta d_L = \Delta c_L + \Delta s_L = \Delta t_L + \Delta b_L = -3. \tag{5.2}$$

From this, we can have processes like



Through the CKM rotation, this generates the process like

$$p + p^c + n^c \to e^+ + \tau^+ + \bar{\nu}_\mu, \tag{5.3}$$

where "c" means the u and d come from the CKM rotation. The cross section of this process is suppressed by the instanton action $e^{-16\pi^2/g_{EW}^2}$ like the one in the amplitude eq. (4.29). In this case, it is

$$e^{-16\pi^2/g_{EW}^2} = e^{-16\pi^2/e^2 \sin^{-2}\theta_W} \simeq 10^{-262}. \tag{5.4}$$

This gives a deuteron lifetime of the order of 10^{218}yr. Does this mean the instanton-type baryon and the lepton number violation can never happen? In fact, all the discussion we just have are in the zero temperature case. The free energy scale comparing to the factor $e^{-16\pi^2/g_{EW}^2}$ is about the EW symmetry breaking scale. In the early universe, before the temperature is lowered down to the EWSB scale, we can get 'over' the free-energy barriers (between different θ-vacua) instead of tunneling through them. This different vacua changing effect is called the 'sphaleron effect'[12], which is important for baryogenesis models such as leptogenesis.

6 Conclusion

We have seen that the penetration between different vacua can be described using instantons. The vacuum structure of a non-abelian gauge theory is related to different homotopy classes of the mapping between the configuration space and the Euclidean spacetime. The different vacua in this case are also connected by instantons.

Besides the solution of the $U(1)$-problem and a tool for the baryon and lepton violation described in this report, the idea of instantons are heavily used in many different aspects of the field theory, such as the strong CP problem, the NSVZ β function and the gaugino mass in SUSY, the model of composite fields, and the tunneling mechanism in the inflation models. The physics hidden in the vacuum structure keeps bringing us new surprise.

[12] A good introduction is in [10].

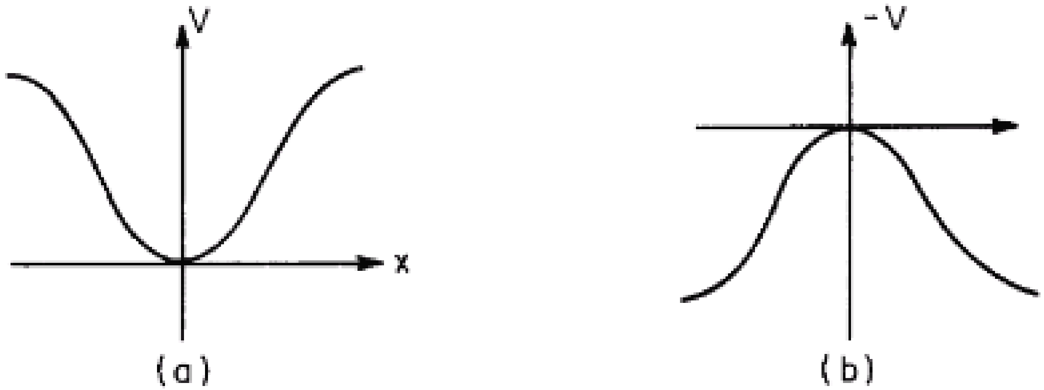


Figure 4

Acknowledgements

Thanks Csaba for giving me this instanton question. I finally make my mind to read Coleman's book. Thanks Yang and Yong-Hui for useful discussions, and Flip for introducing me useful review articles. Finally, even though you do not know me at all, Mr. Coleman and Mr. 't Hooft, you are amazing.

A Euclidean functional integrals

In this appendix we calculate the correlation function $\langle 0|e^{-HT/\hbar}|0\rangle$ which gives eq. (2.12) usgin Euclidean functional integral. The action we have is

$$S = \int_{-T/2}^{T/2} dt \left[\frac{1}{2} \left(\frac{dx}{dt} \right)^2 + V \right], \tag{A.1}$$

with V given in Fig. **??**. Write $x(t)$ into the classical route $\bar{x}(t)$ (with $\delta S/\delta \bar{x} = 0$) plus the eigenfunctions of the second variational derivative $x_n(t)$ (with $\delta^2 S/\delta \bar{x}^2 = \sum_n \lambda_n x_n$), we have

$$x(t) = \bar{x}(t) + \sum_n c_n x_n(t), \tag{A.2}$$

and

$$\int_{-T/2}^{T/2} dt x_n(t) x_m(t) = \delta_{nm}, \qquad [dx] = \prod_n (2\pi\hbar)^{-\frac{1}{2}} dc_n. \tag{A.3}$$

Use this to calculate the correlation function,

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle = N e^{-S(\bar{x})/\hbar} \prod_n \lambda_n^{-\frac{1}{2}} = N e^{-S(\bar{x})/\hbar} [\det(-\partial_t^2 + V''(\bar{x}))]^{-\frac{1}{2}}. \tag{A.4}$$

Here we ignore the term with higher power of $\hbar$ when doing the perturbation. When the particle stays at the vacuum in the classical limit, $\bar{x} = 0$. The correlation function becomes

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle = N[\det(-\partial_t^2 + \omega^2)]^{-\frac{1}{2}}. \tag{A.5}$$

Here we define $\omega^2 \equiv V''(0)$. One can show that for large T,

$$N[\det(-\partial_t^2 + V''(\bar{x}))]^{-\frac{1}{2}} = \left(\frac{\omega}{\pi\hbar}\right)^{\frac{1}{2}} e^{-\omega T/2}, \tag{A.6}$$

this gives the vacuum correlation function eq. (2.12).

B Conventions for the gauge theory

In this appendix we establish notational conventions for the gauge theory.

B.1 Lie algebra

For two matrices T^a amd T^b in a representation of a Lie group, we can always choose them such that $Tr(T^a T^b) \propto \delta^{ab}$. We define the Cartan inner product as

$$(T^a, T^b) = \delta^{ab}. \tag{B.1}$$

For the $SU(2)$ adjoint representations $T^a = -i\sigma^a/2$, we have

$$(T^a, T^b) = -2Tr(T^a T^b). \tag{B.2}$$

For the $SU(2)$ fundamental representations $\phi_1^T = \frac{i}{\sqrt{2}}(1, 0)$ and $\phi_2^T = \frac{i}{\sqrt{2}}(0, 1)$, we have

$$(\phi^a, \phi^b) = -\frac{1}{2}Tr(\phi^a \phi^b). \tag{B.3}$$

For two $SU(n)$ representations t^a and t^b,

$$Tr(T^a T^b) = -C(r)\delta^{ab}, \tag{B.4}$$

where r denotes the representations. For the fundamental rep, $C(n) = 1/2$. For the adjoint rep, $C(adj) = n$.

B.2 Gauge fields

We define the gauge fields A_μ as

$$A_\mu = g A_\mu^a T^a, \tag{B.5}$$

where g is the gauge coupling. The field strength tensor is defined by

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]. \tag{B.6}$$

Pure gauge field theory is defined by the Euclidean action

$$S = \frac{1}{4g^2}\int d^4x (F_{\mu\nu}, F_{\mu\nu}) \equiv \frac{1}{4g^2}\int (F, F). \tag{B.7}$$

B.3 Gauge transform

A gauge transformation is a function, $h(x)$, from Euclidean space into the gauge group, G

$$h(x) = \exp \lambda^a(x) T^a, \tag{B.8}$$

where λs are arbitrary functions. Under such a transform

$$A_\mu \to h A_\mu h^{-1} + h \partial_\mu h^{-1}, \qquad F_{\mu\nu} \to g F_{\mu\nu} g^{-1}. \tag{B.9}$$

If $F_{\mu\nu}$ vanishes,

$$A_\mu = h \partial_\mu h^{-1}. \tag{B.10}$$

C Some tools for the $U(1)$ problem

Here we derive some tools used in the $U(1)$ problem.

C.1 Chiral Ward identity

We want to study a theory of fermions interacting with c-number gauge fields,

$$S = -i \int d^4 x \bar{\psi} (\not{D} - M) \psi. \tag{C.1}$$

For the Green's function with m local multilinear functions of ψ's, the Green's function are defined by

$$\left\langle \phi^{(1)}(x_1) ... \phi^m(x_m) \right\rangle^A = \frac{\int [d\psi][d\bar{\psi}] e^{-S} \phi^{(1)}(x_1) ... \phi^m(x_m)}{\int [d\psi][d\bar{\psi}] e^{-S}}. \tag{C.2}$$

The superscript A is to remind us that we are working in an external gauge field. When having a chiral transform with $\exp(-i\gamma_5 \delta\alpha)$

$$\delta\psi = -i\gamma_5 \psi \delta\alpha, \qquad \delta\bar{\psi} = -i\bar{\psi}\gamma_5 \delta\alpha, \tag{C.3}$$

we have the Ward identity in Schwinger-Dyson equation as

$$\begin{aligned} \partial^\mu \left\langle j^5_\mu(y) \phi^{(1)}(x_1) ... \phi^m(x_m) \right\rangle^A + \left\langle \bar{\psi} M \gamma_5 \psi(y) \phi^{(1)}(x_1) ... \phi^m(x_m) \right\rangle^A & \\ + \delta^{(4)}(y - x_1) \left\langle \partial \phi^{(1)}(x_1)/\partial\alpha ... \phi^m(x_m) \right\rangle^A & \\ + ... + \delta^{(4)}(y - x_m) \left\langle \phi^{(1)}(x_1) ... \partial\phi^m(x_m)/\partial\alpha \right\rangle^A & \\ = -\frac{iC}{8\pi^2} (F(y), \tilde{F}(y)) \left\langle \phi^{(1)}(x_1) ... \phi^m(x_m) \right\rangle^A & . \end{aligned} \tag{C.4}$$

Where the last term comes from the anomaly, C is defined by the

$$Tr(T^a T^b) = -C \delta^{ab}. \tag{C.5}$$

For N $SU(n)$ fundamental fields, $C = N/2$.

Integrating out the 4-space in eq. (C.4), the first term only has boundary contribution and vanishes since there is no massless fields can give a non-vanishing surface term. On the right we can use

$$\int d^4y(F, \tilde{F}) = 32\pi^2\nu. \tag{C.6}$$

We then have

$$2\left\langle \int d^4y\bar{\psi}M\gamma_5\psi(y)\phi^{(1)}(x_1)...\phi^m(x_m)\right\rangle^A + \frac{\partial}{\partial\alpha}\left\langle\phi^{(1)}(x_1)...\phi^m(x_m)\right\rangle^A = -4iC\nu\left\langle\phi^{(1)}(x_1)...\phi^m(x_m)\right\rangle^A. \tag{C.7}$$

We will use this later.

C.2 The sum rule

Now let us prove eq. (4.19). Using eq. (C.7), when there is no ϕ's, we have

$$-2Ni\nu = 2\left\langle\int d^4y\bar{\psi}M\gamma_5\psi(y)\right\rangle^A = \frac{2\int[d\psi][d\bar{\psi}]e^{-S}\int d^4y\bar{\psi}M\gamma_5\psi}{\int[d\psi][d\bar{\psi}]e^{-S}}. \tag{C.8}$$

Writing the fermion into eigenfunctions of $i(\not{D} - M)$, and using the orthogonal relation for the zero and non-zero modes

$$\int d^4y\psi_r^\dagger\gamma_5\psi_r = 0, \text{ with } \lambda_r \neq 0, \qquad \int \mathrm{d^4y}\psi_s^\dagger\gamma_5\psi_s = \chi_s, \text{ with } \lambda_s = 0, \tag{C.9}$$

we can separate the numerator into the zero-mode part with ψ_s and the non-zero-mode part with ψ_r

$$2\int[d\psi_r][d\bar{\psi}_r]e^{-S_{r\neq s}}\int[d\psi_s][d\bar{\psi}_s]e^{\sum_s\bar{\psi}_s(-iM_s)\psi_s}\sum_s\int d^4y\bar{\psi}_sM_s\gamma_5\psi_s. \tag{C.10}$$

The pseudoscalar integral only exists for zero-modes. The first integral gives the usual determinant term $\prod_{r\neq s}(\lambda_r - iM)$. Using eq. (4.18) and $\bar{\psi}_s(-iM_s)\psi_s = \bar{\psi}_s(-iM_s)\gamma^5\gamma^5\psi_s = chi_s\bar{\psi}_s(-iM_s)\gamma^5\psi_s$, we can write the last two integrals into

$$\int[d\psi_s][d\bar{\psi}_s]e^{\sum_s\chi_s\bar{\psi}_s(-iM_s\gamma^5)\psi_s}\sum_s\int d^4y\bar{\psi}_sM_s\gamma_5\psi_s = \sum_s\frac{iM_s}{\chi_s}\frac{\partial}{\partial M_s}\int[d\psi_s][d\bar{\psi}_s]e^{\sum_s\chi_s\bar{\psi}_s(-iM_s\gamma^5)\psi_s},$$
$$= \sum_s\frac{iM_s}{\chi_s}\frac{\partial}{\partial M_s}\int[d\psi_s][d\bar{\psi}_s]e^{\sum_s\bar{\psi}_s(-iM_s)\psi_s} = \sum_s\frac{iM_s}{\chi_s}\frac{\partial}{\partial M_s}\prod(-iM_s) = i\left(\sum_s\chi_s^{-1}\right)\prod_s(-iM_s). \tag{C.11}$$

Combine the denominator term together, the green function becomes

$$-2Ni\nu = \frac{2i\left(\sum_s\chi_s^{-1}\right)\prod_s(-iM_s)\prod_{r\neq s}(\lambda_r - iM)}{\prod_s(-iM_s)\prod_{r\neq s}(\lambda_r - iM)} = 2i\sum_s\chi_s^{-1} = 2i(n_+ - n_-). \tag{C.12}$$

For one quak only, we have $N = 1$. This gives eq. (4.19). In the real QCD case, $N = 2$. This gives eq. (4.34).

C.3 $U(1)_A$ and the θ-vacua transform

In this section we want to proof eq. (4.21). Define the denominator-free Green's function

$$\langle\langle \phi^{(1)}(x_1)...\rangle\rangle^A \equiv \int [d\psi][d\bar{\psi}]e^{-S}\phi^{(1)}.... \tag{C.13}$$

We can get the chiral Ward identity in the same way as deriving eq. (C.7). The only difference now is that $M = 0$ for massless quarks, and $C = \frac{1}{2}$. This gives

$$\left[\frac{\partial}{\partial\alpha} + 2i\nu\right] \langle\langle \phi_1(x_1)...\rangle\rangle^A = 0. \tag{C.14}$$

Fourier transforms the n-vacua into θ-vacua, for a given $|\theta\rangle$, the Green's function of the baby QCD can be written as

$$\langle\theta|\phi_1(x_1)...|\theta\rangle = \frac{\int [dA]e^{-S_g}e^{i\nu\theta}\langle\langle\phi^{(1)}(x_1)...\rangle\rangle^A}{\int [dA]e^{-S_g}e^{i\nu\theta}\langle\langle 1\rangle\rangle^A}, \tag{C.15}$$

where S_g is the gauge-field part of the action. By eq. (C.14), we have eq. (4.21)

$$\left[\frac{\partial}{\partial\alpha} + 2i\nu\right] \langle\theta|\phi_1(x_1)...|\theta\rangle^A = 0. \tag{C.16}$$

References

[1] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[2] Sidney Coleman, "Aspects of symmetry", Cambride University press.

[3] Lewis H.Ryder, "Quantum field theory", Cambride University press.

[4] John Terning, "Modern supersymmetry", Oxford science publications.

[5] G. 't Hooft, Phys. Rept. **142**, 357 (1986).

[6] S. Weinberg, Phys. Rev. D **11**, 3583 (1975).

[7] Ta-Pei Cheng and Ling-Fong Li, "Gauge theory of elementary particle physics," Clarendon Press. Oxford.

[8] J. B. Kogut and L. Susskind, Phys. Rev. D **11**, 3594 (1975).

[9] G. 't Hooft, Phys. Rev. Lett. **37**, 8 (1976); Phys. Rev. D **14**, 3432 (1976) D **18**, 2199 (1978).

[10] Edward W. Kolb and Michael S. Turner, "The early universe", Westview press.