

ANOMALIES IN GAUGE THEORIES*

M.A. SHIFMAN**

*Institute for Theoretical Physics, Bern University, Sidlerstrasse 5, 3012 Bern, Switzerland
and Theoretical Physics Institute, University of Minnesota, 116 Church St. S.E., Minneapolis, MN 55455, USA**

Editor: R. Petronzio

Received December 1990

Contents:

Introduction	343	2.3. Multiloop corrections	365
1. Chiral anomaly in the Schwinger model	344	2.4. Anomalies and external currents	366
1.1. Schwinger model on a circle	344	3. Low-energy theorems; applications to hadronic physics	369
1.2. Dirac sea: the vacuum wave function	346	3.1. The scale Ward identities	370
1.3. Ultraviolet regularization	349	3.2. Conversion of gluons into pions and photons	373
1.4. The theta vacuum	351	3.3. The pion spectrum in the decays $\psi' \rightarrow J/\psi \pi \pi$, $Y' \rightarrow Y \pi \pi$	375
1.5. Two faces of the anomaly	356	4. Conclusions	377
2. Anomalies in quantum chromodynamics	360	References	377
2.1. Classical symmetries	360		
2.2. The axial and scale (dilatation) anomalies	361		

Abstract:

We discuss the scale and chiral anomalies in quantum chromodynamics and their implications for the theory of hadrons. In the first part the physical meaning of the anomaly is demonstrated. To this end the simplest gauge model is considered – the two-dimensional Schwinger model. In this model it is extremely easy to explain which properties of the theory are responsible for the quantum anomaly, and the physical nature of the phenomenon is elucidated. The second part is devoted to the derivation of the anomaly relations within QCD. The subtle question of the multiloop corrections is discussed. The third part presents applications. Starting from the chiral and scale anomalies we obtain low-energy theorems which lead, in turn, to predictions for observable processes. In particular, a proof based on first principles is given of the existence of massless pions, the amplitudes are calculated for the conversion of gluon operators into pions, which can be measured experimentally. Other practical problems are also considered, the solution of which turns out to be possible due to anomalies.

* This review is a revised version of my review paper published in *Usp. Fiz. Nauk* 157 (1989) 561 (in Russian).

**** Work supported in part by Schweizerischer Nationalfonds.**

† Permanent address.

ANOMALIES IN GAUGE THEORIES

M.A. SHIFMAN

*Institute for Theoretical Physics, Bern University, Sidlerstrasse 5, 3012 Bern, Switzerland
and Theoretical Physics Institute, University of Minnesota, 116 Church St. S.E.,
Minneapolis, MN 55455, USA*



NORTH-HOLLAND

Introduction

In this review we will discuss quantum anomalies and the consequences stemming from them for the theory of hadrons. The term “quantum anomalies” in field theory has a concrete narrow meaning. Let us consider an action possessing certain invariance at the classical level. If this invariance cannot be preserved at the quantum level, i.e., taking account of the quantum corrections, such a phenomenon is called a “quantum anomaly”.

The first encounter of the theorists with the quantum anomalies occurred long before the QCD era (the famous $\pi^0 \rightarrow 2\gamma$ puzzle, see e.g. the review in ref. [1]). After the triumph of the gauge theories – the Glashow–Weinberg–Salam model in 1972 and QCD in 1973 – the status of the problem drastically changed: from a relatively local issue it turned into an important and universal theoretical construction occupying one of the central places in modern theory.

There are two types of anomalies; let us call them internal and external. In the first case the symmetry we deal with is the gauge symmetry – the gauge invariance of the classical action is violated at the quantum level. In other words, when quantum corrections (loops) are taken into account, the current with which gauge bosons interact ceases to be conserved.

External anomalies also result in current non-conservation. In this case, however, the anomalous current is not connected with the gauge bosons and corresponds to “external” global symmetries of the classical action.

The presence of the internal anomalies is usually considered to be a disaster for the theory. Such theories cannot be consistently quantized, they are non-renormalizable and self-contradictory. (Let us note, though, a series of articles [2] presenting an attempt to circumvent the problems. We will not deal with these works since the development here is far from completion.) The standard strategy consists of cancelling all internal anomalies by means of a special choice of the fields in the original Lagrangian. For instance, in the Glashow–Weinberg–Salam model the internal triangle anomaly cancels provided that the quarks and leptons from the given generation possess quite definite values of the hypercharge – namely, the standard values – and all the quarks and leptons in the generation take part in the cancellation. The requirement of the absence of internal anomalies imposes constraints on the structure of the gauge models. The constraints can serve (and do serve) as a guideline in model-building. Suffice it to recall that within the superstring approach, the present-day candidate for the “theory of everything”, the cancellation of the internal anomalies leaves only two possible options for the gauge group, SU_{32} and $E_8 \times E_8$ [3].

We will not dwell on these issues. Both topics – quantization of the theories with internal anomalies and cancellation of the internal anomalies in the course of model-building – call for a special discussion.

In this review we will concentrate on external anomalies. The main purpose is to explain the physical meaning of the phenomenon in the simplest and most transparent form and to demonstrate the wide-ranging consequences for the theory of hadrons stemming from the existence of two anomalies in QCD: the chiral anomaly and the scale anomaly. The attention which will be paid to pedagogical aspects and special emphasis on practical applications – both these are points that distinguish the discussion below from numerous excellent reviews on anomalies available in the literature. In particular, I would like to mention the review paper of ref. [4] where technical and mathematical details are fully discussed. One can also find there a representative list of references.

I begin with the simplest known example: two-dimensional quantum electrodynamics (QED). The example is very instructive and nicely illustrates the answers to such questions as:

– which properties of the theory are responsible for the occurrence of the quantum anomalies?

– what is the physical nature of the phenomenon?

The crucial importance of the following two points will be clearly seen:

(i) the presence of the infinite number of degrees of freedom in field theory and the necessity of introducing an ultraviolet cut-off; and

(ii) the clash between two classical symmetries in which only one will turn out to be a “victor” at the quantum level; the one which gains the victory will be preserved at a price of sacrificing the second, less “fortunate” symmetry. It is this latter symmetry that is the quantum anomaly.

The second part of the review is devoted to chiral and conformal anomalies in QCD. Here I present different derivations of the anomalies and discuss some subtle points which are usually skipped in the literature. Finally, the third part is devoted to the low-energy theorems following from the anomalous relations and to applications in hadronic physics. I will show here how one can use the anomalies for predicting such non-trivial quantities as, for instance, the gluon–pion conversion amplitudes.

1. Chiral anomaly in the Schwinger model

1.1. Schwinger model on a circle

Two-dimensional QED with the massless Dirac fermion seems to be the simplest gauge model. The Lagrangian is

$$\mathcal{L} = -(1/4e_0^2)F_{\mu\nu}F_{\mu\nu} + \bar{\psi}i\not{D}\psi, \quad (1.1)$$

where $F_{\mu\nu}$ is the photon field strength tensor,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu, \quad (1.2)$$

e_0 is the gauge coupling constant having the dimension of mass for $D = 2$. Moreover, D_μ is the covariant derivative

$$iD_\mu = i\partial_\mu - A_\mu, \quad (1.3)$$

and ψ is the two-component spinor field. Gamma matrices can be chosen in the following way:

$$\begin{aligned} \gamma^0 &= \sigma_2, & \gamma^1 &= i\sigma_1, & \gamma^5 &= \sigma_3, & \text{in Minkowski space,} \\ \gamma^1 &= \sigma_1, & \gamma^2 &= \sigma_2, & \gamma^5 &= \sigma_3, & \text{in Euclidean space.} \end{aligned} \quad (1.4)$$

Correspondingly, the spinor $\psi_L = \begin{pmatrix} \psi_1 \\ 0 \end{pmatrix}$ will be called left-handed ($\gamma^5\psi_L = \psi_L$), while the spinor $\psi_R = \begin{pmatrix} 0 \\ \psi_2 \end{pmatrix}$ will be called right-handed ($\gamma^5\psi_R = -\psi_R$).

In spite of the considerable simplification compared to the four-dimensional QED, the dynamics of the model (1.1) is still too complicated for our purposes. Indeed, the set of the asymptotic states in this model drastically differs from the fields in the Lagrangian. In the two-dimensional theory the photon, as is well-known, has no transversal degrees of freedom and essentially reduces to the Coulomb interaction. The latter, however, grows linearly with the distance. The linear growth of the Coulomb potential results in the confinement of the charged fermions in the Schwinger model irrespectively of

the value of the coupling constant e_0 . The model (1.1) was even used as a prototype for describing the color confinement in QCD (see e.g. ref. [5]).

In order to simplify the situation further let us do the following. Consider the system described by the Lagrangian (1.1) in a finite spatial domain of length L . If L is small, $e_0 L \ll 1$, the Coulomb interaction never becomes strong and one can actually treat it as a small perturbation. In particular, in the first approximation its effect can be neglected altogether. We impose periodic boundary conditions on the field A_μ and antiperiodic ones on ψ . Thus, the problem to be considered below is the Schwinger model on the circle. Notice that the antiperiodic boundary conditions are imposed on the fermion field for convenience only. As will be seen, any other boundary condition – periodic, for instance – would do as well; nothing would be changed except minor technical details. Thus

$$A(x = -\frac{1}{2}L, t) = A(x = \frac{1}{2}L, t), \quad \psi(x = -\frac{1}{2}L, t) = -\psi(x = \frac{1}{2}L, t). \quad (1.5)$$

Equation (1.5) implies that the fields A and ψ can be expanded in the Fourier modes, $\exp(ikx2\pi/L)$ for the bosons and $\exp[i(k + \frac{1}{2})x(2\pi/L)]$ for the fermions.

Now, let us recall the fact that the Lagrangian (1.1) is invariant under the local gauge transformations

$$\psi \rightarrow e^{i\alpha(x,t)}\psi, \quad A_\mu \rightarrow A_\mu - \partial_\mu \alpha(x, t). \quad (1.6)$$

It is quite evident that all modes for the field A_1 except the zero mode (i.e. $k = 0$) can be gauged away. Indeed, the term of the type $a(t) \exp(ikx2\pi/L)$ in A_1 is gauged away with the aid of the gauge function $\alpha(x, t) = -(ik)^{-1}a(t) \exp(ikx2\pi/L)$. The latter is periodic on the circle and does not violate the conditions (1.5), as it should be. Thus, in the most general case we can treat A_1 as an x -independent constant.

This is not the end of the story, however, since the possibilities provided by gauge invariance are not exhausted yet. There exists another class of admissible gauge transformations with the gauge function which is not periodic in x ,

$$\alpha = (2\pi/L)nx, \quad n = \pm 1, \pm 2, \dots, \quad (1.7)$$

where n is an integer. In spite of the non-periodicity, such a choice of the gauge function is also compatible with the conditions (1.5). This fact is readily verifiable; since $\partial\alpha/\partial x = \text{const.}$ and $\partial\alpha/\partial t = 0$ the periodicity for A_μ is not violated; the analogous assertion is also valid for the phase factor $\exp(i\alpha)$ – the difference of phases at the endpoints of the interval $x \in [-L/2, L/2]$ is equal to $2\pi n$.

As a result, we arrive at the conclusion that the variable A_1 (recall that in the sense of x -dependence A_1 is a constant) should not be considered in the whole interval $(-\infty, \infty)$. The points $A_1, A_1 = \pm 2\pi/L, A_1 = \pm 4\pi/L$, etc. are gauge equivalent and must be identified. The variable A_1 is an independent variable only in the interval $[0, 2\pi/L]$; going beyond these limits we find ourselves in the gauge image of the original interval. Following the commonly accepted terminology we may say that A_1 lives on the circle of length $2\pi/L$.

Moreover, the gauge invariance of the theory is closely interrelated with the conservation law for the electric charge. Indeed, the Lagrangian (1.1) with finite L admits multiplication of the fermion field by a constant phase,

$$\psi \rightarrow e^{i\alpha}\psi, \quad \psi^+ \rightarrow \psi^+ e^{-i\alpha}.$$

Using the standard line of reasoning one easily derives from this phase invariance the conservation of the electric current

$$j_\mu = \bar{\psi} \gamma_\mu \psi, \quad \dot{Q}(t) = 0, \quad Q = \int dx j_0(x, t).$$

(The vanishing of the current divergence stems also directly from the equations of motion.)

Observe that the classical Lagrangian (1.1) is invariant under one more rotation, the global axial transformation

$$\psi \rightarrow e^{i\alpha\gamma^5} \psi, \quad \psi^+ \rightarrow \psi^+ e^{-i\alpha\gamma^5},$$

which multiplies the left- and right-handed fermions by the opposite phases ($\gamma^5 = \sigma_3$). If the axial charge of the left-handed fermions $Q_5 = +1$, for the right-handed fermion $Q_5 = -1$. At the classical level the axial current $j_{\mu 5} = \bar{\psi} \gamma_\mu \gamma_5 \psi$ is conserved just in the same way as the electromagnetic one. The conservation of Q and Q_5 is equivalent to the conservation of the number of the left-handed and right-handed fermions separately. The fact is quite obvious for any Born graph. Indeed, in all such graphs the fermion lines are continuous, the photon emission does not change their chirality, and the number of in-going fermion legs is equal to that of the out-going legs. In the exact answer, however, only the sum of the chiral charges is conserved, only one out of two classical symmetries survives the quantization of the theory.

As will be seen below, the characteristic excitation frequencies for A_1 are of order e while those associated with the fermionic degrees of freedom are of order L^{-1} . Since $eL \ll 1$ the variable A_1 is adiabatic with respect to the fermionic degrees of freedom. In the next section we will analyze in more detail the fermion sector assuming temporarily that A_1 is a fixed (time-independent) quantity.

For our pedagogical purposes we can confine ourselves to the study of the limit $e_0 L \ll 1$. Those readers who would like to know about the solution of the Schwinger model on a circle with arbitrary L should turn to the original publications (e.g., ref. [43]).

1.2. Dirac sea: the vacuum wave function

Following the standard prescription of the adiabatic approximation we freeze the photon field A_μ and consider it as “external”. As for A_0 we can put $A_0 \approx 0$ (the fact that A_0 is actually non-vanishing results in negligible corrections). The difference between these two components lies in the fact that the fluctuations of A_0 are small, while this is not the case for A_1 . The wave function is not localized in A_1 in the vicinity of $A_1 = 0$. It is just this phenomenon – delocalization of the A_1 wave function and the possibility of penetration to large values of A_1 – that will lead to observable manifestations of the chiral anomaly. The $\mu = 0$ component of the photon field is responsible for the Coulomb interaction between the charges; the corresponding effect is of the order $eL \ll 1$ and does not show up in the leading approximation to which we will limit ourselves in the present section.

In the two-dimensional electrodynamics the Dirac equation determining the fermion energy levels has the form

$$[i\partial/\partial t + \sigma_3(i\partial/\partial x - A_1)]\psi = 0. \quad (1.8)$$

(Notice that $A_1 = -A^1$, etc.) For the k th stationary state $\psi \sim \exp(-iE_k t) \psi_k(x)$, and the energy of the k th state is

$$E_k \psi_k = -\sigma_3(i\partial/\partial x - A_1)\psi_k(x). \quad (1.9)$$

Furthermore, the eigenfunctions are proportional to

$$\psi_k \sim \exp[i(k + \frac{1}{2})(2\pi/L)x], \quad k = 0, \pm 1, \pm 2, \dots \quad (1.10)$$

The extra term $\frac{1}{2}(2\pi/L)x$ in the exponent ensures the antiperiodic boundary conditions, see eq. (1.5). As a result, we conclude that the energy of the k th level for the left-handed fermions [the definition is given following eq. (1.4)] is

$$E_{k(L)} = (k + \frac{1}{2})(2\pi/L) + A_1; \quad (1.11a)$$

while for the right-handed fermions

$$E_{k(R)} = -(k + \frac{1}{2})(2\pi/L) - A_1. \quad (1.11b)$$

The energy level dependence on A_1 is displayed in fig. 1. The dashed lines show the behavior of $E_{k(L)}$ and the solid lines are $E_{k(R)}$. At $A_1 = 0$ the energy levels for the left-handed and right-handed fermions are degenerate.

If A_1 increases, the degeneracy is lifted and the levels are split. At the point $A_1 = 2\pi/L$ the structure of the energy levels is precisely the same as for $A_1 = 0$; the degeneracy takes place again. The identity of the points $A_1 = 0$ and $A_1 = 2\pi/L$ is the remnant of the gauge invariance of the original theory (see the discussion above).

We note that the identity is achieved in a non-trivial way; in passing from $A_1 = 0$ to $A_1 = 2\pi/L$ a restructuring of the fermion levels takes place. All left-handed levels are shifted upwards by one interval while all right-handed levels are shifted downwards by the same one interval. This phenomenon, the restructuring of the fermion levels, lies in the basis of the chiral anomaly in the model at hand, as will become clear shortly.

Let us proceed from the one-particle Dirac equation to field theory. The first task is the construction

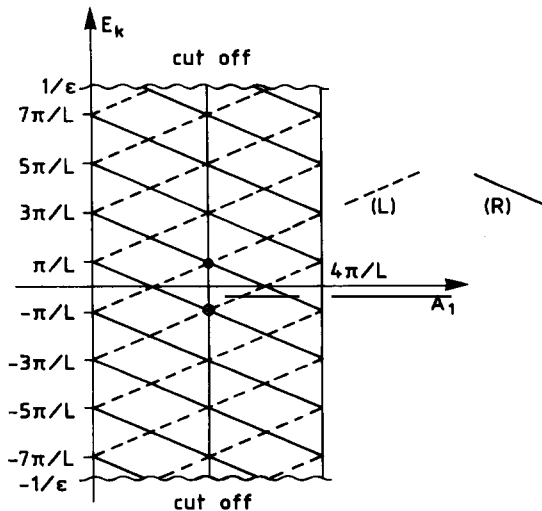


Fig. 1. Fermion energy levels as a function of A_1 .

of the ground state, the vacuum. To this end, following the well-known Dirac prescription we fill up all levels lying in the Dirac sea, leaving all positive-energy levels empty. The following notations will be used below for filled and empty levels with a given k : $|1_{L,R}, k\rangle$ and $|0_{L,R}, k\rangle$, respectively. The subscript L (R) indicates that we deal with the left-handed (right-handed) fermions.

Recall that A_1 is a slowly varying adiabatic variable; the corresponding quantum mechanics will be considered later. At first, the value of A_1 is fixed in the vicinity of zero, $A_1 \cong 0$. Then, the fermion wave function of the vacuum, as seen from fig. 1, reduces to

$$\Psi_{\text{ferm. vac.}} = \left(\prod_{k=-1, -2, \dots} |1_L, k\rangle \right) \left(\prod_{k=1, 2, \dots} |0_L, k\rangle \right) \left(\prod_{k=0, 1, 2, \dots} |1_R, k\rangle \right) \left(\prod_{k=-1, -2, \dots} |0_R, k\rangle \right). \quad (1.12)$$

The Dirac sea, or all negative-energy levels, are completely filled. Now let A_1 increase adiabatically from 0 to $2\pi/L$. The same figure shows that at $A_1 = 2\pi/L$ the wave function (1.12) describes that state which, from the standpoint of the normally filled Dirac sea, contains one left-handed particle and one-right-handed hole (small circles on fig. 1).

Do the quantum numbers of the fermion sector change in the process of the transition from $A_1 = 0$ to $A_1 = 2\pi/L$? Answering this question in the most naive manner, we would say that the appearance of the particle and the hole does not change the electric charge since the electric charges of the particle and the hole are obviously opposite. In other words, the electromagnetic current is conserved. On the other hand, the axial charges of the left-handed particle and the right-handed hole are the same ($Q_5 = 1$) and, hence, in the transition at hand

$$\Delta Q_5 = 2. \quad (1.13)$$

Equation (1.13) can be rewritten as $\Delta Q_5 = (L/\pi) \Delta A_1$. Dividing by Δt , the transition time, we get

$$\dot{Q}_5 = (L/\pi) \dot{A}_1, \quad (1.14)$$

which implies, in turn, that the conserved quantity has the form

$$\int dx \left(j^{05} - \frac{1}{\pi} A_1 \right). \quad (1.15)$$

The current corresponding to the charge (1.15) is obviously,

$$j^{\mu 5} = j^{\mu 5} - (1/\pi) \varepsilon^{\mu\nu} A_\nu, \quad \partial_\mu \tilde{j}^{\mu 5} = 0, \quad \partial_\mu j^{\mu 5} = (1/\pi) \varepsilon^{\mu\nu} \partial_\mu A_\nu, \quad (1.16)$$

where $\varepsilon^{\mu\nu}$ is the antisymmetric tensor, $\varepsilon^{01} = -\varepsilon^{10} = 1$. (Notice that $\varepsilon_{01} = -1$.) The second equality in (1.16) represents the famous axial anomaly in the Schwinger model. We succeeded in deriving it by the “hand-waving” arguments, by inspecting the picture of motion of the fermion levels in the external field $A_1(t)$. It turns out that in this language the chiral anomaly presents an extremely simple and widely known phenomenon; the crossing of the zero point in the energy scale by this or that level (or by a group of levels). The presence of the infinite number of levels and the Dirac “multiparticle” interpretation, according to which the emergence of the filled level from the sea means the appearance

of the particle while the submergence of the empty level into the sea is equivalent to production of a hole, are the most essential elements of the whole construction. With the finite number of levels, when there is no place for such an interpretation, there can be no quantum anomaly.

I would like also to draw the reader's attention to a somewhat different (although intimately related with the previous) aspect of the picture. The fermion levels move parallel to each other through the bulk of the Dirac sea. Therefore, the disappearance of the levels beyond the zero-energy mark occurs simultaneously with the disappearance of the "copies" beyond the ultraviolet cut-off, which is always implicitly present in the field theory (below we will introduce it explicitly). Because of this fact the heuristic derivation of the anomaly given in this section and a more standard treatment based on the ultraviolet regularization are actually one and the same. Often it turns out more convenient to trace just the crossing of the ultraviolet cut-off by the levels from the Dirac sea; and beyond the toy models, in the real theories like QCD, the latter approach becomes an absolute necessity, not a question of convenience, due to the notorious "infrared slavery". The connection between the ultraviolet and infrared interpretations of the anomaly is discussed in more detail in subsections 1.3 and 1.5. The interested reader is referred to the original work [6] where all subtle points are exhaustively analyzed.

1.3. Ultraviolet regularization

In spite of the transparent character of this heuristic derivation almost all of the "evident" points above can be questioned by the careful reader. Indeed, why is the wave function (1.12) the appropriate choice? In what sense is the energy of this state minimal, taking into account the fact that, according to (1.11),

$$E \sim - \sum_{m=0}^{\infty} (m + \frac{1}{2}) \frac{2\pi}{L},$$

and the sum is ill-defined (the series is divergent)? Moreover, it is usually asserted that the quantum anomalies are due to the necessity of the ultraviolet regularization of the theory. If so, why speak of the Dirac sea and the crossing of the zero-energy point by the fermion levels?

Surprising though it is, all these questions are connected with each other. Probably, it will be most convenient to start with the last one. Now I will explain that although the ultraviolet regularization was not even mentioned thus far, actually, it is the key element; more than that, the derivation sketched above tacitly assumes quite a specific regularization.

The fermion levels stretch in the energy scale up to indefinitely large energies (more exactly, $|E|$). The wave function (1.12) describing the fermion sector at $A_1 \cong 0$ contains, in particular, the direct product of the infinitely large number of the filled states $|1_R, k\rangle, |1_L, k\rangle$ with the negative energy. It is clear that such an object – the infinite product – is ill-defined, and one cannot do without this or that regularization in calculating physical quantities. The contribution corresponding to large energies (momenta) should be somehow cut off.

At first sight, it seems it would be sufficient simply to throw away the terms with $|k| > |k|_{\max}$ ($|k|_{\max}$ is a fixed number independent of A_1). This is a regularization, of course, but, clearly enough, the prescription will inevitably lead to a violation of the gauge invariance and the electric charge non-conservation. Indeed, in the gauge theories the momentum p always appears only in the combination $p + A$, and any gauge invariant cut-off should respect this property. In other words, we have the right to limit $p + A$, not p (or, what is the same, k).

In order to preserve the gauge invariance, it is possible and convenient to use the regularization called in the literature the Schwinger, or ε splitting. The regularization will provide a more solid mathematical basis to the heuristic derivation presented above. Instead of the original currents

$$j_\mu = \bar{\psi}(t, x) \gamma_\mu \psi(t, x), \quad j_{\mu 5} = \bar{\psi}(t, x) \gamma_\mu \gamma_5 \psi(t, x), \quad (1.17)$$

we introduce the regularized objects

$$j_\mu^{\text{Reg}} = \bar{\psi}(t, x + \varepsilon) \gamma_\mu \psi(t, x) \exp\left(-i \int_x^{x+\varepsilon} A_1 dx\right), \quad (1.18)$$

$$j_{\mu 5}^{\text{Reg}} = \bar{\psi}(t, x + \varepsilon) \gamma_\mu \gamma_5 \psi(t, x) \exp\left(-i \int_x^{x+\varepsilon} A_1 dx\right).$$

It is implied that $\varepsilon \rightarrow 0$ in the final answer for the physical quantities. At the intermediate stages, however, all computations are performed with fixed ε . The exponential factor in (1.18) ensures the gauge invariance of the “split” currents. Without this factor multiplying $\psi(x, t)$ by an x -dependent phase, $\psi(t, x) \rightarrow \exp[i\alpha(x)] \psi(t, x)$, yields

$$\psi_\alpha^+(t, x + \varepsilon) \psi_\beta(t, x) \rightarrow \exp[-i\alpha(x + \varepsilon) + i\alpha(x)] \psi_\alpha^+(t, x + \varepsilon) \psi_\beta(t, x). \quad (1.19)$$

The gauge transformation of A_1 ($A_1 \rightarrow A_1 - \partial\alpha/\partial x$) compensates for the phase factor in eq. (1.19).

Now, there is no difficulty to calculate the electric and axial charges of the state (1.12) “scientifically”. If

$$Q = \int dx j_0^{\text{Reg}}(t, x), \quad Q_5 = \int dx j_{05}^{\text{Reg}}(t, x), \quad (1.20)$$

then for the vacuum wave function we, evidently, get

$$Q = Q_L + Q_R, \quad Q_5 = Q_L - Q_R, \quad (1.21)$$

$$Q_L = \sum_k \exp\{-i\varepsilon[(k + \frac{1}{2})(2\pi/L) + A_1]\}, \quad (1.22)$$

$$Q_R = \sum_{k'} \exp\{-i\varepsilon[(k' + \frac{1}{2})(2\pi/L) + A_1]\}. \quad (1.23)$$

In the limit $\varepsilon \rightarrow 0$ both charges, Q_L and Q_R , turn into the sum of units – each unit represents one filled level from the Dirac sea. Equations (1.22) and (1.23) once again demonstrate the gauge invariance of the accepted regularization. Indeed, the cut-off suppresses the states with $|p_1 + A_1| \geq \varepsilon^{-1}$. If it were not for the phase factor in eq. (1.18), the suppressing function would not contain the desired combination, $p + A$.

We hasten to add here that although superficially eqs. (1.22) and (1.23) do not differ from each other, actually they do not coincide because the summation runs over different values of k, k' . Just what the particular values are is easy to establish from fig. 1 [see also eq. (1.12)].

Let $|A_1| < \pi/L$. Then in the “left-handed” sea the filled levels have $k = -1, -2, \dots$. In the “right-handed” sea the filled levels correspond to $k = 0, 1, 2, \dots$. Thus, if $|A_1| < \pi/L$ we have

$$Q_L = \sum_{k=-1}^{-\infty} \exp\{-i\varepsilon[(k + \tfrac{1}{2})(2\pi/L) + A_1]\}, \quad Q_R = \sum_{k=0}^{\infty} \exp\{-i\varepsilon[(k + \tfrac{1}{2})(2\pi/L) + A_1]\}. \quad (1.24)$$

Performing the summation and expanding in ε we arrive at

$$(Q_L)_{\text{vac}} = (-i\varepsilon 2\pi/L)^{-1} + (L/2\pi)A_1 + O(\varepsilon), \quad (Q_R)_{\text{vac}} = (i\varepsilon 2\pi/L)^{-1} - (L/2\pi)A_1 + O(\varepsilon), \quad (1.25)$$

We pause here to summarize the results. Equations (1.25) show that under our choice of the vacuum wave function $\Psi_{\text{ferm. vac.}}$ the charge of the vacuum vanishes, $Q = Q_L + Q_R = 0$; there is no time dependence, the charge is conserved. The axial charge consists of two terms: the first term represents an infinitely large *constant* and the second one gives a linear A_1 dependence. In the transition $(A_1 \cong 0) \rightarrow (A_1 \cong 2\pi/L)$ the axial charge changes by two units.

These conclusions are not new for us. We have found just the same from the illustrative picture described above in which the electric and axial charges of the Dirac sea are determined intuitively. Now we learned how to sum up the infinite series of units, $\sum_k 1$, the charges of the “left-handed” and “right-handed” seas, by virtue of the well-defined procedure which automatically cuts off the levels with $|p_1 + A_1| \gtrsim \varepsilon^{-1}$.

The procedure suggests an alternative language for describing the axial charge non-conservation in the transition $(A_1 \cong 0) \rightarrow (A_1 \cong 2\pi/L)$. Previously we thought that the non-conservation is due to the level crossing of the zero-energy point. It is equally correct to say – as we see now – that the non-conservation is explained by the following: one right-handed level from the sea leaves the “fiducial domain” via the lower boundary (the cut-off $-\varepsilon^{-1}$) and one new left-handed level appears in the sea through the same boundary (fig. 1). Both phenomena, though – the crossing of the zero-energy point and the departure (arrival) of the levels via the ultraviolet cut-off – occur simultaneously and represent, actually, two different facets of one and the same anomaly, which admits both, the infrared and ultraviolet interpretations. The connection between the ultraviolet and infrared aspects of the quantum anomaly lies in the basis of the so-called ‘t Hooft consistency conditions, to be discussed in section 2.4.

One last remark concerning the axial charge is in order. Instead of eq. (1.18) one could regularize the axial charge in a different way, so that $\partial_\mu j_{\mu 5} = 0$ and $\Delta Q_5 = 0$. (A nice exercise for the reader!) Under such a regularization, however, the expression for the axial current would not be gauge invariant. Specifically, the conserved axial current, apart from eq. (1.18), includes an extra term $(-1/\pi)\varepsilon^{\mu\nu}A_\mu$ [cf. eq. (1.16)]. As has already been mentioned, there is no regularization ensuring simultaneously the gauge invariance and conservation of $j_{\mu 5}$.

1.4. The theta vacuum

Now, we leave the issue of charges and proceed to the calculation of the fermion-sea energy, the problem which could not be solved at the naive level, without regularization. Fortunately, all necessary elements are already prepared.

The fermion part of the Hamiltonian [cf. eq. (1.8)],

$$H = -\psi^+(t, x) \sigma_3 (\partial/\partial x - A_1) \psi(t, x),$$

reduces after the ε splitting to

$$H^{\text{Reg}} = -\psi^+(t, x + \varepsilon) \sigma_3 \left(i \frac{\partial}{\partial x} - A_1 \right) \psi(t, x) \exp \left(-i \int_x^{x+\varepsilon} A_1 dx \right). \quad (1.26)$$

This formula implies, in turn, the following regularized expression for the energies of the “left-handed” and “right-handed” seas:

$$E_L = \sum_{k=-1}^{-\infty} E_{k(L)} \exp(-i\varepsilon E_{k(L)}), \quad E_R = \sum_{k=0}^{\infty} E_{k(R)} \exp(i\varepsilon E_{k(R)}), \quad (1.27)$$

where the energies of the individual levels $E_{k(L,R)}$ are given in eq. (1.11) and summation runs over all levels with the negative energy. The concrete values of the summation indices in eq. (1.27) correspond to $|A_1| < \pi/L$. Expressions (1.27) have an absolutely obvious meaning: in the limit $\varepsilon \rightarrow 0$ they simply reduce to the sum of the energies of all filled fermion levels from the Dirac sea. The additional exponential factors guarantee the convergence of the sums.

Furthermore, we notice that E_L and E_R can be obtained by differentiating the expressions (1.24) for $Q_{L,R}$ with respect to ε . The latter – the geometrical progression – are trivially summable. Expanding in ε we get

$$E_{\text{sea}} = E_L + E_R = (L/2\pi)(A_1^2 - \pi^2/L^2) + O(\varepsilon) + (\text{a constant independent of } A_1). \quad (1.28)$$

Two remarks are in order here. First, it is instructive to check that the Born–Oppenheimer approximation, accepted from the very beginning, is indeed justified. In other words, let us verify that the dynamics of the variable A_1 is slow in the scale characteristic of the fermion sector.

The effective Lagrangian determining the quantum mechanics of A_1 is

$$\mathcal{L} = (L/2e_0^2) \dot{A}_1^2 - (L/2\pi) A_1^2. \quad (1.29)$$

This is the ordinary harmonic oscillator with the wave function of the ground state,

$$\Psi_0(A_1) = (L/e_0 \pi^{3/2})^{1/4} \exp(-LA_1^2/2e_0\sqrt{\pi}), \quad (1.30)$$

and the level splitting,

$$\omega_A = e_0/\sqrt{\pi}. \quad (1.31)$$

The characteristic frequencies in the fermion sector are $\omega_{\text{ferm}} \sim L^{-1}$. Hence, $\omega_A/\omega_{\text{ferm}} \sim Le_0 \ll 1$.

The second remark concerns the structure of the total vacuum wave function. We have convinced ourselves that

$$\Psi_{\text{vac}} = \Psi_{\text{ferm. vac.}} \Psi_0(A_1) \quad (1.32)$$

is the eigenstate of the Hamiltonian of the Schwinger model on the circle in the Born–Oppenheimer approximation. The wave function (1.32) is quite satisfactory from the point of view of the “small” gauge transformations, i.e., those continuously deformable to the trivial (unit) transformation. [More exactly, eq. (1.32) refers to the specific gauge in which the gauge degrees of freedom associated with A_1 are eliminated and A_1 is independent of x .] This wave function, however, is not invariant under the “large” gauge transformations $A_1 \rightarrow A_1 + (2\pi/L)k$, $k = \pm 1, \pm 2, \dots$

The essence of the situation becomes clear if we return to fig. 1. When A_1 performs small and slow oscillations in the vicinity of zero, the Dirac sea is filled in such a way as shown in eq. (1.12). But A_1 can oscillate as well in the vicinity of the gauge equivalent point $A_1 = 2\pi/L$. In this case, if we do *not* restructure the fermion sector and leave it just as in eq. (1.12), then the configuration of eq. (1.12) is obviously *not* the vacuum – it corresponds to one particle plus one hole. The assertion is confirmed, in particular, by the plot showing the Dirac sea energy as a function of A_1 (fig. 2). In order to get the configuration with the lower energy it is necessary to fill the fermion levels as follows:

$$\prod_{k=-2,-3,\dots} |1_L, k\rangle \prod_{k=-1,0,1,\dots} |1_R, k\rangle$$

[the empty levels are not shown explicitly, cf. eq. (1.12)].

Thus, the Hilbert space is naturally split into distinct sectors corresponding to different structure of the fermion sea. The wave function of the ground state in the n th sector has the form

$$\Psi_n = \left(\prod_{k=-1-n}^{-\infty} |1_L, k\rangle \right) \left(\prod_{k=-n}^{\infty} |1_R, k\rangle \right) \Psi_0 \left(A_1 - \frac{2\pi}{L} n \right), \quad n = 0, \pm 1, \pm 2, \dots \quad (1.33)$$

The organization of the fermion sea is correlated with the position of the “center of oscillation” of A_1 . It is quite evident that Ψ_n and $\Psi_{n'}$ are strictly orthogonal to each other due to the fermion factors if $n \neq n'$.

Is it possible to construct the vacuum wave function invariant under the “large” gauge transformations $A_1 \rightarrow A_1 + (2\pi/L)k$ (with the simultaneous renumbering of the fermion levels)? The answer is yes. Moreover, such a wave function is not unique. It depends on a new hidden parameter θ which is often called the vacuum angle in the literature. Consider the linear combination

$$\Psi_{\theta \text{ vac}} = \sum_n [\exp(in\theta)] \Psi_n. \quad (1.34)$$

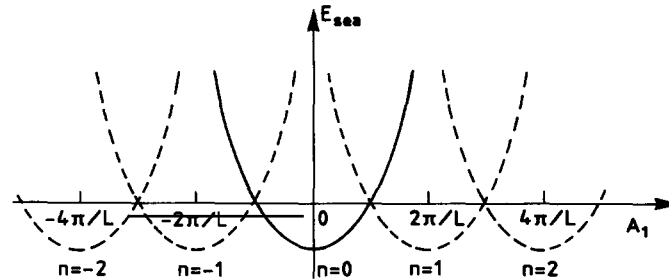


Fig. 2. Energy of the Dirac sea in the Schwinger model on a circle. The solid line corresponds to eq. (1.12). Dashed lines reflect restructuring of the Dirac sea necessary if $|A_1| > \pi/L$.

This linear combination is also the eigenfunction of the Hamiltonian with the lowest energy, just in the same way as Ψ_n . But unlike Ψ_n the “large” gauge transformations leave $\Psi_{\theta \text{ vac}}$ essentially intact. More exactly, under $A_1 \rightarrow A_1 + (2\pi/L)$ the wave function (1.34) is multiplied by $\exp(i\theta)$. This overall phase of the wave function is unobservable; all physical quantities resulting from averaging over the θ -vacuum are invariant under the gauge transformations.

Summarizing, we have now become acquainted with one more notion, the vacuum angle θ , the θ -vacuum, which is absolutely transparent in the Schwinger model on the circle and has a direct analogue in QCD. Notice that the presence of the vacuum angle θ in the wave function is imitated in the Lagrangian language by adding the so-called topological density to the Lagrangian. In the Schwinger model the topological density is

$$\Delta\mathcal{L}_\theta = (\theta/2\pi)\epsilon^{\mu\nu}F_{\mu\nu} = (\theta/\pi)\epsilon^{\mu\nu}\partial_\mu A_\nu. \quad (1.35)$$

The corresponding extra term in the action is the integral over the full derivative: it does not affect the equations of motion and gives a vanishing contribution for any topologically trivial configuration $A(x, t)$. The topological density $\Delta\mathcal{L}_\theta$ shows up only if

$$\int_{-L/2}^{L/2} dx [A_1(x, t = +\infty) - A_1(x, t = -\infty)] = 2\pi k, \quad |k| = 1, 2, \dots$$

The topological properties are mentioned here not by chance. It is very instructive to discuss the topological aspect of the theoretical construction under consideration, the more so because in this point as well there is a direct parallel with QCD.

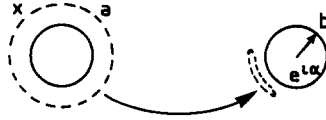
The model at hand possesses the U(1) gauge invariance. The element of the U(1) group, as is well-known, can be written as $\exp(i\alpha)$. Using the gauge freedom one can reduce the fields $A_1(x, t)$ or $\psi(x, t)$ at a given moment of time to a standard form by choosing an appropriate gauge function $\alpha(x)$. (For instance, the standard form of A_1 is $A_1 = \text{const.}$)

Moreover, under our boundary conditions the variable x represents the circle of length L , and, consequently, we deal here with the (continuous) mappings of the circle into the gauge group U(1). The set of the mappings can be divided in classes. The mathematical formula expressing the fact that the mappings are decomposed into classes is

$$\pi_1(\text{U}(1)) = \mathbb{Z}. \quad (1.36)$$

The meaning of eq. (1.36) is very simple. Inside each class all mappings, by definition, can be reduced to each other by continuous deformations. On the other hand, no continuous deformations transform mappings from one class into those belonging to the other class.

When the mappings of the circle on U(1) are considered, the difference between the classes is especially transparent. Actually, in this case the circle is mapped on another circle because $\exp(i\alpha)$ topologically is nothing else than the circle (fig. 3). Assume that we started from a certain point, went around the circle a once and returned to the starting point. In doing so, we simultaneously went around the circle $b, 0, \pm 1, \pm 2$, etc. times. (The negative sign corresponds to circulation in the opposite direction.) The number of windings around the circle b is just the class of the mapping. It is quite clear that all mappings with the given winding number are continuously deformable into each other. On the

Fig. 3. Mapping of the circle in the coordinate space into the $U(1)$ group.

contrary, different winding numbers guarantee that the continuous deformation is impossible. The letter \mathbb{Z} in eq. (1.36) denotes the set of integers and shows that the set of different mapping classes is isomorphic to the set of integers; each class is characterized by an integer having the meaning of the winding number. The mappings corresponding to the zero winding number are called topologically trivial, the others are topologically non-trivial.

This information is sufficient to establish the existence of the vacuum sectors labelled by n ($n = 0, \pm 1, \pm 2, \dots$) for which $(A_\mu)_{\text{vac}} \sim \partial_\mu \alpha_{(n)}$ without any explicit construction like (1.33) [$\alpha_{(n)}$ belongs to the n th class]. The necessity of introducing the vacuum angle θ also stems from this information.

Finally, the last issue to be discussed in connection with the Schwinger model is as follows. Sometimes the question is raised as to why the vacuum wave function cannot be chosen in the form (1.33) with fixed n . The gauge invariance under the “small” (topologically trivial) transformations is preserved which automatically implies the electric charge conservation. What is lost is only the invariance under the “large” (topologically non-trivial) transformations; it seems that there is nothing bad in that. Then, why is it absolutely necessary to pass to $\Psi_{\theta \text{ vac}} = \sum_n \exp(in\theta) \Psi_n$?

The point is that Ψ_n taken as the vacuum wave function violates clusterization – one of the basic properties in field theory which can be traced back to causality and unitarity of the theory. The following is understood by clusterization: the vacuum expectation value of the T product of several local operators must be reducible to the sum over intermediate states including the vacuum intermediate state plus excitations over the *given vacuum*. The violation of the clusterization can be demonstrated explicitly. Consider the two-point function

$$\mathcal{A}(t) = \langle \Psi_n | T \{ O^+(t), O(0) \} | \Psi_n \rangle, \quad O(t) = \int \bar{\psi}(x, t) (1 + \gamma^5) \psi(x, t) dx. \quad (1.37)$$

The operator O changes the axial charge of the state by two units (adds a particle and a hole to the Dirac sea), O^+ returns it back, and, as a result, $\mathcal{A}(t) \neq 0$. Moreover, if $t \rightarrow \infty$ in the Euclidean domain $\mathcal{A}(t) \rightarrow \text{const.}$ (For a concrete calculation see, e.g., ref. [5] based on the bosonization method. In this work the limit $L \rightarrow \infty$ is considered but all relevant expressions can be readily rewritten for finite L .) The fact that $\mathcal{A}(t)$ tends to a non-vanishing constant at $T \rightarrow \infty$ means, according to clusterization, that the operators $\bar{\psi}(1 \pm \gamma^5)\psi$ acquire a non-vanishing vacuum expectation value.

On the other hand, if $|\text{vac}\rangle = |\Psi_n\rangle$ then $\langle \bar{\psi}(1 \pm \gamma^5)\psi \rangle = 0$ for a trivial reason. Indeed, the operator $\bar{\psi}(1 + \gamma^5)\psi$ acting on Ψ_n produces an electron and a hole, and the corresponding state is, obviously orthogonal to Ψ_n^* .

The clusterization property restores itself if from Ψ_n one passes to the θ -vacuum (1.34). In this case there emerges the non-diagonal expectation value

$$\langle \Psi_{n+1}^* | \bar{\psi}(1 \pm \gamma^5)\psi | \Psi_n \rangle \sim L^{-1} \exp(-\pi^{3/2}/Le_0) \exp(i\theta). \quad (1.38)$$

If the line of reasoning appealing to the necessity of clusterization will seem too academic to the

reader, it might be instructive to consider another argument [connected with eq. (1.37) and the subsequent discussion]. Let us ask the question: what will happen if instead of the massless Schwinger model we will consider the model with a small mass, i.e., introduce an extra mass term $\Delta\mathcal{L}_m = -m\bar{\psi}\psi$ in the Lagrangian (1.1)? Naturally, all physical quantities obtained in the massless model will be shifted. It is equally natural to require, however, the shifts to be small for small m , so that there would be no change in the limit $m \rightarrow 0$. Otherwise, we would encounter an unstable situation while we would like to have the mass term as a small perturbation.

But in the presence of the degenerate states (and the states Ψ_n with different n are degenerate) any perturbation is potentially dangerous and can lead to large effects. Just such a disaster occurs, in particular, if $\Delta\mathcal{L}_m$, acting on the vacuum, is non-diagonal.

If we prescribe the states like Ψ_n to be the vacuum then $\Delta\mathcal{L}_m$ will by no means be diagonal, as follows from the discussion after eq. (1.37). This we cannot accept. On the other hand, the mass term is certainly diagonalized in the basis of the wave functions (1.34),

$$\langle \Psi_{\theta' \text{ vac}} | \Delta\mathcal{L}_m | \Psi_{\theta \text{ vac}} \rangle = 0 \quad \text{if} \quad \theta' \neq \theta.$$

1.5. Two faces of the anomaly

In conclusion, it will be extremely useful to discuss the connection between the picture presented above and the more standard derivation of the chiral anomaly in the Schwinger model. This discussion will represent a bridge between the physical picture described above and the standard approach to anomalies in QCD and other gauge theories.

We have already emphasized the double nature of the anomaly which shows up as the infrared effect in the current and the ultraviolet effect in the divergence of the current. The line of reasoning accepted thus far puts more emphasis on the infrared aspect of the problem – the finite “box” served as a natural infrared regularization. The same result for $\partial_\mu j_{\mu 5}$ as in eq. (1.16) could be obtained with no reference to the infrared regularization.

A conventional treatment of the issue is based on the standard Feynman diagram technique. The usual explanation one can find in numerous textbooks connects the anomalies to the ultraviolet divergence of certain Feynman graphs. The assertion of the ultraviolet divergence is valid if one deals directly with $\partial_\mu j^{\mu 5}$. Thus, the emphasis is shifted to the ultraviolet aspect of the anomaly.

Below, first of all, we will sketch the standard derivation. Then we will show that the diagrammatic language used, as a rule, for the analysis of $\partial_\mu j_{\mu 5}$ from the point of view of the ultraviolet regularization can be successfully used for the “infrared” derivation of the anomaly. The fact that the anomalies reveal themselves in the infrared behavior of Feynman graphs is rarely mentioned in the literature, and, hence, deserves a more detailed discussion. The pragmatically oriented reader *can omit this section in the first reading*.

Thus, we would like to demonstrate that

$$\partial_\mu j^{\mu 5} = (1/2\pi) \varepsilon^{\mu\nu} F_{\mu\nu}, \quad (1.39)$$

by considering directly $\partial_\mu j^{\mu 5}$, not $j_{\mu 5}$ as previously. Then we need to bother only about the ultraviolet regularization, and, in particular, the theory can be considered in the infinite space since the finiteness of L does not affect the result coming from the short distances.

One of the convenient methods of the ultraviolet regularization is due to Pauli and Villars. In the

model at hand it reduces to the following. In addition to the original massless fermions in the Lagrangian, heavy regulator fermions are introduced with the mass M_0 ($M_0 \rightarrow \infty$) and the opposite metric. The latter means that each loop of the regulator fermions is supplied by an extra minus sign relatively to the normal fermion loop. The interaction of the regulator fermions with the photons is assumed to be just the same as for the original fermions, and the only difference is the mass. Then the role of the Pauli–Villars fermions in the low-energy processes ($E \ll M_0$) is the introduction of the ultraviolet cut-off in the formally divergent integrals corresponding to the fermion loops. Such a regularization procedure, clearly, automatically guarantees the gauge invariance and the electromagnetic current conservation.

In the model regularized according to Pauli and Villars the axial current has the form

$$j^{\mu 5} = \bar{\psi} \gamma^\mu \gamma^5 \psi + \bar{R} \gamma^\mu \gamma^5 R, \quad (1.40)$$

where R is the fermion regulator. In calculating the divergence of the regularized current the naive equations of motion can be used. Then

$$\partial_\mu j^{\mu 5} = 2iM_0 \bar{R} \gamma^5 R.$$

The divergence is non-vanishing (the axial current is not conserved!), but, as was expected, $\partial_\mu j^{\mu 5}$ contains only the regulator anomalous term.

The last step is contracting the regulator fields in the loop in order to convert $M_0 \bar{R} \gamma^5 R$ in the “normal” light fields in the limit $M_0 \rightarrow \infty$. The relevant diagrams are displayed in fig. 4 where the solid line denotes the standard heavy fermion propagator $(p\gamma - M_0)^{-1}$. The graph in fig. 4a does not depend on the external field. The corresponding contribution to $\partial_\mu j^{\mu 5}$ represents a number which can be set equal to zero. The graph in fig. 4c, with two photon legs, and all others having more legs die off in the limit $M_0 \rightarrow \infty$. The only surviving graph is that of fig. 4b. Calculation of this diagram is trivial,

$$2iM_0 \bar{R} \gamma^5 R \rightarrow (1/\pi) \varepsilon^{\mu\nu} \partial_\mu A_\nu. \quad (1.41)$$

As a result, we reproduce the anomalous relation $\partial_\mu j^{\mu 5} = (1/\pi) \varepsilon^{\mu\nu} \partial_\mu A_\nu$ obtained previously by a different method.

The easiest method allowing one to check eq. (1.41) is, probably, the so-called background field technique. We will not explain it here because the corresponding explanations would lead us far astray. The interested reader is referred to the review paper of ref. [7] where all relevant details are fully

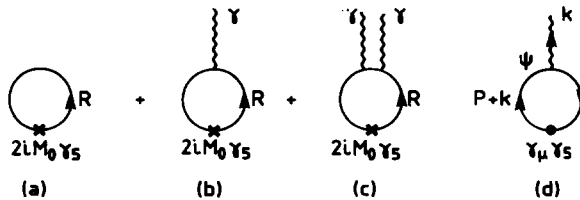


Fig. 4. Diagrammatic representation of the anomaly in the axial current in the Schwinger model. (a), (b), (c): Heavy regulator fields in the divergence of the current. (d): infrared singular anomalous contribution in $\bar{\psi} \gamma^\mu \gamma^5 \psi$.

discussed. Using the notations used in this paper one can write

$$2iM_0\bar{R}\gamma^5 R = -2M_0 \text{Tr}[\gamma^5(\not{P} - M_0)], \quad (1.42)$$

where $\not{P} = P_\mu \gamma^\mu$, P_μ is the generalized momentum operator, and we have taken into account the fact that the minus sign in the fermion loop does not appear for the regulator fields.

Moreover,

$$\begin{aligned} (\not{P} - M_0)^{-1} &= (\not{P} + M_0)(\not{P}\not{P} - M_0^2)^{-1} \\ &= (\not{P} + M_0)(P^2 - \tfrac{1}{2}i\varepsilon^{\mu\nu}F_{\mu\nu}\gamma^5 - M_0^2)^{-1}. \end{aligned} \quad (1.43)$$

The second equality here stems from the definition of the γ matrices (eq. 1.4) and the following property:

$$[P_\mu P_\nu] = -[D_\mu D_\nu] = iF_{\mu\nu}. \quad (1.44)$$

Now, since $M_0 \rightarrow \infty$ the trace in eq. (1.42) can be expanded in the inverse powers of M_0 ,

$$\text{Tr}[\gamma^5(\not{P} - M_0)^{-1}] = \text{Tr}\left[\gamma^5(\not{P} + M_0)\left(\frac{1}{P^2 - M_0^2} + \frac{1}{P^2 - M_0^2} \tfrac{1}{2}i\varepsilon^{\mu\nu}F_{\mu\nu}\gamma^5 \frac{1}{P^2 - M_0^2} + \dots\right)\right]. \quad (1.45)$$

The first term in the expansion vanishes after taking the trace of the γ matrices. The third and all other terms are irrelevant because they vanish in the limit $M_0 \rightarrow \infty$. The only relevant term is the second one where we can substitute the operator $P_\mu = iD_\mu = i\partial_\mu - A_\mu$ by the momentum p_μ since the result is explicitly proportional to the background field $F_{\mu\nu}$, and the chiral anomaly in the Schwinger model is linear in $F_{\mu\nu}$. Then

$$2iM_0\bar{R}\gamma_5 R = -2iM_0^2 \int \frac{d^2 p}{(2\pi)^2} \frac{1}{(p^2 - M_0^2)^2} F_{\mu\nu} \varepsilon^{\mu\nu}.$$

Performing the Wick rotation and integrating over p we arrive at eqs. (1.41) and (1.16). This computation completes the standard derivation of the anomaly. One should have a very rich imagination to be able to see in these formal manipulations the simple physical nature of the phenomenon which has been described above (restructuring of the Fermion sea and the level crossing). Nevertheless, this is the same phenomenon viewed from a different angle – less transparent but more economic since we can get the final result very quickly using the well-developed machinery of the diagram technique, familiar to everybody.

Let us ask the following question: “What is the infrared origin (or infrared face, if you wish) of the anomaly in the diagram language?” To extract the infrared aspect of the anomaly from the Feynman graphs it is necessary to turn back to consideration of the current $j_{\mu 5}$. Our aim is to calculate the matrix element of the current $j_{\mu 5}$ in the background photon field. Unlike $\partial_\mu j^{\mu 5}$ the matrix element $\langle j^{\mu 5} \rangle$ contains an infrared contribution. Because of this it is impossible to consider $\langle j^{\mu 5} \rangle$ for the on-mass-

shell photon, with the momentum $k^2 = 0$. We are forced to introduce “off-shellness” to ensure the infrared regularization (a substitute for finite L , see above). Thus, we will consider the photon field A_μ which does not obey the equations of motion.

General arguments (such as gauge invariance) imply the following expression for the matrix element $\langle j^{\mu 5} \rangle$ stemming from the diagram in fig. 4d:

$$\langle j^{\mu 5} \rangle = \text{const.} (k^\mu / k^2) \varepsilon^{\alpha\beta} F_{\alpha\beta}, \quad (1.46)$$

where the constant in the right-hand side can be determined by an explicit computation of the graph. In principle, there is one more structure with the appropriate dimension and quantum numbers, namely $\varepsilon^{\mu\nu} A_\nu$, but it cannot appear by itself if the gauge invariance is maintained. In other words, one can say that the *local* structure $\varepsilon^{\mu\nu} A_\nu$ can always be eliminated by subtraction of an ultraviolet counterterm.

It is worth noting that, purely kinematically,

$$k^\mu \varepsilon^{\alpha\beta} F_{\alpha\beta} = 2i \varepsilon^{\mu\nu} [k^2 A_\nu - k_\nu (k^\rho A_\rho)]. \quad (1.47)$$

It is seen that in order to distinguish an infrared singular term proportional to k^{-2} from the local term depending on the ultraviolet regularization it is necessary to assume that $kA \neq 0$. The infrared singular term is fixed unambiguously by the diagram in fig. 4d. The easiest way to get it is just to compute this graph in a straightforward way,

$$\langle j^{\mu 5} \rangle = (-1) \int \text{Tr} \left(\gamma^\mu \gamma^5 \frac{i \not{p}}{p^2} (-i) \gamma^\rho \frac{i(\not{p} + \not{k})}{(p+k)^2} \right) A_\rho \frac{d^2 p}{(2\pi)^2}. \quad (1.48)$$

Performing the p integration and disregarding terms non-singular in k^2 we get

$$\int \frac{p^\alpha}{p^2} \frac{(p+k)^\beta}{(p+k)^2} \frac{d^2 p}{(2\pi)^2} \rightarrow \frac{i}{4\pi} \frac{k^\alpha k^\beta}{k^2},$$

which implies, in turn,

$$\langle j^{\mu 5} \rangle_{\text{sing.}} = -(1/4\pi) \text{Tr}(\gamma^5 \gamma^\mu \not{k} \gamma^\rho \not{k}) A_\rho k^{-2} = -(1/\pi) \varepsilon^{\mu\alpha} k_\alpha (kA) k^{-2}.$$

Now, inserting the local term in order to restore the gauge invariance and using eq. (1.47) we arrive at

$$\langle j^{\mu 5} \rangle = (1/2i\pi) (k^\mu / k^2) \varepsilon^{\alpha\beta} F_{\alpha\beta}. \quad (1.49)$$

Taking the divergence is equivalent to multiplication of the right-hand side by ik_μ , and we reproduce, now for the third time, the anomalous relation (1.16).

Let us draw the readers' attention to the pole k^{-2} in eq. (1.49). The emergence of this pole is the manifestation of the infrared nature of the anomaly. We see that it can be derived from this side within the familiar Feynman technique. In the Schwinger model the pole term is not physical since its coefficient is proportional to $\partial_\mu A^\mu$ [cf. eq. (1.47)]. An absolutely similar pole term in QCD leads to far-reaching physical consequences, however, as we will see below.

2. Anomalies in quantum chromodynamics

2.1. Classical symmetries

Before proceeding to quantum anomalies in QCD let us first list the symmetries of the classical action. We will assume that the theory contains n_f massless quarks and temporarily forget about heavy quarks which are inessential in the given context. In the actual situation $n_f = 2$ ($q = u, d$) or $n_f = 3$ ($q = u, d, s$). The corrections due to the small u -, d -, s -quark masses can (and will) be considered separately where necessary.

In the limit $m_q = 0$ – in the literature it is often referred to as the chiral limit – the classical QCD action is

$$S = \int d^4x \mathcal{L}(x), \quad (2.1)$$

$$\mathcal{L} = -\frac{1}{4} G_{\mu\nu}^a G_{\mu\nu}^a + \sum_q (\bar{q}_L i \not{D} q_L + \bar{q}_R i \not{D} q_R), \quad q_{L,R} = \frac{1}{2}(1 \pm \gamma_5)q,$$

where q is the quark field and $G_{\mu\nu}^a$ is the gluon-field strength tensor. The action S is invariant under the following global transformations.

(i) Rotations of quarks of different flavors,

$$q \rightarrow Uq, \quad q = \begin{bmatrix} u \\ d \\ s \end{bmatrix}, \quad (2.2)$$

where U is a 3×3 matrix, arbitrary element of the $SU(3)$ group. Since the left-handed and right-handed quarks enter the Lagrangian (2.1) as separate terms, independent chiral transformations are allowed, $q_L \rightarrow Uq_L$, $q_R \rightarrow U'q_R$ where U and U' are generally speaking different $SU(3)$ matrices. Thus, the QCD Lagrangian actually possesses $SU(3)_L \times SU(3)_R$ symmetry called the chiral flavor invariance.

Sometimes it turns out to be more convenient to form linear combinations from the generators of the $SU(3)_L \times SU(3)_R$ so as to pass from the chiral rotations to the vector and axial transformations. In other words it is equally correct to say that the symmetry group of the classical action (2.1) is $SU(3)_V \times SU(3)_A$.

(ii) The phase $U(1)$ transformations

$$\begin{aligned} \text{vector: } q_L &\rightarrow q_L e^{i\alpha}, \quad q_R \rightarrow q_R e^{i\alpha}; \\ \text{axial: } q_L &\rightarrow q_L e^{i\beta}, \quad q_R \rightarrow q_R e^{-i\beta}. \end{aligned} \quad (2.3)$$

The physical meaning of eqs. (2.2) and (2.3) is obvious: any Born diagram conserves the number of the left-handed and right-handed fermions separately.

(iii) The scale transformations, i.e., dilations of the coordinates with the simultaneous rescaling of the quark and gluon fields in accordance with their normal dimension,

$$A_\mu(x) \rightarrow \lambda A_\mu(\lambda x), \quad q(x) \rightarrow \lambda^{3/2} q(\lambda x). \quad (2.4)$$

The scale invariance obviously stems from the fact that the classical action (2.1) contains no dimensionful constants. This invariance is a part of a larger conformal group. Unfortunately, discussion of the conformal symmetry is beyond the scope of the present review.

At the quantum level the fate of the above symmetries is different. The currents generating the $SU(3)_V \times SU(3)_A$ transformations are conserved even with the ultraviolet regularization switched on. They are anomaly-free.*) The vector $SU(3)$ symmetry corresponding to the current $j_{\mu V}^a = j_{\mu L}^a + j_{\mu R}^a$ ($a = 1, \dots, 8$) is realized linearly. As far as the axial $SU(3)$ subgroup $j_{\mu A}^a = j_{\mu L}^a - j_{\mu R}^a$ is concerned here, the spontaneous breaking of the symmetry takes place, and the given $SU(3)$ subgroup is realized non-linearly, with participation of the octet of the massless Goldstone bosons (π, η, K). Further discussion of the $SU(3)_A$ and its spontaneous breaking is given in section 2.4.

Moreover, the vector $U(1)$ invariance $q \rightarrow q e^{i\alpha}$ also stays a valid anomaly-free symmetry at the quantum level. This symmetry is responsible for the fact that the quark number is constant in any QCD process. Finally, the current generating the axial $U(1)$ transformations and the dilatation current are not conserved at the quantum level due to the anomalies. Below we dwell on this issue.

2.2. The axial and scale (dilatation) anomalies

As was explained above, the concrete form of the anomalous relations can be established without going beyond the perturbation theory, provided an appropriate ultraviolet regularization is chosen. We begin this subsection by a warning referring to the subtle point which causes a lot of confusion in the current literature. Two languages are used for descriptions of one and the same anomalous relations, and many authors do not even realize the distinction between the languages.

Within the first approach one establishes an *operator relation*, say, between the divergence of the axial current and $G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a$ [see eq. (2.9)]. Both, the axial current $j_{\mu A}$ and the product $G_{\mu\nu}^a \tilde{G}_{\mu\nu}^a = \frac{1}{2} \varepsilon^{\mu\nu\rho\sigma} G_{\rho\sigma}^a G_{\mu\nu}^a$ are treated within this procedure as Heisenberg operators of the quantum field theory: $j_{\mu A}$ as the quark operator, $G_{\mu\nu} \tilde{G}_{\mu\nu}$ as the gluon one. In order to convert the operator relations into the observable amplitudes it is necessary to make one more step: calculate, according to the general rules, the matrix elements of the operators figuring in the right-hand and left-hand sides of the anomalous equality. (The words “observable amplitudes” are used here metaphorically. In the present section we will speak only about quark and gluon scattering amplitudes while in reality, of course, only hadronic amplitudes are observable. The same remark refers to the expression “matrix elements”).

Within the second approach one operates directly with the matrix elements. More exactly, usually one fixes an external (or background) gluonic field and determines $\partial_\mu j_{\mu A}$ in this field in this or that way. In the absence of the external field $\partial_\mu j_{\mu A} = 0$. The fact of existence of the anomaly implies that $\partial_\mu j_{\mu A} \neq 0$ and $\partial_\mu j_{\mu A}(x)$ is locally expressible in terms of the external field at the same point, $G_{\mu\nu}(x)$. The analysis of the anomaly in the Schwinger model (section 1) has been undertaken just in this vein.

Although one and the same letters are used in both cases – perhaps, the confusion is due to this tradition – it is quite evident that the Heisenberg operator at the point x and the expression for the background field at the same point are by no means identical objects. Certainly, in the leading order

$$\langle G_{\mu\nu} \tilde{G}_{\mu\nu} \rangle = (G_{\mu\nu} \tilde{G}_{\mu\nu})_{\text{ext}}, \quad \langle G_{\mu\nu} G_{\mu\nu} \rangle = (G_{\mu\nu} G_{\mu\nu})_{\text{ext}}, \quad (2.5)$$

*) Let us emphasize that the assertion about the absence of the anomalies in the flavor-octet currents $j_{\mu V}^a, j_{\mu A}^a$ refers to pure quantum chromodynamics. Including the photons which interact with the quarks according to the standard rules results in the photon anomaly in $j_{\mu A}^a$. Further details are given in section 2.4.

where $\langle \dots \rangle$ denotes in the case at hand averaging over the external gluon field while the subscript “ext” marks the external field. In the next-to-leading order, however, the right-hand side of eq. (2.5) acquires α_s corrections, generally speaking. Therefore if the anomalies are discussed beyond the leading order it is absolutely necessary to specify what particular relations are considered: the operator relations or those for the matrix elements. Only in the one-loop approximation do both versions superficially coincide.

In order to derive the low-energy theorems applicable to physical hadronic processes (see section 3) we will need the quantum anomalies in the operator form. In the remainder of the chapter the term “anomaly” will mean the operator anomalous relation.

Let us proceed now to calculations and begin with the axial anomaly since it is simpler in the technical sense and a close example has been already analyzed in the Schwinger model.

The current generating the axial U(1) transformation is

$$j_{\mu A} = \sum_{q=u,d,s} \bar{q} \gamma_\mu \gamma_5 q. \quad (2.6)$$

Differentiating naively and invoking the equation of motion $\not{D}q = 0$ we get

$$\partial_\mu j_{\mu A} = \sum_q (\bar{q} \not{D} \gamma_5 q - \bar{q} \gamma_5 \not{D} q) = 0.$$

The lesson obtained in the Schwinger model teaches us, however, that this is not the whole story and the conservation of the axial current will evaporate after switching on the ultraviolet regularization. In the Schwinger model on the circle ($eL \ll 1$) we deal with the weak coupling regime and, therefore, can choose any of the alternative lines of reasoning either based on the infrared or on the ultraviolet approaches. In QCD it is rather meaningless to speak about the infrared behavior of the quarks. To make the calculation of the anomaly reliable we must avoid infrared formulation and invoke only the Green functions at short distances. As a result, we are forced to shift the emphasis from the analysis of the current $j_{\mu A}$ and concentrate directly on $\partial_\mu j_{\mu A}$.

Using one of the variants of the ultraviolet regularization, the well-tried ε -splitting, we write

$$j_{\mu A}^{\text{Reg}} = \sum_q \bar{q}(x + \varepsilon) \gamma_\mu \gamma_5 \left[\exp \left(\int_{x-\varepsilon}^{x+\varepsilon} ig A_\rho(y) dy \right) \right] q(x - \varepsilon). \quad (2.7)$$

Then the problem reduces to the quark loop in the background field. Using the equations of motion and expanding the exponential in the braces up to $O(\varepsilon)$ we get

$$\partial_\mu j_{\mu A}^{\text{Reg}} = \sum_q \{ \bar{q}(x + \varepsilon) [-igA(x + \varepsilon)\gamma_5 - \gamma_5 igA(x - \varepsilon) + ig\gamma_\mu \gamma_5 \varepsilon_\beta G_{\mu\beta}(0)] q(x - \varepsilon) \}.$$

The third term in the square brackets contains the gluon-field strength tensor and results from differentiation of the exponential factor. For convenience we have imposed the Fock–Schwinger gauge condition $y_\mu A_\mu(y) = 0$ on the external field (see the review paper, ref. [7]). (This gauge condition is not obligatory, of course.) In this gauge $A_\mu(y) = \frac{1}{2} y_\rho G_{\rho\mu}(0) + \dots$. As usual, $A_\mu = \frac{1}{2} t^a A_\mu^a$ where t^a are the color Gell-Mann matrices.

Contracting the quark lines in the loop we arrive at the following expression:

$$\begin{aligned}
\partial_\mu j_{\mu A}^{\text{Reg}} &= -i g n_f \text{Tr}_{\text{color} + \text{Lorentz}} [-2i \varepsilon_\rho G_{\rho\mu}(0) \gamma_\mu \gamma_5 S(x - \varepsilon, x + \varepsilon)] \\
&= -n_f \frac{1}{2} g^2 G_{\rho\mu}^a \tilde{G}_{\alpha\varphi}^a(0) (\varepsilon_\rho \varepsilon_\alpha / \varepsilon^2) (1/8 \pi^2) \text{Tr}_{\text{Lorentz}} (\gamma_\mu \gamma_5 \gamma_\alpha \gamma_5) \\
&= (n_f \alpha_s / 4\pi) [G_{\alpha\beta}^a \tilde{G}_{\alpha\beta}^a]_{\text{ext}}.
\end{aligned} \tag{2.8}$$

Here $S(y, x)$ is the massless quark propagator in the background field. The expression for this propagator as a series in the background field in the Fock–Schwinger gauge is known (ref. [7]). The terms vanishing in the limit $\varepsilon \rightarrow 0$ are neglected in eq. (2.8). Notice that the quarks propagate only at very short distances $\sim \varepsilon \rightarrow 0$. Equation (2.8) gives the answer for the axial anomaly in the one-loop approximation,

$$\partial_\mu j_{\mu A} = (n_f \alpha_s / 4\pi) G_{\alpha\beta}^a \tilde{G}^{a\alpha\beta}, \quad \tilde{G}^{a\alpha\beta} = \frac{1}{2} \varepsilon^{\alpha\beta\rho\sigma} G_{\rho\sigma}^a. \tag{2.9}$$

This result is readily reproducible within the Pauli–Villars regularization (an exercise for the reader). Higher orders in α_s will be discussed in section 2.3.

Proceed now to the dilatation anomaly. First of all it is instructive to check that the scale transformations are generated by the current

$$j_{\nu D} = x^\mu \theta_{\mu\nu}, \tag{2.10}$$

where $\theta_{\mu\nu}$ is the energy–momentum tensor (symmetric and conserved),

$$\theta_{\mu\nu} = -G_{\mu\alpha}^a G_{\nu\alpha}^a + \frac{1}{4} g_{\mu\nu} G_{\alpha\beta}^a G_{\alpha\beta}^a + \frac{1}{4} i \sum_q [\bar{q}(\gamma_\mu D_\nu + \gamma_\nu D_\mu) q - \bar{q}(\gamma_\mu \tilde{D}_\nu + \gamma_\nu \tilde{D}_\mu) q]. \tag{2.11}$$

The corresponding charge can be represented as

$$D \equiv \int d^3x j_{0D} = tH + \tilde{D}, \quad \tilde{D} = \int d^3x \theta_{i0} x^i, \tag{2.12}$$

where H is the Hamiltonian and

$$[\tilde{D}, p_i] = i p_i, \quad [\tilde{D}, \exp(ip_i \cdot x^i)] = -(x^i p_i) \exp(ip_i \cdot x^i), \quad i = 1, 2, 3.$$

Let $O(x)$ be an arbitrary colorless operator. Then

$$[D, O(x)] = t[H, O(x)] + [\tilde{D}, O(x)] = -it(\partial/\partial t)O(x) + [\tilde{D}, O(x)]. \tag{2.13}$$

Moreover, the commutator $[\tilde{D}, O(x)]$ contains two terms, the first one corresponding to rescaling of x_i and the second one to rescaling of the operator O . Indeed,

$$\begin{aligned}
[\tilde{D}, O(t, x^i)] &= [\tilde{D}, \exp(ip_i x^i) O(t, 0) \exp(-ip_i x^i)] \\
&= [\tilde{D}, \exp(ip_i x^i)] O(t, 0) \exp(-ip_i x^i) + \exp(ip_i x^i) O(t, 0) [\tilde{D}, \exp(-ip_i x^i)] \\
&\quad + \exp(ip_i x^i) [\tilde{D}, O(t, 0)] \exp(-ip_i x^i) \\
&= -ix^i \partial_i O(t, x^i) + \exp(ip_i x^i) [\tilde{D}, O(t, 0)] \exp(-ip_i x^i).
\end{aligned} \tag{2.14}$$

It is easy to understand that the commutator $[\tilde{D}, O(t, 0)]$ should be proportional to $O(t, 0)$, and, hence

$$[D, O(x)] = -i[x^\mu(\partial/\partial x^\mu) O(x) + dO(x)], \quad (2.15)$$

where d is a dimensionless number determined by a particular form of the operator O .

Only the general properties of the quantum field theory have been used thus far. The numerical value of the coefficient d depends on the specific structure of the theory. In QCD d is equal to the normal dimension of the operator O . [For instance, if $O(x) = \mathcal{L}(x)$ where \mathcal{L} is the Lagrangian, $d = 4$.] Equation (2.15) expresses in a mathematical language the change of the units of mass and length. This explains the origin of the name, the scale transformations.

Notice that the reservation about the colorless structure of the operator O is not superfluous. For the colored operators, say, q or A_μ^a , the commutator with D , as is seen from the direct computation, does not reduce to the form (2.15). The reader should not be puzzled by this fact. The complication is due to gauge fixing, and additional terms can be eliminated by an appropriate gauge transformation.

Using the classical equations of motion we get

$$\partial^\mu j_{\mu D} = \theta_\mu^\mu = 0 \quad (2.16)$$

[cf. eq. (2.11)]. Classically the trace of the energy–momentum tensor vanishes provided that $m_q = 0$. At the quantum level $\theta_{\mu\nu}$ is no longer traceless, and the trace θ_μ^μ is given by the dilatation (scale) anomaly. The fact that the scale invariance (2.4) is lost in loops is obvious. Indeed, the invariance (2.4) takes place because there are no dimensionful parameters in the classical action (2.1). Already in one-loop, however, such a parameter inevitably appears in the effective action, the ultraviolet cut-off M_0 .

The M_0 dependence of the effective action is known beforehand. It is very convenient to use this information, seemingly it is the shortest way to calculate θ_μ^μ . In the one-loop approximation the effective action can be written somewhat symbolically as

$$S_{\text{eff}} = -\frac{1}{4} \int d^4x \left(\frac{1}{g_0^2} - \frac{b}{16\pi^2} \ln M_0^2 x_{\text{char}}^2 \right) (G_{\mu\nu}^a G_{\mu\nu}^a)_{\text{ext}} + \dots, \quad (2.17)$$

where $b = \frac{11}{3}N_c - \frac{2}{3}n_f$ is the first coefficient in the Gell-Mann–Low function and the dots stand for the fermion terms. We rescaled temporarily the gluon field, $gA \rightarrow A$, so that the coupling constant figures only as an overall factor in front of G^2 . The variation of the effective action under the transformation (2.4) is

$$\delta S_{\text{eff}} = -((1/32)\pi^2) b G_{\mu\nu}^2 \ln \lambda, \quad (2.18)$$

implying that

$$\partial^\mu j_{\mu D} = -(1/32)\pi^2 b G_{\mu\nu}^2. \quad (2.19)$$

Returning to the standard normalization of the gluon field we finally get

$$\theta_\mu^\mu = -(b\alpha_s/8\pi) G_{\mu\nu}^a G_{\mu\nu}^a + O(\alpha_s^2). \quad (2.20)$$

2.3. Multiloop corrections

Although in practical applications it is quite sufficient to limit oneself to the one-loop expressions for the anomalies (2.9), (2.20) the question of the higher-order corrections, being academic, still deserves a brief discussion. Until recently it was generally believed that the question was totally solved. Namely, the axial anomaly, by the Adler–Bardeen theorem [8], is purely one-loop [i.e., there are no corrections to eq. (2.9)] while the scale anomaly contains the complete QCD β -function in the right-hand side, $\beta(\alpha_s(\mu)) = d\alpha_s(\mu)/d\ln\mu$. In other words, $(-b\alpha_s/8\pi)$ in eq. (2.20) is substituted by $\beta(\alpha_s)/4\alpha_s$ if higher-order corrections are taken into account [9].

Later it became clear, however, that the situation is far from being thus simple and calls for an additional study [10]. Surprising though it is, the flaw in the standard arguments was first revealed not within QCD but in a more complex model, supersymmetric Yang–Mills theory. The minimal model of such a type includes gluons and gluinos, Majorana fermions in the adjoint representation of the color group. The axial current $j_{\mu A}$ (the so-called R current) and the energy–momentum tensor $\theta_{\mu\nu}$ in supersymmetric theories enter one supermultiplet [11] and, consequently, the coefficients in the chiral and dilatation anomalies cannot be different – one-loop for $\partial^\mu j_{\mu A}$ and multiloop for θ_μ^μ .

Being unable to discuss here the multi-loop corrections in detail, I note only that the generally accepted treatment is based on the confusion just mentioned in the beginning of subsection 2.2. The standard derivation of the Adler–Bardeen theorem seems to be valid only provided that the axial anomaly is treated as the operator equality. At the same time the relation

$$\theta_\mu^\mu = [\beta(\alpha_s)/4\alpha_s][G_{\mu\nu}^a G_{\mu\nu}^a]_{\text{ext}} \quad (2.21)$$

takes place only for the matrix elements (cf. the derivation presented at the end of subsection 2.2).

It is quite natural to try to reduce both the anomalies to a unified form, preferably to the operator form. Then $\partial^\mu j_{\mu A}$ is exhausted by the one-loop approximation (2.9), at least, within a certain ultraviolet regularization [12]. As far as θ_μ^μ is concerned, in this case at the moment we, strictly speaking, do not know even the two-loop coefficient in front of the operator $G_{\mu\nu}^a G_{\mu\nu}^a$, to say nothing about higher-order corrections. Calculations existing in the literature should be re-analyzed anew in order to separate the genuinely ultraviolet contributions from those containing an infrared part; the latter should be interpreted as a matrix element. The remainder of the subsection elucidates this assertion and in principle can be omitted by the reader with no damage to the subsequent material.

Calculating the effective action is beyond any doubt the most convenient method for determination of the scale anomaly. The term “effective action” is applied to two somewhat different objects. The first object is the sum of all vacuum loops in the given external field. The functional Γ obtained in this way and depending on the external field is often called the effective action, although another name would be more exact: the generator of the one-particle irreducible (1PI) vertices. The other object is the effective action à la Wilson [13], $S_w(\mu)$, which differs from Γ in the respect that S_w takes into account only the contribution of the loop momenta $p > \mu$ and excludes the infrared domain $p \leq \mu$. If Γ represents a c -number functional of the external field, S_w is the quantum-field operator; it can be considered as the original action with respect to the low-frequency fluctuations with the wavelengths $\leq \mu$. The functional Γ results from taking the matrix element of $\exp[iS_w(\mu)]$.

The standard line of reasoning which leads to eq. (2.21) is based on the analysis of Γ . Just this functional defines the effective coupling constant $g^2(\mu)$,

$$\Gamma = -\frac{1}{4}[1/g^2(\mu)](G_{\mu\nu}^a G_{\mu\nu}^a)_{\text{ext}} + \dots,$$

and differentiating $g^2(\mu)$ over $\ln \mu$ yields the complete β function. If, instead, we would like to get θ_μ^μ in the operator form it is necessary to work with S_W , not Γ . In other words, in calculating the two-loop vacuum graphs in the external field one should carefully separate – and throw away – possible infrared contributions.

In QCD this has not been done yet. The question whether the higher-order coefficients in the operator scale anomaly coincide or not with the higher-order coefficients in the β -function remains open. But in a simpler model, scalar electrodynamics, it is definitely proven [10] that such a coincidence does not take place. Indeed, in the scalar QED [cf. eq. (2.25)]

$$S_W(\mu) = \int d^4x \left[-\frac{1}{32\pi^2} \left(\frac{2\pi}{\alpha_0} + \frac{1}{3} \ln(M_0/\mu) + \frac{\alpha_0}{2\pi} (1 - \frac{1}{2}\xi) \ln(M_0/\mu) \right) F_{\mu\nu} F_{\mu\nu} + Z D_\mu \varphi^* D_\mu \varphi \right] \quad (2.22)$$

where $\alpha = e^2/4\pi$, $F_{\mu\nu}$ is the photon field tensor, φ is the complex scalar massless field and Z is its renormalization constant

$$Z = 1 - (1 + \frac{1}{2}\xi)(\alpha_0/\pi) \ln(M_0/\mu). \quad (2.23)$$

Here M_0 is the ultraviolet cut-off and ξ is the gauge parameter in the photon propagator

$$D_{\mu\nu} = e^2 [-g_{\mu\nu} + (\xi k_\mu k_\nu / k^2) / k^2].$$

Equation (2.22) demonstrates that the two-loop contribution in the operator $F_{\mu\nu}^2$ does not coincide with that in the β -function ($\beta^{(2)} = 1/\pi$). More than that, it is not even gauge invariant!

The explicit gauge invariance is recovered after proceeding from S_W to Γ , i.e., after calculating the photonic matrix element of S_W . Formally, the operator $\int d^4x D_\mu \varphi^* D_\mu \varphi$ vanishes by the equations of motion. One can convince oneself, however [10], that in the external photon field,

$$\langle D_\mu \varphi^* D_\mu \varphi \rangle = (1/64\pi^2)(F_{\mu\nu}^2)_{\text{ext}}, \quad (2.24)$$

because of the infrared contribution.

Combining eqs. (2.22)–(2.24) we arrive at

$$\Gamma = \int d^4x \left[-\frac{1}{32\pi^2} \left(\frac{2\pi}{\alpha_0} + \frac{1}{3} \ln(M_0/\mu) + \frac{\alpha_0}{\pi} \ln(M_0/\mu) \right) (F_{\mu\nu}^2)_{\text{ext}} + \dots \right] \quad (2.25)$$

[cf. eq. (2.22)]. The two-loop coefficient in front of $F_{\mu\nu}^2$ corresponds, as it should, to the two-loop coefficient in the β -function.

2.4. Anomalies and external currents

If in addition to “pure” QCD one introduces the interaction of the quarks with the photons or with other “external” currents, then apart from the axial and scale anomalies (2.9), (2.20), one will find new

anomalies and, moreover, there appear some new terms in the old anomalous relations for $\partial^\mu j_{\mu A}$ and θ_μ^μ . For simplicity let us consider only the photons, leaving aside the possibility of other external currents.

Consider one of the flavor-octet axial currents,

$$a_\mu^3 = \bar{u}\gamma_\mu\gamma_5 u - \bar{d}\gamma_\mu\gamma_5 d, \quad (2.26)$$

with the π^0 quantum numbers. It is quite clear that in pure QCD $\partial^\mu a_\mu^3 = 0$ (see subsection 2.2). With the photons switched on, however, the divergence of a_μ^3 is non-vanishing. From fig. 5 it is easy to understand that

$$\partial^\mu a_\mu^3 = (\alpha/2\pi)N_c(Q_u^2 - Q_d^2)F_{\mu\nu}\tilde{F}_{\mu\nu}, \quad (2.27)$$

where $N_c = 3$ is the number of colors and $Q_u = 2/3$, $Q_d = -1/3$ are the u-, d-quark electric charges. The product $F_{\mu\nu}\tilde{F}_{\mu\nu}$ in the right-hand side of eq. (2.27) is an operator with respect to the photons. If we were interested in the amplitude of the transition into two photons with the momenta $k^{(1)}$, $k^{(2)}$ and polarization vectors $\varepsilon^{(1)}$, $\varepsilon^{(2)}$ we would have

$$F\tilde{F} \rightarrow (-2)F^{(1)}\tilde{F}^{(2)}, \quad F_{\mu\nu}^{(i)} = k_\mu^{(i)}\varepsilon_\nu^{(i)} - k_\nu^{(i)}\varepsilon_\mu^{(i)}.$$

The expression (2.27) has far-reaching consequences. Let us ask the question: “What is the answer for the amplitude of the transition of the current a_μ^3 – just the current, not $\partial^\mu a_\mu^3$ – into two real photons?” In order to find the answer one should first of all perform the kinematical analysis, (i.e., write down all kinematical structures for the amplitude $a_\mu^3 \rightarrow 2\gamma$ with the additional conditions $k^{(1)2} = k^{(2)2} = 0$ plus the gauge invariance with respect to both the photons plus eq. (2.27). Omitting the details of this simple exercise we give the answer,

$$\begin{aligned} \langle 0|a_\mu^3|2\gamma\rangle = & -(iq_\mu/q^2)(4\alpha/\pi)N_c(Q_u^2 - Q_d^2)\varepsilon^{\mu\nu\alpha\beta}k_\mu^{(1)}\varepsilon_\nu^{(1)}k_\alpha^{(2)}\varepsilon_\beta^{(2)} \\ & + \text{a part transversal in } q_\mu, \quad q = k^{(1)} + k^{(2)}. \end{aligned} \quad (2.28)$$

The most striking point is the presence of the pole factor q_μ/q^2 in the right-hand side of eq. (2.28) [14, 15]. It is worth emphasizing that eq. (2.28) is the exact result of QCD valid, in particular, in the limit $q^2 \rightarrow 0$. Technically, the singularity in a_μ^3 at $q^2 \rightarrow 0$ is the consequence of the massless quark propagation in the triangle graph of fig. 5. In the confining theory, however, it is absolutely meaningless to speak about the quarks as physical degrees of freedom in the spectrum. Then, how can one understand the singularity of the amplitude $\langle 0|a_\mu^3|2\gamma\rangle$ in physical terms?

Only one explanation is possible. The physical spectrum should include massless colorless composite

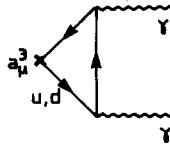


Fig. 5. Photonic anomaly in the axial current in QCD.

particles coupled both to photons and a_μ^3 . The corresponding contribution should saturate the matrix element $\langle 0 | a_\mu^3 | 2\gamma \rangle$, or, more exactly, its singular part. Thus, starting from the massless quarks and assuming the color confinement we succeeded in proving the existence of the massless hadrons. The proof is based on the requirement of matching the singular contributions in a_μ^3 at the level of quarks and at the level of hadrons. This elegant idea belongs to 't Hooft [15] and is called the 't Hooft consistency condition. Let us try to find out which particular hadrons are massless.*¹ In principle, there are two alternative variants, both leading to the pole singularity in a_μ^3 , (i) massless baryons; and (ii) massless pions.

Consider first the scenario (i), $M_B = 0$. In the theory with u, d, s quarks the spin-1/2 baryons form the well-known octet

$$B = (p, n, \Sigma^\pm, \Lambda, \Sigma^0, \Xi^-, \Xi^0). \quad (2.29)$$

Scenario (i) – a very important point – implies that both the vector and axial SU(3) symmetries are realized linearly**¹ and, hence, the baryon–baryon–photon coupling constants and the constants $\langle B | a_\mu^3 | B \rangle$ at zero momentum transfer are fixed (for instance, $\langle \Sigma^+ | a_\mu^3 | \Sigma^+ \rangle = 2\bar{\Sigma} \gamma_\mu \gamma_5 \Sigma$). Calculating the triangle diagram of fig. 6 (more exactly, its pole part) we find that the baryon octet does not contribute to the pole in a_μ^3 due to the cancellations: the proton is cancelled by Ξ^- and Σ^- by Σ^+ . The other baryons from (2.29) decouple from the photon. The immediate conclusion is: scenario (i) is not compatible with the anomalous relation (2.27).

Thus, we are forced to conclude that it is scenario (ii) that is actually realized, and the massless hadron saturating eq. (2.28) is the pion. The 't Hooft consistency condition then looks trivial (fig. 7) and requires only a certain relation between the $\pi^0 \rightarrow a_\mu^3$ amplitude and the $\pi^0 \rightarrow 2\gamma$ coupling constant. The corresponding result has been known for a long time. It is nothing but the famous PCAC prediction for $A(\pi^0 \rightarrow 2\gamma)$ in terms of f_π , the $\pi \rightarrow \mu\nu$ coupling constant ($\langle 0 | a_\mu^3 | \pi^0 \rangle = \sqrt{2} i f_\pi q_\mu$ where q_μ is the pion momentum. For a nice review of PCAC and a representative list of references, see ref. [18]. The existence of the massless pion coupled to the axial current a_μ^3 means the spontaneous breaking of the axial SU(3)_A symmetry, or more exactly, the non-linear Goldstone realization of the SU(3)_A.

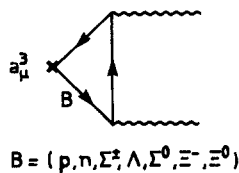


Fig. 6. Massless baryon saturation of the photonic anomaly (unsuccessful attempt).

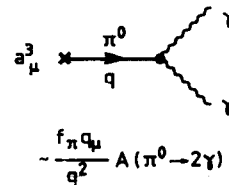


Fig. 7. Saturation in terms of the Goldstone pseudoscalar mesons.

*¹ The arguments given below simplify the actual situation, leaving aside some subtleties. In particular, in discussing the “baryonic” scenario, we consider only the baryon octet (2.29). Thus, it is tacitly assumed that the baryons with other quantum numbers, e.g. $J^P = 1/2^-$, do not contribute in the pole part of the amplitude $\langle 0 | a_\mu^3 | 2\gamma \rangle$. Actually, it is rather difficult to prove that the combined contribution of all other baryons cannot be equal to (2.28) – it turns out to be necessary to invoke additional assumptions (see in this connection refs. [16, 17]). Therefore, although the main line of reasoning demonstrating the spontaneous breaking of the SU(3)_A symmetry is reproduced correctly, I would not like to produce an impression that the derivation of the spontaneous chiral symmetry breaking is completely trivial and stems almost from nothing.

**¹ In QCD one can rigorously show that the spontaneous breaking of the vector SU(3)_V symmetry is impossible [17].

Summarizing, the analysis of the anomalies in QCD with the external currents allows one to get important additional information. In particular, we are able to get such a profound result as the spontaneous breaking of the flavor $SU(3)_A$ group and the presence of the octet of the Goldstone mesons in QCD.

A brief digression of the historical character is in order here. The Goldstone nature of the π , η , K mesons has been exploited in hadronic physics long before the QCD era. The soft-pion technique created in the sixties on the basis of the hypothesis of approximate chiral invariance of strong interactions (in those days that was a brilliant hypothesis indeed) generated a series of elegant predictions in the low-energy pion physics. The Goldberger–Treiman and Adler–Weisberger relations (see, e.g., ref. [18]) seem to be the most typical. The consequences of PCAC proved to be extremely useful in diverse applications. The agreement of the PCAC predictions with the data served as conclusive evidence in favor of the approximate chiral invariance of strong interactions. The latter, in turn, was the key argument used by the fathers of QCD [19] in order to introduce the vector coupling of gluons with quarks. Somewhat later the same ideas allowed one to determine the masses of the current u , d , s quarks [20] and fix the quark condensate.

Returning to the anomalies, we recall that the introduction of the photons (and other “external” currents) apart from creating the new anomalies modifies the old ones in $\partial^\mu j_{\mu A}$ and θ_μ^μ adding some new terms there. We postpone the discussion of the corresponding effects until subsection 3.2.

3. Low-energy theorems; applications to hadronic physics

Low-energy theorems in field theory were invented almost as long ago as field theory itself. Suffice it to recall the Low theorems [21] for the photon bremsstrahlung in the low-frequency limit. As a rule, the low-energy theorems represent relations among amplitudes, or n -point Green functions with different numbers of legs (particles). The relations between the amplitudes emerge as a reflection of some symmetry, exact or anomalous, existing in the theory. For instance, the Low theorems mentioned above are the consequence of the gauge invariance of electromagnetic interactions.

The search for symmetries and constraints they impose on the observable quantities is of special importance in the theory of hadrons. The low-energy theorems, or the Ward identities stemming from the symmetry properties of the theory and independent of the unknown details of the confinement mechanism yield information on physical processes inaccessible by other existing methods. In a purely theoretical aspect they serve as a reference point, a basis for further theoretical constructions.

Apart from the classical PCAC theorems^{*)} in QCD there emerge additional predictions both for the Goldstone meson emission and for some other processes. The first predictions were obtained shortly after the invention of QCD (see, e.g., ref. [22]). The systematic analysis of the subject dates back to refs. [23–26]. Many applications based on relations (2.9), (2.20) were worked out recently. Some of them will be discussed below. The most non-trivial results refer to the amplitudes describing the conversion of gluons into hadrons and photons at large distances. The information obtained in this way supplements our understanding of the hadron structure and that of the QCD vacuum.

This section is devoted to the derivation of the basic low-energy theorems specific to QCD. Their virtues will be demonstrated in a few instructive examples. It is worth noting that the most beautiful

^{*)} Relations stemming from pure PCAC are not discussed here. The corresponding low-energy results, however, are widely used as intermediate building blocks.

consequence of the anomalies in QCD was already discussed in subsection 2.4. Some additional applications not considered in the present section can be found in ref. [27], see also refs. [28–30].

3.1. The scale Ward identities

If $m_{u,d,s} = 0$ the only mass parameter of the theory that appears at the quantum level, is

$$\mu = M_0 \exp(-8\pi^2/bg_0^2), \quad (3.1)$$

where as usual M_0 is the ultraviolet cut-off and g_0 is the corresponding coupling constant, $g_0 = g(M_0)$. Equation (3.1) is a manifestation of the anomaly in the trace of the energy–momentum tensor,

$$\sigma(x) \equiv \theta_\mu^\mu = -(b\alpha_s/8\pi) G_{\mu\nu}^a G_{\mu\nu}^a \quad (3.2)$$

derived in section 2 [see eq. (2.20)].

The classical scale invariance of the action, being lost in loops, still does not vanish without a trace (both in the direct and mathematical meanings of these words). The remnant of this symmetry is a set of the low-energy theorems of the type

$$\lim_{q \rightarrow 0} i \int e^{iqx} dx \langle T\{O(x), \sigma(0)\}_{\text{connected}} \rangle = -d \langle O \rangle \quad (3.3)$$

where O is an arbitrary local operator built from gluons and/or u, d, s quarks, d is its normal dimension. Actually, eq. (3.3) is valid up to $O(m_q)$ terms which are implied but will not be written out explicitly, except in the cases where it is necessary. To make the expressions more concise the subscript “connected” will be omitted in what follows.

A few important concrete examples are

$$i \int dx \langle T\{\sigma(x), \sigma(0)\} \rangle = -4 \langle \sigma \rangle, \quad (3.4)$$

$$i \int dx \langle T\{\bar{q}(x)q(x), \sigma(0)\} \rangle = -3 \langle \bar{q}q \rangle. \quad (3.5)$$

Relations between different Green functions involving $\sigma(x)$ take place not only for the vanishing external momenta. More exactly, it is not necessary to have *all* external momenta equal to zero. Those which correspond to “foreign” operators O can be arbitrary [31].

The derivation of all these formulae is quite analogous, and we will consider the simplest case, eq. (3.3). First, rescale the gluon field, $\bar{G}_{\mu\nu}^a = g_0 G_{\mu\nu}^a$, so as to factor out the g_0 dependence in the Lagrangian, $\mathcal{L} = -(1/4g_0^2)\bar{G}^2$. Then

$$i \int dx \langle T\{O(x), \bar{G}^2(0)\} \rangle = -\frac{d}{d(1/4g_0^2)} \langle O \rangle. \quad (3.6)$$

The fact that the only mass parameter of the theory is given by eq. (3.1) implies

$$\langle O \rangle = \text{const.} [M_0 \exp(-8\pi^2/bg_0^2)]^d. \quad (3.7)$$

Differentiating eq. (3.7) we arrive at eq. (3.3).

The simple derivation sketched above actually should be supplemented by some regularization since in most cases $\langle O \rangle$ is divergent in perturbation theory (a rare exception is the operator $\bar{q}q$ in the chiral limit $m_q \rightarrow 0$). Reference [23] shows that a careful regularization procedure involving a sufficient number of the regulator fields does not alter our conclusions.

Let us proceed now to phenomenological consequences of the scale Ward identities. In the world with no light quarks the lightest (and, hence, stable) particle would be scalar gluonium, σ_g . Study of its properties and, in particular, scattering amplitudes would be an important direction in “experimental physics”. It is highly probable that the σ_g mass is small compared to the characteristic scale in this channel ($\sim 20 \text{ GeV}^2$) [23]. From this point of view the situation resembles that with pions in the real world, $m_\pi^2 \ll \text{GeV}^2$. Of course, the smallness of m_π^2 is parametrical while the suppression of $m_{\sigma_g}^2$ seems to be of a numerical character. In spite of this fact, one can formulate the following problem: construct the low-energy (effective) Lagrangian describing the interaction of σ_g and realizing the scale Ward identities (3.3)–(3.5) just in the same way as the pion chiral Lagrangian realizes the chiral Ward identities at the tree level.

The solution was given in ref. [31] (more complicated cases including electromagnetism are considered in ref. [32]). The very fact that the solution can be found is far from being trivial and crucially depends on the sign of the vacuum energy; the solution is stable only provided that $\varepsilon_{\text{vac}} < 0$. Thus, our task is to find an effective theory for the field $\sigma(x)$ [see eq. (3.2)] in which $\theta_\mu^\mu = \sigma$. Recalling that θ_μ^μ is connected with dilatations, it is quite natural to call $\sigma(x)$ the dilaton field. Notice that the normal dimension of $\sigma(x)$ is equal to 4. Therefore, if the kinetic term in the Lagrangian is

$$\mathcal{L}_{\text{kin}} = \text{const.} \times (\partial_\mu \sigma)^2 \sigma^{3/2}, \quad (3.8)$$

then the integral $\int d^3x \mathcal{L}_{\text{kin}}$ is scale invariant and, hence, \mathcal{L}_{kin} does not contribute to θ_μ^μ . We can concentrate on the potential term $\mathcal{L}_{\text{pot}}(\sigma)$ and choose it so as to ensure the desired equality $\theta_\mu^\mu = \sigma$.

The variation of σ under the scale transformation $x \rightarrow x(1 + \varepsilon)$ is $\sigma \rightarrow \sigma(1 - 4\varepsilon)$. The corresponding variation of the action,

$$\Delta_{\text{scale}} \int d^4x \mathcal{L}_{\text{pot}}(\sigma) = \int \left(4\mathcal{L}_{\text{pot}} - 4\sigma \frac{\delta \mathcal{L}_{\text{pot}}}{\delta \sigma} \right) d^4x, \quad (3.9)$$

according to the general rules is equal to $-\int d^4x \theta_\mu^\mu(x)$. As a result we get

$$4\mathcal{L}_{\text{pot}} - 4\sigma \delta \mathcal{L}_{\text{pot}} / \delta \sigma = -\sigma. \quad (3.10)$$

The solution of this simple equation is

$$\mathcal{L}_{\text{pot}} = \frac{1}{4} \sigma (\ln \sigma + \text{const.}). \quad (3.11)$$

The constants figuring in eqs. (3.8), (3.11) can be expressed in terms of the mass of σ_g and the vacuum energy, m and ε_{vac} , respectively. It is convenient to redefine the field σ , $\sigma = 4\varepsilon_{\text{vac}} \exp \chi$; then the final

answer for the effective Lagrangian takes the form^{*)}

$$\mathcal{L}_{\text{eff}} = -(\varepsilon_{\text{vac}}/m^2)^{1/2} (\partial_\mu \chi)^2 e^{\chi/2} + \varepsilon_{\text{vac}} (\chi - 1) e^\chi. \quad (3.12)$$

Recall that $\varepsilon_{\text{vac}} < 0$ (fig. 8).

Thus far we have actually used only the information coded in eq. (3.4) and other analogous Ward identities. The theorems (3.3) determine the form of any local *colorless* operator O (in pure gluodynamics) in terms of a single constant,

$$O = \langle O \rangle_{\text{vac}} \exp(\tfrac{1}{4} d\chi). \quad (3.13)$$

In particular, $\sigma = \langle \sigma \rangle_{\text{vac}} \exp \chi = 4\varepsilon_{\text{vac}} \exp \chi$. One of the consequences of eq. (3.13) is $\langle O_1 O_2 \rangle = \langle O_1 \rangle \langle O_2 \rangle$. The factorization property will not seem surprising if one recalls that the solution for the effective scalar field theory constructed in this section is valid in the tree approximation, i.e., in the leading order in $1/N_c$. At $N_c \rightarrow \infty$ the vacuum expectation values of the colorless operators indeed do factorize. It is instructive to check that the Lagrangian (3.12) satisfied the general requirements of the multicolor QCD [33, 34],

$$\mathcal{L}_{\text{eff}} = \frac{1}{2} (\partial_\mu \tilde{\chi})^2 - \frac{1}{2!} m^2 \tilde{\chi}^2 + \frac{1}{4} \frac{m}{|\varepsilon_{\text{vac}}|^{1/2}} (\partial_\mu \tilde{\chi})^2 \tilde{\chi} + \cdots - \frac{2}{3!} \frac{m^3}{|\varepsilon_{\text{vac}}|^{1/2}} \tilde{\chi}^3 + \cdots,$$

where we have changed the normalization of the field χ ($\chi \rightarrow \tilde{\chi}$) in order to ensure the standard normalization of the kinetic term and expanded the exponentials. Since $m \sim M_c^0$ and $\varepsilon_{\text{vac}} \sim N_c^2$ the behavior of the interaction vertices is in full accord with the general rules: the cubic constant $\propto N_c^{-1}$, the quartic constant $\propto N_c^{-2}$, etc.

The Lagrangian (3.12) exhaustively describes the interaction of the scalar gluonium at energies $\sim m_\sigma$. Unfortunately, for obvious reasons, we cannot undertake a direct experimental test of the consequences stemming from eq. (3.12). The best we can get is a very rough indirect estimate which simultaneously shows the possible size of the effects induced by the light quarks. Assume that relations analogous to (3.13) are valid not only in pure gluodynamics but in the theory with quarks as well, both for gluon and quark operators. Given this hypothesis – which automatically excludes the scalar quarkonium contribution and $\sigma_q - \sigma_g$ mixing – we readily find the nucleon matrix element of $\bar{u}u + \bar{d}d$.

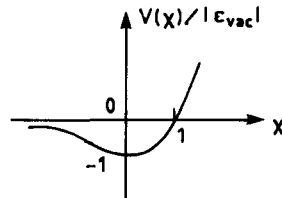


Fig. 8. Potential energy of the dilaton in QCD.

^{*)} The canonical energy-momentum tensor stemming from eq. (3.12) has the trace which does not coincide with (3.9). The reason is that the expression for the energy-momentum tensor of the scalar field is not unique. In order to reproduce the condition $\theta_{\mu\mu} = \sigma$, an extra term should be added to the canonical energy-momentum tensor, namely $-(8\varepsilon_{\text{vac}}/3m^2)(\square g_{\mu\nu} - \partial_\mu \partial_\nu) \exp \chi$. This extra term is symmetric and is conserved by itself.

At the tree level

$$\begin{aligned}\langle N|\bar{u}u + \bar{d}d|N\rangle &= \langle \bar{u}u + \bar{d}d \rangle^{\frac{3}{4}} \langle N|\chi|N\rangle \\ &= \langle \bar{u}u + \bar{d}d \rangle \frac{3}{4} \frac{\langle N|\theta_\mu^\mu|N\rangle}{\langle \theta_\mu^\mu \rangle} = \frac{3}{4} \langle \bar{u}u + \bar{d}d \rangle \frac{m_N}{4\varepsilon_{\text{vac}}} \bar{\psi}_N \psi_N ,\end{aligned}\quad (3.14)$$

where ψ_N is the nucleon wave function, $\langle N|\theta_\mu^\mu|N\rangle = m_N \bar{\psi}_N \psi_N$ and $\langle \theta_\mu^\mu \rangle = 4\varepsilon_{\text{vac}}$. On the other hand, the matrix element $\langle N|\bar{u}u + \bar{d}d|N\rangle$ is more or less known empirically, it constitutes $3\bar{\psi}_N \psi_N$. The theoretical estimate (3.14) yields $\sim 1.8\bar{\psi}_N \psi_N$ provided the standard values of the quark and gluon condensates are substituted. Notice that the signs of the empirical and theoretical estimates coincide while the absolute values differ by a factor ~ 2 . We conclude that the results for the scalar gluonium obtained in pure gluodynamics are valid in the real world up to a factor of 2. On the other hand, the effect due to switching on the light quarks is of order unity.

3.2. Conversion of gluons into pions and photons

Any QCD expert unfamiliar with the low-energy theorems would say that the matrix elements like

$$\langle 0|\alpha_s G^2|2\pi\rangle, \quad \langle 0|\alpha_s G^2|2\gamma\rangle, \quad \langle 0|\alpha_s G\tilde{G}|2\gamma\rangle \quad (3.15)$$

cannot be reliably calculated in the present-day theory since they are determined by large-distance dynamics. The last part of this assertion is true, of course, and still, surprising though it is, the gluon conversion amplitudes (3.15) are unambiguously calculable in the limit of small momenta of the final particles. For instance,

$$\langle 0|(-b\alpha_s/8\pi)G_{\mu\nu}^a G_{\mu\nu}^a|\pi^+(p_1)\pi^-(p_2)\rangle = (p_1 + p_2)^2 + O(p^4). \quad (3.16)$$

The proof is based on the consideration of the matrix element $\langle \pi^+(p_1)\pi^-(p_2)|\theta_{\mu\nu}|0\rangle$ where $\theta_{\mu\nu}$ is the complete regularized energy-momentum tensor in QCD. The most general expression for $\langle \pi^+\pi^-|\theta_{\mu\nu}|0\rangle$ is

$$\langle \pi^+(p_1)\pi^-(p_2)|\theta_{\mu\nu}|0\rangle = Ar_\mu r_\nu + Bq_\mu q_\nu + Cg_{\mu\nu}, \quad r = p_1 - p_2, \quad q = p_1 + p_2, \quad (3.17)$$

where A, B, C are some scalar functions of the four-momenta. For soft pions, however, these functions reduce to constants fixed by the following requirements (in the chiral limit): (i) symmetry and conservation of $\theta_{\mu\nu}$; (ii) normalization condition $\langle \pi^+(p)|\theta_{\mu\nu}|\pi^-(p)\rangle = 2p_\mu p_\nu$; and (iii) neutrality of $\theta_{\mu\nu}$ with respect to the axial charge, $[\theta_{\mu\nu}(x), Q_5^+] = 0$. Points (i) and (ii) are obvious by themselves. Point (iii) is slightly less trivial. Recall, however, that in the chiral limit the axial charge Q_5^+ is a conserved operator and, hence,

$$[Q_5^+, \theta_{00}(x)] = i\partial^\mu a_\mu^+(x) = 0, \quad a_\mu^+ = \bar{u}\gamma_\mu \gamma_5 d.$$

The vanishing of the commutator $[\theta_{\mu\nu}, Q_5^+]$ can also be checked directly since explicit expressions for $\theta_{\mu\nu}$ and Q_5^+ in terms of the fundamental fields are known.

It is convenient to rewrite condition (iii) in the following equivalent form: (iii') $\langle \pi^+ \pi^- | \theta_{\mu\nu} | 0 \rangle \rightarrow 0$ if p_1 or $p_2 \rightarrow 0$. The equivalence of (iii) and (iii') becomes clear if one takes into account that $\langle \pi^+ \pi^- | \theta_{\mu\nu} | 0 \rangle \xrightarrow{p_1 \rightarrow 0} i f_\pi^{-1} \langle \pi^- | [\theta_{\mu\nu}, Q_5^+] | 0 \rangle$.

Combining (i), (ii) and (iii') we arrive at the following result:

$$\langle \pi^+ \pi^- | \theta_{\mu\nu} | 0 \rangle = \frac{1}{2} r_\mu r_\nu - \frac{1}{2} q_\mu q_\nu + \frac{1}{2} q^2 g_{\mu\nu}, \quad (3.18)$$

$$\langle \pi^+ \pi^- | \theta_\nu^\mu | 0 \rangle = q^2. \quad (3.19)$$

Using the scale anomaly (2.20) we reproduce the theorem (3.16).*)

The left-hand side of eq. (3.16) is proportional to α_s while the right-hand side does not contain such a factor. This “disbalance” takes place for the conversion of gluons into photons as well and results in far-reaching consequences.

Without any connection with the hadronic physics it is worth noting that the ideas presented above have been exploited in ref. [35] treating the propagation of the Goldstone bosons in the gravitational background. Equation (3.19) actually means that the corresponding equation does not contain the term proportional to the scalar curvature, the so-called ξ term.

If the pionic matrix element measures the coupling of gluons (in the 0^+ channel) with the “real” quarks, the photonic amplitudes probe the virtual quark loops since there is no direct photon–gluon interaction. Below we will show that

$$\langle 0 | (\alpha_s/4\pi) G_{\alpha\beta}^a \tilde{G}_{\alpha\beta}^a | \gamma(k_1) \gamma(k_2) \rangle = (\alpha/\pi) N_c \langle Q_q^2 \rangle F_{\mu\nu}^{(1)} \tilde{F}_{\mu\nu}^{(2)}, \quad F_{\mu\nu}^{(i)} = k_\mu^{(i)} \varepsilon_\nu^{(i)} - k_\nu^{(i)} \varepsilon_\mu^{(i)}. \quad (3.20)$$

Equation (3.20) is valid in the chiral limit and, besides that, assumes that the momentum transfer $(k_1 + k_2)^2$ is small as compared to the characteristic hadronic scale ~ 1 GeV. The proof of this relation is very simple. In order to derive the theorem (3.20) multiply the amplitude

$$\left\langle 0 \left| \sum_q \bar{q} \gamma_\mu \gamma_5 q \right| \gamma(k_1) \gamma(k_2) \right\rangle$$

by $q_\mu = (k_1 + k_2)_\mu$ and let $q \rightarrow 0$. On one hand, since there are no massless particles in the singlet channel even in the limit $m_q = 0$, we must get zero. On the other hand, the explicit computation of the divergence of the axial current yields

$$n_f \frac{\alpha_s}{4\pi} G_{\alpha\beta}^a \tilde{G}_{\alpha\beta}^a - \frac{\alpha}{\pi} N_c F_{\mu\nu}^{(1)} \tilde{F}_{\mu\nu}^{(2)} \sum_q Q_q^2, \quad (3.21)$$

where both the gluon and photon parts of the chiral anomaly are taken into account. The low-energy theorem for the matrix element

$$\langle 0 | -(b\alpha_s/8\pi) G_{\alpha\beta}^a G_{\alpha\beta}^a | \gamma(k_1) \gamma(k_2) \rangle$$

*) Equation (3.16) is found in ref. [24] without explicit derivation. The derivation proving the absence of ambiguities in the pionic energy–momentum tensor is worked out in ref. [25].

is very peculiar and interesting. A detailed discussion would go beyond the limits of these lectures. The interested reader is referred to ref. [36].

Concluding this subsection, we give without any details the expressions for the conversion of the operator $G\tilde{G}$ into π and η [37–39],

$$\langle 0|(3\alpha_s/4\pi)G_{\mu\nu}^a\tilde{G}_{\mu\nu}^a|\pi^0\rangle = \frac{3}{\sqrt{2}} \frac{m_d - m_u}{m_d + m_u} f_\pi m_\pi^2, \quad \langle 0|(3\alpha_s/4\pi)G_{\mu\nu}^a\tilde{G}_{\mu\nu}^a|\eta\rangle = \sqrt{3/2} f_\pi m_\eta^2.$$

3.3. The pion spectrum in the decays $\psi' \rightarrow J/\psi \pi\pi$, $Y' \rightarrow Y\pi\pi$

The qualitative experimental situation in these decays is as follows. It is known that the pions are emitted predominantly in the S-wave. The contribution of the D-wave pion does not exceed a few percent in the total decay rate. Besides that, the effective amplitude

$$A_{\text{eff}}(q^2) = [(d\Gamma/dg^2)(\text{phase space})^{-1}]^{1/2} = [(S\text{-wave})^2 + \overline{(D\text{-wave})^2}]^{1/2} \quad (3.22)$$

is well-approximated by the linear dependence of the type

$$A_{\text{eff}} = C(q^2 - \delta). \quad (3.23)$$

[The bar in eq. (3.22) denotes averaging over the pion phase space, C and δ in eq. (3.23) are numerical constants, $q = p_1 + p_2$.]

The questions addressed by the theory are:

- Why is the relative contribution of the D-wave so small?
- Could one predict the linear law (3.23) beforehand?
- What does the theory say about C and δ ?

The starting point of the theoretical analysis is the so-called multipole expansion [40–42], which reflects the fact that we are dealing here with the two-step process. At the first step the gluons are emitted by the heavy quarks (quasi)locally^{*} and at the second step conversion of gluons into pions at large distances takes place. As a result, the amplitude factorizes,

$$A(n_1^3S_1 \rightarrow n_1^3S_1 + \pi\pi) = \langle A_{\text{if}} \rangle \langle 0|\alpha_s E^a E^a|\pi\pi\rangle + \text{higher multipoles}, \quad (3.24)$$

where E^a is the chromoelectric field and the coefficient $\langle A_{\text{if}} \rangle$ contains information only about the heavy quarkonium. One can find it using a model or phenomenologically, as is done in ref. [24] where $\langle A_{\text{if}} \rangle_{J/\psi}$ is extracted from $\psi' \rightarrow J/\psi \eta$.

Now we are interested not in the heavy quarkonium structure but in the gluon dynamics at large distances. Therefore, we will treat $\langle A_{\text{if}} \rangle$ as a given number and will concentrate on the pion matrix element. Since our aim is pedagogical we will work in the chiral limit, $m_\pi^2 = 0$.

A trivial but important observation reads

$$\alpha_s E^a E^a = \frac{1}{2} \alpha_s (E^a E^a + H^a H^a) + \frac{1}{2} \alpha_s (E^a E^a - H^a H^a) = \alpha_s \theta_{00}^G + (2\pi/b) \theta_\mu^\mu, \quad (3.25)$$

^{*} The distance between \bar{Q} and Q in the J/ψ and Y systems is noticeably less than 1 fm, the characteristic distance for conventional old hadrons.

where \mathbf{H}^a is the chromomagnetic field and θ_{00}^G is the 00 component of the gluon part of the energy-momentum tensor. Using the theorem (2.20) we get

$$\langle \pi^+ \pi^- | \alpha_s E^a E^a | 0 \rangle = (2\pi/b) q^2 + \langle \pi^+ \pi^- | \alpha_s(\mu) \theta_{00}^G(\mu) | 0 \rangle. \quad (3.26)$$

In the right-hand side we have to indicate the normalization point of the operator $\alpha_s \theta_{00}^G$ since it is not renormalization-group-invariant; μ is of the order of the inverse quarkonium radius. Repeating almost literally the arguments leading to eqs. (3.18), (3.19) one can easily convince oneself that

$$\langle \pi^+(p_1) \pi^-(p_2) | \theta_{\mu\nu}^G | 0 \rangle = \rho_\pi^G(\mu) \frac{1}{2} (r_\mu r_\nu - q_\mu q_\nu) + \left(\frac{11}{36} + \frac{1}{4} \rho_\pi^G(\mu) \right) q^2 g_{\mu\nu}, \quad (3.27)$$

where ρ_π^G is the gluon share of the pion momentum which is more or less known empirically, $\rho_\pi^G(\mu \sim 1 \text{ GeV}) \cong 0.4$.

At this stage it is convenient to rewrite the $\psi' \rightarrow J/\psi \pi\pi$ amplitude (for definiteness we will speak about this decay) in the Lorenz-invariant form,

$$A(\psi' \rightarrow J/\psi \pi^+ \pi^-) = -(\langle A_{if} \rangle / M_{\psi'}^2) (\varepsilon'_\alpha \varepsilon_\alpha) [(2\pi/b) q^2 \mathcal{P}^2 + \langle \pi^+ \pi^- | \alpha_s(\mu) \theta_{\beta\gamma}^G(\mu) | 0 \rangle \mathcal{P}_\beta \mathcal{P}_\gamma], \quad (3.28)$$

where \mathcal{P} is the ψ' momentum. In the ψ' rest frame eq. (3.28) reduces to eq. (3.24). Omitting simple arithmetic we quote the final answer for the amplitude,

$$\begin{aligned} A(\psi' \rightarrow J/\psi \pi^+ \pi^-) &= -\langle A_{if} \rangle (\varepsilon'_\alpha \varepsilon_\alpha) (2\pi/b) \\ &\times \{ [q^2 - \kappa(q\mathcal{P})/\mathcal{P}^2] + \frac{3}{2} \kappa [r_\beta r_\gamma - \frac{1}{3} (q_\beta q_\gamma - q^2 g_{\beta\gamma})] \mathcal{P}_\beta \mathcal{P}_\gamma / \mathcal{P}^2 \}, \\ \kappa &= (b/6\pi) \alpha_s(\mu) \rho_\pi^G(\mu). \end{aligned} \quad (3.29)$$

The term in the first square brackets represents the pure S-wave, while that in the second square brackets represents the D-wave. (This is clearly seen in the $\pi^+ \pi^-$ center of mass system.) Inessential $O(\alpha_s)$ corrections to the S-wave amplitude are omitted in eq. (3.29).

Thus, the $\psi' \rightarrow J/\psi \pi\pi$ amplitude is completely specified. Our theoretical prediction possesses the following properties.

(i) The D-wave contribution in A_{eff} [see eq. (3.22)] relative to that of the S-wave is equal to $\kappa^2/5$. Although numerically κ is rather uncertain – the normalization point μ is rather low and constitutes several hundred MeV for J/ψ and about 1 GeV for Y – still a very strong suppression of the D-wave is undeniable. Indeed, if $\alpha_s(\mu) \sim 0.7$, $\rho_\pi^G(\mu) \sim 0.4$, then

$$\kappa_\psi \sim 0.15 \text{ to } 0.2, \quad (3.30)$$

and $\kappa^2/5 \sim 10^{-2}$. This fact implies that the gluons “like” to convert into the quark states in the 0^+ channel but are reluctant to do that in the 2^+ channel where the mixing parameter is small, $\sim \alpha_s/\pi$.

(ii) Neglecting now the D-wave which is perfectly justified we can write down the amplitude as follows:

$$A(\psi' \rightarrow J/\psi \pi^+ \pi^-) = -\langle A_{if} \rangle (\varepsilon'_\alpha \varepsilon_\alpha) (2\pi/b) [q^2 - \kappa(\Delta M)^2], \quad (3.31)$$

where $\Delta M = M_{\psi'} - M_{J/\psi}$. This result nicely reproduces the empirical linear q^2 dependence (3.23). The intercept in q^2 in eq. (3.31) lies in the vicinity of 0.1 GeV^2 if $\kappa \cong 0.2$. ($\delta_{\psi \text{ exp}}$ is also approximately equal to 0.1 GeV^2 .)

Inspired by this success in the charmonium family we can attempt to find the pion spectrum intercept δ_Y for $Y' \rightarrow Y\pi\pi$. The accuracy of both the multipole and α_s expansions is expected to be much better here because of the smaller size of Y . The μ dependence of ρ_π^G is known to be very weak and the main change is due to a fall-off of α_s , by a factor of two. Moreover, $(\Delta M)_\psi = (\Delta M)_Y$ and hence the theory predicts that $\kappa_Y \sim 0.5\kappa_\psi$ and $\delta_Y \sim 0.5\delta_\psi$.

This prediction probing rather subtle aspects of QCD is confirmed experimentally.

4. Conclusions

In this report I have given a pedagogical review of the anomalies and their role in quantum chromodynamics. I had no intention to discuss multiple new developments of the issue, growing up rapidly especially in connection with the string theory. My task was to explain the simple physical meaning of the phenomenon and to destroy a mysterious attitude about the subject that emerged partially due to the successful and deep mathematization. I also wanted to consider a few examples of applications in the hadronic physics in order to demonstrate their elegance and power.

Acknowledgements

I am extremely grateful to Roger L. Gilson for his generous help in preparing the manuscript for publication.

References

- [1] S.B. Treiman, R. Jackiw and D. Gross, *Lectures on Current Algebra and its Applications* (Princeton Univ. Press, Princeton, 1972).
- [2] L.D. Faddeev, *Phys. Lett. B* 145 (1984) 81;
L.D. Faddeev and S.L. Shatashvili, *Phys. Lett. B* 167 (1986) 225;
R. Jackiw and R. Rajaraman, *Phys. Rev. Lett.* 54 (1985) 1219;
R. Rajaraman, *Phys. Lett. B* 154 (1985) 305; *B* 162 (1985) 148;
R. Rajaraman and J. Lott, *Phys. Lett. B* 165 (1985) 321;
S.L. Shatashvili, *Teor. Mat. Fiz.* 71 (1987) 40.
- [3] M. Green and J. Schwarz, *Phys. Lett. B* 149 (1984) 117.
- [4] A.Yu. Morozov, *Usp. Fiz. Nauk* 150 (1986) 337.
- [5] A. Casher, J. Kogut and L. Susskind, *Phys. Rev. Lett.* 31 (1973) 792;
G. Kogut and L. Susskind, *Phys. Rev. D* 11 (1975) 3594.
- [6] V.N. Gribov, preprint KFKI-1981-66, Budapest (unpublished).
- [7] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Fortschr. Phys.* 32 (1984) 585.
- [8] S.L. Adler and W.A. Bardeen, *Phys. Rev.* 182 (1969) 1517.
- [9] J.S. Collins, L. Duncan and S.D. Joglekar, *Phys. Rev. D* 16 (1977) 438.
- [10] M.A. Shifman and A.I. Vainshtein, *Nucl. Phys. B* 277 (1986) 456.
- [11] S. Ferrara and B. Zumino, *Nucl. Phys. B* 87 (1975) 207.
- [12] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Phys. Lett. B* 57 (1985) 169.
- [13] K. Wilson and J. Kogut, *Phys. Rev.* 12 (1974) 75.
- [14] A.D. Dolgov and V.I. Zakharov, *Nucl. Phys. B* 27 (1971) 525.
- [15] G. 't Hooft, in: *Recent Developments in Gauge Theories*, eds. G. 't Hooft et al. (Plenum Press, New York, 1980) p. 241.

- [16] S. Coleman and E. Witten, *Phys. Rev. Lett.* 45 (1989) 100.
- [17] S. Vafa and E. Witten, *Nucl. Phys. B* 234 (1984) 173.
- [18] A.I. Vainshtein and V.I. Zakharov, *Usp. Fiz. Nauk* 100 (1970) 225.
- [19] Y. Nambu, in: *Preludes in Theoretical Physics*, ed. A. de Shalit (North-Holland, Amsterdam, 1966);
H. Fritzsch and M. Gell-Mann, in: *Proc. XVI Inter. Conf. on High Energy Physics*, Vol. 2 (Batavia, New York, 1972) p. 135;
S. Weinberg, *Phys. Rev. Lett.* 31 (1973) 494.
- [20] H. Leutwyler, *Nucl. Phys. B* 76 (1974) 413;
S. Weinberg, in: *Festschrift for I. I. Rabi*, ed. L. Motz, (New York Academy of Sciences, New York, 1978).
- [21] E.M. Lifshits and L.P. Pitaevsky, *Relativistic Quantum Theory*, Part II (Moscow, Nauka, 1971) sections 141, 142;
A.I. Akhiezer and V.B. Berestetsky, *Quantum Electrodynamics* (Moscow, Nauk, 1969) section 40.2.
- [22] R. Crewther, *Phys. Lett. B* 70 (1977) 349.
- [23] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys. B* 191 (1981) 301.
- [24] M.B. Voloshin and V.I. Zakharov, *Phys. Rev. Lett.* 45 (1980) 688.
- [25] V.A. Novikov and M.A. Shifman, *Z. Phys. C* 8 (1981) 43.
- [26] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys. B* 166 (1980) 493.
- [27] M.A. Shifman, *Usp. Fiz. Nauk.* 157 (1989) 561 [*Sov. Phys. – Usp.* 32 (1989) 289].
- [28] M.B. Voloshin, *Yad. Fiz.* 44 (1986) 738.
- [29] I.I. Balitsky, D.I. Dyakonov and A.V. Yung, *Yad. Fiz.* 35 (1982) 1300; *Z. Phys. C* 33 (1986) 265.
- [30] D. Dyakonov and M. Eides, in: *Proc. XVI Winter School of Physics (LINP, Leningrad, 1981)* p. 123;
D. Dyakonov, in: *Proc. XXI Winter School of Physics (LINP, Leningrad, 1986)* p. 3.
- [31] A.A. Migdal and M.A. Shifman, *Phys. Lett. B* 114 (1982) 445.
- [32] J. Ellis and J. Lanik, *Phys. Lett. B* 150 (1987) 289.
- [33] G. 't Hooft, *Nucl. Phys. B* 72 (1974) 461.
- [34] E. Witten, *Nucl. Phys. B* 160 (1979) 57.
- [35] M.B. Voloshin and A.D. Dolgov, *Yad. Fiz.* 35 (1982) 213.
- [36] H. Leutwyler and M. Shifman, *Phys. Lett. B* 221 (1989) 384.
- [37] D. Gross, S.B. Treiman and F. Wilczek, *Phys. Rev. D* 19 (1979) 2188.
- [38] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Nucl. Phys. B* 165 (1980) 55.
- [39] B.L. Ioffe and M.A. Shifman, *Phys. Lett. B* 95 (1980) 99.
- [40] K. Gottfried, *Phys. Rev. Lett.* 40 (1978) 598.
- [41] M.B. Voloshin, *Nucl. Phys. B* 154 (1979) 365;
M. Peskin, *Nucl. Phys. B* 156 (1979) 365.
- [42] T.M. Yan, *Phys. Rep.* 142 (1986) 357.
- [43] J.E. Hetrick and Y. Hosotani, *Phys. Rev. D* 38 (1988) 2621.