NATURAL SCIENCE REVIEW 3 200705 (2026)
nsr.jinr.int

Symmetries of massless QCD

A. Smilga*[1]

[1] *SUBATECH, Université de Nantes, 4 rue Alfred Kastler, BP 20722, Nantes 44307, France*

Abstract

We present a pedagogical review of certain exact theoretical results concerning the physics of an imaginary world where one quark or more are deprived of their masses.

Keywords: chiral symmetry, conformal symmetry, quantum chromodynamics

DOI: 10.54546/NaturalSciRev.200705

Contents

1. Introduction

Quantum chromodynamics (QCD) is the theory of strong interactions. Its Lagrangian includes fundamental quark and gluon fields. It reads

$$\mathcal{L}_{\rm QCD} = -\frac{1}{2g^2}\,{\rm Tr}\,\{F_{\mu\nu}F^{\mu\nu}\} \;+\; \sum_{f=1}^{6} \bar\psi_f (i\not{\mathcal{D}} - m_f)\psi_f \;+\; \frac{\theta}{32\pi^2}\varepsilon^{\mu\nu\alpha\beta}\,{\rm Tr}\,\{F_{\mu\nu}F_{\alpha\beta}\}, \tag{1.1}$$

*Corresponding author e-mail address: andrei.smilga@subatech.in2p3.fr

Received: 17 April 2026; Revised: 20 May 2026; Accepted: 22 May 2026; Published: 8 June 2026

where

$$\hat{F}_{\mu\nu} = \partial_\mu \hat{A}_\nu - \partial_\nu \hat{A}_\mu - i[\hat{A}_\mu, \hat{A}_\nu] = i[\mathcal{D}_\mu, \mathcal{D}_\nu] \tag{1.2}$$

with $\mathcal{D}_\mu = \partial_\mu - i\hat{A}_\mu$ being gluon field densities belonging to the adjoint representation of the color $SU(3)$ gauge group; ψ_f are the quark fields lying in the fundamental representation of $SU(3)$; g is the strong coupling constant; the experimental restriction on the parameter θ in the last term in (1.1) breaking the P and T symmetries is[1] $|\theta| \lesssim 10^{-10}$.

Now, m_f are the quark masses. Three quarks are heavy: their pole masses (the positions of the poles in the perturbative quark propagators) are $m_c \approx 1.35$ GeV, $m_b \approx 4.8$ GeV, and $m_t \approx 170$ GeV. Three other quarks are relatively light: $m_u \approx 3$ MeV, $m_d \approx 5$ MeV, and $m_s \approx 100$ MeV. m_u, m_d, and m_s are essentially less than the characteristic hadron scale $\mu_{\rm hadr} \sim 500$ MeV. Bearing the latter fact in mind, it is theoretically interesting to inquire what would happen if these masses would be absent. What would be the physics of such an imaginary world?

This question is discussed in the present review. It covers the material well known to experts in a concise and hopefully pedagogical way. It was written on the basis of Lectures 12, 13, 14 of the book [1], but some extra discussions have been added.

We will assume in the following that $\theta = 0$. The physics of an imaginary world with nonzero θ is also rather nontrivial, and we dare address an interested reader to Lectures 15 and 16 of Ref. [1] or to the original papers [2, 3].

Obvious symmetries of QCD, massless or not, are the global Poincaré symmetry and the local gauge symmetry:

$$\begin{aligned} \hat{A}_\mu &\to \Omega(x)\hat{A}_\mu\Omega^\dagger(x) - i\,[\partial_\mu\Omega(x)]\,\Omega^\dagger(x), \\ \psi_f(x) &\to \Omega(x)\psi_f(x) \end{aligned} \tag{1.3}$$

with $\Omega(x) \in SU(3)$.

However, the gauge symmetry is actually *not* a symmetry in the same sense as rotational or Lorentz symmetry are. Namely, in the case of gauge symmetry, we are not allowed to consider states which are not invariant under symmetry transformations; it does not act on the Hilbert space of physical states, which are all gauge singlets annihilated by the generators of gauge transformations. Gauge symmetry exhibits itself only in the Lagrangian formulation, involving some extra unphysical variables which can in principle be disposed of. One can say that gauge symmetry is not a symmetry, but rather a convenient way to describe constrained systems.

We will concentrate here on global symmetries specific for *massless* QCD. These are *conformal symmetry* and chiral symmetries.

2. Conformal symmetry and its breaking

The bare Lagrangian (1.1) of QCD with strictly massless quarks involves no scale. That means that the classical action is invariant under the transformations

$$\left\{ \begin{array}{l} x^\mu \to \lambda x^\mu \\ \hat{A}_\mu \to \lambda^{-1}\hat{A}_\mu \\ \psi_f \to \lambda^{-3/2}\psi_f \end{array} \right. . \tag{2.1}$$

[1] It follows from the restrictions on the observed value of the neutron electric dipole moment.

Here we took into account the fact that the gauge field A_μ has the canonical dimension of mass, and the canonical dimension of the quark field is $[\psi] = m^{3/2}$.

For pure gauge theory, a somewhat stronger form of the scaling symmetry (2.1) holds. Consider the action of Yang–Mills theory on a curved four-dimensional manifold

$$S^{\rm YM}_{\rm curved} \;=\; -\frac{1}{2g_0^2}\,{\rm Tr}\int d^4x\sqrt{-\det\|g\|}\;F_{\mu\nu}F_{\alpha\beta}g^{\alpha\mu}g^{\beta\nu}, \tag{2.2}$$

where g_0^2 is the coupling constant, not to be confused with the metric tensor $g_{\mu\nu}(x)$ [2]. Note now that the action (2.2) is invariant under the local conformal transformations

$$\left\{\begin{array}{lcl} g_{\mu\nu}(x) & \to & \lambda^{-2}(x)g_{\mu\nu}(x) \\ g^{\mu\nu}(x) & \to & \lambda^2(x)g^{\mu\nu}(x) \end{array}\right., \tag{2.3}$$

while the fields and coordinates are not transformed. For a flat metric $g_{\mu\nu}(x) = \eta_{\mu\nu}$ and $\lambda(x) = \lambda$, the transformation (2.3) amounts to a homogeneous scale dilatation and is equivalent to (2.1)[3].

As far as Yang–Mills theory is concerned, the symmetry (2.3) has an important consequence[4]:

$$0 = \frac{\delta S^{\rm YM}_{\rm curved}}{\delta\lambda(x)} \;\propto\; g_{\mu\nu}(x)\frac{\delta S^{\rm YM}_{\rm curved}}{\delta g_{\mu\nu}(x)} = -\frac{1}{2}\Theta^\mu_\mu(x)\sqrt{-\det\|g\|}, \tag{2.6}$$

where $\Theta^{\mu\nu}(x)$ is (the symmetric version of) the energy–momentum tensor:

$$\Theta^{\mu\nu} = \frac{2}{g_0^2}\,{\rm Tr}\left[-\hat F^{\mu\rho}\hat F^\nu{}_\rho \;+\; \frac{g^{\mu\nu}}{4}\hat F_{\rho\sigma}\hat F^{\rho\sigma}\right]. \tag{2.7}$$

To derive (2.7), we used the language of Riemannian geometry, but it holds also in flat space with $g_{\mu\nu}(x) = \eta_{\mu\nu}$ we are primarily interested in. Then the divergence $\partial_\mu\Theta^{\mu\nu}$ vanishes — this is the energy and momentum conservation.

[2]We assume that the reader is familiar with the basic notions of Riemannian geometry. Equation (2.2) is an obvious non-Abelian generalization for the action of an electromagnetic field on a curved background (see, e.g., Ref. [4]).

[3]Note in parentheses that the local conformal invariance is not specific to gauge theories. Some other Weyl-invariant theories exist. Einstein gravity is not Weyl-invariant, but a variant of quadratic gravity, the *Weyl gravity* is. Its action reads

$$S^{\rm Weyl} \;\propto\; \int d^4x\, C^2_{\mu\nu\rho\sigma}\sqrt{-\det\|g\|}, \tag{2.4}$$

where the *Weyl tensor* is

$$C_{\mu\nu\rho\sigma} \;=\; R_{\mu\nu\rho\sigma} - \frac{1}{2}(g_{\mu\rho}R_{\nu\sigma} + g_{\nu\sigma}R_{\mu\rho} - g_{\mu\sigma}R_{\nu\rho} - g_{\nu\rho}R_{\mu\sigma}) + \frac{R}{6}(g_{\mu\rho}g_{\nu\sigma} - g_{\nu\rho}g_{\mu\sigma})$$

($R_{\mu\nu}$ is the Ricci tensor and $R = R^\mu_\mu$ is the scalar curvature).

Also the theory of scalar field coupled to gravity in the following way,

$$S \;\propto\; \int d^4x\left(\frac{1}{12}R\phi^2 + \frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi\right)\sqrt{-\det\|g\|}, \tag{2.5}$$

enjoys Weyl invariance.

[4]Cf. Eq. (94.5) in the book [4].

Note that the expression for $\Theta^{\mu\nu}$ derived in the framework of flat canonical formalism by Noether's method does not coincide in form with (2.7). For example, for pure photodynamics with $\mathcal{L} = -F_{\rho\sigma}F^{\rho\sigma}/4$, one obtains

$$(\Theta^{\mu\nu})^{\rm can} = \frac{\delta\mathcal{L}}{\delta(\partial_\mu A_\rho)}\partial^\nu A_\rho - \eta^{\mu\nu}\mathcal{L} = -F^{\mu\rho}\partial^\nu A_\rho + \frac{\eta^{\mu\nu}}{4}F_{\rho\sigma}F^{\rho\sigma}. \tag{2.8}$$

In contrast to Eq. (2.7), this expression is not symmetric under interchange $\mu \leftrightarrow \nu$ (*à propos*, it is not gauge invariant either). We have

$$(\Theta^{\mu\nu})^{\rm can} = (\Theta^{\mu\nu})^{\rm sym} - F^{\mu\rho}\partial_\rho A^\nu = (\Theta^{\mu\nu})^{\rm sym} - \partial_\rho(F^{\mu\rho}A^\nu),$$

where we used the equation of motion $\partial_\rho F^{\mu\rho} = 0$ of pure photodynamics. As $\partial_\mu\partial_\rho(F^{\mu\rho}A^\nu) = 0$ by (anti)symmetry, both $(\Theta^{\mu\nu})^{\rm can}$ and $(\Theta^{\mu\nu})^{\rm sym}$ are conserved: the forms (2.7) and (2.8) are equivalent.

The relation $\Theta^\mu_\mu = 0$ following from (2.6) can be interpreted as a local conservation law of the dilatation current

$$J^\mu_D \;=\; x_\nu\Theta^{\mu\nu}. \tag{2.9}$$

To understand better the meaning of (2.9), let us derive the canonical expression for the dilatation current for a general theory with Lagrangian $\mathcal{L}(\phi_i, \partial_\mu\phi_i)$, depending on some set of bosonic fields $\phi_i(x)$ with canonical dimension 1 and their derivatives. To this end, consider an infinitesimal scale transformation

$$\left\{ \begin{array}{l} \delta\phi_i = \alpha\phi_i \\ \delta x^\mu = -\alpha x^\mu \end{array} \right. \tag{2.10}$$

($\alpha \ll 1$) and rewrite this in a form where only the fields, but not the coordinates, are transformed. We have

$$\begin{array}{rcl} \delta\phi_i &=& \alpha(x^\nu\partial_\nu\phi_i + \phi_i) \\ \stackrel{\alpha=\rm const}{\Longrightarrow} \quad \delta(\partial_\mu\phi_i) &=& \alpha(x^\nu\partial_\nu\partial_\mu\phi_i + 2\partial_\mu\phi_i), \end{array} \tag{2.11}$$

where the first terms in $\delta\phi_i(x)$, $\delta[\partial_\mu\phi_i(x)]$ compensate for the shift of argument and the second terms reflect the canonical dimension 1 of the fields ϕ_i and the canonical dimension 2 of the fields $\partial_\mu\phi_i$. The variation of the Lagrangian under the global transformation (2.11) is

$$\delta\mathcal{L} = \alpha\left\{x^\nu\partial_\nu\mathcal{L} + \sum_i\left[\frac{\delta\mathcal{L}}{\delta\phi_i}\phi_i + 2\frac{\delta\mathcal{L}}{\delta(\partial_\mu\phi_i)}\partial_\mu\phi_i\right]\right\}. \tag{2.12}$$

For a Lagrangian of canonical dimension 4, the second term is just $4\alpha\mathcal{L}$ and the variation (2.12) boils down to a total derivative $\delta\mathcal{L} = \alpha\partial_\nu(x^\nu\mathcal{L}) \stackrel{\rm def}{=} \alpha\partial_\nu f^\nu$. Assume now that the parameter $\alpha(x)$ is not a constant and calculate the canonical Noether current

$$J^\mu \;=\; \frac{\delta\mathcal{L}}{\delta(\partial_\mu\alpha)} - f^\mu. \tag{2.13}$$

The variation of $\partial_\mu \phi_i$ in Eq. (2.11) acquires an extra term:

$$\delta(\partial_\mu \phi_i) = \alpha(x^\nu \partial_\nu \partial_\mu \phi_i + 2\partial_\mu \phi_i) + \partial_\mu \alpha (x^\nu \partial_\nu \phi_i + \phi_i).$$

Then

$$\frac{\delta \mathcal{L}}{\delta(\partial_\mu \alpha)} = \sum_i \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi_i)} (\phi_i + x^\nu \partial_\nu \phi_i)$$

and

$$(J_D^\mu)^{\rm can} = \sum_i \phi_i \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi_i)} + x_\nu (\Theta^{\mu\nu})^{\rm can}, \tag{2.14}$$

where

$$(\Theta^{\mu\nu})^{\rm can} = \sum_i \frac{\delta \mathcal{L}}{\delta(\partial_\mu \phi_i)} \partial^\nu \phi_i - \eta^{\mu\nu} \mathcal{L}$$

is the canonical energy–momentum tensor. For photodynamics, the expression (2.14) differs from $x_\nu(\Theta^{\mu\nu})^{\rm sym}$ by the term

$$\Delta J_D^\mu = (J_D^\mu)^{\rm can} - x_\nu (\Theta^{\mu\nu})^{\rm sym} = -F^{\mu\rho} A_\rho - x_\nu F^{\mu\rho} \partial_\rho A^\nu. \tag{2.15}$$

Taking into account the equations of motion $\partial_\rho F^{\mu\rho} = 0$, we obtain $\Delta J_D^\mu = -\partial_\rho (x_\nu F^{\mu\rho} A^\nu)$, i.e. the canonical dilatation current (2.14) coincides with the current (2.9) up to a term whose divergence is zero. Thus, one can use the current (2.9) instead of (2.14) in all calculations.

We presented a quite explicit demonstration of this fact for photodynamics, and a similar explicit demonstration for non-Abelian theory (I did not find it in the literature) may be of a certain methodic interest. An underlying reason for this equivalence is the local conformal invariance of the action (2.2) in a curved background.

For a theory involving also massless quarks, an extra term

$$\left(\Theta^\mu_\mu\right)^q = -3i \sum_f \bar\psi_f \not{D} \psi_f \tag{2.16}$$

appears in the trace. It is not zero off-shell, but it is still zero for fields satisfying the classical equations of motion $\not{D}\psi_f = 0$. For massive quarks, scale invariance is lost and $\Theta^\mu_\mu \sim \sum_f m_f \bar\psi_f \psi_f \neq 0$.

Up to now, we were only discussing the classical theory. What happens in the quantum case? One may ask whether conformal symmetry is still present and, in particular, whether the trace of the quantum *operator* corresponding to the classical expression (2.7) for the energy–momentum tensor still vanishes.

First of all, note that the neglection of terms which vanish due to the classical equations of motion is quite justified. The counterpart of the latter in quantum theory is the Heisenberg operator equations of motion which tell us that all the matrix elements of operators like $(i\not{D} - m)\psi$ between the physical states vanish. Still, the answer to the question above is *negative* due to the phenomenon of *dimensional transmutation.*

As is well known, the coupling constants in QED or in QCD *run* are not constant — they depend on a characteristic energy of a given process. In QCD, they go down with the energy:

this is what is called *asymptotic freedom*. In leading logarithmic approximation, the energy dependence of $\alpha_s(E) = g^2(E)/4\pi$ has the form

$$\alpha_s(E) \;=\; \frac{2\pi}{b_0 \ln \frac{E}{\Lambda_{\rm QCD}}}\,, \tag{2.17}$$

where

$$b_0 \;=\; \frac{11N_c - 2N_f}{3} \tag{2.18}$$

(N_c being the number of colors and N_f — the number of light quark flavors) and $\Lambda_{\rm QCD}$ (rather than running $g(E)$ or $\alpha_s(E)$) is the real fundamental constant of quantum chromodynamics. Experiment gives the value $\Lambda_{\rm QCD} \approx 200$ MeV.

According to (2.17), $\alpha_s(\Lambda_{\rm QCD})$ become infinite, but this perturbative formula is not valid when α_s becomes large. One should rather say that at some energy scale less than 1 GeV, the strong interaction becomes really strong so that perturbative calculations are not possible anymore and the physics cannot be described in terms of quarks and gluons. At low energies, the latter are hidden (*confined*) in the colorless hadron states, the lowest such states having the mass of the order of $\Lambda_{\rm QCD}$ multplied by a numerical factor. Even though we cannot prove it theoretically and the corresponding Millennium Prize is not awarded yet, we see that in experiment. What is important for us in this review is the fact that conformal symmetry of pure Yang–Mills theory or of massless QCD is there at the classical level, but this symmetry is broken by quantum effects, it is *anomalous*.

We derived the existence of fundamental intrinsic energy scale in QCD based on the perturbative formula (2.17), which could be derived using operator approach. But a more clear understanding of this phenomenon may be achieved using path integral language. To attribute meaning to path integral symbol, one has to *regularize* it — to replace an infinite-dimensional integral by a finite-dimensional one. The most natural way to do so is to put the theory on a *lattice* of finite size[5]. Such lattice has a scale a — the distance between two adjacent nodes. The corresponding energy scale a^{-1} has the meaning of *ultraviolet regulator* Λ_{UV}. It is an artificial unphysical scale introduced in theory by hand. But it is this ultraviolet regulator which together with the bare coupling constant g_0 entering the QCD Lagrangian (1.1) determines quite physical and important scale $\Lambda_{\rm QCD}$. It follows from Eq. (2.17) that, in the leading logarithmic order,

$$\Lambda_{\rm QCD} = \Lambda_{UV} \exp\left\{-\frac{8\pi^2}{b_0 g_0^2}\right\}. \tag{2.19}$$

And this is what is called dimensional transmutation.

The relation (2.19) takes into account only leading logarithms. An improved version of this relation that includes all perturbative corrections reads[6]

$$\Lambda_{\rm QCD} = \Lambda_{UV} \exp\left\{\int_{g_0^2}^{\infty} \frac{dt}{\beta(t)}\right\}. \tag{2.20}$$

[5]Technical details are given in Sec. 6.

[6]This result follows from the solution of the functional *renormalization group* equations. This issue is elucidated in many textbooks (e.g., in Chapter 12 of the book [5] or in Lecture 9 of the book [1]), where we are addressing the reader.

The *Gell-Mann–Low function* $\beta(t)$ is defined as the logarithmic derivative of the running coupling constant:

$$\beta[g^2(E)] = \frac{dg^2(E)}{d\ \ln E}. \tag{2.21}$$

The first terms of the expansion of $\beta(t)$ are

$$\beta(t) = -\frac{b_0 t^2}{8\pi^2} - \frac{b_1 t^3}{128\pi^4} + \cdots \tag{2.22}$$

with [6]

$$b_1 = \frac{34}{3}N_c^2 - \left(\frac{13}{3}N_c - \frac{1}{N_c}\right) N_f.$$

The absence of scale invariance at the quantum level means that the trace $\hat{\Theta}^\mu_\mu$ of the Heisenberg operator describing the energy–momentum tensor is not zero anymore. Accurately defining what the operator $\hat{\Theta}^{\mu\nu}$ actually means, one can derive:

$$\hat{\Theta}^\mu_\mu \ = \ \frac{\beta(g^2)}{2g^4}\,\mathrm{Tr}\,\{\hat{F}_{\mu\nu}\hat{F}^{\mu\nu}\}. \tag{2.23}$$

This is the renowned *conformal anomaly*. It means that the classically conserved dilatation current (2.9) is no longer conserved in the full quantum theory. Conformal symmetry is broken explicitly by quantum effects. Equation (2.23) is an operator equality, i.e. all matrix elements of the operators on the left- and right-hand sides, taken between some physical states, coincide.

We will derive Eq. (2.23) with path integral methods. Consider a regularized Euclidean Yang–Mills path integral depending on an ultraviolet regulator Λ_0 and a bare coupling constant g_0^2. The scale transformation affects only the regulator: $\delta\Lambda_{UV} = \alpha\Lambda_{UV}$. Due to Eq. (2.20), this brings about the same change of all physical mass scales $\sim \Lambda_{\mathrm{QCD}}$ as the shift of the bare coupling constant $\delta g_0^2 = -\alpha\beta(g_0^2)$, with Λ_{UV} kept fixed. The corresponding modification of the integrand

$$\exp\{-S^E\} = \exp\left\{-(1/2g_0^2)\int \mathrm{Tr}\,\{\hat{F}_{\mu\nu}\hat{F}_{\mu\nu}\}d^4x\right\}$$

in the path integral is then described by the shift

$$\delta S^E = \frac{\alpha\beta(g_0^2)}{g_0^2}S^E. \tag{2.24}$$

On the other hand, the variation of the Minkowski action is

$$\delta S^M/\delta\alpha = \int(\partial_\mu J^\mu_D)d^4x = \int \Theta^\mu_\mu d^4x. \tag{2.25}$$

Comparing this with the Minkowski counterpart of Eq. (2.24), we derive[7]

$$\int \hat{\Theta}^\mu_\mu d^4x = \frac{\beta(g_0^2)}{2g_0^4}\int \mathrm{Tr}\{\hat{F}_{\mu\nu}\hat{F}^{\mu\nu}\}d^4x. \tag{2.26}$$

[7] A note for pundits: in supersymmetric gauge theories, the most natural definition of the Gell-Mann–Low function entering Eq. (2.26) is such that it involves only the leading term and the higher-order corrections vanish. This makes the expression for the conformal anomaly (2.23) similar to that of the *chiral* anomaly, Eq. (3.5), which we will discuss below. (The coefficient in Eq. (3.5) has a geometric interpretation and does not involve a series in g^2.) This is not accidental. In supersymmetric theories, the chiral current and the dilatation current belong to the same supermultiplet and their anomalies are intimately related to each other.

Strictly speaking, this alone does not guarantee that the corresponding *integrands* also coincide. To show this, one has to find the variation of the path integral under a *local* scale transformation (2.3). We should imagine a lattice whose spacing (in physical units) depends on x: $a(x) = a_0[1 + \alpha(x)]$. Somewhat heuristically, we might say that a theory with constant coupling g_0^2 defined on the lattice with x-dependent spacing describes the same physics at distances that are much larger than $a(x)$ as the theory with x-dependent coupling constant $g^2(x) = g_0^2 + \alpha(x)\beta(g_0^2)$ defined on the lattice with constant spacing. Thereby, the variation of the action is

$$\delta S^E \;=\; \frac{\beta(g_0^2)}{2g_0^4} \int \alpha(x)\,\mathrm{Tr}\,\{\hat{F}_{\mu\nu}\hat{F}_{\mu\nu}\} d^4x. \tag{2.27}$$

Varying it over $\alpha(x)$, we derive Eq. (2.23).

The anomaly (2.23) could also be derived by operator methods, which we will not tackle here, but illustrate in the following section.

3. Anomalous chiral symmetry

Consider Yang–Mills theory with just one massless quark. The term $i\bar{\psi}\not{\mathcal{D}}\psi$ in the Lagrangian is invariant under global chiral transformations[8]

$$\delta\psi = -i\alpha\gamma^5\psi, \quad \delta\bar{\psi} = -i\alpha\bar{\psi}\gamma^5. \tag{3.3}$$

The corresponding canonical Noether current,

$$j^{\mu 5} = \bar{\psi}\gamma^\mu\gamma^5\psi, \tag{3.4}$$

is conserved upon applying the equations of motion.

The Lagrangian is also invariant under the symmetry $\delta\psi = i\beta\psi$, but it is just a special case of gauge symmetry which, as was mentioned, is not, in effect, a symmetry and is not a subject of this review.

An important fact is that the symmetry (3.3) exists only in the classical case. The full quantum path integral is *not* invariant under the transformations (3.3). Like it was also the case with the conformal symmetry, this symmetry breaking due to quantum effects can be expressed as an operator identity involving an anomalous divergence:

$$\partial_\mu j^{\mu 5} = -\frac{1}{16\pi^2}\varepsilon^{\alpha\beta\mu\nu}\,\mathrm{Tr}\,\{\hat{F}_{\alpha\beta}\hat{F}_{\mu\nu}\}. \tag{3.5}$$

There are many ways to derive and understand this relation. Historically, it was first derived by purely diagrammatic methods. Here we will concentrate on two other ways which are

[8]We will use in the following the spinor representation where

$$\gamma^0 = \begin{pmatrix} 0 & \mathbb{1} \\ \mathbb{1} & 0 \end{pmatrix}, \quad \gamma^j = \begin{pmatrix} 0 & -\sigma_j \\ \sigma_j & 0 \end{pmatrix}, \tag{3.1}$$

and

$$\gamma^5 \stackrel{\text{def}}{=} i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{pmatrix}. \tag{3.2}$$

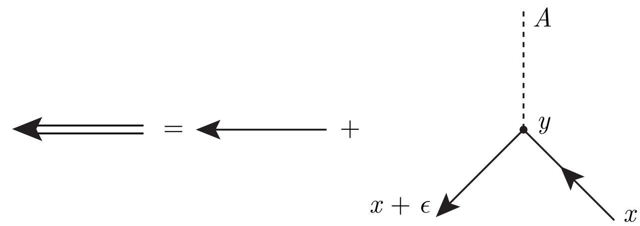

Figure 1. Fermion propagator in an external field.

more modern and more general. First, we derive (3.5) as an operator equality (we skipped an analogous derivation when discussing the conformal anomaly) and, second, we will show that the path integral measure is actually not invariant under chiral transformations, but is modified in such a way that the relation (3.5) is satisfied.

Let us start with the operator derivation and, for simplicity, consider first the Abelian case. To begin with, we need to define a quantum operator corresponding to the classical axial current (3.4). The problem is that one cannot harmlessly multiply field operators $\psi(x)$ at coinciding points. Indeed, e.g., the vacuum average $\langle \psi(x)\bar{\psi}(0)\rangle_0$ behaves as $i\not{x}/(2\pi^2 x^4)$ at small x, and the limit $x \to 0$ is singular.

Following Schwinger, we *define* the axial current operator of QED as

$$j^{\mu 5} = \lim_{\epsilon \to 0} \bar{\psi}(x+\epsilon)\gamma^\mu \gamma^5 \exp\left\{ i \int_x^{x+\epsilon} A_\nu(y) dy^\nu \right\} \psi(x). \tag{3.6}$$

The factor $\exp\left\{ i \int_x^{x+\epsilon} A_\nu(y) dy^\nu \right\}$ is very important and makes the expression gauge-invariant in spite of the different arguments of $\bar{\psi}(x+\epsilon)$ and $\psi(x)$. The limit $\epsilon \to 0$ is taken, assuming averaging over the directions of ϵ^μ (otherwise the current (3.6) would not be a Lorentz vector). Expanding the exponential up to terms $O(\epsilon)$, differentiating the whole expression (3.6) with respect to x^μ, and making use of the operator equations of motion $\not{\mathcal{D}}\psi = \gamma^\mu(\partial_\mu - iA_\mu)\psi = 0$, we obtain

$$\begin{aligned} \partial_\mu j^{\mu 5} \ &= \ \lim_{\epsilon \to 0} \bar{\psi}(x+\epsilon)\left[-i\gamma^\mu A_\mu(x+\epsilon) + i\gamma^\mu A_\mu(x) + \right. \\ &\left. + i\epsilon^\nu \gamma^\mu \partial_\mu A_\nu(x) \right] \gamma^5 \psi(x) = \lim_{\epsilon \to 0} i\epsilon^\nu F_{\mu\nu}(x) \bar{\psi}(x+\epsilon)\gamma^\mu\gamma^5\psi(x). \end{aligned} \tag{3.7}$$

Superficially, it seems to be zero in the limit $\epsilon \to 0$. This is not the case, however. Let us average Eq. (3.7) over a state involving a classical background field $A_\mu(x)$. The fermion Green's function is the free Green's function plus the term describing one insertion of the external field plus terms with multiple insertions (see Figure 1). The free Green's function $\sim \not{\epsilon}/\epsilon^4$ does not contribute in our case because the corresponding spinor trace $\mathrm{Tr}\left\{\gamma^\alpha\gamma^\mu\gamma^5\right\}$ is zero in four dimensions. To calculate the graph with one field insertion, it is convenient to choose the *Fock–Schwinger* [7] (alias, fixed point) gauge $(y-x)^\alpha A_\alpha(y) = 0$, so that[9]

$$A_\alpha(y) = -\frac{1}{2}(y-x)^\beta F_{\alpha\beta} + o(y-x). \tag{3.8}$$

[9]The choice (3.8) where the potential vanishes at the position x is convenient because it assures the translational invariance of Green's function. With an arbitrary choice of the fixed point, it would not be invariant, but the invariance would, of course, be restored for all gauge-invariant quantities.

An explicit calculation (see **Problem 1** below) gives

$$\langle \psi(x)\bar{\psi}(x+\epsilon)\rangle_A = -\frac{i\not\epsilon}{2\pi^2\epsilon^4} - \frac{1}{32\pi^2\epsilon^2}F_{\alpha\beta}(\not\epsilon\gamma^\alpha\gamma^\beta + \gamma^\alpha\gamma^\beta\not\epsilon) + \ldots \tag{3.9}$$

Multiplying this by $-i\epsilon^\nu\gamma^\mu\gamma^5 F_{\mu\nu}$ (the extra minus sign comes from the permutation of anticommuting field operators), taking the trace,

$$\mathrm{Tr}\,\{\gamma^5\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta\} = -4i\varepsilon^{\mu\nu\alpha\beta}$$

($\varepsilon^{0123} = 1$) and averaging over the directions $\epsilon^\nu\epsilon_\rho/\epsilon^2 \to \delta^\nu_\rho/4$, we arrive at the result

$$\partial_\mu j^{\mu 5} \;=\; -\frac{1}{16\pi^2}\varepsilon^{\alpha\beta\mu\nu}F_{\alpha\beta}F_{\mu\nu}. \tag{3.10}$$

Note that the terms describing multiple insertions in the Green's function are less singular in ϵ, and their contribution to $\partial_\mu j^{\mu 5}$ vanishes in the limit $\epsilon \to 0$.

Let us briefly discuss the situation in other even[10] dimensions. In two dimensions, the term with one insertion of the external field in the propagator is $O(\epsilon)$ and does not contribute to the anomaly. But the anomaly is still there. It appears due to the leading term $i\not\epsilon/(2\pi\epsilon^2)$ in the expansion of the fermion Green's function: in two dimensions,

$$\mathrm{Tr}\,\{\gamma^\mu\gamma^\nu\gamma^5\} = 2\varepsilon^{\mu\nu} \;\neq 0 \tag{3.11}$$

with the convention

$$\gamma^0 = \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\; \gamma^1 = i\sigma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\; \gamma^5 = \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{3.12}$$

$\varepsilon^{01} = -\varepsilon_{01} = 1$. Equation (3.7) then gives[11]

$$\partial_\mu j^{\mu 5} = -\frac{1}{2\pi}\varepsilon^{\mu\nu}F_{\mu\nu}. \tag{3.13}$$

In six dimensions, both the leading term and the term linear in the external field do not contribute because the corresponding spinor traces vanish. The anomaly is there due to the term with two field insertions. For $d = 8$, the term with three insertions works, etc. In general, one can derive

$$\partial_\mu j^{\mu 5} = \pm\frac{2}{n!(4\pi)^n}\,\varepsilon^{\mu_1\cdots\mu_{2n}}F_{\mu_1\mu_2}\cdots F_{\mu_{2n-1}\mu_{2n}}, \tag{3.14}$$

where $d = 2n$.

Let us return to four-dimensional QCD, however. We again define the axial current as in Eq. (3.6), and the whole derivation can be repeated. The only difference is that, along with the

[10] If the dimension D is odd, there is no γ^5 matrix (the product $\prod_{\mu=0}^{D-1}\gamma^\mu$ is proportional to the unit matrix), no axial symmetry and no anomaly.

[11] The negative sign in the right-hand side depends on the chosen conventions. With other definitions of γ^5 and/or of the covariant derivative, the sign may be positive — see, e.g., Eq. (19.18) in Ref. [5].

terms coming from differentiating the fermion fields in (3.6), we also have to take into account the term $\sim O(\epsilon)$ in the expansion of the exponential. That will give the contribution

$$\Delta \left(\partial_\mu j^{\mu 5}\right)^{\text{non-Ab}} = \lim_{\epsilon \to 0} \bar\psi(x+\epsilon)\gamma^\mu \gamma^5 \left[\hat A_\mu(x) \int_x^{x+\epsilon} \hat A_\nu(y) dy^\nu - \right.$$
$$\left. - \left(\int_x^{x+\epsilon} \hat A_\nu(y) dy^\nu \right) \hat A_\mu(x) \right] \psi(x) = \lim_{\epsilon \to 0} \epsilon^\nu \bar\psi(x+\epsilon)\gamma^\mu\gamma^5 [\hat A_\mu(x), \hat A_\nu(x)]\psi(x), \quad (3.15)$$

which, being combined with the other terms, just gives the non-Abelian field strength tensor. Also the relation (3.9) still holds with the full non-Abelian $F_{\alpha\beta}$, which is a part of the magics of the Fock–Schwinger gauge method. The only distinction is that the quark Green's function is now a nontrivial matrix not only in spinor, but also in color indices. Thereby, the result (3.5) is reproduced.

Let us now derive the anomaly relation (3.5) with path integral methods [8]. As we have seen, the anomaly appears due to the necessity to regularize theory in the ultraviolet. The politically most correct approach would be to study a path integral regularized by a lattice. However, an accurate definition of the fermionic action and fermionic path integral on the lattice is not so easy. We will address this last issue in Sec. 6 and, for the time being, use instead the more habitual *finite mode* regularization.

Consider an Euclidean path integral for the partition function for QCD with one massless quark flavor. The fermionic part of the integral is

$$\int \prod_x d\bar\psi(x) d\psi(x) \exp\left\{ i \int d^4x \bar\psi \not{\mathcal{D}}^E \psi \right\} \quad (3.16)$$

with Hermitian[12] $\mathcal{D}^E$.

An important remark is that the fermion variables ψ and $\bar\psi$ are *not* assumed here to be related as $\bar\psi = \psi^\dagger \gamma^0$, as was the case for the Heisenberg field operators in the Minkowski space. Neither they are assumed to be complex conjugate to one another. They represent *independent* Grassmann variables in the path integral (3.16) giving the determinant of the Dirac operator, $\det(-i\not{\mathcal{D}}^E)$.

Let us assume that the theory is somehow regularized in the infrared so that the spectrum of the operator $\not{\mathcal{D}}^E$ is discrete. A proper way to think about the relationship between $\bar\psi(x)$ and $\psi(x)$ is to expand $\psi(x)$, $\bar\psi(x)$ over the eigenstates of $\not{\mathcal{D}}^E$:

$$\begin{cases} \psi(x) &= \sum_k c_k u_k(x) \\ \bar\psi(x) &= \sum_k \bar c_k u_k^\dagger(x) \end{cases} \quad (3.19)$$

where $\{u_k(x)\}$ is a complete basis in the corresponding Hilbert space. It is conveniently chosen as a set of eigenfunctions of the Dirac operator: $\not{\mathcal{D}}^E u_k(x) = \lambda_k u_k(x)$; c_k and $\bar c_k$ are independent

[12] We use the convention where Euclidean γ matrices are anti-Hermitian, satisfying

$$\gamma^E_\mu \gamma^E_\nu + \gamma^E_\nu \gamma^E_\mu = -2\delta_{\mu\nu}. \quad (3.17)$$

The matrix γ^5, defined as $\gamma^5 = \gamma^E_0\gamma^E_1\gamma^E_2\gamma^E_3$, is the same as in the Minkowski space so that

$$\mathrm{Tr}\,\{\gamma^5 \gamma^E_\mu \gamma^E_\nu \gamma^E_\alpha \gamma^E_\beta\} = -4\varepsilon_{\mu\nu\alpha\beta}, \qquad \varepsilon_{0123} = 1. \quad (3.18)$$

The Euclidean fermion Lagrangian is $\mathcal{L}_E = -i\bar\psi \not{\mathcal{D}}^E \psi$.

Grassmann integration variables. Then

$$\prod_x d\bar{\psi}(x)d\psi(x) \stackrel{\text{def}}{=} \prod_k d\bar{c}_k dc_k. \tag{3.20}$$

Suppose that the field variables are transformed by an infinitesimal global chiral transformation (3.3). Now, $\psi' = \psi + \delta\psi$ and $\bar{\psi}' = \bar{\psi} + \delta\bar{\psi}$ can again be expanded in the series (3.19). The new expansion coefficients are related to the old ones:

$$c_k' = c_k - i\alpha R_{km} c_m, \qquad \bar{c}_k' = \bar{c}_k - i\alpha \bar{c}_m R_{mk} \tag{3.21}$$

with

$$R_{km} = \int d^4x \; u_k^\dagger(x)\gamma^5 u_m(x). \tag{3.22}$$

The point is that the transformation (3.21) has a nonzero Jacobian. We have[13]

$$\prod_k d\bar{c}_k{}' dc_k' = J^{-2} \prod_k d\bar{c}_k dc_k, \tag{3.24}$$

where

$$J = \det\left(1 - i\alpha A\right) \approx \exp\left\{-i\alpha \sum_k \hat{A}_{kk}\right\}. \tag{3.25}$$

Or otherwise

$$\ln J = -i\alpha \int d^4x \sum_k u_k^\dagger(x)\gamma^5 u_k(x) + o(\alpha). \tag{3.26}$$

One's first (wrong!) impression might be that $\int d^4x \sum_k u_k^\dagger(x)\gamma^5 u_k(x)$ is just zero. Indeed, $\{\not{\mathcal{D}}^E, \gamma^5\} = 0$ and the function $u_k' = \gamma^5 u_k$ is also an eigenfunction of the Dirac operator with eigenvalue $-\lambda_k$. *If* $\lambda_k \neq 0$, $u_k(x)$ and $\gamma^5 u_k(x)$ thereby represent *different* eigenfunctions and the integral $\int d^4x \; u_k^\dagger(x)\gamma^5 u_k(x)$ vanishes.

A nonzero value of (3.26) is due to the fact that, for intricate enough *topologically nontrivial* gauge fields, the spectrum of the Dirac operator involves some number of exact zero modes for which $\gamma^5 u_0(x) = \pm u_0(x)$ (depending on whether the modes are right-handed or left-handed) and their contribution in the sum (3.26) is responsible for the whole effect. A famous theorem of Atiyah and Singer [9] dictates

$$\int d^4x \sum_k u_k^\dagger(x)\gamma^5 u_k(x) = n_R^{(0)} - n_L^{(0)} = q, \tag{3.27}$$

[13] The Jacobian appears in the denominator due to the Berezin rule

$$\int \prod_k d\bar{c}_k dc_k \, e^{-\bar{c}Ac} = \det(A) \tag{3.23}$$

and the requirement for the integral to stay invariant under a variable change.

where $n^{(0)}_{R,L}$ is the number of the right-handed (left-handed) zero modes and q is the *Pontryagin index* of the gauge field configuration[14]:

$$q = \frac{1}{16\pi^2} \int d^4x \; \mathrm{Tr}\, \{\hat{F}_{\mu\nu} \hat{\tilde{F}}_{\mu\nu}\}, \tag{3.31}$$

where $\hat{\tilde{F}}_{\mu\nu} = \frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}\hat{F}_{\alpha\beta}$.

How to prove this theorem? The mathematical proof due to Atiyah and Singer is complicated and is not familiar to the author. We present here a proof, which is adapted to the problem in interest and relies solely on the standard apparatus of quantum mechnics. Consider first a little more simple case where the gauge field is Abelian.

As all nonzero modes contribute zero to the sum, we are allowed to consider instead of (3.27) the sum

$$\int d^4x \sum_k u_k^\dagger(x)\gamma^5 u_k(x)\, e^{-\lambda_k^2/M^2}. \tag{3.32}$$

The finite parameter M brings about the finite mode regularization mentioned above: the contribution of modes with large λ_k^2 is suppressed. To calculate the sum, consider a quantum mechanical problem with the matrix Hamiltonian[15]

$$\hat{H} = (\not{\mathcal{D}})^2 = -[\gamma^E_\mu(\hat{p}_\mu - A_\mu)]^2, \quad \hat{p}_\mu = -i\partial_\mu. \tag{3.33}$$

The sum (3.32) is nothing but $\int d^4x\, \mathrm{Tr}\,\{\gamma^5 \mathcal{K}(x,x;1/M^2)\}$, where $\mathcal{K}$ is the evolution operator of our quantum mechanical system with imaginary time $\beta = 1/M^2$. It is also a matrix. [Note that the phase space of our system is 8-dimensional, involving the coordinates x_μ, p_μ, and the evolution occurs in some unphysical "fifth" time.] The trace is done over the spinor and color indices.

The Euclidean evolution operator can be expressed as a path integral

$$\mathcal{K}(x,x;\beta) \;=\; \int \exp\left\{ i\int_0^\beta \left[p_\mu \frac{dx_\mu}{d\tau} + iH(p,x)\right] d\tau \right\} \prod_\tau \frac{dp_\mu(\tau) dx_\mu(\tau)}{2\pi} \tag{3.34}$$

---

[14] The most familiar configurations with nonzero Pontryagin index $q=1$ are BPST instantons [10]:

$$A^a_\mu = \frac{2\eta^a_{\mu\nu}x_\nu}{x^2+\rho^2}, \tag{3.28}$$

where $\eta^a_{\mu\nu}$ are the so-called 't Hooft symbols

$$\eta^a_{44} = 0, \quad \eta^a_{jk} = \varepsilon_{ajk}, \quad \eta^a_{4j} = -\delta_{aj}, \quad \eta^a_{j4} = \delta_{aj}. \tag{3.29}$$

The field strength correponding to the gauge potential (3.28) is

$$F^a_{\mu\nu} = -\frac{4\rho^2\eta^a_{\mu\nu}}{(x^2+\rho^2)^2}. \tag{3.30}$$

[15] We follow the tradition and use the notation $\hat{X}$ for two rather different purposes. It may mark the color matrix nature of $\hat{X} = X^a t^a$ as for $\hat{A}_\mu$ or its operator nature as for $\hat{H}$ and $\hat{p}_\mu$ in Eq. (3.33). Hopefully the reader will not be confused by that.

with the classical (but still spinor matrix) $H(p,x)$. It runs over the trajectories $x_\mu(\tau)$ and $p_\mu(\tau)$ satisfying the periodic boundary conditions

$$x_\mu(\beta) = x_\mu(0) = x, \qquad p_\mu(\beta) = p_\mu(0). \tag{3.35}$$

Assume now that M is very large, so that β is very small. Bearing in mind (3.35), $x_\mu(\tau)$ and $p_\mu(\tau)$ stay practically constant, the term $\propto dx_\mu/d\tau$ in the exponent can be neglected and the functional integral reduces to an ordinary finite-dimensional integral over momenta:

$$\mathcal{K}(x,x;\beta) \stackrel{\text{small}\,\beta}{\approx} \int \frac{d^4p}{(2\pi)^4}\, e^{-\beta H(p,x)}. \tag{3.36}$$

We derive

$$\int d^4x\, \mathrm{Tr}\,\{\gamma^5 \mathcal{K}(x,x;1/M^2)\} \approx \int \frac{d^4x d^4p}{(2\pi)^4}\, \mathrm{Tr}\left\{\gamma^5 e^{-(\not{\mathcal{D}})^2/M^2}\right\}. \tag{3.37}$$

Using the convention (3.17), giving

$$(\not{\mathcal{D}})^2 \;=\; -\mathcal{D}^2 - (i/2)\gamma^E_\mu \gamma^E_\nu F_{\mu\nu} = (p_\mu - A_\mu)^2 - (i/2)\gamma^E_\mu \gamma^E_\nu F_{\mu\nu},$$

evaluating the trace using (3.18) and shifting the variable of the integration $p_\mu - A_\mu \to p_\mu$, the integral can be easily calculated to leading order in $1/M^2$. The result is the Abelian version of q:

$$q_{\text{Ab}} = \frac{1}{16\pi^2} \int d^4x\, F_{\mu\nu}\tilde{F}_{\mu\nu}. \tag{3.38}$$

The same reasoning works also in the non-Abelian case with the only complication that $(p_\mu - A_\mu)^2$ is now replaced by a color matrix $(p_\mu - A^a_\mu t^a)^2$, and one cannot just shift the momentum variable in the integral.

But actually one can. To see that, consider the integral $\int_{-\infty}^\infty \exp\{-(p-\hat{A})^2\}$, where $\hat{A}$ is some p-independent matrix, and expand it in $\hat{A}$. We obtain

$$\int_{-\infty}^\infty dp\, \exp\{-(p-\hat{A})^2\} = \int_{-\infty}^\infty dp\, e^{-p^2} - \hat{A}\int_{-\infty}^\infty dp \left[\frac{\partial}{\partial p} e^{-p^2}\right] + \frac{1}{2}\hat{A}^2 \int_{-\infty}^\infty dp \left[\frac{\partial^2}{\partial^2 p} e^{-p^2}\right] + \cdots$$

All the terms besides the first one vanish, and the integral does not depend on $\hat{A}$. Thus, the index theorem (3.27) is proven[16].

Substituting (3.25) into (3.24), with bearing in mind (3.27) and (3.31), we now see that the change of the measure under a chiral transformation can be represented as a shift of the effective Euclidean action

$$\delta S^E = -\frac{i\alpha}{16\pi^2}\varepsilon_{\alpha\beta\mu\nu} \int d^4x\, \mathrm{Tr}\,\{\hat{F}_{\alpha\beta}\hat{F}_{\mu\nu}\}. \tag{3.39}$$

[16] By the way, the Atiyah–Singer index for the Dirac operator is identical to the *Witten index* of a certain supersymmetric quantum-mechanical system (with an opposite sign for the manifolds of dimension $4k+2$), but we will not delve into further details here addressing an interested reader to the original papers [11] and to Chapter 15 of the book [12]. The method for deriving Eq. (3.27) presented above coincides in fact with the known derivation of an integral representation for the Witten index due to Cecotti and Girardello [13].

The shift of the Minkowski action is given by the same expression without the prefactor i. In other words, a global chiral transfomation is equivalent to leaving the fermionic fields intact, but shifting instead the parameter θ in the original theory (1.1): $\theta \to \theta - 2\alpha$.

Just like in the case of the conformal anomaly, we have studied up to now only the change of the measure under global symmetry transformations. In the chiral symmetry case, it is not too difficult to find out how the measure changes under local transformations with x-dependent parameters $\alpha(x)$. We have

$$\ln J[\alpha(x)] = -i \int d^4x \ \alpha(x) \ \sum_k u_k^\dagger(x)\, \gamma^5 u_k(x) + o(\alpha). \tag{3.40}$$

Repeating all the steps of the above derivation, we find[17]

$$\lim_{M\to\infty} \sum_k u_k^\dagger(x)\, \gamma^5 u_k(x)\, e^{-\lambda_k^2/M^2} = \frac{1}{32\pi^2} \varepsilon_{\alpha\beta\mu\nu}\, \mathrm{Tr}\, \{\hat F_{\alpha\beta} \hat F_{\mu\nu}\}(x),$$

which gives

$$\delta S^E = -\frac{i}{16\pi^2} \varepsilon_{\alpha\beta\mu\nu} \int d^4x \alpha(x)\, \mathrm{Tr}\, \{\hat F_{\alpha\beta} \hat F_{\mu\nu}\}(x). \tag{3.41}$$

Varying the corresponding shift of the Minkowski action with respect to $\alpha(x)$, we obtain the anomalous divergence of the axial current in accordance with (3.5).

The last comment is that, in real QCD with several light quarks, each flavor contributes on equal footing to the anomaly of the singlet axial current $j^{\mu 5(\mathrm{singl})} = \sum_f \bar\psi_f \gamma^\mu \gamma^5 \psi_f$, and the result (3.5) is just multiplied by N_f.

Problem 1. Using the Fock–Schwinger gauge (3.8), derive the expression (3.9) for the fermion Green's function.

Solution. With the gauge choice (3.8), the Green's function depends only on the difference of coordinates $x - (x+\epsilon) = -\epsilon$. We can set $x = 0$ and write

$$\langle \psi(0) \bar\psi(\epsilon) \rangle_A = G_0(-\epsilon) + \int d^4y\, G_0(y-\epsilon) i\gamma^\alpha A_\alpha(y) G_0(-y), \tag{3.42}$$

where $G_0(u)$ is the free fermion Green's function:

$$G_0(x) = \langle \psi(x) \bar\psi(0) \rangle_0 \ = \int \frac{d^4p}{(2\pi)^4}\, e^{-ipx} G(p).$$

Substituting here $A_\alpha(y) = -\dfrac{1}{2} y^\beta F_{\alpha\beta}$ and going over into momentum space, we obtain

$$G_A(p) = \ \frac{i\not p}{p^2} - \frac{F_{\alpha\beta}}{2} \frac{\not p}{p^2} \gamma^\alpha \left(\frac{\partial}{\partial p_\beta} \frac{\not p}{p^2} \right) \ = \frac{i\not p}{p^2} - \frac{F_{\alpha\beta}}{4p^4} (\not p \gamma^\alpha \gamma^\beta + \gamma^\alpha \gamma^\beta \not p). \tag{3.43}$$

Its Fourier image gives Eq. (3.9).

[17] Note that this is now a *local* quantity not directly related to the global properties of a gauge field configuration like a nonzero net topological charge q and the presence of fermion zero modes.

Problem 2. For the instanton field configuration (3.28), show that the function

$$u_0(x) = \frac{\rho}{\pi(x^2+\rho^2)^{3/2}} \begin{pmatrix} \varepsilon_{i\alpha} \\ 0 \end{pmatrix} \tag{3.44}$$

($i = 1,2$ is the color and $\alpha = 1,2$ is the spinor index) represents an exact normalized right-handed zero mode of the Dirac equation [14]).

Solution. It is convenient to express the Euclidean γ matrices as

$$\gamma_\mu^E = \begin{pmatrix} 0 & -\sigma_\mu^\dagger \\ \sigma_\mu & 0 \end{pmatrix} \tag{3.45}$$

with $\sigma_\mu = (i, \boldsymbol{\sigma})$. It is easy to check that such matrices satisfy the properties (3.17) and (3.18).

The matrices σ_μ satisfy the relations

$$\begin{aligned} \sigma_\mu^\dagger \sigma_\nu + \sigma_\nu^\dagger \sigma_\mu &= \sigma_\mu \sigma_\nu^\dagger + \sigma_\nu \sigma_\mu^\dagger = 2\delta_{\mu\nu}, \\ \sigma_\mu^\dagger \sigma_\nu - \sigma_\nu^\dagger \sigma_\mu &= 2i\eta^a_{\mu\nu}\sigma^a, \\ \sigma_\mu \sigma_\nu^\dagger - \sigma_\nu \sigma_\mu^\dagger &= 2i\bar\eta^a_{\mu\nu}\sigma^a. \end{aligned} \tag{3.46}$$

Note also the corollaries

$$\sigma_\mu^\dagger \sigma_\nu \sigma_\mu^\dagger = -2\sigma_\nu^\dagger, \quad \sigma_\mu \sigma_\nu^\dagger \sigma_\mu = -2\sigma_\nu.$$

Let us first convince ourselves that only the right-handed solutions [satisfying $\gamma^5 u = u$ with $\gamma^5 = \text{diag}\,(\mathbb{1}, -\mathbb{1})$, as defined in Eq. (3.2)] to the equation $\not{\mathcal{D}}^E u = 0$ are admissible. Indeed, a left-handed solution should satisfy

$$\not{\mathcal{D}}^E u_L = \gamma_\mu^E \mathcal{D}_\mu \begin{pmatrix} 0 \\ u_L \end{pmatrix} = 0$$

giving $\sigma_\mu^\dagger \mathcal{D}_\mu u_L = 0$. Act upon this with the operator $\sigma_\nu \mathcal{D}_\nu$. We obtain

$$\begin{aligned} \sigma_\nu \mathcal{D}_\nu \sigma_\mu^\dagger \mathcal{D}_\mu &= (\mathcal{D}_\mu)^2 + \frac{1}{4}[\sigma_\nu\sigma_\mu^\dagger - \sigma_\mu\sigma_\nu^\dagger][\mathcal{D}_\nu, \mathcal{D}_\mu] = \\ &= (\mathcal{D}_\mu)^2 - (\bar\eta^a_{\nu\mu}\sigma^a)(\eta^b_{\nu\mu}\sigma^b)\frac{\rho^2}{(x^2+\rho^2)^2} = (\mathcal{D}_\mu)^2, \end{aligned} \tag{3.47}$$

where we have substituted the instanton field strength $[\mathcal{D}_\nu, \mathcal{D}_\mu] = -iF_{\nu\mu}$ from Eq. (3.30) and used the property $\bar\eta^a_{\nu\mu}\eta^b_{\nu\mu} = 0$. But the operator $-(\mathcal{D}_\mu)^2 = (i\mathcal{D}_\mu)^2$ is a sum of squares of Hermitian operators and is thus positive definite. The equation $\mathcal{D}^2 u_L = 0$ and hence the equation $\sigma_\mu^\dagger \mathcal{D}_\mu u_L = 0$ have no solutions.

For the right-handed spinors, the zero mode equation has the form

$$(\sigma_\mu)_{\alpha\beta}\left[\partial_\mu - \frac{i\eta^a_{\mu\nu}x_\nu\sigma^a}{x^2+\rho^2}\right]_{ij} u^R_{j\beta}(x) = 0. \tag{3.48}$$

Using the relations (3.46) and the property $\sigma_\mu^T = -\sigma_2\sigma_\mu^\dagger\sigma_2$, we can rewrite this equation as

$$\left[\partial_\mu - \frac{x_\nu}{2(x^2+\rho^2)}(\sigma_\mu^\dagger\sigma_\nu - \sigma_\nu^\dagger\sigma_\mu)\right] u\sigma_2\sigma_\mu^\dagger = 0. \tag{3.49}$$

The natural ansatz $u_{i\alpha} = (\sigma_2)_{i\alpha} G(x^2) = -i\varepsilon_{i\alpha} G(x^2)$ satisfies the equation provided

$$2G'(x^2) + \frac{3G(x^2)}{x^2+\rho^2} = 0 \quad \text{and hence} \quad G(x^2) = \frac{C}{(x^2+\rho^2)^{3/2}}.$$

The particular value of C in (3.44) normalizes the solution to one:

$$\int d^4x \, u^\dagger_{i\alpha} u_{i\alpha} = 1.$$

Due to the index theorem (3.27), and since left-handed solutions are absent, the instanton field does not admit other zero mode solutions.

4. Nonsinglet chiral symmetry and its spontaneous breaking

A theory with several flavors of massless quarks enjoys, besides the singlet axial symmetry $\delta\psi_f = -i\alpha\gamma^5\psi_f$ (which, as we have seen, it *does* not actually enjoy), a set of flavor-nonsinglet symmetries:

$$\delta\psi_f = -i\alpha_a [t^a \psi]_f \tag{4.1}$$

and

$$\delta\psi_f = -i\beta_a \gamma^5 [t^a \psi]_f, \tag{4.2}$$

where t^a are the generators of the flavor $SU(N_f)$ group normalized in a standard way,

$$\mathrm{Tr}\,\{t^a t^b\} \;=\; \frac{1}{2}\delta^{ab}.$$

The symmetry (4.1) is the ordinary isotopic symmetry. It is still present even if the quarks are endowed with a mass (of the same magnitude for all flavors). The symmetry (4.2) holds only in massless theory. The corresponding Noether currents are

$$(j^\mu)^a = \bar\psi t^a \gamma^\mu \psi, \qquad (j^{\mu 5})^a = \bar\psi t^a \gamma^\mu \gamma^5 \psi. \tag{4.3}$$

They are not anomalous, i.e. they are duly conserved not only at the classical level, but also in full quantum theory. To describe a finite element of the symmetry group, it is convenient to define left-handed and right-handed quark fields:

$$\psi_{L,R} = \frac{1}{2}(1 \mp \gamma^5)\psi, \qquad \bar\psi_{L,R} = \frac{1}{2}\bar\psi(1 \pm \gamma^5). \tag{4.4}$$

One can easily see that the Lagrangian of massless QCD is invariant under the transformations

$$\psi_L \to V_L \psi_L, \qquad \psi_R \to V_R \psi_R, \tag{4.5}$$

where V_L and V_R are two different $U(N_f)$ matrices. The singlet axial transformations with $V_L = V_R^* = e^{i\phi}$ are anomalous by the same token as in the theory with only one quark flavor. Therefore, the true fermionic symmetry group of massless QCD is

$$\mathcal{G} = SU_L(N_f) \times SU_R(N_f) \times U_V(1). \tag{4.6}$$

A fundamental *experimental fact* is that the symmetry (4.6) is actually *spontaneously broken*, which means that the vacuum state is not invariant under the action of the group $\mathcal{G}$. The symmetry $\mathcal{G}$ is, however, not broken completely. The vacuum is still invariant under the transformations with $V_L = V_R$, generated by the vector isotopic currents[18]. Thereby, the pattern of breaking is

$$SU_L(N_f) \times SU_R(N_f) \to SU_V(N_f). \tag{4.7}$$

Spontaneous breaking of the axial symmetry shows up in the appearance of nonzero vacuum expectation values,

$$\Sigma^{fg} \;=\; \langle \psi_L^f \bar{\psi}_R^g \rangle_0 \tag{4.8}$$

(the *quark condensate* matrix). Nonbreaking of the vector symmetry implies that the matrix order parameter (4.8) can be brought into the form

$$\Sigma^{fg} = \frac{1}{2} \Sigma \, \delta^{fg} \tag{4.9}$$

by the group transformations (4.5). This means that a generic condensate matrix Σ^{fg} is a unitary $SU(N_f)$ matrix multiplied by Σ.

In general, Σ could be any complex number. It can be made real by a global $U_A(1)$ rotation which, according to Eq. (3.39) (with the factor N_f), amounts to a shift of the vacuum angle θ. In other words, in the theory with quarks, the physics does not depend on the parameter θ and the phase of Σ separately, but only on their combination $\theta - N_f \arg(\Sigma)$. The earlier mentioned fact that the experimental value of θ is very small actually refers to *this* particular combination. It is convenient then to choose Σ real and positive and $\theta = 0$. From experiment[19] we know that $\Sigma \approx (250 \text{ MeV})^3$ with about 30% uncertainty.

By *Nambu–Goldstone theorem* [17], if a global continuous symmetry is broken spontaneously, purely massless particles called *Goldstone bosons* appear in the spectrum. The simplest example for this is the theory of a complex scalar field $\phi(x)$ with a "Mexican hat" potential $\lambda(\bar{\phi}\phi - \mu^2)^2$. The Lagrangian is invariant with respect to global phase rotations $\phi(x) \to e^{i\alpha}\phi(x)$, but the vacuum state (say, the state with $\langle \phi \rangle_{\text{vac}} = \mu$) is characterized by a particular value of the phase playing the role of an order parameter. The point is that a field configuration involving small fluctuations of the phase $\phi(x) = \mu e^{i\alpha(x)}$, $\alpha(x) \ll 1$, still has zero potential energy, while the kinetic energy is proportional to $\int d^4x (\partial_\mu \alpha)^2$. Quantizing this effective Lagrangian gives massless particles.

In general, the number of Goldstone particles is equal to the dimension of the vector space formed by the generators of the gauge group which are "broken", i.e. act nontrivially on the vacuum state. By construction, for any such generator J, a Goldstone branch in the spectrum $|\alpha_J\rangle_p$ exists, so that the matrix elements $\langle 0|J|\alpha_J\rangle_p$ are not zero. In our case, the breaking (4.7) is associated with the appearance of $N_f^2 - 1$ Goldstone particles [$N_f^2 - 1$ is the dimension of the original group $\mathcal{G}$ minus the dimension of the residual group $U_V(N_f)$]. As it is the axial symmetry which is broken, the Goldstone particles are pseudoscalars.

[18] An exact *Vafa–Witten theorem* [19] that the vector symmetry cannot be broken spontaneously in QCD will be proven in Sec. 7.

[19] From the Gell-Mann–Oakes–Renner relation (5.7), from QCD sum rules [15], and from lattice numerical simulations [16].

An analogy with an ordinary iron bar (with one of its magnetic domains) is very instructive here. The Hamiltonian of a ferromagnet is rotationally invariant. Spontaneous magnetization signals the spontaneous breaking of the rotational invariance $SO(3)$ down to $SO(2)$. The direction $\mathbf{n}$ of the magnetization $\langle \mathbf{M} \rangle = M_0 \mathbf{n}$, with $\mathbf{n}^2 = 1$, is arbitrary. Rotating the reference frame, we can choose, say, $\mathbf{n} = (0, 0, 1)$ [cf. Eq. (4.9)]. Fluctuations of the vector $\mathbf{n}$ in space and time are described by $3 - 1 = 2$ parameters and correspond to magnon massless excitations.

5. Effective chiral Lagrangian

It is a fundamental and important fact that spontaneous breaking of a continuous symmetry not only leads to massless Goldstone particles, but also fixes the *interactions* of the latter at low energies. To see this, let us first recall that the Goldstone field describes fluctuations of the order parameter: $\Sigma^{fg} \to \Sigma^{fg}(x) = \frac{1}{2}\Sigma U(x)$ with $U(x) \in SU(N_f)$. It is convenient to express $U(x)$ as an exponential

$$U(x) = \exp\left\{\frac{2i\phi^a(x)t^a}{F_\pi}\right\}, \tag{5.1}$$

where $\phi^a(x)$ are the physical meson fields [so that $\phi^a(x) = 0$ corresponds to the vacuum (4.9)] and F_π is a constant carrying dimension of mass.

In massless QCD, Goldstone particles are massless whereas all other states in the physical spectrum have nonzero mass. Therefore, we are in the *Born–Oppenheimer* situation: there are two distinct energy scales and one can write down an effective Lagrangian depending only on *slow* Goldstone fields, with the *fast* degrees of freedom corresponding to all other particles being integrated out[20].

To fix the exact form of this Lagrangian, note that the transformations (4.5) are realized at the level of the effective Lagrangian as $U \to V_L U V_R^\dagger$. Any scalar function depending on U and invariant under this symmetry is also a function of $U^\dagger U = 1$, i.e. it is just a constant. There is only one invariant structure involving two derivatives:

$$\mathcal{L}^{(2)}_{\text{eff}} = \frac{F_\pi^2}{4}\,\text{Tr}\,\left\{\partial_\mu U \partial^\mu U^\dagger\right\}. \tag{5.2}$$

Take $N_f = 2$. The perturbative expansion of Eq. (5.2) in powers of ϕ reads

$$\mathcal{L}^{(2)}_{\text{eff}} = \frac{1}{2}(\partial_\mu \phi^a)^2 + \frac{1}{6F_\pi^2}\left[(\phi^a \partial_\mu \phi^a)^2 - (\phi^a\phi^a)(\partial_\mu \phi^b)^2\right] + \cdots \tag{5.3}$$

Also for $N_f \geqslant 3$, we have, on top of the standard kinetic term, a quartic term involving two derivatives, with somewhat more complicated group structure. We see that the symmetry dictates rather specific interactions between the Goldstone bosons. They do not interact at the S wave level, which means that the amplitude of their scattering vanishes at zero momenta, but the strength of interaction grows rapidly with energy.

[20] In the pioneer paper [18], Born and Oppenheimer found the spectrum of the hydrogen molecule by solving first the Schrödinger equation involving fast electron degrees of freedom and analysing then the effective Hamiltonian that depended on the positions and momenta of heavy protons.

Equation (5.2) describes the effective chiral Lagrangian to leading order. The first corrections involve 4 derivatives and there are 3 different invariant functions:

$$\mathcal{L}_{\text{eff}}^{(4)} = L_1 \text{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger\right\}^2 + L_2 \text{Tr}\left\{\partial_\mu U \partial_\nu U^\dagger\right\} \text{Tr}\left\{\partial^\mu U \partial^\nu U^\dagger\right\} + L_3 \text{Tr}\left\{\partial_\mu U \partial^\mu U^\dagger \partial_\nu U \partial^\nu U^\dagger\right\} \qquad (5.4)$$

(only 2 linearly independent structures are left for $N_f = 2$).

The relevant Born–Oppenheimer expansion parameter is

$$\kappa_{\text{chir}} \sim p^{\text{char}}/F_\pi.$$

When $\kappa_{\text{chir}} \sim 1$ (in practice, one should rather take $\kappa_{\text{chir}} \sim 2\pi$), the Born–Oppenheimer approximation, as well as the whole effective Lagrangian approach, breaks down, and non-Goldstone degrees of freedom become important. The physical meaning of F_π is thus clarified. It characterizes the gap in the spectrum and sets a scale below which massive degrees of freedom can be disregarded.

Let us discuss the real world now. The Lagrangian of real QCD (1.1) is not invariant under the axial symmetry transformations just because quarks have nonzero masses. The symmetry (4.6) is still very much relevant for QCD because *some* of the quarks happen to be very light. This is especially so for u and d quarks whose masses — $m_u \approx 3$ MeV and $m_d \approx 5$ MeV — are much smaller than the characteristic hadron scale $\mu_{\text{hadr}} \approx .5$ GeV: the symmetry (4.6) is almost there!

Spontaneous breaking of an exact $SU_L(2) \times SU_R(2)$ symmetry would lead to the existence of three strictly massless Goldstone bosons. As the symmetry is not quite exact, the Goldstone particles have a small mass. However, their mass M goes to zero in the chiral limit $m_{u,d} \to 0$. Indeed, trading the mass term

$$-m_u \bar{u}u - m_d \bar{d}d = m_u(u_L \bar{u}_R + u_R \bar{u}_L) + m_d(d_L \bar{d}_R + d_R \bar{d}_L) \qquad (5.5)$$

in the QCD Lagrangian for the contribution[21]

$$\mathcal{L}_{\text{eff}}^{(m)} = \Sigma \, \text{Re}\left[\text{Tr}\left\{\mathcal{M}U\right\}\right] \qquad (5.6)$$

[$\mathcal{M}$ is the quark mass matrix which is chosen here in the form $\mathcal{M} = \text{diag}(m_u, m_d)$ with real m_u, m_d; do not confond $\mathcal{M}$ with M !] in the effective chiral Lagrangian and expanding Eq. (5.6) in ϕ^a, we obtain the *Gell-Mann–Oakes–Renner relation* [20]:

$$F_\pi^2 M^2 = (m_u + m_d)\Sigma + O(m_q^2). \qquad (5.7)$$

These light pseudo-Goldstone particles are well known to experimentalists. They are nothing but the pions. Experimentally, $F_\pi \approx 93$ MeV. The constant F_π appears also in the matrix element $\langle \text{vac}\, | A_\mu^+ | \pi \rangle = i\sqrt{2} F_\pi p_\mu^\pi$ of the axial current $A_\mu^+ = \bar{d}\gamma_\mu \gamma^5 u$ and determines the charged pion decay rate. The gap between the pseudo-Goldstone sector and the massive sector in QCD is of order of the mass of ρ meson, $M_\rho \approx 8F_\pi$.

In the real world, there is also a third relatively light quark — the strange quark. Its mass, $m_s \approx 100$ MeV, is still small enough for the symmetry (4.6) to make sense. Thus, QCD enjoys

[21] Equation (5.6) is the leading chiral-noninvariant contribution in $\mathcal{L}_{\text{eff}}$. Also terms of higher order in $\mathcal{M}$, as well as terms of first order in $\mathcal{M}$ but involving derivatives of U, are allowed.

an approximate $SU_L(3)\times SU_R(3)$ symmetry which is broken spontaneously with the appearance of the quark condensate (4.9) and also explicitly due to nonzero quark masses. As the latter are relatively small, one can build up a Born–Oppenheimer expansion (alias, *chiral perturbation theory* [21]) over the small parameters $p^{\rm char}/\mu_{\rm hadr}$ and $m_q/\mu_{\rm hadr}$. The spectrum of QCD includes 8 light pseudo-Goldstone mesons, the well-known pseudoscalar octet (π, K, η).

Exploiting the symmetry (4.6) allows one to obtain many nontrivial predictions for the properties of these mesons, but this, as well as many other wonderful achievements in describing the physics of hadrons through QCD, is beyond the scope of our review.

6. QCD on Euclidean lattice

To attribute meaning to the path integral symbol, one should regularize it — replace an infinite-dimensional integral by a finite-dimensional one. The most natural way to do so is to make space-time discrete and replace a continuum of points on $\mathbb{R}^4$ by a set of nodes of hypercubic lattice. A finite distance a between adjacent nodes provides ultraviolet regularization. A finite size L of this lattice provides infrared regularization. The properties of physically interesting continuum field theory can be inferred by studying the limit $a \to 0$, $L \to \infty$. It is all possible to do in Euclidean space where the measure in the path integral $\sim \exp\{-S^E\}$ is positive definite and the lattice integral is well defined. In Minkowski space with the measure $\sim \exp\{iS^M\}$, a rigorous definition of path integral that would satisfy mathematicians is not available though some practical numerical calculations are still possible [22]. In our review, we will stay in Euclidean space, however.

We describe first how to define path integral on Euclidean lattice for pure Yang–Mills theory while keeping gauge invariance. Then we include fermions and tackle the problem of chiral symmetry in the lattice version of QCD. Our report will necessarily be rather succinct. A reader who wants to learn it in more detail is invited to consult an excellent book [23].

6.1. *Pure gauge theory*

Thus, we introduce a four-dimensional hypercubic lattice. The nodes of the lattice are labelled by integer 4-vectors n_α. Let us define on each *link* of the lattice a unitary matrix $U_{n,n+e_\mu} = U^\dagger_{n+e_\mu,n} \in SU(N)$, where $e_1 = (1,0,0,0)$, etc. For each *plaquette* (a two-dimensional face) of the lattice labelled by its corner n_α with all the components less or equal than the corresponding components for other three corners, and by two directional vectors e_μ, e_ν (see Figure 2), we define [24]

$$W_{n,\mu\nu} = \frac{1}{N}\,\mathrm{Re\,Tr}\,\left\{U_{n,n+e_\nu}U_{n+e_\nu,n+e_\mu+e_\nu}U_{n+e_\mu+e_\nu,n+e_\mu}U_{n+e_\mu,n}\right\}. \tag{6.1}$$

(For $SU(2)$, we need not take the real part in Eq. (6.1) as the trace of any $SU(2)$ matrix is real.) Consider the integral

$$Z_{\rm lat} = \int \prod_{\rm links} \mathcal{D}U_{\rm link}\, \exp\left\{-\frac{2N}{g^2}\sum_{\rm plaq}(1 - W_{\rm plaq})\right\}, \tag{6.2}$$

where $\mathcal{D}U$ is the Haar measure on the group. We are going to show that the exponent in Eq. (6.2) is a correct discrete approximation to the Euclidean action of continuum Yang–Mills

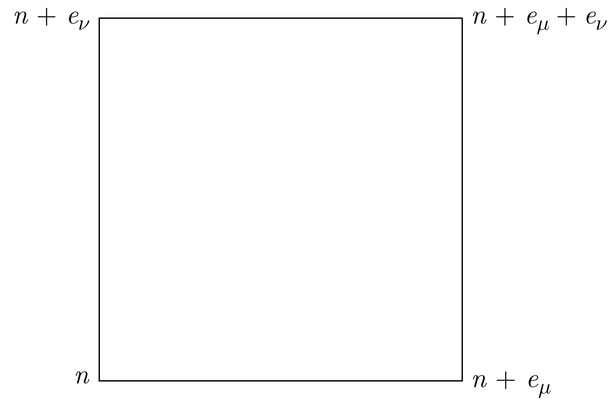

Figure 2. An elementary plaquette.

theory and hence the integral (6.2) is a reasonable discrete approximation of the path integral

$$Z_{\text{cont}} = \int \prod_{\mathbf{x},\tau} dA^a_\mu(\mathbf{x},\tau) \, \exp\left\{-\frac{1}{2g^2}\int d^4x \, \text{Tr}\,\{\hat{F}_{\mu\nu}\hat{F}_{\mu\nu}\}\right\}. \tag{6.3}$$

We assume the actual spacing of our Euclidean lattice a be small compared to the characteristic scale Λ_{QCD} of the theory. Let us associate $U_{n,n+e_\mu}$ with the parallel transporter $\exp\{-ia\hat{A}_\mu e_\mu\}$ along the corresponding link where $\hat{A}_\mu$ may be defined at the point n. Then the quantity under the trace in Eq. (6.1) is nothing but the *Wilson loop* on the closed contour C around the plaquette[22]:

$$W_{n,\mu\nu} = \frac{1}{N}\,\text{Re}\,\text{Tr}\left[P\,\exp\left\{-i\oint_C \hat{A}_\mu(\xi)\xi_\mu\right\}\right]. \tag{6.4}$$

For small contours, this reduces to

$$W_{n,\mu\nu} = 1 - \frac{a^4}{2N}\,\text{Tr}\,\{\hat{F}_{\mu\nu}\hat{F}_{\mu\nu}\} + O(a^6). \tag{6.5}$$

Thus, we see that, indeed,

$$S_{\text{lat}} \to \frac{2N}{g^2}a^4 \sum_n \frac{1}{2}\sum_{\mu\nu} \frac{1}{2N}\,\text{Tr}\,\{\hat{F}^2_{\mu\nu}\} + O(a^6) \to \int d^4x \, \frac{1}{2g^2}\,\text{Tr}\,\{\hat{F}^2_{\mu\nu}\}$$

in the continuum limit (the identity $\sum_{\text{plaq}} = \sum_n \sum_{\mu>\nu} = \frac{1}{2}\sum_n\sum_{\mu\nu}$ was used).

The integral (6.2) is invariant under transformations

$$U_{n,n+e_\mu} \to \Omega_n U_{n,n+e_\mu}\Omega^\dagger_{n+e_\mu}, \tag{6.6}$$

where $\{\Omega_n\}$ is a set of unitary matrices defined in the nodes of the lattice. This is the lattice version of gauge invariance (1.3) of continuum theory.

The integral in (6.2) involves a discrete but still infinite number of variables. To make it finite, our lattice should have finite size in both spatial and Euclidean time directions. In

[22]The symbol P means a path-ordered product.

practice, it is convenient to implement this by imposing periodic boundary conditions on the matrix link variables:

$$U_{n+L_\alpha e_\alpha, n+e_\mu+L_\alpha e_\alpha} \;=\; U_{n,n+e_\mu}, \tag{6.7}$$

where the set of integers $L_\alpha = \{L_x, L_y, L_z, L_\tau\}$ characterizes the size of the lattice (the number of nodes in the corresponding direction); no summation over α is assumed. With the conditions (6.7), the theory is effectively defined on a discrete four-dimensional torus.

Toroidal boundary conditions are easier to handle than boundary conditions with rigid walls: finite size effects, which are present in all practical numerical calculations are less prominent for the torus. For these boundary effects not to be important, the physical length of the torus aL_μ should be larger than the characteristic scale of the theory, while a should be kept much smaller. In practical calculations, "much" usually means at most 4–5 times. Calculating the path integral (6.2) numerically on an asymmetric lattice with $L_x = L_y = L_z \gg L_\tau$ in the regime where boundary effects due to finite Euclidean time extension are important, while effects due to a finite spatial size are not, one finds the partition function of the system at finite temperature $T = 1/(aL_\tau)$. The properties of the vacuum wave functional are explored in calculations on large symmetric lattices.

Problem 3. Derive (6.5).

Solution. Consider a plaquette in the plane $(\mu\nu)$. We have

$$\begin{aligned}
&P\,\exp\left\{-i\oint_{\square_{\mu\nu}} \hat{A}_\rho(\xi)\xi_\rho\right\} \approx \left[1 - ia\hat{A}_\mu(n) - \frac{a^2}{2}\hat{A}_\mu(n)^2\right] \times \\
&\times \left[1 - ia\hat{A}_\nu(n+e_\mu) - \frac{a^2}{2}\hat{A}_\nu(n+e_\mu)^2\right]\left[1 + ia\hat{A}_\mu(n+e_\mu+e_\nu) - \frac{a^2}{2}\hat{A}_\mu(n+e_\mu+e_\nu)^2\right] \times \\
&\times \left[1 + ia\hat{A}_\nu(n+e_\nu) - \frac{a^2}{2}\hat{A}_\nu(n+e_\nu)^2\right] = \\
&= 1 - ia^2(\partial_\mu\hat{A}_\nu - \partial_\nu\hat{A}_\mu)(n) - a^2[\hat{A}_\mu, \hat{A}_\nu](n) + O(a^3) = 1 - ia^2\hat{F}_{\mu\nu}(n) + O(a^3).
\end{aligned} \tag{6.8}$$

Plugging that in (6.4), we arrive at (6.5).

6.2. Including quarks: first try

Quantum chromodynamics involves quarks and, to define the path integral in QCD, we need to handle fermionic fields on the lattice. Let us define to this end Grassmann variables $\bar\psi_n$, ψ_n in the nodes of the lattice for each quark flavor (color and Lorentz indices are not displayed). As a first and natural guess, let us write the extra term in the action as follows:

$$\begin{aligned}
S^{\text{ferm.lat.}} = -\frac{ia^3}{2}\sum_{n,\mu}[\bar\psi_n U_{n,n+e_\mu}\gamma^E_\mu \psi_{n+e_\mu} - \bar\psi_n U_{n,n-e_\mu}\gamma^E_\mu\psi_{n-e_\mu}] + \\
+ ma^4\sum_n \bar\psi_n\psi_n.
\end{aligned} \tag{6.9}$$

This expression is invariant under gauge transformations (6.6) complemented with $\psi_n \to \Omega_n\psi_n$. We see that the action (6.9) reproduces the action

$$S^E_{\text{ferm}} = \int d^4x[-i\bar\psi\gamma^E_\mu(\partial_\mu - i\hat{A}_\mu)\,\psi + m\bar\psi\psi] \tag{6.10}$$

in the continuum limit. Indeed, for free fermions

$$-\frac{ia^3}{2}\sum_{n,\mu}\bar{\psi}_n\gamma^E_\mu[\psi_{n+e_\mu}-\psi_{n-e_\mu}] \to -i\int d^4x\,\bar{\psi}\gamma^E_\mu\partial_\mu\psi.$$

Expanding $U_{n,n+e_\mu} \equiv 1 - ia\hat{A}_\mu + O(a^2)$, we also restore the interaction term, and the last term in Eq. (6.9) turns into the continuum mass term.

Equation (6.9) is called the "naïve lattice fermion action", and we have to say that if the reader was convinced by the above reasoning that, in the continuum limit, it goes over to Eq. (6.10), s/he was naïve, too. Our implicit assumption was that the fermion fields ψ_n depend on the lattice node n in a smooth manner, so that the finite difference $\psi_{n+e_\mu} - \psi_{n-e_\mu}$ goes over to the continuum derivative. It turns out, however, that fermion field configurations which behave as $\psi_n \sim (-1)^{n_1}$ or $\psi_n \sim (-1)^{n_2+n_4}$ and change significantly at the microscopic lattice scale are equally important for the lattice path integral. After carefully performing continuum limit, these wildly oscillating modes give rise to 15 extra light fermion species with the same mass, the so-called *doublers*.

To understand it, consider first free massless fermions. Let

$$(\partial^+_\mu\psi)_n = \frac{1}{a}[\psi_{n+e_\mu}-\psi_n], \quad (\partial^-_\mu\psi)_n = \frac{1}{a}[\psi_n-\psi_{n-e_\mu}] \tag{6.11}$$

be the forward and backward lattice derivative operators. The naïve lattice counterpart of the free massless Dirac operator, which enters the path integral (6.20), is[23]

$$\mathfrak{D}^0_{\text{free}} = -\frac{i}{2}\gamma^E_\mu(\partial^+_\mu + \partial^-_\mu). \tag{6.13}$$

The eigenfunctions of $\mathfrak{D}^0_{\text{free}}$ are characterized by the Euclidean 4-momentum p_μ,

$$\psi^{(p)}_n = C_p e^{iap_\mu n_\mu}, \tag{6.14}$$

where C_p is a constant Grassmann bispinor. The eigenvalue equation $\mathfrak{D}^0_{\text{free}}\psi^{(p)}_n = -i\lambda_p\psi^{(p)}_n$ implies

$$\left[\frac{1}{a}\gamma_\mu\,\sin(ap_\mu)\right]C_p \;=\; -i\lambda_p C_p \tag{6.15}$$

with

$$\lambda_p = \pm\frac{1}{a}\sqrt{\sum_\mu \sin^2(ap_\mu)}. \tag{6.16}$$

[23] In the continuum limit, $\mathfrak{D}^0_{\text{free}}$ goes over to $-i\gamma^E_\mu\partial_\mu$. It is anti-Hermitian, and its eigenvalues are purely imaginary. The full continuum action of a massive Euclidean fermion in gauge background was written in (6.10):

$$S^E_{\text{ferm}} = \int d^4x\,\bar{\psi}(m - i\not{D})\psi. \tag{6.12}$$

When $ap_\mu \ll 1$, we reproduce the continuum massless fermions with the spectrum $\lambda_p = \pm\sqrt{p_\mu^2}$. Each eigenvalue (6.16) is doubly degenerate due to two possible polarizations. The eigenfunctions with negative λ_p are obtained from the ones with positive λ_p by multiplication by γ^5. Indeed, $\mathfrak{D}^0_{\rm free}$ anticommutes with γ^5. Let $\mathfrak{D}^0_{\rm free} u = \lambda u$. Then $\mathfrak{D}^0_{\rm free}(\gamma^5 u) = -\gamma^5 \mathfrak{D}^0_{\rm free} u = -\lambda\gamma^5 u$.

Note, however, that the lattice Dirac equation (6.15) has an *additional* discrete symmetry[24] $(Z_2)^4$:

$$\hat{Q}_\mu \psi_n = (-1)^{n_\mu} \gamma^E_\mu \gamma^5 \psi_n, \qquad \hat{Q}_\mu \bar\psi_n = (-1)^{n_\mu+1} \bar\psi_n \gamma^5 \gamma^E_\mu. \tag{6.17}$$

For any eigenfunction $\psi_n^{(p)}$, the function $\hat{Q}_\mu \psi_n^{(p)}$ is also an eigenfunction of $\mathfrak{D}^0_{\rm free}$ with the same eigenvalue λ_p. The operators $\hat{Q}_\mu$ commute with $\mathfrak{D}^0_{\rm free}$ and anticommute with each other:

$$\hat{Q}_\mu \hat{Q}_\mu + \hat{Q}_\mu \hat{Q}_\mu \ = \ 2\delta_{\mu\nu}.$$

The functions

$$\psi_n^{(p)}, \quad \hat{Q}_\mu \psi_n^{(p)}, \quad \hat{Q}_{[\mu}\hat{Q}_{\nu]}\psi_n^{(p)}, \quad \hat{Q}_{[\mu}\hat{Q}_\nu Q_{\lambda]}\psi_n^{(p)}, \quad \hat{Q}_{[\mu}\hat{Q}_\nu Q_\lambda Q_{\rho]}\psi_n^{(p)} \tag{6.18}$$

form a degenerate 16-plet.

In the free case, each eigenstate of the naïve Dirac operator is not just 16-fold, but 32-fold degenerate due to polarizations. In the interacting case (on a generic gauge field background), polarization is not a good quantum number, but the 16-fold degeneracy of all eigenstates of $\mathfrak{D}$ still holds. The naïve lattice Dirac operator in Eq. (6.9), which can be written in the form $\mathfrak{D}^0 = -\frac{i}{2}\gamma^E_\mu(\mathcal{D}^+_\mu + \mathcal{D}^-_\mu)$, where

$$\begin{aligned}
(\mathcal{D}^+_\mu \psi)_n &= \frac{1}{a}\left(\psi_{n+e_\mu} U_{n,n+e_\mu} - \psi_n\right), \\
(\mathcal{D}^-_\mu \psi)_n &= \frac{1}{a}\left(\psi_n - \psi_{n-e_\mu} U_{n,n-e_\mu}\right)
\end{aligned} \tag{6.19}$$

are the covariant lattice forward and backward derivatives, still enjoys the symmetries (6.17).

We see that if an eigenstate ψ_n changes smoothly from node to node, its 15 doublers wildly oscillate on the microscopic lattice spacing scale. We might call these modes "unphysical", but they would not listen to us and contaminate with a vengeance any numerical lattice calculation we might wish to do. Some way to get rid of them should be suggested, otherwise QCD, the theory involving only 6 quarks with different masses, would not be operationally defined.

Simple-minded modifications of this naïve action, which leave only one light fermion for each flavor in the continuum limit, do not respect the chiral invariance of the theory. It is a serious problem, and its solution is not easy. At the end of the day, we will see, however, that all these difficulties can be overcome and a good consistent definition of the QCD path integral exists.

The action (6.9), as well as its sophistications to be discussed soon, are bilinear in $\bar\psi_n$, ψ_n. The fermionic part of the path integral has the form

$$\int \prod_i d\bar\psi_i d\psi_i \, \exp\{-\mathfrak{D}_{ij}\bar\psi_i\psi_j\}, \tag{6.20}$$

[24] See **Problem 4** at the end of this section. One can observe that $\hat{Q}_\mu\bar\psi_n$ and $\hat{Q}_\mu\psi_n$ stay mutually conjugated if $\bar\psi_n$ and ψ_n were. This feature is not in fact necessary, bearing in mind that $\bar\psi_n$ and ψ_n should be treated as independent variables in Euclidean formulation, but it is still convenient.

where $i \equiv (n, \alpha)$, with α marking both the color and Lorentz spinor index, and $\mathfrak{D}_{ij}$ is a matrix turning into the Euclidean Dirac operator $i\not{\mathcal{D}}^E - m$ in the continuum limit. The Grassmann integral gives the determinant, $\det \|\mathfrak{D}\|$. Thus, the full path integral for QCD has the form

$$Z = \int \prod_{\text{links}} \mathcal{D}U_{\text{link}} \prod_f \det \|\mathfrak{D}_f\| \, \exp\left\{-\frac{2N}{g^2}\sum_{\text{plaq}}(1 - W_{\text{plaq}})\right\}, \tag{6.21}$$

where $\prod_f$ runs over all quark flavors.

Numerical calculations of the integrals like (6.21) are technically very difficult: not only one has to do a multidimensional integral, but also the *integrand* becomes very complicated involving the determinant of a large matrix. With modern computers and clever algorithms, such calculations are still possible.

Problem 4. Verify that $\mathfrak{D}^0$ is invariant under the transformations (6.17).

Solution. The naïve lattice Euclidean fermion action reads

$$S^E_{\text{ferm}} = \frac{1}{2a}\sum_{n\mu} \bar\psi_n \gamma^E_\mu (U_{n,n+e_\mu}\psi_{n+e_\mu} - U_{n,n-e_\mu}\psi_{n-e_\mu}). \tag{6.22}$$

Consider the term corresponding to a particular node $n = (0,0,0,0)$ of the lattice and look how the variables ψ_n (call it ψ_0) and ψ_{n+e_μ} (call them ψ_μ) transform under the action of $\hat{Q}_1$. The definition (6.17) gives

$$\begin{aligned} &\hat{Q}_1\psi_{0,2,3,4} = \gamma^E_1\gamma^5\psi_{0,2,3,4}, \qquad &\hat{Q}_1\bar\psi_{0,2,3,4} = -\bar\psi_{0,2,3,4}\gamma^5\gamma^E_1, \\ &\hat{Q}_1\psi_1 = -\gamma^E_1\gamma^5\psi_1, \qquad &\hat{Q}_1\bar\psi_1 = \bar\psi_1\gamma^5\gamma^E_1. \end{aligned}$$

Then $\bar\psi_0\gamma_1\psi_1$ transforms as

$$\bar\psi_0\gamma^E_1\psi_1 \to (-\bar\psi_0\gamma^5\gamma^E_1)\gamma^E_1(-\gamma^E_1\gamma^5\psi_1) = -\bar\psi_0\gamma^5\gamma^E_1\gamma^5\psi_1 = \bar\psi_0\gamma^E_1\psi_1,$$

i.e. it does *not* transform [the property (3.17) was used]. Similarly for the other terms:

$$\bar\psi_0\gamma^E_2\psi_2 \to (-\bar\psi_0\gamma^5\gamma^E_1)\gamma^E_2(\gamma^E_1\gamma^5\psi_2) = -\bar\psi_0\gamma^5\gamma^E_2\gamma^5\psi_2 = \bar\psi_0\gamma^E_2\psi_2,$$

and the same for $\bar\psi_0\gamma^E_3\psi_3$ and $\bar\psi_0\gamma^E_4\psi_4$.

6.3. Nielsen–Ninomiya's no-go theorem

The problem to construct a viable lattice counterpart of the continuum Dirac operator is not at all simple. Let us look for some other lattice Dirac operator $\mathfrak{D} \neq \mathfrak{D}^0$ satisfying the following four natural conditions:

1. At distances much larger than the lattice spacing a, $\mathfrak{D} \to -i\not{\mathcal{D}}^E$, giving rise to a massless fermion in the continuum limit[25].

2. All the modes of $\mathfrak{D}$ not associated with the latter are of order $1/a$ (no doublers!).

[25] Adding a finite mass term to $\mathfrak{D}$ provides no difficulties [see Eq. (6.9)].

3. $\mathfrak{D}$ is local. In other words, the matrix elements $\mathfrak{D}_{nn'}$ decay exponentially fast at large distances $|n - n'| \gg 1$.

4. Chiral symmetries (3.3), (4.2) of the massless fermionic action are not broken by the regularization explicitly. [The singlet axial symmetry (3.3) is eventually going to be broken due to noninvariance of the fermionic measure, but we require the absence of *explicit* breaking in the regularized action.] This seems to imply the condition $\mathfrak{D}\gamma^5 + \gamma^5\mathfrak{D} = 0$.

The no-go theorem due to Nielsen and Ninomiya [25] tells us, however, that such a $\mathfrak{D}$ *does not exist*. To understand it, consider first the free fermion case. The momentum p_μ is then a good quantum number, and the Dirac operator in the momentum representation has the form

$$\mathfrak{D}(p) = \gamma^E_\mu F_\mu(p) + G(p). \tag{6.23}$$

Condition 4 tells us that $G(p) = 0$. Condition 1 implies that $F_\mu(p) = p_\mu + O(ap^2)$ for $ap_\mu \ll 1$. Now, $F_\mu(p)$ are four periodic functions of their four arguments p_μ with period $2\pi/a$. They thus realize a smooth map $T^4 \to R^4$. A look at Figure 3 can convince the reader that the point $F_\mu = 0$ on R^4 where our torus touches it always has at least one more pre-image. Their intuition would not betray them: this statement can be proven in a rigorous mathematical manner. Basically, it follows from the fact that the degree of a map $T^d \to R^d$ is zero, which means that [26]

$$\sum_{\substack{\text{pre-images}\\ \text{of } P}} \text{sign}\,[\det \|\partial_\nu F_\mu(p)\|] = 0. \tag{6.24}$$

As the Jacobian of the mapping $p_\mu \to F_\mu(p)$ is equal to 1 at the point $p_\mu = 0$, Eq. (6.24) implies that some other pre-images of zero, i.e. some other solutions of the equation system $F_\mu(p) = 0$ should be present (one can have just one extra solution as in Figure 3 or more: for the "round upright torus" $F_\mu(p) = \sin(ap_\mu)/a$, there are $2^d - 1$ extra solutions). And that means the presence of doublers in contradiction with condition 2.

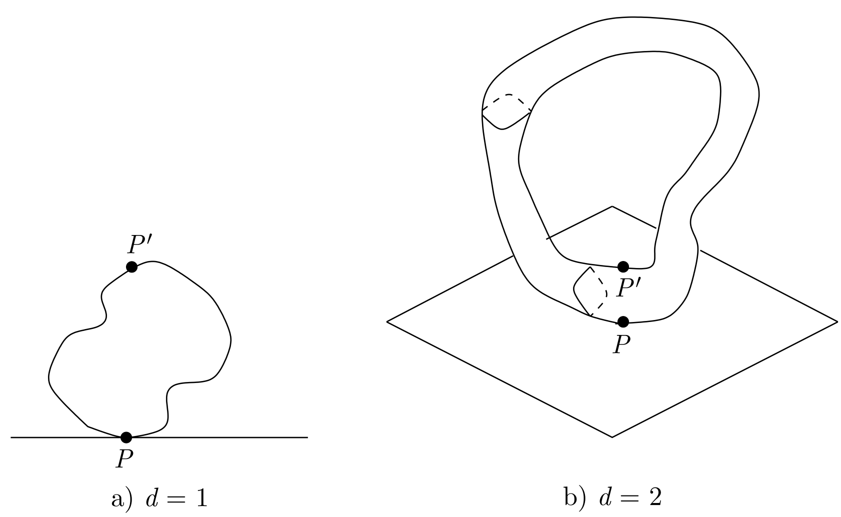

Figure 3. Nielsen–Ninomiya theorem. P and P' are the different zeros of the lattice Dirac operator.

The only remaining possibility is that $F_\mu(p)$ are not continuous. Besides being ugly, it also contradicts condition 3: the matrix elements $\mathfrak{D}_{nn'} = \mathfrak{D}(n-n') = \gamma_\mu^E \tilde{F}_\mu(n-n')$ are actually the Fourier coefficients of the periodic function $\gamma_\mu^E F_\mu(p)$. If the latter is discontinuous, the Fourier coefficients cannot decay faster than $1/|n-n'|$ (otherwise, the Fourier series would converge uniformly on the torus $p_\mu \in \left[0, \frac{2\pi}{a}\right)$ and the sum of such a series would be continuous).

We have proven that a lattice Dirac operator satisfying the above four conditions cannot be found for free fermions, but this also means that it cannot be found for QCD: any $\mathfrak{D}$ having this property should satisfy our list for any smooth set of link variables and, in particular, for the set $U_{\text{each link}} = 1$ corresponding to free theory.

6.4. Ways to go □ The Ginsparg–Wilson way

If we still want to build up a lattice version of QCD, we have to relax at least one of our four conditions. Conditions 1 and 2 are, however, indispensible: a lattice theory where they do not hold just has nothing to do with QCD. Therefore, either locality or manifest chiral invariance of the lattice action should be abandoned.

One of the possible procedures is that only *one mode* of each degenerate 16-plet of $\mathfrak{D}^0$ is taken into account in the fermionic determinant and in the spectral decomposition of fermion Green's functions,

$$\langle \psi_n \bar{\psi}_{n'} \rangle = \sum_k \frac{u_k(n) u_k^\dagger(n')}{m - i\lambda_k}, \tag{6.25}$$

where $u_k(n)$ describe the k-th eigenmode of $\mathfrak{D}^0$ as a function of the node. This amounts to choosing the lattice Dirac operator in the form $(\mathfrak{D}^0)^{1/16}$, which is not local. A similar method is sometimes used in practical lattice calculations, but besides purely technical inconveniences it is unsatisfactory from a philosophical viewpoint: we *would* like to have a local lattice approximation for a local field theory.

But then the chiral invariance like (3.3) is necessarily lost. Though renouncing chiral invariance is also not desirable — when regularizing the theory, we should try to preserve as much of its symmetries as possible — it can still be considered the least of evils. After all, lattice regularization also does not preserve other important symmetries of the continuum action — translational and Lorentz invariances. They are restored only in the limit $a \to 0$. One can allow the same for the chiral invariance.

Two ways of chiral noninvariant lattice regularization have been known for some time and used in practical calculations: (i) *Wilson* fermions [24] and (ii) *Kogut–Susskind* or *staggered* fermions [27]. We will describe here the first method which consists in adding to $\mathfrak{D}^0$ the term $-\frac{ra}{2}\mathcal{D}_\mu^+\mathcal{D}_\mu^- = -\frac{ra}{2}\mathcal{D}_\mu^-\mathcal{D}_\mu^+$ with the covariant lattice derivatives defined in (6.19). Thus, the Wilson–Dirac operator is defined as [23]:

$$\mathfrak{D}^W = -\frac{i}{2}\gamma_\mu^E \left(\mathcal{D}_\mu^+ + \mathcal{D}_\mu^-\right) - \frac{ra}{2}\mathcal{D}_\mu^+\mathcal{D}_\mu^- \tag{6.26}$$

with

$$(\mathcal{D}_\mu^+\mathcal{D}_\mu^-\psi)_n = \frac{1}{a^2}\sum_{\mu=1}^{4}\left(U_{n,n+e_\mu}\psi_{n+e_\mu} + U_{n,n-e_\mu}\psi_{n-e_\mu} - 2\psi_n\right). \tag{6.27}$$

Here r is an arbitrary nonzero positive constant.

The first term in Eq. (6.26) (call it iB) is anti-Hermitian, while the second term (call it A) is Hermitian. The property

$$\gamma^5 \mathfrak{D}^W \gamma^5 \;=\; (\mathfrak{D}^W)^\dagger \tag{6.28}$$

holds. A and B do not commute for a generic gauge field configuration and cannot be simultaneously diagonalized. A generic matrix $\mathfrak{D}^W = A + iB$ still can be diagonalized, but the corresponding transformation matrix is not unitary and the eigenvectors of $\mathfrak{D}^W$ are not orthogonal to each other. This might be not so nice, but it does not prevent one to determine the spectrum of $\mathfrak{D}^W$ (the roots of the characteristic equation) and compute the determinant of $\mathfrak{D}^W + m$ which enters the lattice approximation of the partition function of QCD.

The essential properties of $\mathfrak{D}^W$ can be understood by studying the case of free fermions. The situation is much simpler here than in the interacting case. The Hermitian and anti-Hermitian parts of $\mathfrak{D}^W_{\rm free}$ can now be simultaneously diagonalized, and this can be done explicitly. We have $\mathcal{D}^\pm_\mu \to \partial^\pm_\mu$ and

$$(\partial^+_\mu \partial^-_\mu \psi)_n = \frac{1}{a^2} \sum_\mu \big(\psi_{n+e_\mu} + \psi_{n-e_\mu} - 2\psi_n\big),$$

which is the lattice Laplacian. Passing to the momentum representation, we obtain

$$\mathfrak{D}^W_{\rm free}(p) = \frac{1}{a} \gamma^E_\mu \, \sin{(ap_\mu)} + \frac{2r}{a} \sum_\mu \sin^2\left(\frac{ap_\mu}{2}\right). \tag{6.29}$$

The second term has the form of a momentum-dependent mass. For small $p_\mu \ll 1/a$, it can be neglected and the continuum massless Dirac operator is reproduced. In contrast to $\mathfrak{D}^0$, the operator (6.29) is not anti-Hermitian, and its eigenvalues are complex. What is important is that, for p_μ that are not small, the absolute values of the eigenvalues of $\mathfrak{D}^W_{\rm free}$,

$$-i\lambda^W_p = \pm \frac{i}{a} \sqrt{\sum_\mu \sin^2(ap_\mu)} + \frac{2r}{a} \sum_\mu \sin^2\left(\frac{ap_\mu}{2}\right), \tag{6.30}$$

are of order $1/a$. The doublers disappear. At $p_\mu = \left(\frac{\pi}{a}, 0, 0, 0\right)$, the eigenvalue (6.30) is $\frac{2r}{a}$; at $p_\mu = \left(0, \frac{\pi}{a}, 0, \frac{\pi}{a}\right)$, it is $\frac{4r}{a}$, etc.

The chiral symmetry is broken, however, and it is messy. In principle, when the continuum limit $a \to 0$ is taken, the effects due to the breaking of γ^5 invariance must be suppressed, but in this particular problem, the continuum limit with restoration of chiral symmetry is rather slow to reach, and, though most experts think that it is eventually reached, this has not been shown quite rigorously. In particular, it is difficult to make the pions light. In practical calculations, this is achieved by introducing a large bare quark mass of order $g^2(a)/a$ and fine-tuning it so that the effects due to two chiral noninvariant terms — the Wilson term and the bare quark term — would cancel each other. Needless to say, this is a rather artificial and unaesthetic procedure.

As we see, this Nielsen–Ninomiya puzzle represented a complicated logical knot that resisted attempts to untie it. A remarkable observation [28, 29] was that the best strategy here was to follow the example of the Alexander the Great and just cut it through! The adequate sword was forged back in 1982 by Ginsparg and Wilson [30]. They suggested to consider the lattice Dirac operators satisfying the relation

$$\gamma^5 \mathfrak{D} + \mathfrak{D}\gamma^5 = a\mathfrak{D}\gamma^5\mathfrak{D}. \tag{6.31}$$

The anticommutator $\{\mathfrak{D}, \gamma^5\}$ does not vanish which means that the lattice action is not invariant with respect to the chiral transformations

$$\delta\psi_n = -i\alpha\gamma^5\psi_n, \quad \delta\bar{\psi}_n = -i\alpha\bar{\psi}_n\gamma^5, \tag{6.32}$$

a lattice Euclidean counterpart of Eq. (3.3).

It took 16 years to realize that the lattice fermion action

$$S_F = a^4 \sum_{nn'} \bar{\psi}_n \mathfrak{D}_{nn'} \psi_{n'} \tag{6.33}$$

(color and spinor indices being suppressed), with $\mathfrak{D}$ satisfying the relation (6.31), is invariant with respect to the following transformations[26]:

$$\begin{aligned} \delta\psi &= -i\alpha\gamma^5 \left[1 - \frac{1}{2}a\mathfrak{D}\right]\psi, \\ \delta\bar{\psi} &= -i\alpha\bar{\psi}\left[1 - \frac{1}{2}a\mathfrak{D}\right]\gamma^5. \end{aligned} \tag{6.34}$$

If $\mathfrak{D}$ is local (in the sense of condition 3 in the Nielsen–Ninomiya list), Eq. (6.34) is as good a lattice approximation of the continuous chiral symmetry (3.3) as the trivial (6.32). In particular, the pions would automatically be light (massless in the chiral limit), and no fine tuning is required. In addition, condition 4 above is no longer satisfied and one can now hope to find a local $\mathfrak{D}$ not involving doublers. The problem is still not trivial: as we will see later (from the solution of **Problem 5**), not all solutions of the Ginsparg–Wilson relation (6.31) eliminate the doublers. The simplest solution that *does* was suggested by Neuberger in Ref. [28]. It has the form[27]

$$\mathfrak{D} = \frac{1}{a}\left[1 - A\frac{1}{\sqrt{A^\dagger A}}\right], \tag{6.35}$$

where $A = 1 - a\mathfrak{D}^W$ and $\mathfrak{D}^W$ is the Wilson–Dirac operator (6.26) with $r > 1/2$. The Hermitian matrix $A^\dagger A$ has only positive eigenvalues. Out of many possible square roots, we choose one where the eigenvalues are also positive.

The operator (6.35) satisfies the Ginsparg–Wilson relation. To see that, represent $a\mathfrak{D} = 1 - V$. Note that the operator V is unitary:

$$\begin{aligned} &VV^\dagger = A(A^\dagger A)^{-1/2}(A^\dagger A)^{-1/2}A^\dagger = A[A^{-1}(A^\dagger)^{-1}]A^\dagger = \mathbb{1}; \\ &V^\dagger V = (A^\dagger A)^{-1/2}A^\dagger A(A^\dagger A)^{-1/2} = \mathbb{1}. \end{aligned}$$

[26] In this case, $\delta\bar{\psi} \neq (\delta\psi)^\dagger$ not only because of the presence of i in the right-hand sides, but also due to the fact that $\mathfrak{D}$ is not Hermitian.

[27] This construction is closely related to the so-called *overlap representation* of the Dirac operator [31]. But we will not go into further details here.

Note also that[28]

$$\gamma^5 V \gamma^5 = V^\dagger. \tag{6.36}$$

Compare

$$\gamma^5(1-V) + (1-V)\gamma^5 \quad \text{with} \quad (1-V)\gamma^5(1-V).$$

The left-hand side and right-hand side coincide iff $\gamma^5 = V\gamma^5 V$ or $V\gamma^5 V\gamma^5 = \mathbb{1}$. But that is true due to (6.36) and unitarity of V.

It follows from (6.31) that, in contrast to $\mathfrak{D}^W$, $\mathfrak{D}$ satisfies the property

$$[\mathfrak{D}, \mathfrak{D}^\dagger] = \mathfrak{D}\gamma^5\mathfrak{D}\gamma^5 - \gamma^5\mathfrak{D}\gamma^5\mathfrak{D} \;=\; \frac{1}{a}\left[(\mathfrak{D}\gamma^5 + \gamma^5\mathfrak{D})\gamma^5 - \gamma^5(\mathfrak{D}\gamma^5 + \gamma^5\mathfrak{D})\right] = 0,$$

allowing for a simultaneous diagonalization of its Hermitian and anti-Hermitian part.

In particular, for $r = 1$ and for the free fermions, we have

$$a\mathfrak{D}(p) = 1 - \frac{1 - 2\sum_\mu \sin^2\left(\frac{ap_\mu}{2}\right) - \sum_\mu \gamma_\mu \sin(ap_\mu)}{\left[1 + 8\sum_{\mu<\nu} \sin^2\left(\frac{ap_\mu}{2}\right)\, \sin^2\left(\frac{ap_\nu}{2}\right)\right]^{1/2}}. \tag{6.37}$$

The eigenvalues of (6.37) are different from zero provided $p_\mu \neq 0$. In particular, for former doublers,

$$p_\mu = \left(\frac{\pi}{a}, 0, 0, 0\right), \quad p_\mu = \left(\frac{\pi}{a}, \frac{\pi}{a}, 0, 0\right), \quad p_\mu = \left(\frac{\pi}{a}, \frac{\pi}{a}, \frac{\pi}{a}, 0\right),$$

and

$$p_\mu = \left(\frac{\pi}{a}, \frac{\pi}{a}, \frac{\pi}{a}, \frac{\pi}{a}\right),$$

the eigenvalues $-i\lambda$ of $\mathfrak{D}$ are all equal to $2/a$. The doublers are absent.

A second look at Eq. (6.37) reveals a beautiful feature displayed in Figure 4: the eigenvalues of $\mathfrak{D}$ lie on the circle

$$(\mathrm{Re}\,\lambda)^2 + \left(\mathrm{Im}\,\lambda - \frac{1}{a}\right)^2 = \frac{1}{a^2}. \tag{6.38}$$

This property holds also in the interacting case. As was mentioned, the operator V is unitary, i.e. its eigenvalues lie on the circle $\{e^{i\phi}\}$. And the eigenvalues of $\mathfrak{D} = (1-V)/a$ lie on the shifted circle in Figure 4.

The function (6.37) has singularities associated with the square root, but they all occur at complex values of p_μ. It is analytic on the torus $p_\mu \in \left[0, \frac{2\pi}{a}\right)$, which means that its Fourier image decays exponentially at large distances. The Dirac operator thus constructed is local. In the interacting case, it stays local if the gauge field is smooth enough, i.e. if the link variables $U_{n,n+e_\mu}$ are sufficiently close to 1.

As was mentioned, the singlet axial symmetry (3.3) is anomalous, which shows up in the noninvariance of the fermionic measure. The measure $\prod_n d\bar\psi_n d\psi_n$ is obviously invariant, however, with respect to the ultralocal transformations (6.32). This follows from the fact that

[28] It is basically a collorary of (6.28). It follows from (6.28) immediately for the free fermions when $A^\dagger A = AA^\dagger$, but this property also holds in the interacting case (see **Problem 6**).

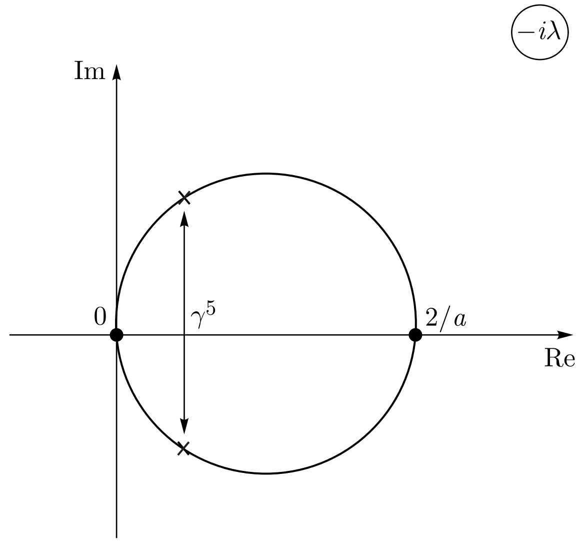

Figure 4. The circle of eigenvalues for Neuberger's operator. The eigenmodes with the eigenvalues marked by crosses are related by a γ^5 transformation.

$\mathrm{Tr}\,\{\gamma^5\} = 0$. On the other hand, Eq. (3.26) relates the modification of the measure to the operator trace of γ^5,

$$\mathrm{Tr}\,\{\gamma^5\} = \int d^4x \sum_k u_k^\dagger(x)\gamma^5 u_k(x), \tag{6.39}$$

which is thus also zero. For the naïve Dirac operator, only the zero modes contribute to the sum (6.39), which means that the number of the right-handed and the left-handed zero modes of $\mathfrak{D}^0$ should be equal.

This also follows from the fact that $\mathfrak{D}^0$ commutes with $\hat{Q}_\mu$ defined in Eq. (6.17). Any smooth eigenfunction $\psi_n^{(p)}$ of $\mathfrak{D}^0$ is accompanied by 15 doublers. It is not difficult to see that if $\psi_n^{(p)}$ is, say, right-handed, the eigenfunctions $\hat{Q}_{[\mu}\hat{Q}_{\nu]}\psi_n^{(p)}$ and $\hat{Q}_{[\mu}\hat{Q}_\nu\hat{Q}_\lambda\hat{Q}_{\rho]}\psi_n^{(p)}$ are also right-handed, whereas $\hat{Q}_\mu\psi_n^{(p)}$ and $\hat{Q}_{[\mu}\hat{Q}_\nu\hat{Q}_{\lambda]}\psi_n^{(p)}$ are left-handed. In particular, instead of a single right-handed zero mode in (the lattice approximation for) the instanton background, we have a degenerate 16-plet with 8 right-handed and 8 left-handed modes which are not necessarily zero modes anymore.

The vanishing of the index of $\mathfrak{D}^0$ is closely related to the identity (6.24). Indeed, the sign of the Jacobian $\det\|\partial_\nu F_\mu(p)\|$ describes the orientation of the neighbourhood of a pre-image $P_i \in T^4$ with respect to the orientation of the tangent space R^4 into which it is mapped. This orientation is obviously related to chirality.

The absence of anomaly is one of the deseases of the naïve lattice Dirac operator. For the operators of Ginsparg–Wilson type, the situation is different. The chiral symmetry is now implemented as in Eq. (6.34) and, generically, the measure is *not* invariant with respect to these transformations. We have, instead of Eq. (3.26),

$$\ln J = -i\alpha\left[n_R^{(0)} - n_L^{(0)}\right] = -i\alpha\,\mathrm{Tr}\left\{\gamma^5\left(1 - \frac{1}{2}a\mathfrak{D}\right)\right\}. \tag{6.40}$$

Even though $\mathrm{Tr}\,\{\gamma^5\} = 0$, $\mathrm{Tr}\,\{\gamma^5\mathfrak{D}\}$ need not vanish and the anomaly is there.

It is instructive to see how a nonzero operator trace on the right side of Eq. (6.40) is realized in the basis including the eigenstates of $\mathfrak{D}$. First, it is still true that, for any eigenfunction u_k of $\mathfrak{D}$, $\gamma^5 u_k$ is also an eigenfunction. However, in contrast to the continuum or naïve lattice Dirac operators where the eigenvalues of $\gamma^5 u_k$ and u_k were related as $\lambda'_k = -\lambda_k$, in the Ginsparg–Wilson case, the corresponding relation is $\lambda'_k = -\lambda^*_k$. Complex conjugation appears! Thus, for almost all eigenstates on the circle in Figure 4, u_k and $\gamma^5 u_k$ have *different* eigenvalues and are orthogonal to each other. Their contribution to the trace vanishes. The only exception are two points on the circle, $\lambda = 0$ and $\lambda = 2i/a$, where $\lambda_k = \lambda'_k$ and the eigenstates can have a definite chirality. But the "heavy doublers" with $-i\lambda = \frac{2}{a}$ obviously give zero contribution in Eq. (6.40). Only zero modes of $\mathfrak{D}$ are relevant. We have derived the *lattice index theorem* [32]:

$$n^0_R(\mathfrak{D}) - n^0_L(\mathfrak{D}) = -\frac{1}{2}\,\mathrm{Tr}\,\{\gamma^5 a\mathfrak{D}\}. \tag{6.41}$$

The right side of Eq. (6.41) is a functional depending of the link variables $\{U\}$. By definition, it is given by a sum over all lattice nodes of some complicated expression, which goes over to the topological charge (3.31) in the continuum limit[29].

Problem 5. Consider the free Wilson–Dirac operator (6.29) with arbitrary r and analyze the corresponding Neuberger's operator (6.35). Show that only the values $r > 1/2$ are admissible.

Solution. As was mentioned above, the eigenvalues of the free Wilson–Dirac operator (6.26) which correspond to the naïve doublers are $-i\lambda^W = 2rl/a$, where l is the number of nonzero components π/a of the momenta p_μ. The corresponding eigenvalues of the operator (6.35) are

$$-i\lambda = \frac{1}{a}\left[1 - \frac{1-2rl}{|1-2rl|}\right]. \tag{6.42}$$

We see that, if $r < 1/2$, four doublers with $l = 1$ become massless again, and that is not what we want.

For $r < 1/8$, *all* the doublers stay massless. Note that the operator (6.35) satisfies the Ginsparg–Wilson relation for any r, which means that its eigenvalues still lie on the circle in Figure 4. But only a part of the circle is covered.

Massless doublers may appear also for r slightly exceeding $1/2$ if gauge fields are present.

Problem 6. Derive the property (6.36).

Solution. Consider an analytic function $f(A^\dagger A)$ and its Taylor expansion $f(y) = a_0 + a_1 y + \ldots$. Then

$$\begin{aligned}\gamma^5 A f(A^\dagger A)\gamma^5 \; &= \gamma^5 A(a_0 + a_1 A^\dagger A + a_2 A^\dagger A A^\dagger A + \cdots)\,\gamma^5 = \\ = \;& a_0 A^\dagger + a_1 A^\dagger A A^\dagger + a_2 A^\dagger A A^\dagger A A^\dagger + \cdots \\ = \;& f(A^\dagger A)A^\dagger = [A f(A^\dagger A)]^\dagger.\end{aligned} \tag{6.43}$$

The reader may complain that $f(y) = 1/\sqrt{y}$ is not analytic at $y = 0$. True, but one can expand it in Taylor series at any other point.

[29] To the best of our knowledge, this expression has never been explicitly written out. It would be an interesting problem to do so.

7. Aspects of chiral symmetry

We now abandon the lattice and will discuss in this section various aspects of chiral symmetry in the continuum limit. Sometimes, we will think in terms of quarks, of the symmetries (3.3) and (4.2), and of the order parameter (4.8) associated with the spontaneous breaking of the flavor-nonsinglet symmetry. Sometimes, we will describe the system in terms of the pseudo-Goldstone degrees of freedom and the effective Lagrangians (5.2), (5.4), and (5.6). Sometimes, we will confront the two languages using the philosophy of the *quark–hadron duality* — that is how many results discussed in this section, a bunch of beautiful *exact theorems of QCD*, will be obtained.

7.1. QCD inequalities □ The Vafa–Witten theorem

As was discussed above, the octet of pseudoscalar mesons (π, K, η) can be interpreted as that of the pseudo-Goldstone particles appearing due to the spontaneous chiral symmetry breaking according to the pattern (4.7) in the massless limit. This is the reason why the pseudoscalar mesons are lighter than those with other quantum numbers. It is interesting that the latter statement can be formulated as an exact theorem of QCD without any reference to the (experimental!) fact that the chiral symmetry is broken.

Consider a QCD-like theory with at least two quark flavors and assume that these quarks (call them u and d) have equal masses $m_u = m_d = m$. Consider a set of Euclidean correlators,

$$C_\Gamma(x,y) = \langle J^{\bar u d}(x) J^{\bar d u}(y)\rangle_{\rm vac}, \tag{7.1}$$

where $J^{\bar u d}(x)$ are flavor-changing bilinear quark currents, $J^{\bar u d} = \bar u \Gamma d$, with Hermitian

$$\Gamma = 1,\ \gamma^5,\ i\gamma^E_\mu,\ \gamma^E_\mu\gamma^5 \text{ and } i\sigma_{\mu\nu} = \frac{i}{2}(\gamma^E_\mu\gamma^E_\nu - \gamma^E_\nu\gamma^E_\mu).$$

At large distances, the correlators (7.1) decay exponentially

$$C_\Gamma(x,y) \propto \exp\{-M_\Gamma|x-y|\}, \tag{7.2}$$

where M_Γ is the mass of the lowest meson state in the corresponding channel[30].

On the other hand, the correlators (7.1) of the quark currents can be expressed as

$$C_\Gamma(x,y) = -Z^{-1}\int d\mu_A \,{\rm Tr}\, \{\Gamma \mathcal{G}_A(x,y)\Gamma\mathcal{G}_A(y,x)\}, \tag{7.3}$$

where

$$d\mu_A = \prod_{x\mu a} dA^a_\mu(x) \prod_f \det \|i\not{\mathcal{D}}^E - m_f\| \, \exp\left\{-\frac{1}{4g^2}\int F^a_{\mu\nu}F^a_{\mu\nu}d^4x\right\} \tag{7.4}$$

is the standard QCD measure and $\mathcal{G}_A(x,y)$ is the Euclidean Green's function of the u and d quarks in a given gauge field background. Note that, when writing down Eq. (7.3), we used the fact that $J^{\bar u d}$ is not a singlet in flavor [otherwise, the disconnected contribution

$$\propto {\rm Tr}\,\{\Gamma\mathcal{G}_A(x,x)\}\,{\rm Tr}\,\{\Gamma\mathcal{G}_A(y,y)\}$$

[30] We assume here that the quarks are confined, otherwise the whole discussion is pointless.

would appear on the right-hand side]. In addition, the assumption $m_u = m_d$ was made [otherwise, we would have two different Green's functions $\mathcal{G}_A^u(x,y) \neq \mathcal{G}_A^d(x,y)$].

An important nontrivial relation

$$\gamma^5 \mathcal{G}_A(x,y)\gamma^5 = \mathcal{G}_A^\dagger(y,x) \tag{7.5}$$

holds. To understand it, write the spectral decomposition for $\mathcal{G}_A(x,y)$,

$$\mathcal{G}_A(x,y) = \langle \psi(x)\bar{\psi}(y)\rangle^A = \sum_k \frac{u_k(x)u_k^\dagger(y)}{m - i\lambda_k}, \tag{7.6}$$

where $u_k(x)$ are the eigenfunctions of the Dirac operator [cf. Eq. (6.25)]. As we mentioned earlier, for any such eigenfunction with eigenvalue λ_k, $\gamma^5 u_k$ is also an eigenfuction with eigenvalue $-\lambda_k$. This allows us to write

$$\begin{aligned} \gamma^5 \mathcal{G}_A(x,y)\gamma^5 &= \sum_k \frac{[\gamma^5 u_k(x)][\gamma^5 u_k(y)]^\dagger}{m - i\lambda_k} = \sum_p \frac{u_p(x)u_p^\dagger(y)}{m + i\lambda_p} = \\ &= \left[\sum_p \frac{u_p(y)u_p^\dagger(x)}{m - i\lambda_p}\right]^\dagger = \mathcal{G}_A^\dagger(y,x), \end{aligned} \tag{7.7}$$

as annonced. We see that the pseudoscalar correlator

$$\sim \left\langle \mathrm{Tr}\,\{\gamma^5 \mathcal{G}_A(x,y)\gamma^5 \mathcal{G}_A(y,x)\}\right\rangle = \left\langle \mathrm{Tr}\,\{|\mathcal{G}_A(x,y)|^2\}\right\rangle$$

plays a distinguished role — it represents an absolute upper bound for any other such correlator. The fastest way to show this is to expand the 4×4 matrix $\mathcal{G}_A(x,y)$ over the full basis

$$\mathcal{G}_A(x,y) = s(x,y) + \gamma^5 p(x,y) + i\gamma_\mu^E v_\mu(x,y) + \gamma_\mu^E \gamma^5 a_\mu(x,y) + \frac{1}{2} i\sigma_{\mu\nu} t_{\mu\nu}(x,y). \tag{7.8}$$

Then

$$\begin{aligned} -\frac{1}{4} C_{\gamma^5}^A(x,y) &= \frac{1}{4}\,\mathrm{Tr}\,\{\gamma^5 \mathcal{G}_A(x,y)\gamma^5 \mathcal{G}_A(y,x)\} = \frac{1}{4}\,\mathrm{Tr}\,\{|\mathcal{G}_A(x,y)|^2\} = \\ &= |s|^2 + |p|^2 + |v_\mu|^2 + |a_\mu|^2 + \frac{1}{2}|t_{\mu\nu}|^2, \end{aligned} \tag{7.9}$$

but, say,

$$\begin{aligned} -\frac{1}{4} C_{\mathbf{1}}^A(x,y) &= \frac{1}{4}\,\mathrm{Tr}\,\{\mathcal{G}_A(x,y)\mathcal{G}_A(y,x)\} = \frac{1}{4}\,\mathrm{Tr}\,\{\mathcal{G}_A(x,y)\gamma^5 \mathcal{G}_A^\dagger(x,y)\gamma^5\} = \\ &= |s|^2 + |p|^2 - |v_\mu|^2 - |a_\mu|^2 + \frac{1}{2}|t_{\mu\nu}|^2. \end{aligned} \tag{7.10}$$

The inequalities

$$|C_{\gamma^5}^A(x,y)| \geqslant |C_\Gamma^A(x,y)| \tag{7.11}$$

in any given gauge background, the positivity of the measure (7.4) in Eq. (7.3), and the asymptotics (7.2) imply that the mass M_{PS} of the lightest pseudoscalar meson in the $\bar{u}d$ channel is less or may be equal to the masses M_S, M_V, M_A, M_T of the lightest scalar, vector, axial, and tensor states[31].

Let us emphasize again that this statement can be justified only in the theory with the positive measure (7.4) (on the other hand, in the theory with nonzero vacuum angle $\theta \neq 0$, the measure *is* not positive and pseudoscalar states *need* not to be the lightest), with equal quark masses, and only for those states that are not flavor-singlet. For flavor-singlet states, it need not be true. Consider, for example, the theory with just one quark of a large mass. Then the lowest meson states would be made of gluons and would know nothing about quarks. The lowest glueball state is believed to be scalar rather than pseudoscalar.

In Sec. 6, we have mentioned already the Vafa–Witten theorem [19] saying that the vector isotopic symmetry is not broken spontaneously in QCD. Now we are ready to prove it. Indeed, if such a breaking occured, the massless Goldstone scalar particles would appear in the spectrum. The inequality $M_{PS} \leqslant M_S$ implies that a massless *pseudoscalar* particle would also exist. But the theory with $m_u = m_d \neq 0$ (which duly enjoys the exact isovector symmetry, a possible spontaneous breaking of which is under discussion now) has no exact axial isotopic symmetry, and there are *no reasons* for the massless pseudoscalar state to exist. So, it does not exist, hence the massless scalar does not exist either, and the isovector symmetry is not broken.

Many more inequalities of this kind (e.g., $M_N \geqslant M_\pi$ or $M_{\pi^+} \geqslant M_{\pi^0}$) can be formulated, but their proof relies on some extra assumptions and even though the assumptions are very natural, the status of these results is a little bit less solid. We address the reader to Ref. [33] for a nice review.

7.2. Euclidean Dirac spectral density

Consider the Euclidean Dirac operator $\not{\mathcal{D}}^E$ in a given gauge field background $A_\mu(x)$. We assume that the system is placed in a finite 4-volume so that the spectrum of $\not{\mathcal{D}}^E$ is discrete. Let $\{\lambda_k\}$ be the background-dependent set of eigenvalues of $\not{\mathcal{D}}^E$. The *spectral density*[32] is defined as follows:

$$\rho(\lambda) = \left\langle \frac{1}{V} \sum_k \delta\left(\lambda - \lambda_k[A_\mu(x)]\right) \right\rangle, \tag{7.12}$$

where the average is done with the weight function (7.4). The γ^5 symmetry of the spectrum implies that $\rho(\lambda)$ is an even function of λ.

In contrast to solids or nuclei, the spectral density (7.12) is defined in Euclidean space and seems to have no direct physical meaning. There is, however, a set of remarkable identities which relate the spectral density of the Euclidean Dirac operator to physical observables.

[31] We were a little bit sloppy here. The inequalities (7.11) hold, strictly speaking, only for unrenormalized correlators, not for the renormalized ones. However, renormalization only brings about multiplicative factors, which do not depend on distance. Thus, sending $|x - y|$ to infinity *before* the limit $\Lambda_{UV} \to \infty$ is done, we can ensure that the inequalities (7.11) for renormalized correlators at large distances are fulfilled.

[32] The notion of spectral density and the definition (7.12) are also widely used in condensed matter physics and nuclear physics. It is especially useful if a system is disordered or involves elements of disorder like it is the case for electron spectra in most solids or for energy levels in complicated nuclei. It makes sense also for ordered systems (such as metals). In this case, rather than averaging over stochastic external field, one averages over some interval of eigenvalues $\Delta\lambda$ much larger than the characteristic level spacing, but much less than the characteristic scale of λ on which $\rho(\lambda)$ is essentially changed.

The simplest such identity relates the spectral density at "zero virtuality" $\lambda = 0$ to the quark condensate.

To derive it, set $x = y$ in the spectral decomposition (7.6), integrate it over $\frac{1}{V} d^4x$, and perform the averaging over the gauge fields with weight (7.4). In view of the definitions (4.4), (4.8), (4.9), and (7.12), using the symmetry $\rho(-\lambda) = \rho(\lambda)$, and assuming the reality of Σ, we obtain

$$\Sigma = \left\langle \sum_k \frac{1}{m - i\lambda_k} \right\rangle = \int_{-\infty}^{\infty} \frac{\rho(\lambda) d\lambda}{m - i\lambda} = 2m \int_0^{\infty} \frac{\rho(\lambda) d\lambda}{\lambda^2 + m^2}. \tag{7.13}$$

To better understand this formula, let us first look at what happens for free fermions. As there is no physical dimensionful scale in this case [remember that λ_k in Eq. (7.12) are eigenvalues of the *massless* Dirac operator], $\rho(\lambda) = C|\lambda|^3$ on dimensional grounds. By counting the eigenvalues of the free Dirac operator,

$$\lambda(n_\alpha) = \pm \frac{2\pi}{L} \sqrt{\sum_\alpha \left(n_\alpha + \frac{1}{2} \right)^2} \tag{7.14}$$

in the 4D ball $1/L \ll |\lambda| < \Lambda$ [antiperiodic boundary conditions for the fermions in all four directions are chosen, n_α are integer, and each level (7.14) involves an extra $2N_c$-fold degeneracy], it is not difficult to determine $C = N_c/(4\pi^2)$. Thus,

$$\rho^{\text{free}}(\lambda) = \frac{N_c}{4\pi^2} |\lambda|^3. \tag{7.15}$$

In the interacting theory, the spectral density behaves as $\rho(\lambda) \propto |\lambda|^3$ for $|\lambda|$ much greater than the characteristic hadron scale μ_{hadr}, so that interaction is weak due to asymptotic freedom. To be more precise, the power $|\lambda|^3$ is multiplied (in the leading logarithmic order) by the anomalous dimension factor

$$\sim \left[\frac{\alpha_s(|\lambda|)}{\alpha_s(\mu)} \right]^{\gamma/b_0}, \tag{7.16}$$

where $\mu \sim \mu_{\text{hadr}}$ is the normalization point and $b_0 = 11 - 2N_f/3$ was defined in Eq. (2.18). According to the calculations in Ref. [34], $\gamma = 16$.

We see that the integral in Eq. (7.13) diverges quadratically in the ultraviolet. The same result can be obtained directly by calculating the fermion bubble graph in the momentum representation

$$\langle \psi(0)\bar{\psi}(0) \rangle = \int \frac{d^4 p_E}{(2\pi)^4} \text{Tr} \left\{ \frac{\not{p}_E + m}{p_E^2 + m^2} \right\} \propto m\Lambda_{UV}^2. \tag{7.17}$$

Thus, strictly speaking, the formula (7.13) does not make much sense as it stands. Note, however, that even though the (purely perturbative) contribution (7.17) diverges in the ultraviolet, it vanishes in the chiral limit $m \to 0$. The whole point is that in QCD the integral (7.13) acquires an additional *nonperturbative* contribution coming from the region of small λ which survives in the "continuum chiral thermodynamic limit" (*first* $V \to \infty$, *then* $m \to 0$, and only then the ultraviolet cutoff is lifted $\Lambda_{UV} \to \infty$). The fact that chiral symmetry is broken spontaneously *means* that the vacuum expectation value $\langle \psi(0)\bar{\psi}(0) \rangle$ is nonzero in this particular limit.

Obviously, the necessary condition for the condensate to develop is $\rho(0) \neq 0$. Neglecting all terms which vanish in the continuum chiral thermodynamic limit defined above, we finally obtain the famous *Banks–Casher relation* [35]:

$$\langle \psi(0)\bar{\psi}(0) \rangle_{\rm vac} \equiv \Sigma = \pi\rho(0). \tag{7.18}$$

Note that the result does not depend on flavor, which tells us again that the flavor vector symmetry is not broken.

Not only $\rho(0)$, but also the form of $\rho(\lambda)$ at small $\lambda \ll \mu_{\rm hadr}$ can be determined [36]. Consider the theory with $N_f \geqslant 2$ light quarks of common mass m. Let us study the integrated correlator

$$\int d^4x \, \langle S^a(x) S^b(0) \rangle = \frac{1}{V} \int d^4x d^4y \, \langle S^a(x) S^b(y) \rangle, \tag{7.19}$$

where $S^a(x) = \bar{\psi}(x) t^a \psi(x)$ and t^a is the generator of the $SU(N_f)$ flavor group. Fix a particular gluon background and define

$$C^{ab}\big|_A = -\frac{1}{V} \int d^4x d^4y \, {\rm Tr} \left\{ t^a \mathcal{G}_A(x,y) t^b \mathcal{G}_A(y,x) \right\}. \tag{7.20}$$

Substitute here the spectral decomposition (7.6) for $\mathcal{G}_A(x,y)$, do the integration, and perform the averaging over the gluon fields trading the sum over eigenvalues for the integral over the spectral density (7.12). We obtain

$$\begin{aligned} C^{ab} &= -\frac{\delta^{ab}}{2V} \left\langle \sum_k \frac{1}{(m - i\lambda_k)^2} \right\rangle = -\frac{\delta^{ab}}{2} \int_{-\infty}^{\infty} \frac{\rho(\lambda) d\lambda}{(m - i\lambda)^2} = \\ &= -\delta^{ab} \int_0^\infty \frac{\rho(\lambda)(m^2 - \lambda^2)}{(m^2 + \lambda^2)^2} d\lambda, \end{aligned} \tag{7.21}$$

where the property $\rho(-\lambda) = \rho(\lambda)$ was used.

On the other hand, the same correlator can be saturated by physical states, among which the lightest pseudo-Goldstone states play a distinguished role. Consider the 1-loop graph in Figure 5 describing the contribution of the 2-goldstone intermediate state $\sim \int \langle 0|S^a|\phi^c\phi^d\rangle\langle\phi^c\phi^d|S^b|0\rangle$ (obviously, the one-particle state does not contribute, because the pseudo-Goldstone mesons are pseudoscalars while $S^a(x)$ is scalar). To calculate it, we need to know the vertex $\langle 0|S^a|\phi^c\phi^d\rangle$, which can be determined via the generating functional of QCD involving scalar sources u^a coupled to the current S^a. Adding the source term $u^a S^a$ to the Lagrangian amounts to adding $u^a t^a$ to the quark mass matrix $\mathcal{M}$. The latter also enters the mass term (5.6) in the effective Lagrangian. Expanding U up to second order in ϕ^a and varying $\mathcal{L}^{(m)}_{\rm eff}$ with respect to u^a, we derive

$$\langle 0|S^a|\phi^c\phi^d\rangle = -\frac{\Sigma}{F_\pi^2} d^{abc}, \tag{7.22}$$

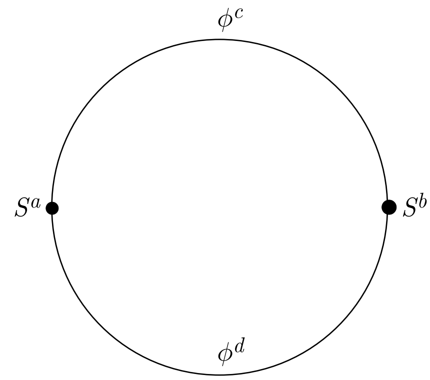

Figure 5. Pseudo-Goldstone loop in scalar correlator.

with d^{abc} being a symmetric group tensor:

$$d^{abc} = 2\mathrm{Tr}\,\{t^a(t^bt^c + t^ct^b)\}. \tag{7.23}$$

The vertex is nonzero only for three or more flavors.

Now we can calculate the graph in Figure 5. Actually, we cannot because the integral diverges logarithmically in the ultraviolet, but anyway the effective theory is not valid at high momenta (technically, the divergence is absorbed into local counterterms of higher order in $p^{\rm char}$ and m). Only the infrared-sensitive part of the integral is relevant. A simple calculation using the identity

$$d^{abc}d^{abd} = \frac{N_f^2-4}{N_f}\delta^{cd}$$

gives

$$\left(C^{ab}\right)^{\rm infrared} = -\frac{N_f^2-4}{32\pi^2 N_f}\left(\frac{\Sigma}{F_\pi^2}\right)^2\delta^{ab}\,\ln\frac{M_\phi^2}{\mu_{\rm hadr}^2}. \tag{7.24}$$

Now compare it with Eq. (7.21). Note first of all that the constant part $\rho(0)$ does not contribute here,

$$\int_0^\infty \frac{m^2-\lambda^2}{(m^2+\lambda^2)^2}d\lambda = 0.$$

Thus, only the difference $\rho(\lambda)-\rho(0)$ is relevant. It is easy to see that, in order to reproduce the singularity $\sim \ln M_\phi^2 \sim \ln m$, we should have $\rho(\lambda)-\rho(0) = C|\lambda|$ at small $|\lambda|$. Substituting it in Eq. (7.21) and comparing the coefficient of $\ln m$ with the coefficient of $\ln M_\phi^2$ in Eq. (7.24), we finally obtain [36]:

$$\rho(\lambda) = \frac{\Sigma}{\pi} + \frac{N_f^2-4}{32\pi^2 N_f}\left(\frac{\Sigma}{F_\pi^2}\right)^2|\lambda| + o(\lambda). \tag{7.25}$$

The behavior is smooth in the theory with two light flavors, but for $N_f \geqslant 3$, $\rho(\lambda)$ has a nonanalytic "dip" at $\lambda = 0$. Physically, it is rather natural that the greater the number of flavors, the stronger the suppression of $\rho(0)$ is. The determinant factor in the measure (7.4) punishes small eigenvalues, and the larger N_f is, the more important this factor becomes. By "analytic continuation" of this argument, one should expect a nonanalytic bump rather than a nonanalytic dip at $\lambda = 0$ in the case $N_f = 1$. Indeed, Eq. (7.25) displays such a bump. One should not forget, of course, that the whole derivation was based on the effective chiral Lagrangian approach and does not directly apply to the case $N_f = 1$. Some additional, more elaborate reasoning based on the random matrix model shows, however, that a bump at $N_f = 1$, as predicted by Eq. (7.25), is there [37]. The existence of the bump was also confirmed by a numerical calculation in the instanton liquid model [38] (see the plots in Fig. 1 there).

7.3. Infrared face of anomaly

As was explained in Sec. 3, the symmetry (3.3) of the classical theory involving massless fermions is broken down by quantum effects. The $U_A(1)$ breaking is introduced by an ultraviolet regularization [so that the measure in the functional integral is not $U_A(1)$ invariant] and the effects due to this breaking do not go away in the limit $\Lambda_{UV} \to \infty$. This is the ultraviolet face of the anomaly. The latter also has, however, a different, infrared face: one can understand its origin exclusively in terms of the low energy dynamics of the theory [39]. We have already seen how the anomaly is related to the presence of fermion zero modes in an Euclidean topologically nontrivial gauge background. Let us now find out what happens in Minkowski space.

Let us discuss the massless QED first. Consider the correlator

$$T^H_{\mu\nu}(q) = i \int \langle T\{j_{\mu 5}(x) j_\nu(0)\}\rangle_{\mathbf{H}}\, e^{iq\cdot x} d^4x, \tag{7.26}$$

where $j_\nu = \bar\psi\gamma_\nu\psi$, $j_{\mu5} = \bar\psi\gamma_\mu\gamma^5\psi$, and the averaging is performed in the presence of an external homogeneous magnetic field $\mathbf{H}$. The correlator can be calculated perturbatively as a series in e^2 and $\mathbf{H}$. The vacuum correlator $\langle T\{j_{\mu5}(x)j_\nu(0)\}\rangle_{\rm vac}$ vanishes and the expansion in $\mathbf{H}$ starts with the linear term. The latter is given by the graphs in Figure 6, where the wavy line signals the modification of the electron propagator due to the external field to leading order,

$$\Delta G(p) = e\epsilon_{ijk} H_k \frac{\gamma_i(\not p - m)\gamma_j}{2(p^2 - m^2)^2}. \tag{7.27}$$

In the massless limit, it coincides with the second term in Eq. (3.43) (in this subsection, devoted to infrared perturbative aspect of the anomaly, we do not include the charge e in the definition of the field). Actually, the graphs in Figure 6 describe the 3-point correlator

$$T_{\mu\nu\alpha}(q,k) = i \int \langle T\{j_{\mu5}(x) j_\nu(0) j_\alpha(y)\}\rangle\, e^{iq\cdot x - ik\cdot y} d^4x d^4y \tag{7.28}$$

in a special kinematics: α is spacelike and $k = (0, \mathbf{k} \to \mathbf{0})$. This kinematics is somewhat simpler and more instructive theoretically than the standard symmetric kinematics $k^2 = (q-k)^2 = 0$ used in Ref. [39] (which, on the other hand, is better adapted to describe the phenomenology of the decay $\pi^0 \to \gamma\gamma$). To make things still simpler, we assume that the vectors $\mathbf{q}$ and $\mathbf{H}$ are parallel and direct them along the 3rd axis. Calculating the integral with, say, the Pauli–Villars regularization method[33], we obtain

$$T^H_{\mu\nu}(q) = \frac{eH}{2\pi^2} \frac{q_\mu \tilde\varepsilon_{\nu\alpha} q^\alpha}{q^2}, \tag{7.29}$$

[33]Calculational details will be given later in the solution of **Problem 7** for the two-dimensional case, which is somewhat simpler.

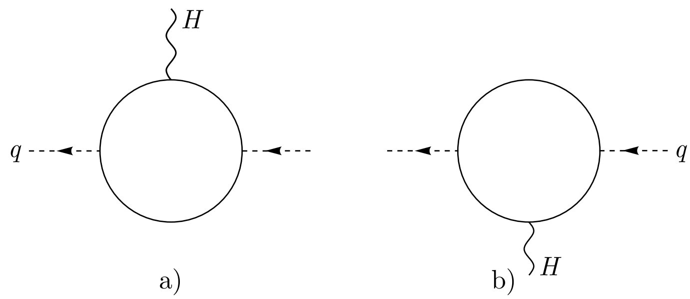

Figure 6. Anomalous triangle in diangle kinematics.

with $H = |\mathbf{H}|$ and $\tilde{\varepsilon}_{\nu\alpha}$ living in the two-dimensional (03)-space so that $\tilde{\varepsilon}_{03} = -\tilde{\varepsilon}_{30} = -1$, and $\tilde{\varepsilon}_{\perp\alpha} = 0$.

The amplitude (7.29) satisfies the property $q^\nu T^H_{\mu\nu} = 0$, which reflects the conservation of the vector current. On the other hand,

$$q^\mu T^H_{\mu\nu} = \frac{eH}{2\pi^2}\tilde{\varepsilon}_{\nu\alpha} q^\alpha \neq 0, \tag{7.30}$$

and this is a manifestation of the anomaly (3.10). In fact, Eq. (7.30) means that the average of the operator $\partial^\mu j_{\mu 5}$ in the presence of a magnetic *and* an electric field $E_i = \partial_0 A_i - \partial_i A_0$ is equal to[34]

$$\langle \partial^\mu j_{\mu 5}\rangle = -\frac{e^2}{2\pi^2}\,\mathbf{E}\cdot\mathbf{H}, \tag{7.31}$$

which is the QED version of Eq. (3.5). The important observation is that the amplitude (7.29) is singular at $q^2 = 0$, and this singularity can only be explained by the presence of *massless particles* in the spectrum. It is very instructive to see what happens if electrons are endowed with a small mass, which explicitly breaks the $U_A(1)$ invariance and also smears out the singularity in Eq. (7.29). The direct calculation of Im $T^{H,m}_{\mu\nu}$ by the graphs in Figure 6 with nonzero mass gives[35]

$$\text{Im}\; T^{H,m}_{\mu\nu} = -\frac{eH}{\pi}\frac{m^2\theta(q^2-4m^2)}{\sqrt{q^2(q^2-4m^2)}}\frac{q_\mu\tilde{\varepsilon}_{\nu\alpha}q^\alpha}{q^2} \;\xrightarrow{m\to 0}\; -\frac{eH}{2\pi}\delta(q^2)q_\mu\tilde{\varepsilon}_{\nu\alpha}q^\alpha. \tag{7.32}$$

This means that anomalous nonconservation of the axial charge in massless QED is associated with the creation of massless e^+e^- pairs of zero energy in the presence of electric and magnetic fields with $\mathbf{E}\cdot\mathbf{H} \neq 0$. These pairs carry nonzero axial charge. If $\mathbf{E}$ and $\mathbf{H}$ are constant and homogeneous, the pairs are created all the time and everywhere. If the fields die out fast enough at spatial infinity and also in the limits $t \to \pm\infty$, the number of created pairs is finite and coincides with the change of the axial charge which, according to Eq. (7.31), is equal to [41]:

$$\Delta Q_5 = -\frac{e^2}{2\pi^2}\int d^3x dt\; \mathbf{E}\cdot\mathbf{H}. \tag{7.33}$$

[34] To derive the operator relation (7.31) out of (7.30), it is more convenient again to treat a little more simple 2D case first and then generalize to four dimensions. We will do so on the next page.

[35] The appearance of the factor m^2 in the numerator of the middle term in Eq. (7.32) can be heuristically understood as follows. This numerator represents a product of the amplitude of production of $q\bar{q}$ pair by the axial source and the amplitude of transition of this pair into two photons.

If the quarks produced by the axial source were strictly massless, their total spin would be 1. But then they could not go into two photons: $\gamma\gamma$ system cannot have total anglular momentum 1 [40].

Problem 7. Discuss a diagrammatic interpretation of the axial anomaly (3.13) in the 2-dimensional QED (*Schwinger model*).

Solution. The right side of Eq. (3.13) is linear in field, which means that the anomaly shows up in the vacuum 2-point correlator

$$T^{\rm 2D}_{\mu\nu}(q) = i\int \langle T\{j_{\mu 5}(x) j_\nu(0)\}\rangle\, e^{iq\cdot x} d^2x. \tag{7.34}$$

Thus, instead of the graphs in Figure 6, we have just one anomalous diangle. In spite of the fact that the corresponding Feynman integral converges in the ultraviolet (it diverges logarithmically by power counting, but the leading term $\propto \mathrm{Tr}\,\{\gamma_\mu\gamma^5 \not{p}\gamma_\nu \not{p}\}$ vanishes after integration), it still has to be regularized (otherwise, the property $q^\nu T_{\mu\nu} = 0$ required by gauge invariance would not hold). We will choose Pauli–Villars regularization. Subtracting the heavy fermion loop, we obtain for small $m^2 \ll |q^2|$ (not forgetting the factor -1 due to the fermion loop) the expression

$$\begin{aligned} T_{\mu\nu}(q) = -i\int \frac{d^2p}{(2\pi)^2}\Bigg[\mathrm{Tr}\left\{\gamma_\mu\gamma^5\frac{i\not{p}}{p^2}\gamma_\nu\frac{i(\not{p}+\not{q})}{(p+q)^2}\right\} + \\ + \frac{M^2}{(p^2-M^2)^2}\,\mathrm{Tr}\,\{\gamma_\mu\gamma^5\gamma_\nu\}\Bigg]. \end{aligned} \tag{7.35}$$

The calculation gives

$$T_{\mu\nu}(q) = \frac{1}{2\pi}\left[\frac{(\varepsilon_{\mu\alpha}q_\nu + \varepsilon_{\nu\alpha}q_\mu)q^\alpha}{q^2} - \varepsilon_{\mu\nu}\right] = \frac{q_\mu\varepsilon_{\nu\alpha}q^\alpha}{\pi q^2}, \tag{7.36}$$

where we used the convention $\varepsilon^{01} = -\varepsilon_{01} = 1$, the property $\eta_{\beta\alpha}\varepsilon_{\mu\nu} + \eta_{\beta\mu}\varepsilon_{\nu\alpha} + \eta_{\beta\nu}\varepsilon_{\alpha\mu} = 0$ and chose the 2D gamma matrices in the form (3.12), giving[36]

$$\begin{aligned} \mathrm{Tr}\,\{\gamma_\mu\gamma_\nu\gamma^5\} &= 2\varepsilon_{\mu\nu}, \\ \mathrm{Tr}\,\{\gamma_\mu\gamma_\nu\gamma_\alpha\gamma_\beta\gamma^5\} &= \varepsilon_{\mu\nu}\,\eta_{\alpha\beta} + \varepsilon_{\alpha\beta}\,\eta_{\mu\nu} + \varepsilon_{\alpha\nu}\,\eta_{\mu\beta} + \varepsilon_{\mu\beta}\,\eta_{\alpha\nu}. \end{aligned} \tag{7.37}$$

Consider now the average value $\langle j_{\mu 5}\rangle_e$ in the presense of the background electric field $E = -\partial_1 A_0$. It is described by the diagram in Figure 7, which gives

$$\langle j_{\mu 5}\rangle_E = -iT_{\mu\nu}(q)(ie)A^\nu(q) = \frac{eq_\mu\varepsilon_{\nu\alpha}q^\alpha A^\nu(q)}{\pi q^2}. \tag{7.38}$$

Multiplying this by q^μ and going over into coordinate space, we derive

$$\langle \partial^\mu j_{\mu 5}\rangle_E = \frac{e}{\pi}\varepsilon^{\nu\alpha}\partial_\alpha A_\nu = -\frac{E}{\pi} = -\frac{e}{2\pi}\varepsilon^{\alpha\nu}F_{\alpha\nu}. \tag{7.39}$$

Absorbing the charge e in the definition of the field, we arrive at (3.13).

Going back to four dimensions, one can observe that the result (7.29) coincides with Eq. (7.36) up to the change $\varepsilon_{\nu\alpha} \to \tilde{\varepsilon}_{\nu\alpha}$ and with an extra factor $eH/2\pi$. The relations (7.30)–(7.32) also have their exact two-dimensional counterparts.

[36]Note that $\gamma^\alpha\gamma_\beta\gamma_\alpha = 0$ in two dimensions.

In particular, the anomaly relation (7.31) can be derived out of (7.30) in the same way as we derived (3.13) out of (7.36). To this end, it is sufficient to consider the kinematics when $\mathbf{E}\|\mathbf{H}$ and both fields are directed along the 3rd axis. Then on the place of (7.39), we obtain (7.31).

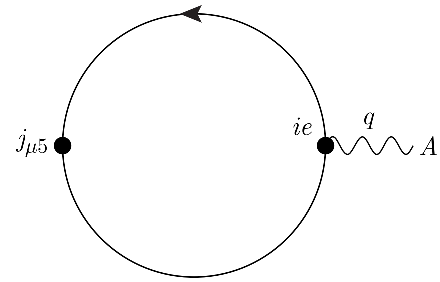

Figure 7. $\langle j_{\mu 5}\rangle_E$ in the Schwinger model.

The physical interpretations in 4D and 2D are practically identical: an external electromagnetic field with nonzero $\int dtd^3x\, \mathbf{E}\cdot\mathbf{H}$ in 4D or an electric field with nonzero $\int dtdx\, E$ in 2D bring about the change of axial charge associated with the creation of soft massless fermion–antifermion pairs.

As a side remark, note that in two dimensions, vector and axial currents are related as $j_{\mu 5} = -\varepsilon_{\mu\nu} j^\nu$, and the singularity of the correlator (7.34) means also the singularity of the vector polarization operator $\Pi_{\mu\nu} \propto (\eta_{\mu\nu} - q_\mu q_\nu/q^2)$, i.e. nonvanishing $\Pi(0)$. As a result, the Schwinger photon acquires the mass,

$$\mu^2 = g^2/\pi. \tag{7.40}$$

Actually, the only physical states in the massless Schwinger model are *free bosons* with mass (7.40). And that means that the fermion fields entering the Lagrangian of the model are confined!

7.4. Chiral symmetry breaking and confinement

Look again at Eq. (3.5). The axial current entering the left-hand side is an external current in a sense that no dynamical field is coupled directly to $j^{\mu 5}$. But the fields entering on the right-hand side are dynamical gluon fields present in the QCD Lagrangian.

In the chiral (left-right asymmetric) gauge theories like the standard electroweak model, both vector and axial currents are coupled directly to the physical gauge fields. Anomalous divergence of such current would mean explicit breaking of the gauge invariance, which is not nice. Therefore, in chiral theories, one should always take care that such purely *internal anomalies* would cancel out at the end of the day. In the Standard Model, they do.

Let us discuss, however, *external* anomalies in QCD which are not related to breaking gauge symmetry but only mean that certain correlators involving external currents are singular.

As a simplest nontrivial example, consider the theory with two massless flavors and look at the correlator

$$K^{ab,\mathcal{H}}_{\mu\nu}(q) = i \int \langle T\{j^a_{\mu 5}(x) j^b_\nu(0)\}\rangle_{\mathcal{H}}\, e^{iq\cdot x} d^4x, \tag{7.41}$$

where a, b are flavor indices and $\mathcal{H}$ is the external flavor-singlet "magnetic field"[37]. The correlator (7.41) is nothing but a three-point vacuum expectation value (7.28) in the kinematics where one of the external momenta associated with the vector current is set to zero.

The one-loop calculation of the corresponding graph displays a singularity,

$$K^{ab,\mathcal{H}}_{\mu\nu}(q) = -\frac{\mathcal{H}}{2\pi^2} \frac{q_\mu \tilde{\varepsilon}_{\nu\alpha} q^\alpha}{q^2} \times N_c \times \frac{1}{2}\delta^{ab}, \tag{7.42}$$

[37]The quotation marks distinguish $\mathcal{H}$, which is an isosinglet that couples to $B/3$, B being the baryon charge, from the physical magnetic field, which has the matrix structure diag $(2/3, -1/3)$ and is a mixture of isotriplet and isosinglet. But we are not interested in dynamics of electromagnetic or weak interactions here. In QCD proper, all color-singlet currents are external. $\mathcal{H}$ is just a source of such vector flavor-singlet current.

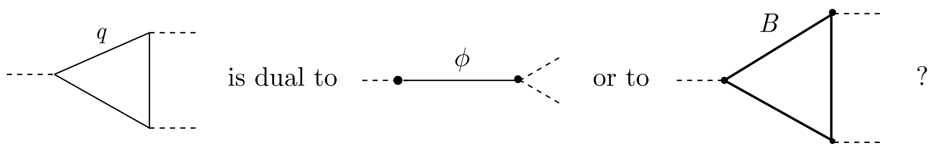

Figure 8. Saturating the external anomaly.

where the last two factors come from the color and flavor trace. The imaginary part of this amplitude is also singular, $\sim \delta(q^2)$, which can be related to the masslessness of quarks. However, the quarks (in contrast to electrons in QED) do not exist as physical particles due to confinement, and one can ask where does the singularity in $K_{\mu\nu}^{ab,\mathcal{H}}(q) \propto 1/q^2$ come from? This is a good question, and the answer is still better: the singularity $\sim 1/q^2$ comes from the propagator of a massless Goldstone boson, which appears due to the spontaneous chiral symmetry breaking and which is directly coupled to the axial current $j_{\mu 5}$ (see the middle graph in Figure 8).

Let us ask now: can one reproduce the singularity in Eq. (7.42) *without* Goldstone bosons and without spontaneous chiral symmetry breaking, but in some other way?

As far as the theory with two light quarks is concerned, the answer is positive: the singularity of the correlator above can be reproduced, in principle, if *massless baryons* are present. Proton and neutron represent, like quarks, a flavor $SU(2)$ doublet. There are $N_c = 3$ quark doublets and only one baryon doublet $|P\rangle = |uud\rangle$ and $|N\rangle = |udd\rangle$. The absence of the overall N_c factor is compensated, however, by the fact that the baryon charge of the nucleon is 3 times larger than that of the quark, and the vertex involving the "magnetic field" $\mathcal{H}$ is 3 times larger for baryons.

Thus, this purely algebraic *anomaly matching* argument due to 't Hooft [42] does not rule out a dynamical scenario where the physical spectrum in the theory with just two massless quark flavors would not involve massless pions but, instead, the massless proton and neutron. It is rather remarkable that, in the theories with $N_f \geqslant 3$, the scenario with massless baryons *is* ruled out. Suppose that, instead of the octet of massless Goldstone fields, we have an octet of massless baryons. The contribution of the corresponding triangle graph in (7.41) would have the same structure as in Eq. (7.42), but with the factor

$$\mathrm{Tr}\,\{T^a T^b\} = C_8 \delta^{ab} \tag{7.43}$$

instead of $\delta^{ab}/2$, where T^a are the flavor generators in the octet representation. To find the *Dynkin index* C_8 of the octet representation, it is sufficient to assume that $a = b = 1$ and decompose the octet with respect to the $SU(2)$ flavor subgroup: $\mathbf{8} = \mathbf{3} + \mathbf{2} + \mathbf{2} + \mathbf{1}$. The contribution of each doublet to C_8 is 1/2 and the contribution of the triplet is 2. Adding it together, we obtain $C_8 = 3$, which is not the same as 1/2, and the required result (7.42) is *not* reproduced. Also, a massless decuplet and all other possible color-singlet baryon representations would give the coefficient in front of the singularity much larger than that in Eq. (7.42), and the anomaly matching condition would not be fulfilled. Therefore, massless baryons do not exist[38].

We have arrived at a remarkable result. In QCD with three massless quarks, the assumption of confinement, allowing the existence of only colorless states in the physical spectrum, *and* the anomaly matching condition lead to the conclusion that massless Goldstone states *should* appear

[38] To be quite precise, one could, in principle, saturate the anomaly with *several* massless baryon multiplets with positive and negative baryon charges. This possibility is so unaesthetic, however, that it can be rejected simply by that reason. In addition, it is ruled out by a careful algebraic analysis [43].

and chiral symmetry *must* be spontaneously broken. If the latter is not true, the only possibility to saturate the anomaly is to assume that massless quarks still exist as physical states in the spectrum and there is no confinement!

In the real world, confinement and spontaneous chiral symmetry breaking in the limit of massless u, d, and s quarks are experimental facts. Whether or not these phenomena take place in hypothetic theories with $N_f \geqslant 4$ is an open question. It is quite possible that starting, say, from $N_f \stackrel{?}{=} 6$, the small eigenvalues in the Dirac operator spectrum are punished by the determinant factor so strongly that $\rho(0) = 0$ and, in view of the Banks–Casher relation (7.18), the quark condensate vanishes and the symmetry is not broken. The anomaly matching argument tells then that there is no confinement in this case[39].

We have called this result — that the chiral symmetry breaking and confinement go together — remarkable. It is also somewhat mysterious. Even though we do not understand well dynamical reasons for confinement to occur, we still expect that *the same* mechanism which works for the theory with $N_f = 3$ works also for the theory with $N_f = 2$. But 't Hooft's argument works only for $N_f \geqslant 3\ldots$

Acknowledgements

I am indebted to R. Narayanan and M. Zubkov for illuminating discussions.

References

[1] A. Smilga, Lectures on Quantum Chromodynamics, World Scientific, 2001.
[2] H. Leutwyler and A. Smilga, Spectrum of Dirac operator and role of winding number in QCD, Phys. Rev. D 46 (1992) 5607.
[3] A. V. Smilga, QCD at $\theta \sim \pi$, Phys. Rev. D 59 (1999) 114021.
[4] L. D. Landau and E. M. Lifshitz, The Classical Theory of Fields, Pergamon, Oxford, 1980.
[5] M. E. Peskin and D. V. Schroeder, An Introduction to Quantum Field Theory, Addison–Wesley, 1995.
[6] W. E. Caswell, Asymptotic behavior of non-Abelian gauge theories to two-loop order, Phys. Rev. Lett. 33 (1974) 244;
D. R. T. Jones, Two loop diagrams in Yang–Mills theory, Nucl. Phys. B 75 (1974) 531.
[7] V. Fock, Proper time in classical and quantum mechanics, Sov. Phys. 12 (1937) 404;
J. Schwinger, On gauge invariance and vacuum polarization, Phys. Rev. 82 (1951) 664.
[8] K. Fujikawa, Path integral for gauge theories with fermions, Phys. Rev. D 21 (1980) 2848.
[9] M. F. Atiyah and I. M. Singer, The index of elliptic operators, Annals Math. 87 (1968) 484, 546; 93 (1971) 119, 139.
[10] A. A. Belavin, A. M. Polyakov, A. S. Schwartz, and Yu. S. Tyupkin, Pseudoparticle solutions of the Yang–Mills equations, Phys. Lett. B 59 (1975) 85.
[11] L. Alvarez-Gaumé, Supersymmetry and the Atiyah–Singer index theorem, Commun. Math. Phys. 90 (1983) 161;

[39] We know, of course, that if the number of the quark flavors is *very* high $N_f > 16$, the asymptotic freedom is lost and we cannot expect confinement. In addition, we are almost sure that quarks and gluons are not confined at $N_f = 16$ or $N_f = 15$, in which case the theory is asymptotically free, but has the infrared fixed point at a small value of α_s, and the coupling constant never grows large. The argument about suppression of the small Dirac eigenvalues and the results of some numerical simulations indicate that confinement might be lost at a smaller value of N_f, not just $N_f = 15$.

D. Friedan and O. Windey, Supersymmetric derivation of the Atiyah–Singer index and the chiral anomaly, Nucl. Phys. B 235 (1984) 395.

[12] A. V. Smilga, Geometry Meets Supersymmetry, World Scientific, 2026.

[13] S. Cecotti and L. Girardello, Functional measure, topology and dynamical supersymmetry breaking, Phys. Lett. B 110 (1982) 69;
L. Girardello, C. Imbimbo, and S. Mukhi, On constant configurations and the evaluation of the Witten index, Phys. Lett. B 132 (1983) 69.

[14] G. 't Hooft, Computation of the quantum effects due to a four-dimensional pseudoparticle, Phys. Rev. D 14 (1976) 3432.

[15] M. Shifman (Ed.), Vacuum Structure and QCD Sum Rules: Reprints, North Holland, Amsterdam, 1992;
[For a review, see Ref. [1], Chapter 17.]

[16] C. McNeile et al., Direct determination of the strange and light quark condensates from full lattice QCD, Phys. Rev. D 87 (2013) 034503.

[17] Y. Nambu, Quasiparticles and gauge invariance in the theory of superconductivity, Phys. Rev. 117 (1960) 648;
J. Goldstone, Field theories with superconductor solutions, Nuovo Cimento 19 (1961) 154.

[18] M. Born and R. Oppenheimer, Zur Quantentheorie der Moleken, Annalen der Physik 20 (1927) 457.

[19] C. Vafa and E. Witten, Restrictions on symmetry breaking in vector-like gauge theories, Nucl. Phys. B 234 (1984) 173.

[20] M. Gell-Mann, R. J. Oakes, and B. Renner, Behavior of current divergences under $SU(3) \times SU(3)$, Phys. Rev. 175 (1968) 2195.

[21] H. Leutwyler, Chiral perturbation theory, Scholarpedia 7 (2012) 8708.

[22] S. I. Blinnikov and N. V. Nikitin, Approximations to path integrals and spectra of quantum systems. arXiv:physics/0309060.

[23] C. Gattringer and C. B. Lang, Quantum Chromodynamics on the Lattice, Springer, 2010.

[24] K. G. Wilson, Confinement of quarks, Phys. Rev. D 10 (1974) 2445.

[25] H. B. Nielsen and N. Ninomiya, No go theorem for regularizing chiral fermions, Phys. Lett. B 105 (1981) 219.

[26] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, Modern Geometry — Methods and Applications, Part 2, Springer-Verlag, 1984, p. 13.

[27] S. Kogut and L. Susskind, Hamiltonian formulation of Wilson's lattice gauge theories, Phys. Rev. D 11 (1975) 395.

[28] H. Neuberger, Exactly mass quarks on the lattice, Phys. Lett. B 417 (1998) 141.

[29] M. Lüscher, The index theorem in QCD with a finite cut-off, Phys. Lett. B 428 (1998) 342.

[30] P. H. Ginsparg and K. G. Wilson, A remnant of chiral symmetry on the lattice, Phys. Rev. D 25 (1982) 2649.

[31] H. Neuberger and R. Naranayan, A construction of lattice chiral gauge theories, Nucl. Phys. B 443 (1995) 305.

[32] P. Hasenfratz, V. Laliena, and F. Niedermayer, Exact chiral symmetry on the lattice and the Ginsparg–Wilson relation, Phys. Lett. B 427 (1998) 125.

[33] S. Nussinov and M. A. Lampert, QCD inequalities, Phys. Rep. 362 (2002) 193.

[34] J.-L. Kneur and A. Neveu, The chiral condensate from renormalization group optimized perturbation, Phys. Rev. D 92 (2015) 074027.

[35] T. Banks and A. Casher, Chiral symmetry breaking in confining theories, Nucl. Phys. B 169 (1980) 103.

[36] A. Smilga and J. Stern, On the spectral density of Euclidean Dirac operator in QCD, Phys. Lett. B 318 (1993) 531.

[37] J. Osborn, D. Toublan, and J. Verbaarschot, From chiral random matrix theory to chiral perturbation theory, Nucl. Phys. B 540 (1999) 317.

[38] J. Verbaarschot, Chiral random matrix theory and the spectrum of the Dirac operator near zero virtuality, Acta Phys. Polon. B 25 (1994) 133.

[39] A. D. Dolgov and V. I. Zakharov, On conservation of the axial current in massless electrodynamics, Nucl. Phys. B 27 (1971) 525.

[40] V. B. Berestetskii, E. M. Lifshitz, and L. P. Pitaevskii, Quantum Electrodynamics, §9, Butterworth-Heinemann, 1982.

[41] J. Ambjorn, J. Greensite, and C. Peterson, The axial anomaly and the lattice Dirac sea, Nucl. Phys. B 221 (1983) 381.

[42] G. 't Hooft, Naturalness, Chiral Symmetry and Spontaneous Chiral Symmetry Breaking, in: [G. 't Hooft (Ed.), Recent Developments in Gauge Theories, Plenum Press, 1980.]

[43] Ling-Xiao Xu, To break or not to break: A review of a no-go theorem on chiral symmetry breaking in QCD-like theories, Int. J. Mod. Phys. A 41 (2026) 2630001 and references therein.