TWO-DIMENSIONAL SIGMA MODELS: MODELLING NON-PERTURBATIVE EFFECTS IN QUANTUM CHROMODYNAMICS

V.A. NOVIKOV, M.A. SHIFMAN, A.I. VAINSHTEIN and V.I. ZAKHAROV

Institute of Theoretical and Experimental Physics, Moscow, 117259, USSR

and

Institute for Nuclear Physics, Novosibirsk, 630090, USSR

NORTH-HOLLAND – AMSTERDAM

PHYSICS REPORTS (Review Section of Physics Letters) 116, No. 3 (1984) 103–171. North-Holland, Amsterdam

TWO-DIMENSIONAL SIGMA MODELS: MODELLING NON-PERTURBATIVE EFFECTS IN QUANTUM CHROMODYNAMICS

V.A. NOVIKOV, M.A. SHIFMAN, A.I. VAINSHTEIN and V.I. ZAKHAROV

Institute of Theoretical and Experimental Physics, Moscow, 117259, USSR

and

Institute for Nuclear Physics, Novosibirsk, 630090, USSR

Received 14 May 1984

Contents:

Abstract:

This review is devoted to the discussion of the parallel existing between four-dimensional gauge theories and two-dimensional sigma models. We use sigma models as a laboratory allowing us to investigate such issues as the operator product expansion beyond perturbation theory, vacuum condensates, low-energy theorems and other non-perturbative aspects. All these questions are intensively discussed in the current literature, and we give a critical analysis of the situation. In particular, it is explained that, contrary to recent claims, one can define the operator product expansion beyond perturbation theory in a perfectly consistent way, with no ambiguities.

The second part of the review represents a detailed discussion of the supersymmetric O(3) sigma model. After a simple description of the model we concentrate on instantons. The instanton-based method for calculating the exact Gell-Mann–Low function and bifermionic condensates is described. An analogue of this method has been previously used by us in four-dimensional Yang–Mills theories. Here we try to elucidate all aspects of the method in simplified conditions. The basic points are: (i) the instanton measure from purely classical analysis; (ii) a non-renormalization theorem in self-dual external fields; (iii) existence of vacuum condensates and their compatibility with supersymmetry.

Pursuing pedagogical purposes we use much space for technical details and computations.

Single orders for this issue

PHYSICS REPORTS (Review Section of Physics Letters) 116, No. 3 (1984) 103–171.

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 42.00, postage included.

0 370-1573/84/$20.70

1. Introduction

Most concentrated efforts are now invested in non-Abelian Yang–Mills theories. The dynamics inherent to these theories is rich and unusual. Suffice it to mention such phenomena as confinement of colour, dimensional transmutation, spontaneous breaking of chiral symmetry, etc. Unfortunately, in spite of considerable progress, we are still far from a complete understanding of these phenomena. It has become clear that they are somehow related to a complicated vacuum structure. However, the investigation of this vacuum structure turned out to be a notoriously difficult task.

To get a new insight into the problem theorists often simplify the original theory. For instance, they neglect fermions, or consider the large N limit, or even – quite a radical step – substitute the original Yang–Mills theory by some simple model which only partly imitates the characteristic features of non-Abelian dynamics. This, of course, shifts the emphasis away from reality, but, simultaneously, gives more freedom for theoretical manoeuvre and reveals new aspects of the theory such as hidden parameters and symmetries. Moreover, such toy models serve sometimes as perfect theoretical laboratories to test methods and approaches developed for solving actual problems in actual physics.

As noted first by Polyakov [1], there exists a deeply rooted analogy between four-dimensional Yang–Mills theories and two-dimensional sigma models. Indeed, sigma models, first proposed 25 years ago [2], served to implement the idea of spontaneous chiral symmetry breaking in times preceding the discovery of QCD. Two years after the discovery of asymptotic freedom in QCD it was shown [1] that the coupling constant of O(N) sigma models in two dimensions, falls off at short distances as well. As for non-trivial topological solutions of classical field equations in QCD and in the O(3) sigma model, they were obtained simultaneously [3, 4]. Moreover, instantons in the SU(N) generalizations of the sigma model (the so called CP(N-1) models) were found for all N [5]. In the large N limit both theories, O(N) and CP($N-1$), were exactly solved [1, 6, 7]. As a result it became possible to understand in the language of two-dimensional models, the role of instantons in the formation of the physical spectrum, to formulate the U(1) problem, to obtain a massless vector gauge field as a bound state of original fields, and so on [7].

Motivated by these observations we will discuss here some non-perturbative effects in the framework of sigma models. More specifically, we concentrate on problems of current interest in QCD, such as the status of the Wilson operator expansion (OPE) outside perturbation theory, vacuum condensates, low-energy theorems, instantons. In many cases interest in the corresponding problems in QCD stems from the so-called QCD sum rules [8]. It might be worth emphasizing from the very beginning that the paper adds very little, if at all, to the understanding of the dynamics of two-dimensional models. Instead we will use sigma models as a safe theoretical framework to analyse under simplified conditions some problems of practical importance – and dispute – in QCD.

This review is organized as follows. We start in section 2 with a review of the basic features of O(N) sigma models that sets up the framework for all subsequent considerations and, hopefully, makes the paper self-contained.

In section 3 we proceed to the discussion of the Wilson operator expansion (OPE) [9]. We touch upon several interrelated but not identical aspects of the problem: the mathematical formulation of the OPE, its physical meaning, and, finally, non-perturbative vacuum expectations of local operators. Among other things, we emphasize that consistent OPE necessarily calls for the introduction of an auxiliary parameter, the normalization point μ of the operators considered. Once this is done, both the coefficient functions and the vacuum condensates are unambiguously defined. Our overall conclusion is that the OPE is well defined outside perturbation theory. Moreover, sigma models with large N bear a close resemblance to QCD as far as numericals are concerned. Namely, the μ dependence of vacuum condensates is weak (under a reasonable choice of μ) and can be neglected numerically to a valid approximation. It might be worth noting that the status of OPE has been recently the subject of some discussion and controversy [10–13]. We present an analysis of the problems as raised in refs. [11, 12] and try to explain how the puzzles mentioned there are resolved.

The next section is devoted to the issue of anomaly in the trace of the energy-momentum tensor. By using the exact solution of the O(N) model in the large N limit we demonstrate that the anomaly determines the masses of physical particles. A similar relation has been asserted to hold within QCD [14].

In section 5.1 we discuss low-energy theorems which relate low-energy scattering amplitudes to the non-perturbative vacuum expectation values of some operators. We demonstrate that the theorems do reproduce the exact answer known in the sigma models for large N. This is all the more gratifying since, in QCD, similar theorems serve as an important and sometimes unique source of dynamical information [15].

The same section treats the so called U(1) problem [16] in CP($N-1$) models. In QCD this problem is a key issue in understanding the mechanism for the generation of the η' mass. On the technical side, one studies the low-energy behaviour of correlation functions induced by operators of topological charge density. We analyse a similar correlation function in the sigma model and demonstrate that the general picture developed in QCD is confirmed.

Finally, we turn in section 6 to the supersymmetric O(3) sigma model proposed in refs. [17, 18]. The construction and basic properties of the model are first reviewed. We then make use of instanton calculus to evaluate the exact Gell-Mann–Low function for this theory. To a great extent the derivation runs parallel to that of ref. [19] where the complete beta function of the supersymmetric Yang–Mills theory has been found. However, there are some novel points as well and, what is more important for our present purposes, the consideration of two-dimensional models allows us to elucidate some points (the pedagogical aspect is essential).

2. The O(N) model in the large N limit

In this section we give a brief review of the O(N) sigma model and its generalizations. These models have been exhaustively studied and their solutions in the large N limit are described for instance in refs. [6, 7]*. We do not assume, however, the reader to be familiar with these papers and try to summarize the obtained results.

The O(N) sigma model in $(1+1)$ space-time is a theory of N fields σ^a ($a = 1, \ldots, N$) defined on the unit sphere:

$$\sigma^a(x)\,\sigma^a(x) = 1\,. \tag{1}$$

They transform according to the vector representation of the group O(N). The Lagrangian is chosen to be of the following form**

* An excellent introduction to CP ($N-1$) model ($N \to \infty$) is given by Coleman [20].

** If not stated otherwise, all expressions refer to Euclidean space-time.

$$\mathscr{L} = \frac{N}{2f} (\partial_\mu \sigma^a(x))\, (\partial_\mu \sigma^a(x)) \,. \tag{2}$$

At first sight this is the theory with no interaction. This is, of course, not so. One can easily convince oneself that, upon solving the constraint (1) with respect to one of the field components we arrive at a non-trivial interaction between the remaining components. Within the framework of perturbation theory it is described by vertices with 4, 6, 8 etc. legs. The quantity f in eq. (2) plays the role of the coupling constant. The exact solution is, however, not exhausted by perturbation theory. Thus, for $N = 3$ (and only for $N = 3$) the classical field equations admit solutions with finite action – the instantons [3] (see section 6). Moreover, non-perturbative effects take place for $N \neq 3$ as well, i.e. they do not depend on the existence of instantons. We shall demonstrate this by considering the large N limit ($N \to \infty$).

Before plunging into solving the theory in this limit it is convenient to change the normalization of the σ field and to account for the constraint (1) by virtue of a Lagrange multiplier $\alpha(x)$. The action S_{E}, and the generating functional for the Green functions $Z_{\mathrm{E}}[J]$ in Euclidean space-time, can be written as

$$S_{\mathrm{E}}[\sigma, \alpha] = \frac{1}{2} \int \mathrm{d}^2 x \left\{ \partial_\mu \sigma^a(x)\, \partial_\mu \sigma^a(x) - \frac{\alpha(x)}{\sqrt{N}} \left(\sigma^a \sigma^a(x) - \frac{N}{f} \right) \right\} \tag{3}$$

and

$$Z_{\mathrm{E}}[J] = \int \prod_x \mathrm{D}\sigma(x)\, \mathrm{D}\alpha(x) \exp\left\{ -S_{\mathrm{E}} + \int \mathrm{d}^2 x\, J^a(x)\, \sigma^a(x) \right\}. \tag{4}$$

The $N^{-1/2}$ factor in front of $\alpha(x)$ is chosen for convenience. The action (3) is bilinear in $\sigma(x)$, and, therefore, the functional integral over σ is readily calculable:

$$\begin{aligned} Z_{\mathrm{E}}[J] &= \int \prod_x \mathrm{D}\alpha(x) \exp\left\{ -S_{\mathrm{eff}} + \frac{1}{2} \int \mathrm{d}^2 x\, J^a(x) \left[\frac{1}{-\partial^2 + \alpha(x)/\sqrt{N}} J^a \right](x) \right\}, \\ S_{\mathrm{eff}} &= \frac{N}{2} \operatorname{Tr} \ln\left[-\partial^2 + \frac{\alpha(x)}{\sqrt{N}} \right] - \int \mathrm{d}^2 x \frac{\sqrt{N}}{2f} \alpha(x) \,. \end{aligned} \tag{5}$$

Here $1/(-\partial^2 + \alpha(x)/\sqrt{N})$ is a symbolic notation for the Green function of the operator $[-\partial^2 + \alpha(x)/\sqrt{N}]$. The crucial point is that for $Z_{\mathrm{E}}[J]$ there exists a stationary point in $\alpha(x)$. As a result, the remaining functional integral can be readily done using the saddle-point technique. Because of Lorentz invariance, the stationary value of $\alpha(x)$ (if it exists) is actually independent of x. Let us denote this constant by $\sqrt{N}\, m^2$. Then

$$\alpha(x) = \sqrt{N}\, m^2 + \alpha_{\mathrm{qu}}(x) \,.$$

Deviations $\alpha_{\mathrm{qu}}(x)$ from the stationary point ($\alpha_{\mathrm{c}} = \sqrt{N}\, m^2$) describe quantum fluctuations of the α field. We expand S_{eff} in α_{qu} assuming the fluctuations to be small. Then

$$S_{\text{eff}} = \frac{N}{2}\operatorname{Tr}\ln(-\partial^2 + m^2) - \int d^2x\, \frac{m^2 N}{2f} + \frac{N}{2}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\operatorname{Tr}\left[\frac{1}{-\partial^2 + m^2}\frac{\alpha_{\text{qu}}}{\sqrt{N}}\right]^k - \frac{\sqrt{N}}{2f}\int d^2x\, \alpha_{\text{qu}}\,. \tag{6}$$

The first two terms in the expansion are inessential constants, and they will be omitted for a while. We next transform the term linear in $\alpha_{\text{qu}}(x)$:

$$\frac{\sqrt{N}}{2}\operatorname{Tr}\left[\frac{1}{-\partial^2 + m^2}\alpha_{\text{qu}}\right] \stackrel{\text{def}}{=} \frac{\sqrt{N}}{2}\int d^2x\,\langle x|\frac{1}{-\partial^2 + m^2}|x\rangle\,\alpha_{\text{qu}}(x)$$
$$= \frac{\sqrt{N}}{2}\langle 0|\frac{1}{-\partial^2 + m^2}|0\rangle\int d^2x\,\alpha_{\text{qu}}(x) = \int \frac{d^2p}{(2\pi)^2}\frac{1}{p^2 + m^2}\frac{\sqrt{N}}{2}\int d^2x\,\alpha_{\text{qu}}(x)\,.$$

The integral over momenta which emerges here calls for a regularization. To this end we introduce a cut-off momentum M_0,

$$\int \frac{d^2p}{(2\pi)^2}\frac{1}{p^2 + m^2} \to \frac{1}{4\pi}\ln\frac{M_0^2}{m^2}\,.$$

Moreover, for the procedure to be consistent with the existence of the saddle-point, the expansion of the effective action (6) should contain no term linear in α_{qu}. This is indeed the case provided that m satisfies the relation

$$\frac{1}{f} = \frac{1}{4\pi}\ln\frac{M_0^2}{m^2}\,.$$

Of course, one can introduce an effective coupling constant $f(\mu)$ instead of the bare one f, and then

$$\frac{1}{f(\mu)} = \frac{1}{4\pi}\ln\frac{\mu^2}{m^2}\,. \tag{7}$$

Two comments on the meaning of eq. (7) are in order. First, eq. (7) implies that the theory is asymptotically free. Indeed, for fixed m and $M_0 \to \infty$ the bare coupling constant vanishes, $f \to 0$. Second, it demonstrates the so-called dimensional transmutation, which corresponds to the occurrence of a mass parameter m whose dependence on the coupling constant is non-analytical:

$$m^2 = M_0^2\exp(-4\pi/f)\,.$$

Let us now turn to the next terms in the expansion in α_{qu} of S_{eff}. The bilinear term

$$S_{\text{eff}}^{(2)} = -\frac{1}{4}\operatorname{Tr}\left[\frac{1}{-\partial^2 + m^2}\alpha_{\text{qu}}\right]^2 = -\frac{1}{4}\int d^2x\, d^2y\,\alpha_{\text{qu}}(x)\,\Gamma(x-y)\,\alpha_{\text{qu}}(y)$$

describes the propagation of α "particles". Their propagator, evidently, reduces to

$$D^{(\alpha)}(p) = -2/\Gamma(p)\,,$$

where $\Gamma(p)$ is the Fourier transform of $\Gamma(x-y)$,

$$\Gamma(p) = A(p) \equiv \int \frac{\mathrm{d}^2 q}{(2\pi)^2} \frac{1}{(q^2+m^2)\,[(p+q)^2+m^2]} = \frac{1}{2\pi} \frac{1}{\sqrt{p^2(p^2+4m^2)}} \ln \frac{\sqrt{p^2+4m^2}+\sqrt{p^2}}{\sqrt{p^2+4m^2}-\sqrt{p^2}}\,. \tag{8}$$

The function $D^{(\alpha)}(p)$ has, however, no poles in p^2 and only a cut starting at $p^2 = -4m^2$; therefore, strictly speaking, the α field does not correspond to any real particle.

Knowing the propagator $D^{(\alpha)}(p)$ one can easily calculate $Z[J]$ with the help of perturbation theory. It is convenient to formulate the result in terms of Feynman diagrams (fig. 1). The theory describes the propagation of N massive particles with Green function $D^{ab}(p) = \delta^{ab}/(p^2+m^2)$ (fig. 1a), the propagation of the α "particle" with Green function $D^{(\alpha)}(p) = -2/A(p)$ (fig. 1b) and also their interaction, characterized by the vertex $\Gamma_{ab} = -(1/\sqrt{N})\delta^{ab}$ (fig. 1c). In the leading N approximation graphs like those depicted in fig. 2 are already accounted for in $D^{(\alpha)}(p)$. To avoid double counting one should then not include them explicitly. The same is true for graphs of the tadpole type (fig. 3).

The perturbation theory constructed above essentially differs from that appearing in the original formulation of the model (1), (2). First and foremost, the former incorporates explicitly the leading non-perturbative effect, the generation of a mass for the σ particles (and an increase in their number, from $N-1$ up to N). Second, the structure of the corresponding Feynman diagrams becomes much simpler and this is also an important advantage. Indeed, since the α "particle" is a singlet with respect to the O(N) group, each term of the $1/N$ expansion is determined now by a finite number of Feynman graphs. In particular, the leading term is fixed by the tree diagrams alone (see above).

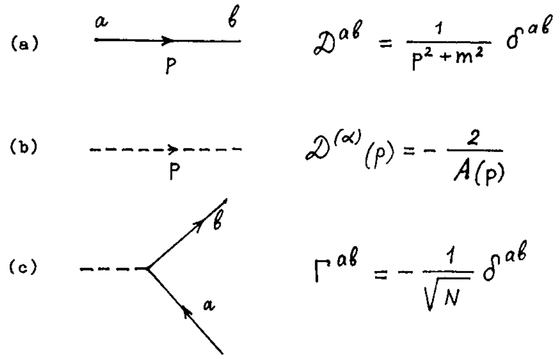


Fig. 1. Feynman rules in the O(N) sigma model.

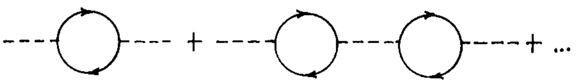


Fig. 2. One-particle reducable diagrams which are accounted for in the α propagator.

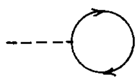


Fig. 3. Tadpoles, not to be included in the Feynman diagrams for σ and α fields.

3. Operator expansion and vacuum condensates

In this section we address ourselves to the central issue of the present review, namely, the status of the operator expansion (OPE) beyond perturbation theory.

The general idea behind the OPE in asymptotically free field theories is as follows. Because of asymptotic freedom small size fluctuations are well under theoretical control. First, there are perturbative fluctuations whose effect is characterized by a small running coupling constant. Second, there exist, as a rule, classical solutions and quantum fluctuations around them. Their effect is exponential in the inverse coupling constant. However, it can be calculated exactly, if needed, as long as the size of the fluctuations remains small.

Thus, the first step in constructing OPE is to integrate over the small size fluctuations explicitly. The result of the integration is a kind of effective Lagrangian, or a set of terms with various numbers of external legs. This is, however, not the final answer for the physical amplitude. To find it we need to account as well for large-scale fluctuations. With respect to the large-scale fluctuations the effective Lagrangian obtained at the first step materializes as a set of operators. The matrix elements of these operators depend upon large distance dynamics. For instance, in QCD, they are generally speaking, not known. On the other hand, in sigma models (at least in the large N limit) all relevant matrix elements are calculable (see below).

We have so far introduced one mass scale, μ. Namely, we assume that an explicit integration over fluctuations of sizes $\rho < \mu^{-1}$ is performed. The mass scale μ must be chosen in such a way that the running coupling constant is small at this mass scale.

If we have no other mass parameter – for instance, if we deal with a vacuum-to-vacuum transition – then the procedure sketched above is of little practical value since all infinite series of matrix elements must be summed up in order to get the physical amplitude.

Therefore, one usually considers a case where there is another large external scale, say, the momentum transfer, $q \gg \mu$. Then, the coefficient functions for the various operators as determined at the first stage – through integration over small-size fluctuations – are inverse powers of (q/μ). One can then cut the series off keeping only the first few terms.

Clearly enough, the choice of μ is not unique. As far as the coupling constant is small, μ can be varied. What happens then? The physical amplitudes do not depend on μ at all. Thus by changing μ we just redistribute the contributions between the matrix elements and the coefficient functions. The matrix elements include, among other things, effects of perturbative fluctuations with size $\rho > \mu^{-1}$. Thus, for practical purposes it is desirable to choose μ as small as possible – within the allowed domain. Then the relative weight of "trivial", i.e. perturbative contributions, and "non-trivial", i.e. large-scale non-perturbative contributions, is tilted in favour of the latter.

In the case of QCD it turns out to be possible to choose μ in such a way that the vacuum matrix elements are grossly dominated [8, 21] by non-perturbative fluctuations. For practical purposes one can keep the μ dependence only in log factors associated with anomalous dimensions and neglect the perturbative contribution to the matrix elements. In other words, to a good approximation, the effect of perturbation theory is absorbed into the coefficient functions while the non-perturbative effects are accounted through the non-trivial vacuum expectation values of various operators. This simplified procedure constitutes the basis of the so-called QCD sum rule method, which has been successfully applied over the past five years to the study of the hadronic spectrum.

This numerical situation is certainly not universal and suited for all field theories. For example, there is no point to discuss non-perturbative effects in QED. The whole business with the OPE in this case is the redistribution of perturbation theory between coefficient functions and matrix elements.

Despite the phenomenological success of the QCD sum rules, there exist a few questions which call for further theoretical consideration. For example, non-perturbative effects are exponential in the inverse coupling constant. Is it then consistent to deal with them without summing up the whole perturbative series?

We will study below some of these problems within O(N) sigma models at large N. The advantage of these models is that they can be solved explicitly and one can confront expectations based on OPE with the exact answer.

It is amazing to find out that sigma models do imitate the basic features of QCD. Namely, in the leading $1/N$ approximation one can forget about the μ dependence of the matrix elements. On the other hand, to deal with more subtle questions, arising in next orders in $1/N$, one must adhere to the general procedure and introduce the normalization point μ.

It is worth mentioning here that OPE in two-dimensional (or simple four-dimensional) models has been studied in recent papers [10–13]. The authors of several publications have arrived at a conclusion that OPE is invalid or, at least, that there exist serious difficulties in the formulation of a consistent procedure. Although, later on, some of them have changed their standpoint, we feel that some questions which have not yet received due treatment do remain. The most common and serious mistake is that one forgets about the necessity of introducing a normalization point. We share with David [10] his conclusion on the validity and simplicity of the OPE in the leading $1/N$ order.

3.1. *Vacuum condensates in the* O(N) *sigma model in the leading* $1/N$ *approximation*

In this section we start a systematic discussion of problems relevant to OPE in the O(N) sigma models. In particular, in this section we identify non-perturbative contributions to vacuum matrix elements.

In section 2 it was shown that the α field develops a non-vanishing vacuum expectation value. Namely, to leading order

$$\langle 0|\alpha(x)|0\rangle = \sqrt{N}\, m^2 = \sqrt{N}\, M_0^2 \exp(-4\pi/f)\,.$$

Quantum corrections to this expression are suppressed as powers of $1/N$, and will be discussed later on. Here the only important matter is that in no way they can destroy the condensate. Condensation of the α field automatically entails vacuum condensates for other fields. In particular, due to the equations of motion, we have:

$$(\partial_\mu \sigma)^2 = -\sigma\, \partial^2 \sigma = -\sigma \frac{\alpha}{\sqrt{N}} \sigma = -\alpha \frac{\sqrt{N}}{f}\,,$$

and, therefore,

$$\langle 0|f(\partial_\mu \sigma)^2|0\rangle = -Nm^2\,. \tag{9}$$

The vacuum expectation values of the products $[f(\partial_\mu \sigma)^2]^k$ are also non-vanishing and, in the leading $1/N$ approximation, they factorize [22]

$$\langle 0|[f(\partial_\mu \sigma)^2]^k|0\rangle = [-Nm^2]^k\,. \tag{10}$$

At first sight eq. (9) seems strange since, in Euclidean space-time, $f(\partial_\mu\sigma)^2$ is a positive-definite operator while its vacuum matrix element happens to be negative. To explain this paradox it is worth recalling that the operator $(\partial_\mu\sigma)^2$ is actually a singular object and calls for an accurate definition. Generally speaking, regularization destroys the positivity condition. Let us comment on this assertion in more detail.

To set forth the problem in more contrasting terms we start from a rough calculation which reduces to the following. The product of field $\partial_\mu\sigma(x)\,\partial_\mu\sigma(x)$ is substituted by the corresponding Green function taken at coinciding points, i.e.

$$\langle 0|f(\partial_\mu\sigma)^2|0\rangle = f N \int \frac{\mathrm{d}^2 p}{(2\pi)^2}\frac{p^2}{p^2+m^2}\,. \tag{11}$$

Cutting the integral at the upper limit of integration, $p^2_{\max} = \mu^2$, we get

$$\begin{aligned}\langle 0|f(\mu)\,(\partial_\alpha\sigma)^2|_\mu|0\rangle &= \frac{f(\mu)N}{4\pi}\left(\mu^2 - m^2 \ln\frac{\mu^2}{m^2}\right) = \frac{f(\mu)\,N}{4\pi}\,\mu^2 - Nm^2\\ &= \frac{f(\mu)\,N}{4\pi}\,\mu^2 - N\mu^2 \mathrm{e}^{-4\pi/f}\,.\end{aligned} \tag{12}$$

Corrections to this relation which are associated with the exchange of the α "particle" are proportional to $(1/N)$ to some positive power. Therefore, calculation of the $(\partial_\mu\sigma)^2$ condensate in the leading $1/N$ approximation begins and ends just at the first step. The result is representable as a sum of two terms. The first one is regular in f and can be obtained in perturbation theory. This corresponds to the original formulation of the theory with massless σ particles. The second term is entirely due to the phenomenon of dimensional transmutation and generation of the mass gap. This term is non-analytical in the coupling constant and explicitly non-perturbative in nature. The origins of these two terms are quite different and they cannot be confused with each other. There is no problem with the sign of the vacuum matrix element since the total sum (12) is obviously positive.

In QCD, naively defined vacuum expectation values of local operators are also representable as sums of infinite terms, regular in the coupling constant (they are proportional to $(\mu^2)^{d/2}$), together with finite terms, associated with non-perturbative effects. Moreover, in 4-dimensional gauge theories, the number of regular terms is infinite, and the question that faces us is then: "how can one separate a finite non-perturbative contribution from an infinite number of infinitely large terms?" Pragmatically, the question becomes even more acute in the light of recent attempts to find vacuum condensates from lattice Monte Carlo simulations of gluodynamics [23]. Further complications are due to the fact that the series of regular terms is usually factorially divergent. Sometimes one can even hear assertion according to which an adequate method of summation of divergent series will reproduce the non-perturbative contribution.

We see that the 2-dimensional sigma model in the leading $1/N$ order is much simpler than QCD and, still, the example considered is quite instructive. It definitely shows that the non-perturbative contribution can by no means emerge from summing terms which are regular in the coupling constant. As a matter of fact there are no terms to be summed over since, in this approximation, $\langle 0|f(\partial_\mu\sigma)^2|0\rangle$ contains a single perturbative term. Moreover, we will introduce a regularization which will ensure the absence of any perturbation theory contribution to this vacuum condensate.

To this end, we must give a formal definition of the singular product $\partial_\mu\sigma(x)\,\partial_\mu\sigma(x)$. As one of the possible definitions let us accept the following

$$\langle 0|\partial_\mu\sigma^a(0)\,\partial_\mu\sigma^a(0)|0\rangle = \lim_{x\to 0}\langle 0|\mathrm{T}\{\partial_\mu\sigma^a(x)\,\partial_\mu\sigma^a(0)\}|0\rangle \tag{13}$$

where the symbol T denotes the Dyson T-ordering. This ordering is defined in Minkowski space-time and, therefore, the matrix elements in eq. (13) are to be understood as referred to Minkowski space. Then

$$\langle 0|\mathrm{T}\{\partial_\mu\sigma^a(x)\,\partial_\mu\sigma^b(y)\}|0\rangle = \frac{\partial}{\partial x_\mu}\frac{\partial}{\partial y_\mu}\langle 0|\mathrm{T}\{\sigma^a(x)\,\sigma^b(y)\}|0\rangle + \delta(x_0-y_0)\,[\partial_0\sigma^a(x)\,\sigma^b(y)]\,.$$

Using the canonical commutation relation we get

$$\begin{aligned}\langle 0|\mathrm{T}\{\partial_\mu\sigma^a\,(x)\,\partial_\mu\sigma^b\,(y)\}|0\rangle &= \delta^{ab}\int\frac{\mathrm{d}^2p}{(2\pi)^2}\,\mathrm{e}^{-\mathrm{i}p(x-y)}\frac{\mathrm{i}p^2}{p^2-m^2} - \mathrm{i}\delta^{ab}\delta^{(2)}(x-y)\\ &\underset{x\to y}{=} \delta^{ab}\left\{\mathrm{i}\,\delta^{(2)}(x-y) + \frac{m^2}{4\pi}\ln\frac{1}{m^2|x-y|^2} + \mathrm{O}(|x-y|^2\ln|x-y|)\right\} - \mathrm{i}\,\delta^{ab}\,\delta^{(2)}(x-y)\,.\end{aligned} \tag{14}$$

The quadratically divergent term proportional to μ^2 in eq. (12) corresponds to the delta function in eq. (14), $\delta^{(2)}(x-y)$, taken at the coinciding points ($x=y$). As compared with the naive calculation outlined above there appears an extra commutator term which cancels the $\delta^{(2)}(x-y)$ term (or, what is just the same, μ^2). As a result, the vacuum expectation value (in Minkowski space) turns out to be

$$\langle 0|f\,\partial_\mu\sigma^a\,\partial_\mu\sigma^a|0\rangle = \lim_{x\to 0} f(|x|)\,\langle 0|\mathrm{T}\{\partial_\mu\sigma^a(x)\,\partial_\mu\sigma^a(0)\}|0\rangle = Nm^2\,,$$

$$f(|x|) \equiv 4\pi/\ln(1/m^2|x|^2)\,,$$

in complete correspondence with the Euclidean matrix element (9). The adopted regularization procedure (see eq. (13)) makes the operator $f(\partial_\mu\sigma)^2$ well-defined but it simultaneously eliminates its positive definiteness so that the negative sign in eq. (9) should not cause any surprise.

One might doubt that the regularization proposed above is the best one and look for a more apt regularization. In particular, it seems more natural to define $\langle 0|f(\partial_\mu\sigma)^2|0\rangle$ in terms of a functional integral. More specifically, one may introduce the generating functional

$$\begin{aligned}Z_{\mathrm{E}}[J,\varphi] = \int\prod_x \mathrm{D}\sigma(x)\,\mathrm{D}\alpha(x)\exp\Big\{-\frac{1}{2}\int\mathrm{d}^2x\left[(1+\varphi)\,(\partial_\mu\sigma)^2 + \frac{\alpha(x)}{\sqrt{N}}\left(\sigma^2(x)-\frac{N}{f}\right)\right]\\ + \int\mathrm{d}^2x\,J^a(x)\,\sigma^a(x)\Big\}\end{aligned} \tag{15}$$

and calculate the vacuum average using the conventional formula

$$-\tfrac{1}{2}\langle 0|(\partial_\mu\sigma)^2|0\rangle = \frac{1}{Z_{\mathrm{E}}[J,\varphi]}\frac{\delta}{\delta\varphi}Z_{\mathrm{E}}[J,\varphi]\bigg|_{\substack{J=0\\ \varphi=0}}. \tag{16}$$

In the operator language such an averaging corresponds usually to the Wick T-product in which (in contradistinction to the Dyson T-product) one can freely transpose the symbols of derivatives and T-ordering without adding commutators. Therefore it seems, at first sight, that the regularization procedure (16) would bring us back to the naive calculations (11)–(12).

This is not so. As well known from non-relativistic quantum mechanics (see, for instance, the book of Feynman and Hibbs [24]) variation over a source coupled to kinetic energy is a rather subtle business. In appendix A it is shown that the consistent treatment of the functional integral actually leads to an answer identical to eq. (14), i.e. to the result stemming from the Dyson T-product. Thus, the perturbative contribution in the vacuum expectation value $\langle 0|f(\partial_\mu\sigma)^2|0\rangle$ vanishes, at least in the leading N approximation.

3.2. The operator expansion in the limit $N\to\infty$

Here we shall prove that the naive OPE is valid in the limit $N\to\infty$. But at first a few words about perturbation theory in O(N) models are in order. The interrelation between perturbative and non-perturbative effects in the model is, in a sense, opposite to what happens in the Higgs model. Perturbation theory in O(N) models can be developed in several distinct ways. We can, for instance, solve the additional condition with respect to one of the components, say, σ_N

$$\sigma_N = \sqrt{N/f - \sigma^i\sigma^i}\,, \qquad i = 1, 2, \ldots, N-1$$

and substitute it in the original Lagrangian

$$\begin{aligned}\mathcal{L} &= \frac{1}{2}\left\{\partial_\mu\sigma^i\,\partial_\mu\sigma^i + \left(\frac{N}{f} - \sigma^i\sigma^i\right)^{-1}(\sigma^k\,\partial_\mu\sigma^k)(\sigma^l\,\partial_\mu\sigma^l)\right\}\\ &= \tfrac{1}{2}\,\partial_\mu\sigma^i\left\{\delta^{ik} + \frac{f}{N}\sigma^i\sigma^k\sum_{n=0}^{\infty}\left(\frac{f}{N}\sigma^2\right)^n\right\}\partial_\mu\sigma^k\,.\end{aligned} \tag{17}$$

If the coupling constant f is small, the second term in eq. (17) can be treated as a perturbation. Then, to zeroth order in the coupling constant, the vacuum expectation value of the Nth field component is non-vanishing

$$\langle 0|\sigma_N|0\rangle = \sqrt{N/f}$$

while oscillations of the orthogonal components σ_i correspond to Goldstone bosons

$$\mathcal{L}^{(0)} = \tfrac{1}{2}\,\partial_\mu\sigma^i\,\partial_\mu\sigma^i\,.$$

From the discussion in section 2 it is clear that the exact solution has nothing to do with the perturbative scenario: the vacuum turns out to be non-degenerate and there are no Goldstone bosons at all. The spectrum consists of N massive particles (this corresponds to one particle more than the number of independent dynamical fields in the Lagrangian (17)). Therefore, perturbation theory as sketched above is constructed over a spontaneously-broken vacuum while the exact solution corresponds to the restoration of the O(N) symmetry*. It seems interesting to check whether one can reproduce the results of the exact solution using the operator expansion in this quite non-trivial situation.

Before turning directly to the operator expansion we must ascertain the values of matrix elements of various operators over the physical vacuum. The results of section 2 imply that

$$\langle f\sigma^i\sigma^k\rangle = \delta^{ik}$$
$$\langle f\,\partial_\mu\sigma^i\,\partial_\nu\sigma^k\rangle = -g_{\mu\nu}\,\delta^{ik}\,m^2/2\,, \qquad i, k = 1, \ldots, N-1\,.$$

Moreover, since the σ field is dimensionless, the number of operators of any given dimension is infinitely large, and even the calculation of a single power correction in two-dimensional theories requires the summation of infinitely many matrix elements. These matrix elements factorize in the leading N approximation, for instance

$$\langle \underbrace{\sigma^2\cdot\sigma^2\cdots\sigma^2}_{n}\,(\partial_\mu\sigma)^2\rangle = \langle\sigma^2\rangle^n\,\langle(\partial_\mu\sigma)^2\rangle\,.$$

The factorization property means, in particular, that the vacuum expectation values of operators with O(N) indices contracted in an unfavourable way are small. For example:

$$\langle\sigma^i\,\partial_\mu\sigma^i\,\sigma^k\,\partial_\mu\sigma^k\rangle = \langle\sigma^i\sigma^k\rangle\,\langle\partial_\mu\sigma^i\,\partial_\mu\sigma^k\rangle = \frac{1}{N}\langle\sigma^2\rangle\,\langle(\partial\sigma)^2\rangle\,;$$

and, because of this, the analysis simplifies greatly.

To circumvent some technical complications we shall deal, however, not with the σ_i fields but introduce instead new variables ϕ_i ** such that [6a]

$$\sigma_i = \phi_i\Big/\left(1+\frac{f}{4N}\,\phi_k\phi_k\right).$$

In terms of these variables the action acquires the form standard for non-linear chiral models,

$$\mathcal{L} = \frac{\frac{1}{2}\partial_\mu\phi^i\,\partial_\mu\phi^i}{[1+(f/4\pi)\phi^2]^2}\,.$$

At the same time, the set of relations obtained for the vacuum mean values of the σ fields reduces to

$$\langle 0|\tfrac{1}{4}f\phi^i\phi^k|0\rangle = \delta^{ik}\,, \qquad \langle 0|\tfrac{1}{4}f\,\partial_\mu\phi^i\,\partial_\nu\phi^k|0\rangle = -\tfrac{1}{2}m^2 g_{\mu\nu}\delta^{ik}\,,$$

together with a factorization property for more complex operators.

* A detailed analysis of the O(N) symmetry from this point of view is carried out in ref. [6].
** Appendix B demonstrates how one can operate directly with the original fields σ_i.

In further computations we shall need, as an input, the operator expansion for the ϕ field propagator. The latter is readily constructed. The most straightforward and convenient method for that purpose is to use the external field technique. We split ϕ into two pieces,

$$\phi^i = c^i + q^i ,$$

where c^i is a c-number external field, satisfying the classical equations of motion, while q^i stand for a small (quantum) fluctuation over the classical background. The part of the Lagrangian bilinear in q^i describes the propagation of waves in the external field $c(x)$:

$$\mathcal{L}^{(2)} = \frac{1}{2(1+\lambda c^2)^2}\left\{(\partial_\mu q)^2 - \frac{2\lambda(\partial_\mu c)^2}{1+\lambda c^2}\, q^2 - 8\lambda\,\frac{(\partial_\mu c^i)c^k}{1+\lambda c^2}\,(\partial_\mu q^i)\, q^k + 12\lambda^2\,\frac{(\partial_\mu c)^2 c^i c^k}{(1+\lambda c^2)^2}\, q^i q^k\right\},$$

where $\lambda = f/4N$. This expression for $\mathcal{L}^{(2)}$ implies that the correlation function $(-\mathrm{i})\langle 0|\mathrm{T}\{q^i(x)\, q^k(y)\}|0\rangle$ is representable in the form

$$D^{ik}(x, y) = \langle x|\{[\mathcal{P}^2 + M^2(c)]\,\delta^{ik} + \mathrm{i}A^{ik}_\mu(c)\,\mathcal{P}_\mu + B^{ik}(c)\}^{-1}(1+\lambda c^2)^2|y\rangle . \tag{18}$$

Here $|x\rangle$ denotes the eigenstate of the coordinate operator, $\hat{X}|x\rangle = x|x\rangle$, (hereafter carets over letters mark operators), $\langle x|y\rangle = \delta(x-y)$, $\hat{\mathcal{P}}_\mu$ is the momentum operator, $\langle x|\hat{\mathcal{P}}_\mu|y\rangle = -\mathrm{i}\,\partial_\mu\delta(x-y)$. Moreover, the coefficients A, B, M are functions of the classical field $c(\hat{X})$

$$M^2(c) = -2\lambda(\partial_\mu c)^2/(1+\lambda c^2) ;$$

$$A^{ik}_\mu(c) = 4\lambda[c\,\partial_\mu c\,\delta^{ik} + (c^k\,\partial_\mu c^i - c^i\,\partial_\mu c^k)]/(1+\lambda c^2) ;$$

$$B^{ik}(c) = 12\lambda^2\frac{(\partial_\mu c)^2\, c^i c^k}{(1+\lambda c^2)^2} + 4\lambda(1+\lambda c^2)^2\,\partial_\mu\left[\frac{c^k\,\partial_\mu c^i}{(1+\lambda c^2)^3}\right].$$

It is helpful to rewrite the propagator in the momentum representation:

$$D^{ik}(q) = \int \mathrm{d}^2x\;\mathrm{e}^{-\mathrm{i}qx}\,\langle x|D^{ik}|0\rangle = \int \mathrm{d}^2x\;\langle x|\mathrm{e}^{-\mathrm{i}q\hat{X}}\,D^{ik}\,\mathrm{e}^{\mathrm{i}q\hat{X}}|0\rangle$$

$$= \int \mathrm{d}^2x\;\langle x|\frac{1}{[(\hat{\mathcal{P}}+q)^2 + M^2]\,\delta^{ik} \mp \mathrm{i}A^{ik}_\mu\,(\hat{\mathcal{P}}+q)_\mu + B^{ik}}|0\rangle . \tag{19}$$

In deriving this formula we have exploited some evident relations, of the type

$$\mathrm{e}^{\mathrm{i}q\hat{X}}\,|y\rangle = \mathrm{e}^{\mathrm{i}qy}\,|y\rangle ; \qquad \mathrm{e}^{-\mathrm{i}q\hat{X}}\,\frac{1}{\hat{\mathcal{P}}_\mu}\,\mathrm{e}^{\mathrm{i}q\hat{X}} = \frac{1}{\hat{\mathcal{P}}_\mu \mp q_\mu} ; \qquad \text{etc.}$$

If the external momentum q^2 is much larger than all other dimensional parameters, we can expand $D^{ik}(q)$ in a series in $1/q^2$:

$$D^{ik}(q) = [1+\lambda\, c^2(0)]^2 \left\{\frac{1}{q^2}\delta^{ik} - \frac{\mathrm{i}q_\mu}{q^4} A^{ik}_\mu \right.$$
$$\left. + \frac{1}{q^4}[-M^2\delta^{ik} + \partial_\mu A^{ik}_\mu - B^{ik}] - \frac{q_\mu q_\nu}{q^6}(2\partial_\nu A^{ik}_\mu + A^{il}_\mu A^{kl}_\nu) + \cdots \right\}. \tag{20}$$

Here the following relations are used:

$$\int \mathrm{d}^2x\, \langle x|F[c(\hat{X})]|0\rangle = F[c(0)] \int \mathrm{d}^2x\, \langle x|0\rangle = F[c(0)]\,,$$

$$\int \mathrm{d}^2x\, \langle x|\hat{\mathscr{P}}_\mu F[c(\hat{X})]|0\rangle = F[c(0)] \int \mathrm{d}^2x\, \mathrm{i}\partial_\mu\, \delta^{(2)}(x) = 0\,,$$

$$F[c(\hat{X})]\,\hat{\mathscr{P}}_\mu = \hat{\mathscr{P}}_\mu\, F[c(\hat{X})] - \mathrm{i}\partial_\mu\, F[c(\hat{X})]\,,$$

where $F[c(\hat{X})]$ is an arbitrary functional of the field $c(\hat{X})$.

The expansion (20) evidently realizes (in the tree approximation) the operator expansion for the two-point function

$$D^{ik}(q) = -\mathrm{i} \int \mathrm{d}^2x\, \mathrm{e}^{-\mathrm{i}qx}\, \mathrm{T}\{\phi^i(x)\, \phi^k(0)\}\,,$$

provided that the functions $M^2(c)$, $A(c)$, $B(c)$ are identified with the operators $M^2(\phi)$, $A(\phi)$, $B(\phi)$. Next, we sandwich eq. (20) between the bra and ket states of the physical vacuum. The rules for calculation of vacuum averages as formulated above, allow us to find all matrix elements, namely:

$$\langle 0|(1+\lambda\phi^2)|0\rangle = 4\,,$$

$$\langle 0|M^2[\phi]|0\rangle = m^2\,,$$

$$\langle 0|A^{ik}_\mu[\phi]|0\rangle = 0\,,$$

$$\langle 0|A_\mu A_\nu|0\rangle \sim \langle 0|B|0\rangle \sim \frac{1}{N} m^2\,.$$

Thus, for large Euclidean momenta $q^2 \gg m^2$ the ϕ field propagator behaves as

$$D^{ik}(q) \underset{q^2\to\infty}{\sim} 4\delta^{ik}\frac{1}{q^2}\left(1 - \frac{m^2}{q^2} + \cdots\right).$$

One can check that the first power correction indeed corresponds to a one-particle excitation with mass m. We shall turn, however, to another problem – the main subject of this section – and analyse the correlation function $S(q^2)$

$$S(q^2) = (-\mathrm{i}) \int \mathrm{d}^2x\, \mathrm{e}^{-\mathrm{i}qx}\, T\{j_s(x), j_s(0)\}\,, \tag{21}$$

where

$$j_s = 2f\mathscr{L}(x) = 4N\lambda\,(\partial_\mu\phi)^2/(1+\lambda\phi^2)^2\,.$$

Let us concentrate first on tree diagrams. At this level the source j_s is expanded up to terms linear in the quantum field q^i. Since $c^i(x)$ is a solution of the classical field equations the expression for the source $j_s(x)$ can be written as a full derivative,

$$j_s = -8N\lambda\,\partial_\mu\left[\frac{\partial_\mu c^i}{(1+\lambda c^2)^2}\,q^i\right].$$

The two-point function (21) can then be rewritten as

$$S^{(0)}(q^2) = (8N\lambda)^2\,q_\alpha q_\beta \int \mathrm{d}^2x\,\langle x|\,\frac{1}{(1+\lambda c^2)^2}\,\partial_\alpha c^i\,\{[(\hat{\mathscr{P}}+q)^2+M^2]\,\delta^{ik}$$
$$+\,\mathrm{i}\,A_\mu^{ik}(\hat{\mathscr{P}}+q)_\mu + B^{ik}\}^{-1}\,\partial_\beta c^k\,|0\rangle\,.$$

Expanding now $S(q^2)$ in $1/q^2$, and transposing the $\hat{\mathscr{P}}_\mu$ operators to the left-most position (by consecutive commuting), we arrive at an explicit expression for $S(q^2)$ in an external field:

$$\hat{S}^{(0)}(q) = \frac{64N^2\lambda^2 q_\alpha q_\beta}{(1+\lambda c^2)^2}\left\{\partial_\alpha c^i\,\partial_\beta c^i\left(\frac{1}{q^2}-\frac{M^2}{q^4}\right)-\frac{1}{q^4}[\partial_\alpha c^i B^{ik}\,\partial_\beta c^k - \partial_\alpha c^i\,\partial^2\partial_\beta c^k\,\delta^{ik} - \partial_\alpha c^i A_\mu^{ik}\,\partial_\mu\partial_\beta c^k]\right.$$
$$\left.+\,\frac{q_\mu q_\nu}{q^6}\,\partial_\alpha c^i\,[-4\partial_\mu\partial_\nu\delta^{ik} + 2(\partial_\mu A_\nu^{ik}) + 2A_\mu^{ik}\partial_\nu - (A_\mu A_\nu)^{ik}]\,\partial_\beta c^k + \mathrm{O}(q^{-6})\right\}.$$

This formula describes a class of graphs displayed in fig. 4. To get the operator expansion we substitute $c^i \to \phi^i$. After averaging over the vacuum state this cumbersome expression simplifies greatly and we are left with the first bracket alone:

$$\langle 0|\hat{S}^{(0)}(q)|0\rangle = S^{(0)}(q) = -64N^2\lambda^2 q_\alpha q_\beta\,\frac{1}{8}\,g_{\alpha\beta}\,\frac{m^2}{\lambda q^2}\left(1-\frac{m^2}{q^2}\right) = -8N^2\lambda m^2\left(1-\frac{m^2}{q^2}\right). \tag{22}$$

The next element of the procedure is the calculation of the loop of fig. 5 in an external field. To this end, the source j_s must be expanded to second order in $q^i(x)$:

$$j_s = 8N\lambda\,\mathscr{L}^{(2)}(q)\,.$$

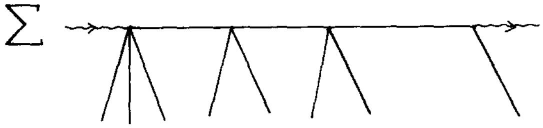


Fig. 4. The tree approximation for the two-point function (21) in an external field.

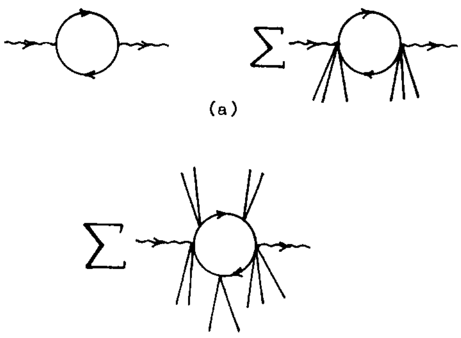


Fig. 5. One-loop diagrams for the two-point function (21) in an external field.

The one-loop calculation is actually no more complicated than the tree calculation but the intermediate formulae are, however, very cumbersome, and we only quote here the answer which emerges after vacuum averaging:

$$S^{(1)}(q^2) = -\frac{q^2 N f^2}{4\pi} \ln \frac{M_0^2}{q^2} \left(1 + \frac{2m^2}{q^2} - \frac{2m^4}{q^4} + \cdots\right) + \mathrm{O}(f^3)\,.$$

As a result, the sum of the tree graphs and of the one-loop graphs reduces to

$$S^{(0)} + S^{(1)} = -\frac{q^2 N f^2}{4\pi} \ln \frac{M_0^2}{q^2} - 2N\left[f + \frac{f^2}{4\pi} \ln \frac{M_0^2}{q^2}\right]\left(m^2 - \frac{m^4}{q^2} + \cdots\right). \tag{23}$$

The terms in the square brackets can be grouped together to give the effective coupling constant

$$f(q^2) = f + \frac{f^2}{4\pi} \ln \frac{M_0^2}{q^2} + \cdots = \frac{4\pi}{\ln(q^2/m^2)}\,,$$

in full accordance with renormalization group arguments*. Let us discuss now the first term which corresponds to a bare loop in perturbation theory. Our aim is to get a renormalization-group-improved expression instead of $\ln(M_0^2/q^2)$.

The point is that the quantities containing a superfluous ("external") logarithm with respect to logarithms governed by the renormalization group must be treated in a special way. The correct recipe is the following. Prior to the final integration one must express the integrand in terms of the effective coupling constant and anomalous dimensions and, only after that, one can perform the integration. In our case this recipe implies

$$q^2 \int_{q^2}^{\infty} \frac{dp^2}{p^2} \frac{N}{4\pi} f^2(p^2) = q^2 \frac{N}{4\pi} \int_{q^2}^{\infty} \frac{dp^2}{p^2} \left(\frac{4\pi}{\ln(p^2/m^2)}\right)^2 = q^2 \frac{4\pi N}{\ln(q^2/m^2)}\,.$$

* Of course, one should convince oneself that the operators at hand have vanishing anomalous dimensions.

The final answer for the two-point function S is representable in the form

$$S(q^2) \underset{q^2\to\infty}{=} -\frac{4\pi N q^2}{\ln(q^2/m^2)}\left[1+\frac{2m^2}{q^2}-\frac{2m^4}{q^4}+\cdots\right](1+\mathrm{O}(f(q^2)))\,. \tag{24}$$

Here all leading log terms of the type $(f\ln(\Lambda^2/q^2))^n$ are summed over. However, the next-to-leading terms $f(f\ln(\Lambda^2/q^2))^n$ are not included. Therefore, strictly speaking, the argument of the logarithm in eq. (24) stays unfixed ($\log(q^2/m^2)$ or $\log(q^2/2m^2)$, or something else?) Anyhow, eq. (24) completes the calculation of the operator product expansion to leading order in $1/N$.

Let us now examine the exact expression for the same quantity. From the equations of motions it is clear that the correlation function at hand coincides (up to a normalization factor) with the propagator for the α field. Namely,

$$S(q^2) = N\,D^{(\alpha)}(q^2) = -N\,4\pi\sqrt{q^2(q^2+4m^2)}\Big/\ln\frac{\sqrt{q^2+4m^2}+\sqrt{q^2}}{\sqrt{q^2+4m^2}-\sqrt{q^2}}\,. \tag{25}$$

Expanding eq. (25) in m^2/q^2 we get for large q^2

$$N\,D^{(\alpha)}(q^2) = -\frac{4\pi N}{\ln(q^2/m^2)+2m^2/q^2+\cdots}\,q^2\left(1+\frac{1}{2}\frac{4m^2}{q^2}-\frac{1}{8}\frac{16m^4}{q^4}+\cdots\right)$$

$$\approx -\frac{4\pi N}{\ln(q^2/m^2)}\,q^2\left(1+\frac{2m^2}{q^2}-\frac{2m^4}{q^4}+\cdots\right)\left[1+\mathrm{O}\left(\frac{1}{\ln(q^2/m^2)}\right)\right].$$

In other words, the OPE-based calculations coincide with the exact result to an approximation where terms of order $q^2(m^2/q^2)^n(\ln q^2)^{-1}$ are included but where one neglects terms suppressed by extra powers of $(\log q^2)^{-1}$. In principle, the latter ones are easily calculable, order by order, and the check of the operator expansion can be continued further. However, it does not seem to be necessary. We hope that the example considered above demonstrates unambiguously that the operator expansion is in perfect correspondence with the exact calculation.

3.3. *The operator expansion beyond the leading approximation*

In our discussion of OPE in the preceding subsection we did not explicitely introduce an intermediate normalization point μ^2. The reason was that, to leading $1/N$ order, there is no dependence on μ^2, neither in the vacuum expectation values of composite operators nor in the coefficient functions. Indeed, the vacuum expectation value of the operator $\alpha(x)=f(\partial_\nu\sigma)^2$, normalized at the point μ^2, can be defined in the following way (see section 3.1):

$$\langle 0|f(\mu)\,(\partial_\alpha\sigma)^2|_\mu|0\rangle = f(\mu)\int\limits_{\mathrm{Eucl.}\,p<\mu}\frac{\mathrm{d}^2p}{(2\pi)^2}\,D(p)\,,$$

where

$$D(p)=\int\mathrm{d}^2x\;\mathrm{e}^{\mathrm{i}p(x-y)}\,\langle 0|\mathrm{T}\{\partial_\alpha\sigma^a(x),\,\partial_\alpha\sigma^a(y)\}|0\rangle$$

$$=\int\mathrm{d}^2x\;\mathrm{e}^{\mathrm{i}p(x-y)}\left\{-\mathrm{i}\,\delta^{(2)}(x-y)+\frac{\partial}{\partial x_\alpha}\frac{\partial}{\partial y_\alpha}\langle 0|\mathrm{T}\{\sigma^a(x)\,\sigma^a(y)\}|0\rangle\right\}=\mathrm{i}\,\frac{m^2}{p^2-m^2}\,, \tag{26}$$

and

$$f(\mu) = 4\pi/\ln(\mu^2/m^2).$$

As a result

$$\langle 0|f(\mu)\,(\partial_\alpha\sigma)^2|_\mu|0\rangle = f(\mu)\int\limits_0^{\mu^2} \frac{dp^2}{4\pi}\frac{m^2}{p^2+m^2} = f(\mu)\frac{m^2}{4\pi}\ln\frac{\mu^2}{m^2} = m^2,$$

and this does not depend on μ^2.

Since the vacuum expectation values of all other operators factorize to leading $1/N$ order and reduce to some powers of $\langle f(\partial\sigma)^2\rangle$, they do not depend on μ^2 either.

However, this is not the general case. Indeed, even in the previous example the operator $(\partial_\alpha\sigma)^2|_\mu$ depends on μ^2 through the log factor. As is well known, this logarithmic dependence on μ^2 can be removed by an appropriate multiplicative redifinition of the operator.

But, generally spreaking, matrix elements can depend on the normalization point μ^2 according to some power of μ^2. A few years ago we have shown [21] that instanton effects in QCD do induce such a power dependence of the vacuum expectation values and of the coefficient functions.

In this section we demonstrate that, in the σ model, this power dependence does emerge in the next-to-leading order in $1/N$. Indeed, let us consider the matrix element of the operator $\alpha^2|_\mu$. By definition

$$\langle\alpha^2|_\mu\rangle = \int\limits_{\mathrm{Eucl.}\,p<\mu} \frac{d^2p}{(2\pi)^2}\tilde{D}^{(\alpha)}(p),$$

where

$$\tilde{D}^{(\alpha)}(p) = \mathrm{i}\int d^2x\, e^{ipx}\,\langle 0|\mathrm{T}\{\alpha(x),\,\alpha(0)\}|0\rangle = (2\pi)^2\,\delta^{(2)}(p)\,\langle\alpha\rangle^2 + D^{(\alpha)}(p). \tag{27}$$

The first term in eq. (27) with $\delta^{(2)}(p)$ corresponds to the factorized piece in the vacuum expectation value of the operator α^2. To leading order in $1/N$, the answer for $\langle\alpha\rangle$ is known,

$$\langle\alpha\rangle = \sqrt{N}\,m^2.$$

To next-to-leading order, there are $1/N$ corrections to $\langle\alpha\rangle$ but, at this moment, the factorized term $\langle\alpha\rangle^2$ is not interesting to us.

As for the non-factorized (connected) part in the vacuum expectation value of $\mathrm{T}\{\alpha(x)\,\alpha(0)\}$, it has an extra factor $1/N$ as compared with $\langle\alpha\rangle^2$. Thus, in order to calculate the non-factorized contribution $\langle\alpha^2\rangle^{\mathrm{n.f.}}$ in the next-to-leading order, it is sufficient to know $D^{(\alpha)}(p)$ to leading order:

$$D^{(\alpha)}(p)|_{\mathrm{Eucl.}} = -4\pi\sqrt{p^2(p^2+4m^2)}\Big/\ln\frac{\sqrt{p^2+4m^2}+\sqrt{p^2}}{\sqrt{p^2+4m^2}-\sqrt{p^2}}.$$

Moreover, it is convenient to use the substitution:

$$x = \left[\sqrt{1+\frac{p^2}{4m^2}} + \sqrt{\frac{p^2}{4m^2}}\right]^4 .$$

As a result,

$$\begin{aligned} \langle \alpha^2|_\mu\rangle^{\text{n.f.}} &= (-m^4) \int\limits_1^{A(\mu)} \frac{dx}{x^2}(x-1)^2 \frac{1}{\ln x}, \\ A(\mu) &= \left[\sqrt{1+\frac{\mu^2}{4m^2}} + \sqrt{\frac{\mu^2}{4m^2}}\right]^4 . \end{aligned} \tag{28}$$

The value of this integral can be written in terms of the special function $\mathrm{Ei}(x)$

$$\langle \alpha^2|_\mu\rangle^{\text{n.f.}} = (-m^4)\{\mathrm{Ei}[\ln A(\mu)] + \mathrm{Ei}[-\ln A(\mu)] - 2\ln\ln A(\mu) - 2\mathbb{C}\} \tag{29}$$

where $\mathbb{C} = 0.57721\ldots$ is the Euler constant.

To compare this exact formula with the operator product expansion we should consider large values of μ^2, namely $\mu^2 \gg m^2$. Indeed, only for such μ^2 the coefficients in OPE are calculable using a perturbation theory expansion with a small expansion parameter, $f(\mu)/4\pi = 1/\ln(\mu^2/m^2)$. Of course, all operators in this OPE should be normalized at the same point μ^2. In this limit eq. (29) takes the form

$$\langle \alpha^2|_\mu\rangle^{\text{n.f.}} = -2\mu^4[e^{-L}\,\mathrm{Ei}(L)] - \frac{4}{L}\mu^2 m^2 + 2m^4\left[\mathbb{C} + \ln L - \frac{1}{L} + \frac{4}{L^2}\right] + O\left(\frac{m^6}{\mu^2}\right), \tag{30}$$

where

$$L = \ln\frac{\mu^4}{m^4} \gg 1;\ e^{-L}\,\mathrm{Ei}[L] \underset{L\to\infty}{\approx} \sum_{n=0}^{\infty} \frac{n!}{L^{n+1}} .$$

To interpret various terms in eq. (30) let us compare it with the general OPE for the T-product of two renorm-invariant currents $j(x)$, $j(0)$:

$$\mathrm{i}\int d^2x\, e^{ipx}\, \mathrm{T}\{j(x)\,j(0)\} = C_0(p^2,\mu^2)\mathbb{1} + C_1(p^2,\mu^2)\,\alpha(0)|_\mu + C_2(p^2,\mu^2)\,(\alpha(0)\,\alpha(0))|_\mu + \cdots \tag{31}$$

where the coefficients C_i depend on μ^2 and p^2. As explained above, there is actually no dependence on μ^2 to leading order in $1/N$, neither in the coefficients C_i nor in the matrix elements of the operators. To next-to-leading order the non-trivial dependence on μ^2 in the matrix element $\langle \alpha^2|_\mu\rangle^{\text{n.f.}}$ (see eq. (30)) is to be compensated by a similar dependence on μ^2 appearing in other terms in the right-hand side of eq. (31). (The vacuum-to-vacuum matrix element of the operator equality (31) has no dependence on μ^2 if the current $j(x)$ is renorm-invariant.)

Let us check this cancellation in more detail. It is clear that the term $\sim\mu^4$ in $\langle \alpha^2|_\mu\rangle^{\text{n.f.}}$ should be compensated by appropriate terms in $C_0(p^2, \mu^2)\langle 0|\mathbb{1}|0\rangle$. We have not performed the calculation explicitly but it is quite evident that such a dependence on μ^2 does arise in $C_0(p^2, \mu^2)$ through perturbative diagrams since, only the region with internal line off-shellness $\geq \mu^2$, is accounted for in $C_0(p^2, \mu^2)$. Moreover, the term $m^2\mu^2$ in the vacuum expectation function $\langle \alpha^2 \rangle^{\text{n.f.}}$ must be compensated by similar terms in the vacuum expectation value of $C_1(p^2, \mu^2)\alpha|_\mu$. In this case, the μ^2 factor comes from the perturbative contribution to the coefficient $C_1(p^2, \mu^2)$ and the m^2 factor comes from the non-perturbative contribution to the matrix element $\langle 0|\alpha|_\mu|0\rangle$. Finally, there are a few terms $\sim m^4$.* It is quite clear that $m^4/\ln(\mu^4/m^4)$, $m^4/\ln^2(\mu^4/m^4)$, and $m^4 \ln \ln(\mu^4/m^4)$ represent the anomalous dimension of the operator $\alpha^2|_\mu$ to next-to-leading order in $1/N$. Indeed, if some operator $\mathcal{O}(\mu)$ has a non-vanishing anomalous dimension γ to next-to-leading order in $1/N$, then

$$\mathcal{O}(\mu) = C(\mu, \tilde{\mu})\,\mathcal{O}(\tilde{\mu})$$

where

$$\begin{aligned} C(\mu, \tilde{\mu}) &\equiv \exp \int_{f(\tilde{\mu})}^{f(\mu)} \mathrm{d}f \frac{\gamma(f)}{\beta(f)} = \exp \frac{1}{N} \int_{f(\tilde{\mu})}^{f(\mu)} \mathrm{d}f \left(\frac{a_{-1}}{f} + a_0 + a_1 f + \cdots\right) \\ &= (\mu \text{ independent factor}) \times \exp\left\{\frac{1}{N}\left(a_{-1} \ln f(\mu) + a_0 f(\mu) + \tfrac{1}{2} a_1 f^2(\mu) + \cdots\right)\right\} \\ &= (\mu \text{ independent factor}) \left\{1 + \frac{1}{N}\left(a_{-1} \ln f(\mu) + a_0 f(\mu) + \tfrac{1}{2} a_1 f^2(\mu) + \cdots\right)\right\}, \end{aligned}$$

$$f(\mu) = 4\pi/\ln(\mu^2/m^2)\,.$$

Comparing this general formula with eq. (30), we can identify the coefficients a_{-1}, a_0, a_1 associated with the anomalous dimension of α^2 to

$$a_{-1} = -2\,, \qquad a_0 = -1/4\pi\,, \qquad a_1 = 1/4\pi^2\,.$$

As usual, this logarithmic dependence on μ^2 in the matrix element is compensated by an analogous dependence present in the coefficient function $C_2(p^2, \mu^2)$.

Thus, at the next order in $1/N$, the coefficient functions in OPE depend on μ^2 in a rather non-trivial way (in particular, $C_i(p^2, \mu^2)$ contain terms $\sim \mu^2$, μ^4). However, in the vacuum matrix element of an operator equality it is possible to rearrange the μ-dependent terms in such a way that the total answer does not depend on μ at all.

Let us note that the $1/N$ perturbative corrections to the coefficient $C_0(p^2, \mu^2)$ contain an infinite number of terms $\sim (f/4\pi)^k$, in contrast to what happens in the leading $1/N$ approximation. This series diverges badly, a feature which is rather common to field theories (factorial divergency).

* The coefficients in front of the power terms (μ^4, μ^2) depend, generally speaking, on the procedure used for separating large and small momenta (subtraction procedure). We have used above a cut-off by virtue of a step function. Another cut-off would give a different answer. Moreover, for instance the dimensional regularization does not give such power terms at all (see below). One should choose a certain procedure and follow it in all the calculations of matrix elements and coefficient functions. The log μ terms are independent of regularization, at least in the leading log approximation.

Thus, we have convinced ourselves that to next order in $1/N$, a normalization point μ^2 has to be introduced and, therefore, that the matrix elements of composite operators receive contributions from both "true non-perturbative" fluctuations and from perturbative fluctuations. If this phenomenon had been present already to leading order, the study of non-perturbative effects using OPE would be much more difficult (but, even in this case, the formal validity of OPE is beyond any doubt).

A similar situation occurs in QCD. Namely, from a general point of view, the intermediate scale μ^2 is a necessary element of OPE but there exists a region of μ where, on the one hand, the QCD coupling constant $\alpha_s(\mu)$ is small enough and, on the other hand, the dependence on μ of the vacuum expectation value is numerically negligible. (Some estimates can be found in ref. [8].)

To conclude this section we would like to compare the results outlined above with those of David, who was the first to consider the matrix element of the operator α^2 beyond leading order [10, 11].

According to ref. [11] there is no unambiguous way to define vacuum expectation values of composite operators in the next-to-leading approximation and, therefore, the mathematical status of OPE beyond perturbation theory is unclear. Our general conclusion is just the opposite.

If one is willing to go into a more detailed comparison of ref. [11] and of the present approach, a few hints might be helpful. Ref. [11] exploits dimensional regularizations. For this reason there are no terms proportional to μ^4 or $m^2\mu^2$ in the matrix element of α^2 (compare with eq. (30)), only the terms proportional to m^4 are kept in ref. [11]. As we have explained above, the μ^2 dependence of the m^4 term is associated with the anomalous dimension of the operator α^2 in the next-to-leading $1/N$ approximation. Although in ref. [11] the normalization point μ is not introduced explicitly, it is clear that the $1/\epsilon$ term in the dimensional regularization procedure of ref. [11] plays the same role as $\ln(\mu^2/m^2)$ in our presentation. Then, the limit $\epsilon \to 0$ is equivalent to the limit $\ln(\mu^2/m^2) \to \infty$ and, as a result the terms $\sim m^4/\ln(\mu^2/m^2)$ and $m^4/\ln^2(\mu^2/m^2)$ are unseen in ref. [11]. The situation with leading effect due to the anomalous dimension, namely, $\sim \ln\ln(\mu^2/m^2)$, is more interesting. Indeed, it is easy to check that the factor $\ln\ln(\mu^2/m^2)$ corresponds to the $\ln\epsilon$ term in the dimensional-regularization (DR) scheme

$$\ln\ln\frac{\mu^2}{m^2} = \int^{\mu^2} \frac{dp^2}{p^2}\,\frac{1}{\ln(p^2/m^2)} \to \int^{\infty} \frac{dp^2}{p^2}\,(p^2)^{-\epsilon}\,\frac{1}{\ln(p^2/m^2)} \underset{\epsilon\to 0}{\simeq} m^{2\epsilon}[-\ln\epsilon - C + O(\epsilon)]\,.$$

Following the usual prescription of the DR scheme, David has tried to calculate the limit $\epsilon \to 0$ for $\langle\alpha^2\rangle|_\epsilon$ and has observed that this limit depends on the way ϵ tends to zero. This was the ground for the claims that the vacuum expectation value of α^2 cannot be defined unambiguously. It is true that $\ln\epsilon$ has a branch point at $\epsilon = 0$ and therefore that $\ln(\epsilon + i0) \neq \ln(\epsilon - i0)$. But the validity of OPE cannot suffer from this fact. Only renorm-invariant products $C_i(\mu)\langle \mathcal{O}_i(\mu)\rangle$ have a physical meaning and thus do not depend on the auxiliary parameter μ^2 or ϵ. The majority of composite operators $\mathcal{O}_i$ does depend on μ and this dependence cannot be eliminated by any subtraction procedure. If one still wishes to get rid of the $\log\log\mu$ (or $\log\epsilon$) terms, one may multiply the operator $\mathcal{O}_i$ by the inverse anomalous-dimension factor $\exp(-\int df\,\gamma(f)/\beta(f)) = 1 - (2/N)\ln\ln(\mu^4/m^4)$. Once this is done, the product $\exp(-\int df\,\gamma(f)/\beta(f))\,\mathcal{O}$ is invariant under variation of the normalization point and the limit $\mu \to \infty$ (or $\epsilon \to 0$) becomes meaningful and unambiguous. Thus, the puzzle of ref. [11] is actually an artifact of the procedure used in this paper.

It is worth noting that there are a few other recent papers (e.g. [12]) which claim difficulties for the OPE beyond perturbation theory. We observe that, in each case, the real cause of the trouble is actually one and the same, i.e. the non-introduction of the μ^2 scale. Once this is done, there are generally speaking, no problems with OPE.

A detailed comparative analysis of OPE in simple models and in QCD is given in ref. [25].

4. Anomalies in the trace of the energy-momentum tensor

In QCD the question of the influence of non-perturbative fluctuations on the vacuum energy density $\epsilon_{\rm vac}$ is of great interest. The reader may be acquainted with works [26] where an attempt has been made to investigate the formation of a bag starting directly from instantons which correspond to the only non-perturbative fluctuations known in QCD. It has been assumed [26] that $\epsilon_{\rm vac} < 0$ while, inside all hadrons, a phase transition takes place and the energy of the vacuum fluctuations vanishes (the energy is measured from its perturbative value; just this non-perturbative contribution is important for hadronic physics).

If so, the volume energy density inside hadrons is higher than in the "pure" vacuum by an amount $|\epsilon_{\rm vac}|$. In the bag model language, the volume energy density is described phenomenologically by the bag constant B. Accepting the hypothesis of ref. [26], one may conclude that meson masses m are proportional to $|\epsilon_{\rm vac}|^{1/4}$. On the other hand, the vacuum energy is proportional to the gluon condensate [8]. Indeed, Lorentz invariance implies

$$\langle 0|\theta_{\mu\nu}|0\rangle = \delta_{\mu\nu}\,\epsilon_{\rm vac}$$

where $\theta_{\mu\nu}$ stands for the energy-momentum tensor. Accounting for the conformal anomaly we arrive at

$$\epsilon_{\rm vac} = \frac{1}{4}\langle 0|\theta_{\mu\mu}|0\rangle = \langle 0| -\frac{b_0\alpha_s}{32\pi}\, G^{\alpha}_{\mu\nu}G^{\alpha}_{\mu\nu}|0\rangle \tag{32}$$

where b_0 is the first coefficient of the Gell-Mann–Low function, $b_0 = \frac{11}{3}N_c - \frac{2}{3}N_f$.

It turns out however that, substituting the numerical value of the gluon condensate extracted from the QCD sum rules, we get a number for the vacuum energy $|\epsilon_{\rm vac}|$ exceeding the generally accepted value of B by a factor of 10 to 20 [27, 15]. Moreover, eq. (32) shows that, in multicolour chromodynamics, the non-perturbative contribution to $|\epsilon_{\rm vac}|$ grows as N_c^2. On the other hand, meson masses should be essentially N_c independent.

These facts have led to the following conclusion: the vacuum energy and the difference between the energies outside and inside a hadron are distinct quantities and, moreover, $|\epsilon_{\rm vac}| \gg B$ [27, 15].

Of course, in the O(N) sigma model there is no need to introduce a bag and to calculate particle masses in terms of B. The masses are known directly. In the exactly solvable model we would like to check that the non-perturbative contribution can be distinguished against a background of divergences, that it is determined by the anomaly, etc. . . .

At first, we shall calculate the vacuum energy starting from the expression (6) for the effective action $S_{\rm eff}$. The latter is strongly divergent and calls for a regularization. The point-splitting regularization accepted in section 3.1 does not seem to be convenient in the case at hand and we shall switch here to the Pauli–Villars method. Then eq. (6) can be rewritten as:

$$\begin{gathered} S^{\rm R}_{\rm eff} = \frac{N}{2}\sum_{i=0}^{2} C_i \,{\rm Tr}\ln\left[-\partial^2 + m_i^2 + \frac{\alpha(x)}{\sqrt{N}}\right] - \frac{\sqrt{N}}{2f}\int {\rm d}^2x\,\alpha(x)\,; \\ \sum_{i=0}^{2} C_i = 0\,, \qquad \sum_{i=0}^{2} C_i m_i^2 = 0\,; \qquad C_0 = 1\,, \qquad m_0 = 0\,. \end{gathered} \tag{33}$$

The coefficients C_1 and C_2 are

$$C_1 = \frac{m_2^2}{m_1^2 - m_2^2}; \qquad C_2 = -\frac{m_1^2}{m_1^2 - m_2^2}.$$

The regulator masses $m_{1,2}$ are to be put to infinity at the very end.

The vacuum energy density turns out to be

$$\begin{aligned} V_2 \epsilon_{\text{vac}} = S_{\text{eff}} &= \frac{N}{2} \sum_{i=0}^{2} C_i \, \text{Tr} \ln(-\partial^2 + m_i^2 + m^2) - \frac{N}{2f} m^2 V_2 \\ &= \frac{N}{8\pi} \frac{m_1^2 m_2^2}{m_1^2 - m_2^2} \ln \frac{m_2^2 + m^2}{m_1^2 + m^2} \cdot V_2, \end{aligned} \tag{34}$$

where $V_2 = \int d^2x = L \cdot T$. We have used the following expression for the coupling constant

$$\frac{1}{f} = \frac{1}{4\pi} \left\{ \ln \frac{m_1^2 + m^2}{m^2} + \frac{m_1^2}{m_1^2 - m_2^2} \ln \frac{m_2^2 + m^2}{m_1^2 + m^2} \right\}$$

which is obtained from the extremality condition on the action (33), at the point $\alpha = \sqrt{N}\, m^2$.

The equation for the vacuum energy found above is, unfortunately, too complicated. To clarify its meaning let us rewrite eq. (34) for the special case of equal regulator masses $m_2^2 = M^2$, $m_1^2 = M^2 x$ and $x \to 1$. Then the vacuum energy can be written as a sum of two terms

$$\epsilon_{\text{vac}} = \frac{N}{8\pi} (-M^2 + m^2). \tag{35}$$

It is now quite clear that the first term is connected with perturbative fluctuations in the vacuum while the second one is due to non-perturbative fluctuations.

Indeed, the perturbative calculations correspond to a non-stable vacuum with $\alpha = 0$, and

$$\epsilon_{\text{vac}}^{\text{p.t.}} V_2 = S_{\text{eff}}^{\text{R}}|_{\alpha=0} = \frac{N}{8\pi} \frac{m_2^2 m_1^2}{m_1^2 - m_2^2} \ln \frac{m_2^2}{m_1^2} \xrightarrow[\substack{m_1^2 = M^2 x \\ m_2^2 = M^2 \\ x \to 1}]{} \frac{N}{8\pi} (-M^2).$$

Therefore the non-perturbative contribution to the vacuum energy reduces to

$$\epsilon_{\text{vac}} - \epsilon_{\text{vac}}^{\text{p.t.}} = \frac{N}{8\pi} m^2.$$

Now we would like to rederive this result using the conformal anomaly for the energy-momentum tensor $\theta_{\mu\nu}$. The regularized experssion for $\theta_{\mu\nu}^{\text{R}}$, corresponding to the action (33), is

$$\theta_{\mu\nu}^{\text{R}} = \sum_{i=0}^{2} C_i \left\{ -\partial_\mu \sigma_i^a \, \partial_\nu \sigma_i^a + \tfrac{1}{2} \delta_{\mu\nu} \left[\partial_\alpha \sigma_i^a \, \partial_\alpha \sigma_i^a + m_i^2 \sigma_i^a \sigma_i^a + \frac{\alpha}{\sqrt{N}} \sigma_i^a \sigma_i^a \right] \right\} - \tfrac{1}{2} \delta_{\mu\nu} \frac{\sqrt{N}}{f} \alpha,$$

where $\sigma^a_{1,2}$ stand for the regulator fields, $\sigma^a_0 = \sigma^a$. The trace of the regularized tensor $\theta^{\mathrm{R}}_{\mu\nu}$ is non-vanishing

$$\theta^{\mathrm{R}}_{\mu\mu} = \sum_{i=0}^{2} C_i \left\{ m_i^2 \sigma_i^a \sigma_i^a + \frac{\alpha}{\sqrt{N}} \sigma_i^a \sigma_i^a \right\} - \frac{\sqrt{N}}{f} \alpha .$$

First of all, let us check that the matrix element of $\theta^{\mathrm{R}}_{\mu\mu}$ over the true vacuum state reduces to ϵ_{vac}. Indeed

$$\langle 0|\theta^{\mathrm{R}}_{\mu\mu}|0\rangle = \sum_{i=0}^{2} C_i m_i^2 \langle 0|\sigma_i^a \sigma_i^a|0\rangle = \frac{1}{4\pi} \frac{m_1^2 m_2^2}{m_1^2 - m_2^2} \ln \frac{m_2^2 + m^2}{m_1^2 + m^2} = 2\epsilon_{\mathrm{vac}} ,$$

which is in full correspondence with the general expression

$$\epsilon_{\mathrm{vac}} = \tfrac{1}{2} \langle 0|\theta_{\mu\mu}|0\rangle .$$

We are now in a position to examine the hypothesis accepted in QCD, according to which the *non-perturbative* piece of ϵ_{vac} is equal to the *non-perturbative* contribution in the anomaly for $\theta_{\mu\mu}$.

Indeed, due to the "heavy" regulator fields the trace of the normalized energy-momentum tensor is non-zero. Moreover, we usually deal with matrix elements of $\theta^{\mathrm{R}}_{\mu\mu}$ over "light" states or (what is the same) with expectation values of $\theta^{\mathrm{R}}_{\mu\mu}$ in "light" external fields $\alpha^{\mathrm{ext}}(x)$ and $\sigma^{\mathrm{ext}}(x)$. In the case at hand there are only two non-vanishing terms of this type

$$\langle \theta^{\mathrm{R}}_{\mu\mu} \rangle = -\frac{N}{4\pi} M^2 + \frac{\sqrt{N}}{4\pi} \alpha^{\mathrm{ext}} + \mathrm{O}\left(\frac{1}{M^2}\right) .$$

It is instructive to rewrite this equality in the operator form

$$\theta^{\mathrm{R}}_{\mu\mu} \underset{M\to\infty}{=} C_{\mathrm{I}} \mathbb{1} + C_\alpha \alpha \; ; \qquad C_{\mathrm{I}} = -\frac{N}{4\pi} M^2 ; \qquad C_\alpha = \frac{\sqrt{N}}{4\pi} .$$

In the approximation considered, the coefficient functions are determined by perturbation theory and non-perturbative effects are expected to be hidden in the matrix elements of operators.

Therefore the coefficient C_{I} of the unit operator has to represent the whole perturbative contribution to the vacuum energy. We see that this is really the case,

$$C_{\mathrm{I}} = -\frac{N}{4\pi} M^2 = 2\epsilon^{\mathrm{p.t.}}_{\mathrm{vac}} .$$

Since we have already checked that $\frac{1}{2}\langle 0|\theta^{\mathrm{R}}_{\mu\mu}|0\rangle$ reduces to the total vacuum energy, the non-perturbative contribution to ϵ_{vac} is exhausted by the non-perturbative part of the anomaly. This can be seen explicitly

$$C_\alpha \langle 0|\alpha|0\rangle = \left(\frac{\sqrt{N}}{4\pi}\right) (\sqrt{N}\, m^2) = \frac{N}{4\pi} m^2 = 2\epsilon^{\mathrm{n.p.t.}}_{\mathrm{vac}} . \tag{36}$$

Thus, our hypothesis is confirmed.

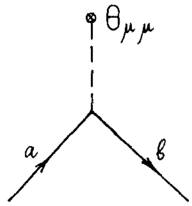


Fig. 6. The average value of $\theta_{\mu\mu}$ over the one-particle state.

Notice also that the ratio of the σ-particle masses to the non-perturbative part of $|\epsilon_{\rm vac}|$ is small,

$$m^2/|\epsilon_{\rm vac}|_{\rm n.p.t.} = \frac{8\pi}{N} \xrightarrow[N\to\infty]{} 0\,.$$

Moreover, we shall now check that, if there is no explicit mass parameter in the Lagrangian, the anomaly in $\theta_{\mu\mu}$ fixes the mass of the particle. It is worth recalling that, in QCD, the assertion does not seem to be quite trivial. It is difficult to imagine that, say, the whole nucleon mass is associated with gluons. However, in the chiral limit [14]

$$2m_N^2 = \langle N|\theta_{\mu\mu}|N\rangle = \langle N|-\frac{b_0\alpha_s}{8\pi}G^2|N\rangle\,.$$

Of course, this result is a strict consequence of the equations of motion. Nevertheless, it is in conflict with intuition and we would like to convince the reader of its validity by examining explicit examples which admit an exact solution.

The O(N) sigma model gives us such an opportunity. Let us calculate (with the help of the standard Feynman rules) the matrix element of the anomaly $\theta_{\mu\mu} = (\sqrt{N}/4\pi)\alpha$ between the states corresponding to σ particles. A single glance at fig. 6 is sufficient to grasp that

$$\langle\sigma^a|\theta_{\mu\mu}|\sigma^b\rangle = \langle\sigma^a|\frac{\sqrt{N}}{4\pi}\alpha|\sigma^b\rangle = \left(\frac{\sqrt{N}}{4\pi}\right)\left(-\frac{\delta^{ab}}{\sqrt{N}}\right)D^{(\alpha)}(0)$$

where $D^{(\alpha)}(0)$ is the α field propagator. The expression for $D^{(\alpha)}(p^2)$ is quoted in eq. (8), and

$$D^{(\alpha)}(0) = -8\pi m^2\,.$$

As a result,

$$\langle\sigma^a|\theta_{\mu\mu}|\sigma^b\rangle = \delta^{ab}\,2m^2\,. \tag{37}$$

Thus, the mechanical mass of the σ particles is indeed determined by the conformal anomaly.

5. Low-energy theorems

Low-energy theorems in field theory were invented almost as long ago as field theory itself. Suffice it to recall the Low theorems (e.g. [28]) for photon bremsstrahlung and for photon scattering off hadrons in the low-frequency limit.

As a rule, the low-energy theorems reduce to certain relations between amplitudes with different numbers of soft external lines (particles). Such relations appear as a reflection of some symmetry – exact or approximate – which exist in the theory. For instance, the Low theorems mentioned above stem from the gauge invariance of electromagnetic interactions.

Search for symmetries and for the constraints which they impose on various matrix elements is of special importance in QCD. Indeed, the corresponding results which are based on the general properties of the theory are independent of our ignorance of the confining dynamics and often yield unique pieces of information, mostly on glueball physics, which are inaccessible in any other way. They are also valuable in their purely theoretical aspect, serving as a spring-board for new constructions and investigations (see e.g. [15]).

Apart from the well-known classical PCAC theorems there exists a set of low-energy relations specific to QCD. As a matter of fact, these relations realize the Ward identities which reflect the following properties of the theory

(i) scale invariance at the classical level; $\theta_{\mu\mu} \neq 0$ only due to the quantum anomaly;

(ii) invariance under simultaneous γ_5 rotations of all quark fields, $q \to \exp(\mathrm{i}(\alpha/2)\gamma_5)q$, accompanied by a compensating transformation of the so-called θ term ($\theta \to \theta + \alpha$).

Both points have parallels in sigma models and we shall examine them in the remainder of this section.

5.1. Scale Ward identities

If the quark masses are switched off and there is no mass parameter in the Lagrangian, one can get the following elegant relation [29]

$$\lim_{q\to 0} \mathrm{i} \int \mathrm{d}x\, \mathrm{e}^{\mathrm{i}qx}\, \langle 0|\mathrm{T}\{\tau(x)\, \tau(0)\}|0\rangle_{\text{connected}} = -d_\mathrm{n}\langle 0|\tau|0\rangle \tag{38}$$

where $\tau(x)$ is the trace of the energy-momentum tensor, $\tau(x) = \theta_{\mu\mu} = (\beta(\alpha_\mathrm{s})/4\alpha_\mathrm{s})G^a_{\mu\nu}G^a_{\mu\nu}$, and d_n denotes the (normal) dimension of the operator τ (in chromodynamics $d_\mathrm{n} = 4$). A derivation of this formula, as well as its generalizations, are presented in refs. [29, 15, 30]. The perturbative contribution is assumed to be subtracted from both the right- and left-hand sides of eq. (38).

The scale Ward identity in the O(N) sigma model superficially looks just the same. The concrete form of $\tau(x)$ is, of course, different,

$$\tau = (\theta_{\mu\mu})_{\sigma\ \text{model}} = -\frac{f}{4\pi}(\partial_\mu\sigma)^2 = \frac{\sqrt{N}}{4\pi}\,\alpha(x)$$

and d_n is equal to 2, not 4. The proof is straightforward.

Consider the correlation function of two τ's, normalized in the following way

$$\Sigma(0) = \int \mathrm{d}^2x\, \langle 0|\mathrm{T}\{f(\partial_\mu\sigma(x))^2\,,\, f(\partial_\nu\sigma(0))^2\}|0\rangle \tag{39}$$

(the Euclidean notation is used here and below). First, we redefine the field variables, $\sigma = (1/\sqrt{f})\tilde{\sigma}$, $\alpha = \tilde{\alpha}$ (actually, return to the original definition, see eq. (2)), so that the action goes into

$$S_E = \frac{1}{2f}\int d^2x \left\{(\partial_\mu\tilde{\sigma})^2 + \frac{\tilde{\alpha}(x)}{\sqrt{N}}(\tilde{\sigma}^2 - N)\right\}.$$

Varying the generating functional Z_E with respect to $(-1/2f)$ we, evidently, get the vacuum expectation value $\langle(\partial_\mu\tilde{\sigma})^2\rangle$:

$$\frac{1}{V_2}\frac{\delta Z_E}{\delta(-1/2f)} = \langle f(\partial_\mu\sigma)^2\rangle = -Nm^2 = -NM_0^2\,e^{-4\pi/f}.$$

As discussed in detail in appendix A (see also section 3.1), under such a definition, the perturbative contribution in the vacuum expectation value vanishes provided that one takes into account the fact that the measure of the functional integration depends on the source ($1/f$ in the case at hand).

The second variation of Z_E yields $\Sigma(0)$:

$$\frac{1}{V_2}\frac{\delta^2 Z_E}{[\delta(-1/2f)]^2} = \int d^2x\,\langle 0|T\{(\partial_\mu\tilde{\sigma}(x))^2, (\partial_\nu\tilde{\sigma}(0))^2\}|0\rangle = \Sigma(0).$$

On the other hand, differentiating the explicit expression for m^2 over $(-1/2f)$ we arrive at

$$\frac{d}{d(-1/2f)}m^2 = 8\pi m^2.$$

Hence, $\Sigma(0) = -8\pi Nm^2$, and this completes the proof of the theorem.

It is interesting to notice that, in the O(N) sigma model, the perturbation theory contribution to $\Sigma(0)$ automatically vanishes after accounting for the f dependence of the measure of integration over the σ fields. This fact has been already used above.

We are now able to check in a direct way that the low-energy theorem thus obtained is indeed correct. Suffice it to notice that $\Sigma(q^2) \equiv ND^{(\alpha)}(q^2)$ and that the propagator $D^{(\alpha)}(q^2)$ is known in the leading N approximation. Taking the limit $q^2/m^2 \to 0$ in eq. (8) we get

$$D^{(\alpha)}(0) = -8\pi m^2 \qquad \text{and} \qquad \Sigma(0) = -8\pi Nm^2,$$

in perfect agreement with eq. (38).

5.2. The $U(1)$ problem in QCD and in $CP(N-1)$ models

The so-called topological charge

$$Q = \frac{g^2}{32\pi^2}G\tilde{G}$$

plays an important role in quantum chromodynamics. As for O(N) sigma models with $N \neq 3$, such a notion cannot be introduced (there is no non-trivial topology). Still we would like to have a simple toy example that would allow us to test various relations for topological charge derived in QCD. To this end we turn to a new class of sigma models, namely, to CP($N-1$) models.

First of all, a few words on their structure. These models describe N complex scalar fields σ^a ($a = 1, \dots, N$), "living" on a sphere

$$\bar{\sigma}^a \sigma^a = N/2f$$

(the bar stands for complex conjugation), and an auxiliary real field A_μ. The action is usually chosen in the following way [5, 6]:

$$S_{\mathrm{E}} = \int \mathrm{d}^2 x \left\{ \overline{\mathscr{D}_\mu \sigma^a} \, \mathscr{D}_\mu \sigma^a + \frac{\alpha(x)}{\sqrt{N}} \left(\bar{\sigma}^a \sigma^a - \frac{N}{2f} \right) \right\}, \qquad \mathscr{D}_\mu = \partial_\mu + \frac{\mathrm{i}}{\sqrt{N}} A_\mu . \tag{40}$$

This expression is obviously invariant under global SU(N) transformations of the σ^a fields. There is an extra – no less evident – symmetry of S_{E}, namely, a local U(1) symmetry. Really, the action of the CP($N-1$) model differs from that of scalar electrodynamics only by the absence of the kinetic term for the vector field A_μ. The latter is gauge invariant by itself.

Since the action (40) contains no derivatives of A_μ it can be fully eliminated, writing:

$$A_\mu = \mathrm{i} \frac{f}{\sqrt{N}} (\bar{\sigma}^a \, \vec{\partial}_\mu \sigma^a - \bar{\sigma}^a \, \overleftarrow{\partial}_\mu \sigma^a) \equiv \mathrm{i} \frac{f}{\sqrt{N}} \bar{\sigma}^a \, \overleftrightarrow{\partial}_\mu \sigma^a .$$

This relation is a consequence of the equations of motion.

For any values of N, CP($N-1$) models admit instanton solutions [5] (for a nice review see the book [31]). We shall return to this ussue in section 6, and all that we need here is the integral determining the topological charge,

$$Q = \frac{1}{\underline{2\pi\sqrt{N}}} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \, \partial_\mu A_\nu \equiv \frac{\mathrm{i} f}{\pi N} \int \mathrm{d}^2 x \, \epsilon_{\mu\nu} \, \overline{\mathscr{D}_\mu \sigma^a} \, \mathscr{D}_\nu \sigma^a .$$

It is rather obvious that Q is fixed by the asymptotic behaviour of the field. For configurations with finite action, Q necessarily takes integer values labelling distinct topological classes of fields.

The CP($N-1$) sigma models can be solved in the limit $N \to \infty$ just in the same vein as it was done with the O(N) models [6, 7]. Leaving aside the corresponding derivation we formulate the Feynman rules for excitations over the physical vacuum.

The propagation functions of the σ and α fields, as well as the vertex function Γ^{ab}, describing their interaction, do not differ from the case of the O(N) models and have the form

$$D^{ab}(p^2) = \langle \bar{\sigma}^a(p) \, \sigma^b(-p) \rangle = \frac{\delta^{ab}}{p^2 + m^2}, \qquad D^{(\alpha)}(p^2) = \langle \alpha(p) \, \alpha(-p) \rangle \equiv -A^{-1}(p), \qquad \Gamma^{ab} = -\frac{1}{\sqrt{N}} \delta^{ab} . \tag{41}$$

The auxiliary vector field A_μ becomes a dynamical field (the effective Lagrangian acquires the corresponding kinetic term). Its propagation function in the Lorentz gauge is

$$D_{\mu\nu}(p) = [\delta_{\mu\nu} - p_\mu p_\nu/p^2]\, D_A(p^2)\,,$$
$$D_A^{-1} = (p^2 + 4m^2)\, A(p) - 1/\pi\,. \tag{42}$$

One can easily convince oneself that $D_A(p^2)$ has a pole at $p^2 = 0$. Thus, A_μ indeed becomes a gauge field, a "photon". It interacts with σ particles according to the standard rules provided that one ascribes to the σ particles the "charge" $-1/\sqrt{N}$.

In the CP($N-1$) models, just as in the O(N) models, there is a condensate of the α field. In other words, the matrix elements of α over the physical vacuum is non-vanishing,

$$\langle 0|\alpha(x)|0\rangle = \langle 0| - \frac{2f}{\sqrt{N}}\,\overline{(\partial_\mu\sigma^a)}\,(\partial_\mu\sigma^a)|0\rangle = \sqrt{N}\, m^2\,.$$

We shall not repeat, however, the analysis of section 3, nor shall we discuss the consequences of this analysis. Instead, we shall dwell on the topological charge.

At first, let us recall why the operator of topological charge density seems to be so important in QCD. Introducing the source

$$j_\mathrm{P} = \frac{3\alpha_\mathrm{s}}{4\pi}\, G^a_{\mu\nu}\tilde{G}^a_{\mu\nu}\,,$$

and considering the two-point function

$$P(q^2) = \mathrm{i}\int \mathrm{d}^4x\; \mathrm{e}^{\mathrm{i}qx}\, \langle 0|\mathrm{T}\{j_\mathrm{P}(x)\, j_\mathrm{P}(0)\}|0\rangle\,, \tag{43}$$

we convince ourselves that the value $P(q^2=0)$ is, generally speaking, non-vanishing in pure gluodynamics. Following Witten [32] we can write that it is proportional to the second derivative of the vacuum energy density, over the vacuum angle θ. If we switch on light quarks two new effects will emerge. First of all, the correlation function $P(q^2)$ will vanish at $q^2 = 0$ because j_P becomes the full derivative of a gauge invariant operator,

$$j_\mathrm{P} = \partial_\mu j^5_\mu\,, \qquad j^5_\mu = \sum_{q=u,d,s} \bar{q}\gamma_\mu\gamma_5 q\,.$$

Second, there appears a light pseudoscalar state in the spectrum, the η' meson, which gives a contribution in the correlation function $P(q^2)$. Explaining the first effect by the presence of the η' meson we arrive at the relation

$$f^2_{\eta'} m^2_{\eta'} = -P(0)|_{\text{pure gluodynamics}}\,,$$
$$\langle 0|j^5_\mu|\eta'\rangle = \mathrm{i} f_{\eta'} p_\mu\,. \tag{44}$$

Thus, the solution of the famous U(1) problem [16] is directly associated with the value of $P(q^2=0)$ in pure gluodynamics.

In ref. [15] we have assumed that the both two-point functions, (44) and its scalar counterpart,

$$\mathrm{i}\int \mathrm{d}x\, \mathrm{e}^{\mathrm{i}qx}\, \langle 0|\mathrm{T}\left\{\frac{3\alpha_s}{4\pi}G^2(x), \frac{3\alpha_s}{4\pi}G^2(0)\right\}|0\rangle$$

are in pure gluodynamics basically determined for low q^2 by field configurations with definite duality. If so,

$$-P(0) \simeq \mathrm{i}\int \mathrm{d}^4x\, \langle 0|\mathrm{T}\left\{\frac{3\alpha_s}{4\pi}G^2(x), \frac{3\alpha_s}{4\pi}G^2(0)\right\} = \frac{18}{b_0}\left\langle\frac{\alpha_s}{\pi}G^2\right\rangle \tag{45}$$

where we have used the low-energy theorem (38). As a result, the following mass formula emerges [15]

$$f_{\eta'}^2 m_{\eta'}^2 \simeq \frac{18}{b_0}\langle 0|\frac{\alpha_s}{\pi}G^\alpha_{\mu\nu}G^a_{\mu\nu}|0\rangle\,.$$

Empirically it is quite successfull, at least within the numerical uncertainties existing at present.

Within the framework of the CP($N-1$) model, one can calculate $P(q^2=0)$ in the absence of quarks and then introduce fermions and find massless pseudoscalars. Finally, one can convince oneself that a SU(N)-singlet meson acquires a mass in agreement with eq. (44). This work has been carried out in ref. [7]. Here we shall concentrate on a relation analogous to (45). The questions to be answered are: "Is there any parallel with QCD in this point?" and "If there is, how far does it extend?"

The correct relative normalization of the scalar and of the pseudoscalar currents in CP($N-1$) models is fixed by the self-duality equation

$$\mathcal{D}_\mu\sigma^a = \pm\mathrm{i}\epsilon_{\mu\nu}\mathcal{D}_\nu\sigma^a\,.$$

Hence, we may choose, for instance,

$$j_s = \overline{\mathcal{D}_\mu\sigma^a}\,\mathcal{D}_\mu\sigma^a\,,$$
$$j_p = \pm\mathrm{i}\epsilon_{\mu\nu}\overline{\mathcal{D}_\mu\sigma^a}\,\mathcal{D}_\nu\sigma^a\,.$$

These combinations are related with the α and A_μ fields by virtue of the equations of motion

$$j_s(x) = \tfrac{1}{2}\sqrt{N}\,\alpha(x)\,,$$

$$j_p(x) = \tfrac{1}{2}\sqrt{N}\,\epsilon_{\mu\nu}\,\partial_\mu A_\nu(x)\,.$$

It follows that the two-point functions induced by j_s and j_p reduce to propagation functions of the corresponding fields. In Euclidean space-time we have

$$S(q^2) = \int \mathrm{d}^2x\, \mathrm{e}^{-\mathrm{i}qx}\,\langle \mathrm{T}\{j_s(x)\,j_s(0)\}\rangle = \tfrac{1}{4}N\,D^{(\alpha)}(q^2)\,,$$

$$P(q^2)=\int \mathrm{d}^2x\, \mathrm{e}^{-\mathrm{i}qx}\, \langle \mathrm{T}\{j_\mathrm{p}(x)\, j_\mathrm{p}(0)\}\rangle = \tfrac{1}{4}N(q^2\,\delta_{\mu\nu}-q_\mu q_\nu)\, D^{(A)}_{\mu\nu} = \tfrac{1}{4}Nq^2\, D^{(A)}(q^2)\,.$$

Although $D^{(A)}_{\mu\nu}(q)$ is a gauge-dependent quantity, the correlation function P is, of course, gauge invariant. Introducing explicit expressions for $D^{(\alpha)}$ and $D^{(A)}$ we get (at $q^2=0$):

$$\begin{aligned} &S(0) = -N\pi m^2\,, \\ &P(0) = 3N\pi m^2\,, \\ &S(0)/P(0) = -\tfrac{1}{3}\,. \end{aligned} \tag{46}$$

We see that $S(0)$ and $P(0)$ are rather close to each other, however, the sign of the ratio $S(0)/P(0)$ is opposite to the one found in the case of QCD. (Notice that relations (45), referring to QCD, are written in Minkowski space, while eq. (46) corresponds to Euclidean space. Going back to Minkowski space changes the sign of P.) This fact – the difference in signs – is rather striking but it could be foreseen from the very beginning. Indeed, there is a deep difference between QCD and the model considered. Namely, the non-perturbative contribution to the vacuum energy is negative in QCD and positive in sigma models:

$$[\epsilon_\mathrm{vac}]_\mathrm{QCD} = -\langle 0|\frac{b_0\alpha_\mathrm{s}}{32\pi}\, G^a_{\mu\nu}G^a_{\mu\nu}|0\rangle < 0\,,$$

$$[\epsilon_\mathrm{vac}]_{\sigma\ \mathrm{model}} = \frac{N}{8\pi}\, m^2 > 0\,.$$

Therefore, the values of the scalar correlators at $q^2=0$ are also different:

$$S(0)_\mathrm{QCD} > 0\,, \qquad S(0)_{\mathrm{CP}(N-1)} < 0\,.$$

On the other hand, the value of $P(0)_{(\text{no fermions})}$ is connected with the contribution of a real physical particle and, hence, is unambiguously fixed in both cases.

Summarizing, our hypothesis on the dominance of definite-duality fluctuations is literally invalid in the CP($N-1$) models. On the other hand we see that the ratio $S(0)/P(0)$ is still of order unity. Therefore, the prediction of ref. [15] for $P(0)_\mathrm{QCD}$ may be considered to be a good estimate.

6. Supersymmetric sigma models

Supersymmetric (SUSY) theories [33] (for an exhaustive review see [34] or the books [35]) have attracted much attention in recent years. While their ultimate relevance to Nature has not been established yet the great impact which they have produced on theoretical physics is indisputable. The aspect which has been investigated most thoroughly, and which exhibited some novel and striking features, is the SUSY perturbative expansion. Indeed the very first applications of SUSY models revealed remarkable cancellations among various graphs. This line of development culminated in the derivation of non-renormalization theorems which, in some cases, prove the vanishing of all radiative corrections. The well-known example of this kind is the non-renormalization of the so called F terms. The mere cancellation of all radiative corrections in certain mass terms is at the origin of the hope that SUSY models will eventually resolve the mass hierarchy problem of weak interactions.

While perturbation theory exhausts the content of weak coupling models, SUSY theories with strong interactions pose further problems. In view of the simplicity of the perturbative series all the non-trivial dynamics becomes in such theories the realm of non-perturbative physics. In particular, non-perturbative effects are a central issue for supersymmetric composite models of elementary particles.

At first sight, and basing ourselves on experience gained with ordinary QCD, there is little hope for a substantial progress in the field of non-perturbative effects. Indeed, a decade of intense studies in QCD has actually left us with instantons as the single example of non-perturbative fluctuations fully understood theoretically. It follows that to manage non-perturbative effects in QCD, one has at present to rely heavily on instanton based models of the physical vacuum. While there exist a few well-educated and advanced attempts of this kind, decisive arguments showing why instantons should be the ultimate answer for the vacuum wave function are still lacking.

Quite unexpectedly, a recent development indicates that instantons can be used in a very different fashion in supersymmetric theories. Thus, the very existence of instantons as a mathematical construction (i.e. the existence of instantons of arbitrary small size) is sufficient for the calculation of the Gell-Mann–Low function to all orders in the coupling constant. Moreover, one can determine some vacuum condensates. Loosely speaking, the trick is to combine instanton calculus with some general properties of the theory such as renormalizability or the existence of Ward identities.

In more detail, in the case of the β function, one addresses oneself to the consideration of the vacuum energy in the presence of an instanton of arbitrary small size. One can then prove a general theorem stating that radiative corrections to the vacuum energy cancel in the presence of a self-dual external field (instanton). The theorem shows a close resemblance with the well-known theorem on the vanishing of all radiative corrections to the vacuum energy in SUSY perturbation theory. However, the general rule has an exception. Namely, the zero modes should be considered separately and they do contribute to the (differential) vacuum energy. Thus, one finds an exact and non-vanishing result for some physical quantity. It is quite straightforward then to convert this calculation into the calculation of the β function.

In the case of vacuum condensates, one starts with an amplitude which is dominated by instantons at short distances. The Ward identities allow then to extend the result obtained to the case of any distance, and this fixes the vacuum condensate.

The purpose of this section is to present again these ideas and to elaborate upon them using the simpler example offered by a two-dimensional SUSY σ model. In the first part of the review we have had already the chance to argue that σ models look very similar to 4-dimensional gauge theories. This section provides a fresh support to this statement. The material is organized in the following form. First we investigate classical solutions, or instantons, in the O(3) sigma model [4] and introduce its supersymmetric extension [17, 18] to show how one can calculate the exact Gell-Mann–Low function (to all orders in the coupling constant) in supersymmetric models [19, 64, 53]. Our main purpose is to present as simple a survey as possible for a method proposed in ref. [19] in the context of supersymmetric gluodynamics, and to prove the validity of assertions which seem far less obvious in gluodynamics. We make no apologies for concentrating on this concrete aspect of supersymmetry and instanton calculus – other aspects are discussed in detail elsewhere.

The literature on supersymmetry is very rich. Even limiting ourselves to the case of two-dimensional

SUSY models we are still left with a very large number of papers. We quote only papers directly related to the content of the present review, and our list of references thus does not pretend to be complete in any way.

6.1. Supersymmetric generalization of the sigma model

We skip the description of the ordinary O(3) sigma model in the hope that the reader has already a general idea from section 2. The specific properties of this model are reviewed in the book [31].

The most straightforward procedure for constructing the supersymmetric variant seems to be the following [17, 18]. Instead of the real field, one introduces a real superfield N^a

$$N^a(x, \theta) = \sigma^a(x) + \bar{\theta}\psi^a(x) + \tfrac{1}{2}\bar{\theta}\theta F^a(x) \tag{47}$$

where θ is a two-component Majorana (real) spinor, ψ^a is a fermion field and F^a is an auxilary boson field. Unlike in the four-dimensional theory, the Majorana representation in two dimensions is realizable both in Minkowski and Euclidean spaces. Therefore, we can write down the Lagrangian in terms of N^a for both versions of the theory. For definiteness let us consider first the pseudo-Euclidean version [17] (for a treatment in Euclidean space see ref. [18]). The gamma matrices are chosen as

$$\gamma^0 = \sigma_2\,, \qquad \gamma^1 = \mathrm{i}\sigma_1\,; \tag{48}$$

they are purely imaginary and satisfy the standard algebra of gamma matrices. In eq. (47) $\bar{\theta} = \theta\gamma^0$. The action of the supersymmetric model is a direct generalization of the action $\int (\partial_\mu\sigma^a)^2\, \mathrm{d}^2x$:

$$S = \frac{1}{2g^2}\int \mathrm{d}^2x\, \tfrac{1}{2}\mathrm{d}^2\theta\, \epsilon^{\alpha\beta}\, \mathcal{D}_\alpha N^a \mathcal{D}_\beta N^a \tag{49}$$

where $\mathcal{D}_\alpha$ is the so-called supercovariant derivative,

$$\mathcal{D}_\alpha = \partial/\partial\bar{\theta}_\alpha - \mathrm{i}(\gamma^\mu\theta)_\alpha\partial_\mu \tag{50}$$

and the coupling constant is denoted by g^2 in order to distinguish it from the constant f of the non-supersymmetric variant (see section 2).

The constraint $\sigma^a\sigma^a = 1$ now becomes

$$N^a(x, \theta)\, N^a(x, \theta) = 1\,. \tag{51}$$

The reader will readily convince himself/herself that, in the component form, the supersymmetric model reduces to

$$S = \frac{1}{2g^2}\int \mathrm{d}^2x\, [(\partial_\mu\sigma^a)^2 + \bar{\psi}^a\, \mathrm{i}\gamma^\mu\, \partial_\mu\psi^a + F^2]\,. \tag{52}$$

Moreover, expansion of eq. (51) in θ yields three constraints

$$\sigma^2 = 1\,, \qquad \sigma\psi = 0\,, \qquad \sigma F = \tfrac{1}{2}\bar{\psi}\psi\,. \tag{53}$$

As expected, derivatives from the F field are absent and this implies in turn that the F field can be eliminated altogether by virtue of the equations of motion.

Derivation of these equations in the case at hand calls for special care since we must take the constraints into account. The simplest way to do this is to introduce corresponding Lagrange multipliers. Leaving this problem as an exercise we present here only the final answer

$$\begin{cases} (-\delta^{ab} + \sigma^a \sigma^b)\, \partial^2 \sigma^b - \sigma^b \bar{\psi}^a\, \mathrm{i}\gamma^\mu\, \partial_\mu \psi^b = 0\,, \\ (\delta^{ab} - \sigma^a \sigma^b)\, \mathrm{i}\gamma^\mu\, \partial_\mu \psi^b + \tfrac{1}{2}(\bar{\psi}\psi)\, \psi^a = 0\,, \\ F^a = \tfrac{1}{2}(\bar{\psi}\psi)\, \sigma^a\,. \end{cases} \tag{54}$$

In terms of the physical fields the action becomes

$$S = \frac{1}{g^2} \int \mathrm{d}^2 x\, [\tfrac{1}{2}(\partial_\mu \sigma^a)^2 + \tfrac{1}{2}\bar{\psi}^a\, \mathrm{i}\not{\partial}\psi^a + \tfrac{1}{8}(\bar{\psi}\psi)^2]\,. \tag{55}$$

We pause here to make a few comments about the explicit and "hidden" symmetries of the action (55). First of all, the global O(3) symmetry is built in. There are three generators corresponding to rotations in isospace around the first, second and the third axes. Moreover, the use of the superfield formalism guarantees supersymmetry, i.e. the invariance of the action with respect to the transformations

$$\delta\sigma^a = \bar{\epsilon}\psi^a\,, \qquad \delta\psi^a = -\mathrm{i}\gamma^\mu \epsilon\, \partial_\mu \sigma^a\,. \tag{56}$$

Indeed,

$$\delta S = \frac{1}{g^2} \int \mathrm{d}^2 x \left\{ -\partial_\mu ((\partial_\mu \sigma^a)(\bar{\epsilon}\psi^a)) - \frac{\mathrm{i}}{2}(\bar{\psi}\psi)(\bar{\psi}^a \gamma_\mu \epsilon)\, \partial_\mu \sigma^a \right\}$$

and the second term reduces to a full derivative because of the constraint $\psi^a \sigma^a = 0$.

The corresponding conserved supercurrent is

$$S^\mu = (\partial_\lambda \sigma^a)\, \gamma^\lambda \gamma^\mu \psi^a\,. \tag{57}$$

It is instructive to check its conservation explicitly

$$\begin{aligned} \partial_\mu S^\mu &= (\partial_\lambda \sigma^a)\, \gamma^\lambda\, \overrightarrow{\not{\partial}}\psi^a + \partial_\lambda \sigma^a\, \gamma^\lambda\, \overleftarrow{\not{\partial}}\psi^a \\ &= \frac{\mathrm{i}}{2}(\partial_\lambda \sigma^a)\, \gamma^\lambda\, \psi^a\, (\bar{\psi}\psi) - \sigma^b\, (\bar{\psi}^a\, \mathrm{i}\not{\partial}\psi^b)\, \psi^a \\ &= -\frac{\mathrm{i}}{2}(\sigma^a \not{\partial}\psi^a)(\bar{\psi}\psi) + \frac{\mathrm{i}}{2}(\sigma^b \not{\partial}\psi^b)(\bar{\psi}\psi) = 0\,. \end{aligned} \tag{58}$$

Here we have used eqs. (53), (54) and the fact that $(\bar{\psi}^a \gamma^\mu \psi^b)\, \psi^a = -\tfrac{1}{2}(\bar{\psi}^a \psi^a)\, \gamma^\mu \psi^b$.

The supercurrent S^μ exists for an arbitrary group O(N), and not only for O(3). What is more surprising, the O(3) model (and only this one) possesses an extended supersymmetry algebra. Namely, in this case one can find an extra conserved supercurrent [17]

$$\tilde{S}^\mu = \epsilon^{abc}\sigma^a \, \partial_\nu \sigma^b \, \gamma^\nu \, \gamma^\mu \, \psi^c \, . \tag{59}$$

The general reason for the appearance of an $n = 2$ superalgebra in the O(3) model was elucidated by Zumino [36].

Below, the extended superalgebra will be realized explicitly in terms of complex chiral superfields.

In addition, there is a conserved vector current

$$J^\mu = -\frac{\mathrm{i}}{2g^2}\, \epsilon^{abc}\, \sigma^a \, \bar{\psi}^b \, \gamma^\mu \, \psi^c \tag{60}$$

and its chiral partner, axial-vector current

$$J^{\mu 5} = -\frac{\mathrm{i}}{2g^2}\, \epsilon^{abc}\, \sigma^a \, \bar{\psi}^b \, \gamma^\mu \, \gamma^5 \, \psi^c \, . \tag{61}$$

Both involve the ϵ symbol and hence have no analogues for $N > 3$. Notice that, at this point, the parallel with supersymmetric gluodynamics is not absolute since, in the latter theory, there is no colour-singlet conserved vector current.

The action is invariant under transformations generated by the currents (57), (59), (60), the corresponding symmetries being exact and unbroken, even at the quantum level. As for the chiral symmetry, it is valid only at the classical level and it is destroyed by a quantum anomaly, just in the same way as in QCD.

Let us discuss this issue in more detail since its treatment in the literature is rather fragmentary and since, on the other hand, the derivation of the anomaly is not as trivial as it might seem at first sight.

(i) Naive conservation.

Differentiating $J^{\mu 5}$ gives

$$\begin{aligned} \partial_\mu J^{\mu 5} &= -\frac{\mathrm{i}}{2g^2}\, \epsilon^{abc} \{ (\partial_\mu \sigma^a)\, \bar{\psi}^b \gamma^\mu \gamma^5 \psi^c + \sigma^a \, (\bar{\psi}^b \overleftarrow{\not{\partial}})\, \gamma^5 \psi^c - \sigma^a \bar{\psi}^b \gamma^5 \, \not{\partial} \psi^c \} \\ &= -\mathrm{i}\, \frac{\epsilon^{abc}}{2g^2} \{ \partial_\mu \sigma^a \bar{\psi}^b \gamma^\mu \gamma^5 \psi^c - \mathrm{i} \sigma^a \bar{\psi}^b \gamma^5 \psi^c (\bar{\psi}\psi) \} \, . \end{aligned} \tag{62}$$

Moreover, since $\sigma^2 = 1$ the vectors $\partial_\mu \sigma^\alpha$ and σ^a are orthogonal to each other in isospace; we have

$$\partial_\mu \sigma^a (x) = \epsilon^{ade} \sigma^d (x) \, \nu^e (x)$$

where $\nu^e(x)$ is an auxiliary isovector. Substituting this relation in the right-hand side of eq. (62), and recalling that $\sigma\psi = 0$, we find that the first term vanishes.

The second term is identically zero because of the properties of the Grassmann numbers. Indeed, let us fix the axes in isospace in such a way that, at a given point x, the vector $\boldsymbol{\sigma}(x)$ reduces to $(0, 0, \sigma^3)$.

Then the constraint $\sigma^a\psi^a = 0$ implies $\psi^3(x) = 0$ and we are left with

$$\bar{\psi}^1\gamma^5\psi^2(\bar{\psi}^1\psi^1 + \bar{\psi}^2\psi^2).$$

Since we are dealing with real two-component spinors

$$\psi^1_\alpha\psi^1_\beta\psi^1_\gamma \equiv 0, \qquad \psi^2_\alpha\psi^2_\beta\psi^2_\gamma \equiv 0.$$

(ii) Axial anomaly.

As it is well-known in QCD, formal manipulations with the equations of motion may lead to a wrong conclusion since the theory must be first regularized and some of the symmetries of the classical action can then disappear. Such a situation is actually realized with the axial current. There exist various methods of regularization – however, the simplest one in the case at hand is the introduction of an auxiliary heavy fermion field R (with mass M_0)

$$J^{\mu 5}_{\mathrm{REG}} = -\frac{\mathrm{i}}{2g^2}\epsilon^{abc}\sigma^a\{\bar{\psi}^b\gamma^\mu\gamma^5\psi^c - \bar{R}^b\gamma^\mu\gamma^5R^c\}.$$

The divergence of the current is no more vanishing but, instead,

$$\partial_\mu J^{\mu 5}_{\mathrm{REG}} = \frac{\mathrm{i}}{2g^2}2\mathrm{i}M_0\epsilon^{abc}\sigma^a\bar{R}^b\gamma^5R^c. \tag{63}$$

The remainder of the derivation does not basically differ from that of the triangle anomaly in QCD (the anomaly is however now diangle, as discussed later). We start by inspecting the matrix elements of the operator $M_0\bar{R}\gamma^5R$ between all states containing only light fields, and we sort them out trying to identify a matrix element which does not vanish with $M_0 \to \infty$. If the search is successful the anomaly does exist. What particular matrix element is relevant? Simple dimensional arguments will prompt us the answer.

It is technically convenient to consider the propagation of the fermion R in an external (background) σ field. Then

$$\partial_\mu J^{\mu 5}(x) = -\frac{\mathrm{i}}{2g^2}2\mathrm{i}M_0\epsilon^{abc}\sigma^a(x)\,\mathrm{i}\,\mathrm{Tr}_{\mathrm{L}}[\gamma^5G^{cb}(x,x)] \tag{64}$$

where $G(x, y)$ is the R propagator in the external field defined by

$$\frac{1}{g^2}[(\mathrm{i}\partial_\mu\delta^{ab} + A^{ab}_\mu)\gamma^\mu - M_0\delta^{ab}]\,G^{bc}(x,y) = \delta^{ac}\,\delta(x-y) \tag{65}$$

$$A^{ab}_\mu = \mathrm{i}\,\sigma^a\,\partial_\mu\sigma^b \qquad \text{(cf. eq. (54))}$$

and Tr_{L} means taking the trace over the Lorentz (spinor) indices.

Now, the last effort. The Green function is determined by expanding $G(x, y)$ in a series in A^{ab}_μ. It is sufficient to keep only the term proportional to $(\partial_\alpha A^{ab}_\mu(x))(x-y)$. (Indeed, the term proportional to $A^{ab}_\mu(x)$ would give zero being contracted with the ϵ symbol in eq. (64); $A^{ab}_\mu(x)\,A^{bc}_\nu(x) \equiv 0$; all other terms of the expansion are of a higher dimensionality and therefore their contribution vanishes with $M_0 \to \infty$).

Thus, we are actually dealing with a diangle anomaly (fig. 7). Playing a bit with eq. (64) (for details see the review paper [37]) yields

$$\partial_\mu J^{\mu 5} = \frac{1}{2\pi} \epsilon^{\mu\rho} \epsilon^{abc} \sigma^a \partial_\mu \sigma^b \partial_\rho \sigma^c . \tag{66}$$

Instead of dealing with the original fields σ^a and ψ^a it is more convenient in many respects to work with unconstrained fields – and this is the last point to be discussed in this subsection. Since the σ fields "live" on a unit sphere there is a very natural way of introducing new independent variables, namely, stereographic projection [4] (fig. 8). We trade three constrained fields σ^1, σ^2, σ^3 ($\boldsymbol{\sigma}^2 = 1$) for two independent variables φ_1, φ_2, the corresponding transformation law being clear from fig. 8:

$$\sigma^1 = \frac{2\varphi_1}{\varphi_1^2 + \varphi_2^2 + 1}, \qquad \sigma^2 = \frac{2\varphi_2}{\varphi_1^2 + \varphi_2^2 + 1}, \qquad \sigma^3 = \frac{1 - \varphi_1^2 - \varphi_2^2}{\varphi_1^2 + \varphi_2^2 + 1}. \tag{67}$$

Then φ_1 and φ_2 are combined in a complex field φ,

$$\varphi = \operatorname{Re} \varphi + \mathrm{i} \operatorname{Im} \varphi \equiv \varphi_1 + \mathrm{i}\varphi_2 . \tag{68}$$

The fermionic analogue of eq. (67) is readily derived by virtue of supersymmetry,

$$\begin{aligned}
\psi^1 &= \frac{2 \operatorname{Re} \Psi}{1 + \varphi\varphi^*} - \frac{2 \operatorname{Re} \varphi \,[\varphi^* \Psi + \text{h.c.}]}{(1 + \varphi\varphi^*)^2}, \\
\psi^2 &= \frac{2 \operatorname{Im} \Psi}{1 + \varphi\varphi^*} - \frac{2 \operatorname{Im} \varphi \,[\varphi^* \Psi + \text{h.c.}]}{(1 + \varphi\varphi^*)^2}, \\
\psi^3 &= -2 \frac{[\varphi^* \Psi + \text{h.c.}]}{(1 + \varphi\varphi^*)^2},
\end{aligned} \tag{69}$$

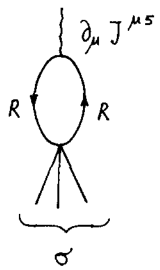


Fig. 7. Heavy-fermion loop determining the anomaly in the divergence of the axial-vector current $J^{\mu 5}$. (For a definition of $J^{\mu 5}$ see eqs. (61), (63).)

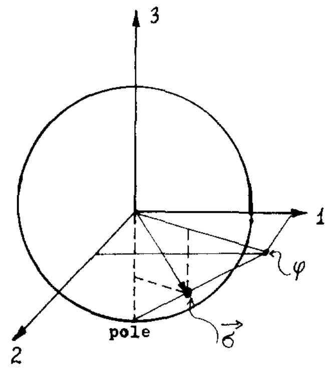


Fig. 8. The unit sphere $\boldsymbol{\sigma}^2 = 1$ is mapped into a plane (φ_1, φ_2) by means of a stereographic projection.

where Ψ is a complex two-component spinor field, which is the counterpartner of φ. Notice that the parametrization (67), (69) automatically yields $\sigma^a\sigma^a = 1$, $\sigma^a\psi^a = 0$. It is worth emphasizing once more that neither φ nor Ψ possess isotopic indices.

For completeness we give here also the formulae for the inverse transformation, which are however rather trivial

$$\varphi = \frac{\sigma^1 + i\sigma^2}{1+\sigma^3}, \qquad \Psi = \frac{\psi^1 + i\psi^2}{1+\sigma^3} - \frac{\sigma^1 + i\sigma^2}{\underline{(1+\sigma^3)^2}}\psi^3 . \tag{70}$$

The next step is to rewrite the Lagrangian (55) in terms of the new fields φ and Ψ. This is a simple but rather lengthy algebraic exercise resulting in the following Lagrangian [38]

$$\mathcal{L} = \frac{2}{g^2}\frac{1}{\chi^2}\Big\{\partial_\mu\varphi^+\,\partial_\mu\varphi + \frac{i}{2}(\bar\Psi\gamma^\mu\,\partial_\mu\Psi - (\partial_\mu\bar\Psi)\,\gamma^\mu\Psi)$$
$$-\frac{i}{\chi}\bar\Psi\gamma^\mu\Psi(\varphi^+\,\partial_\mu\varphi - \varphi\,\partial_\mu\varphi^+) - \frac{1}{2\chi^2}(\Psi^{\bar\alpha}\Psi^{\bar\beta}\epsilon_{\bar\alpha\bar\beta})^+(\Psi^\alpha\Psi^\beta\epsilon_{\alpha\beta})\Big\} \tag{71}$$

where $\chi \equiv 1 + \varphi^+\varphi$.

In deriving eq. (71) we have used eqs. (67), (69) and the explicit form of the gamma matrices (see eq. (48)). More exactly, we have exploited such general properties as $(\mathrm{Re}\,\Psi)\,(\gamma^0\gamma^\mu)\,\mathrm{Re}\,\Psi = 0$, etc.

Now, we would like to return to a fact first mentioned in ref. [17], namely that the superfield formalism used above ensures an $n = 1$ supersymmetry; there are, however, actually two conserved supercurrents (eqs. (57), (59)) and the O(3) model possesses an extended ($n = 2$) supersymmetry. In the original formulation of the model, eq. (49), the conservation of $\tilde S_\mu$ is implicit. After introducing the fields φ and Ψ we are able to realize the full supersymmetry explicitly [36].

First of all we construct a complex chiral superfield Φ

$$\Phi = \varphi(x_{\mathrm{ch}}) + \sqrt{2}\,\epsilon_{\alpha\beta}\theta^\alpha\,\Psi^\beta(x_{\mathrm{ch}}) + \epsilon_{\alpha\beta}\theta^\alpha\theta^\beta\,F(x_{\mathrm{ch}}) \tag{72}$$

where F is an auxiliary boson field, and all functions φ, Ψ, F depend on the chiral argument,

$$(x^\mu)_{\mathrm{ch}} = x^\mu + i\bar\theta\gamma^\mu\theta . \tag{73}$$

Notice that the parameter θ figuring in this expression is now a complex two-component spinor ($\bar\theta$ and θ are independent Grassmann variables), while the original superfield (47) depends on the Majorana spinor θ. Thus, we double the number of fermionic coordinates*. The action can be rewritten in terms of the superfield Φ [36]

$$S = \frac{1}{2g^2}\int d^2x\; d^2\theta\; d^2\bar\theta\,\ln(1 + \Phi^+\Phi) . \tag{74}$$

Given the answer, one can just check that, after integration over θ and $\bar\theta$ according to the standard

* One and the same letter, θ, denotes the real and complex spinor – we do not dare to invent a new notation for the fermionic coordinate. We hope that this will cause no confusion; the meaning of θ is always clear from the context.

rules of the Grassmann algebra, we come back (up to full derivatives) to the Lagrangian (71).

Actually, the existence of an $n = 2$ supersymmetry and its explicit realization (see eq. (74)) is not specific to O(3) sigma model. As it was noted by Zumino [36], the phenomenon is of a general nature. In the O(3) model the σ fields "live" on a sphere. Moreover, the two-dimensional sphere is a complex Kähler manifold (for a pedagogic discussion see [39, 40]). It turns out that for any Kähler manifold the corresponding sigma models possess an $n = 2$ supersymmetry [36]. Moreover, starting from general considerations, Zumino could obtain in this case an expression for the action which is analogous to eq. (74).

The price we have payed for having an explicit realization of the $n = 2$ supersymmetry is rather high. The O(3) symmetry transparent in eq. (49) is now partly hidden. Rotations of N^a around the third axis (in isospace) correspond to the following infinitesimal transformations of Φ

$$\Phi \to \Phi + \mathrm{i}\delta \cdot \Phi, \qquad \Phi^+ \to \Phi^+ - \mathrm{i}\delta \cdot \Phi^+$$

where δ is a real parameter. The action (74) is evidently invariant under such a phase rotation. As to the two other isotopic rotations of N^a (around the first and the second axes), the corresponding symmetry manifests itself in the invariance of the action (74) under the transformations

$$\Phi \to \varepsilon + \varepsilon^* \Phi^2, \qquad \Phi^+ \to \varepsilon^* + \varepsilon(\Phi^+)^2$$

with a complex parameter ε.

6.2. Instantons

The configurational space in which the σ fields are defined is topologically equivalent (after Wick rotation and compactification*) to a two-dimensional sphere. On the other hand, in the O(3) model the σ fields "live" on the same two-dimensional sphere, $(\sigma^1)^2 + (\sigma^2)^2 + (\sigma^3)^2 = 1$. Thus, there should exist topologically non-equivalent classes of field configurations corresponding to topologically distinct mappings of the two spheres on each other. It is also clear that for an O(N) group with $N > 3$ all mappings $S^2 \to S^{N-1}$ are topologically equivalent to the trivial one and, hence, instantons are absent.

The expression for the topological charge q in the Euclidean O(3) sigma model is adduced in the pioneering paper [4]. It reads:

$$q = \frac{1}{8\pi} \int \mathrm{d}^2 x\, \epsilon^{abc} \epsilon_{\mu\nu} \sigma^a\, \partial_\mu \sigma^b\, \partial_\nu \sigma^c\,. \tag{75}$$

One can easily convince oneself that the integral (75) indeed meets all necessary requirements. To illustrate this fact we shall show that, once the asymptotics of σ is fixed, variations of σ do not change q. Indeed, keeping the terms linear in $\delta\sigma$ we get

* The term "Wick rotation" means that we are dealing with Euclidean space. The term "compactification" roughly speaking means that all infinitely distant points are identified with a single point. In other words, we are considering only such configurations $\sigma^a(x)$ that have a unique limit as $|x| \to \infty$, irrespectively of the direction followed.

$$
\begin{aligned}
\delta q &= \frac{1}{8\pi} \int \mathrm{d}^2x\, \epsilon^{abc}\epsilon_{\mu\nu}\{(\delta\sigma^a)\,\partial_\mu\sigma^b\,\partial_\nu\sigma^c + 2\sigma^a\,\partial_\mu(\delta\sigma^b)\,\partial_\nu\sigma^c\} \\
&= \frac{1}{8\pi} \int \mathrm{d}^2x\, \epsilon^{abc}\epsilon_{\mu\nu}\{2\partial_\mu[\sigma^a\,(\delta\sigma^b)\,\partial_\nu\sigma^c] + 3(\delta\sigma^a)\,\partial_\mu\sigma^b\,\partial_\nu\sigma^c\}\,. \qquad (76)
\end{aligned}
$$

Here the first term is a full derivative and hence reduces to a surface integral. This integral in turn vanishes because the distortions of σ are localized, by assumption, $\delta\sigma \to 0$ if $|x| \to \infty$. The second term is equal to zero by itself. Indeed, the constraint $\sigma^a\sigma^a = 1$ implies that $\sigma^a\,\delta\sigma^a = 0$, or $\delta\sigma^a(x) = \epsilon^{ade}\,\sigma^d(x)\,\nu^e(x)$ where $\nu^e(x)$ is some isovector. Substituting this relation yields

$$
\epsilon^{abc}\epsilon_{\mu\nu}\epsilon^{ade}\sigma^d\nu^e\,\partial_\mu\sigma^b\,\partial_\nu\sigma^c = 2\epsilon_{\mu\nu}[\nu^c\,(\partial_\mu\boldsymbol{\sigma}^2)\,\partial_\nu\sigma^c] = 0\,.
$$

Thus, q is constant under continuous variations of σ, a topological characteristic that labels the classes of the fields. The normalization factor $(8\pi)^{-1}$ is chosen [4] in such a way that q takes only integer values 0, ±1, ±2, etc. The trivial (vacuum) field configuration $\boldsymbol{\sigma}(x) = \text{const.}$ evidently corresponds to $q = 0$.

If $q \geq 0$ it is convenient to rewrite the action in the following identical form

$$
S = \frac{1}{4g^2} \int \mathrm{d}^2x\, [\partial_\mu\sigma^a + \epsilon^{abc}\sigma^b\epsilon_{\mu\nu}\,\partial_\nu\sigma^c]^2 + \frac{1}{2g^2}\,q\,. \qquad (77)
$$

One immediately sees that local minima of the action are obtained when

$$
\partial_\mu\sigma^a = -\epsilon^{abc}\sigma^b\epsilon_{\mu\nu}\,\partial_\nu\sigma^c\,. \qquad (78)
$$

This is the duality equation in the O(3) sigma model. Its solutions, instantons, satisfy the classical equations of motion as well. Anti-instantons satisfy the same equation but with a plus sign in the right-hand side.

Due to O(3) invariance, the existence of a solution with asymptotics $\sigma^a(x)|_{|x|\to\infty} = \sigma^a_{\mathrm{ASY}}$ would imply that there exist rotated solutions with any asymptotics, compatible with the constraint $\sigma^a\sigma^a = 1$. If $\sigma^3 \to 1$, $\sigma^{1,2} \to 0$ ($|x| \to \infty$) we shall refer to these boundary conditions as to the standard ones.

The one-instanton solution satisfying the standard boundary conditions has the form [4]

$$
\begin{aligned}
&\sigma^1 = \frac{2((x-x_0)y)}{(x-x_0)^2 + y^2}\,, \qquad \sigma^2 = \frac{2[(x-x_0)\times y]}{(x-x_0)^2 + y^2}\,, \\
&\sigma^3 = \frac{(x-x_0)^2 - y^2}{(x-x_0)^2 + y^2}\,, \qquad ([a \times b] \equiv a_1b_2 - a_2b_1)\,,
\end{aligned} \qquad (79)
$$

where x_0 and y are arbitrary 2D vectors, x_0 plays the role of the instanton centre (two collective coordinates), $|y| \equiv \rho$ is the instanton size, and, finally, the direction of y fixes the instanton orientation with respect to the third axis in isotopic space. Thus, eq. (79) contains four collective coordinates. In principle, one can rotate this solution around the first and/or the second axes in isotopic space, producing other solutions with "non-standard" asymptotics. This would clearly bring in two new collective coordinates.

It is worth noticing that

$$(\partial_\mu \sigma^a(x))^2 = 8y^2/[(x-x_0)^2 + y^2]^2\,,$$

and the one-instanton action is

$$S_0 = \frac{1}{2g^2}\int \mathrm{d}^2x\,(\partial_\mu\sigma^a(x))^2_{\rm inst} = \frac{4\pi}{g^2}\,. \tag{80}$$

Moreover, the topological charge corresponding to eq. (32) is

$$q = \frac{1}{8\pi}\int \epsilon^{abc}\,(\sigma^a(x)\,\partial_\mu\sigma^b(x)\,\partial_\nu\sigma^c(x))_{\rm inst}\,\epsilon^{\mu\nu} = 1\,. \tag{81}$$

Combining the definition of the topological charge, eq. (75), with the axial anomaly, eq. (66), we see that (just like in QCD) the instanton necessarily generates fermionic zero modes. More exactly, in the one-instanton transition $\Delta Q_5 = 4$ (Q_5 stands for the axial charge). In other words, each instanton is accompanied by four fermion legs. This fact will be important for us later.

Let us now show how one can easily find the solution (79) and other multi-instanton solutions starting from the duality equation. In terms of the complex field φ eq. (78) reduces to [4]

$$\frac{\partial}{\partial\bar z}\,\varphi(z,\bar z) = 0 \tag{82}$$

where

$$z = x_1 + \mathrm{i}x_2\,, \qquad \bar z = x_1 - \mathrm{i}x_2\,.$$

Any function that depends only on z satisfies this equality. Moreover, the classical field evidently must have no essential singularities, and, hence, $\varphi(z)$ must be an analytic function of z. However, poles are admissible – a pole at some point $\tilde z$ simply means that $\sigma^3 \to -1$ at $x_1 \to \mathrm{Re}\,\tilde z$, $x_2 \to \mathrm{Im}\,\tilde z$.

The only analytic function with no poles is $\varphi(x) \equiv$ const., the trivial vacuum solution. Analytic functions with one pole correspond to a one-instanton solution, $q = 1$, with two poles to two-instanton solution, $q = 2$, etc.

For one instanton the most general parametrization is

$$\varphi_{\rm inst} = C + y/(z - z_0) \tag{83}$$

where C, y, z_0 are three independent complex constants (six real collective coordinates). Using eqs. (67) one can express $(\sigma^a)_{\rm inst}$ in terms of $\varphi_{\rm inst}$ and check that the standard boundary condition corresponds to $C = 0$. Moreover, z_0 is just the position of the instanton centre, and y has the same meaning as in eq. (79); $|y|$ plays the role of the instanton radius ρ, while the phase of y is the angle of rotation around the third axis in isospace. Adding C we rotate the instanton around the first and/or the second axis.

Just like in QCD, differentiation with respect to the collective coordinates yields the zero modes, bosonic zero modes in the case at hand.* Two complex modes which will be of interest are

$$(\varphi_0)_1 \sim y/(z-z_0)^2\,, \qquad (\varphi_0)_2 \sim 1/(z-z_0)\,. \tag{84}$$

Their geometrical meaning is quite simple: we have four real modes, two of them are associated with translations, one with dilatations and the last one with rotations around the third axis in isotopic space (the latter do not change the standard asymptotics).

If the asymptotics of the pseudoparticle field were not fixed, there would arise two extra zero modes corresponding to rotations around the first and second axes. However, simultaneously the trivial amplitude corresponding to the zero topological charge would acquire the same two zero modes – the phenomenon having no parallel in QCD. Moreover, since the one-instanton amplitude is always normalized to perturbation theory, $\sigma^a(x) = \text{const.}$ (see below) the effect of the two extra zero modes would cancel out and we may disregard them from the very beginning.

Thus far our discussion of instantons was independent of fermions. As well-known from QCD the most drastic effect of massless fermions is the appearance of fermionic zero modes which totally suppress the instanton amplitude in the absence of external fermionic sources.

The model we are considering is no exception. Moreover, as was mentioned above, the number of the fermion zero modes is known beforehand, from the consideration of the axial anomaly. For instance, for the instanton we expect four (real) zero modes.

Let us dwell on this issue. First of all we must perform an analytical continuation of the Lagrangian (71) to Euclidean space. The procedure is quite trivial for the bosonic piece and is slightly less trivial for the fermionic one. A detailed discussion with all relevant definitions can be found in the review paper [42] and we give here directly the final answer

$$\mathscr{L}_{\mathrm{E}} = \frac{2}{g^2\chi^2}\left\{\partial_\mu\varphi^+\,\partial_\mu\varphi - \frac{\mathrm{i}}{2}(\Psi^+\gamma_\mu^{\mathrm{E}}\,\partial_\mu\Psi - (\partial_\mu\Psi^+)\,\gamma_\mu^{\mathrm{E}}\Psi) + \frac{\mathrm{i}}{\chi}\,\Psi^+\gamma_\mu^{\mathrm{E}}\Psi(\varphi^+\,\partial_\mu\varphi - \varphi\,\partial_\mu\varphi^+) + \frac{1}{\chi^2}(\Psi^+\Psi)^2\right\} \tag{85}$$

where, in Euclidean space, Ψ^+ and Ψ are to be treated as independent variables, $\chi = 1+\varphi^+\varphi$, and γ_1^{E}, γ_2^{E}, the Euclidean gamma matrices, satisfy the standard algebra $\frac{1}{2}\{\gamma_\mu^{\mathrm{E}}\gamma_\nu^{\mathrm{E}}\} = \delta_{\mu\nu}$. A convenient set of γ's is

$$\gamma_\mu^{\mathrm{E}} = \sigma_\mu\,, \qquad \mu = 1, 2\,.$$

Notice that since Ψ and Ψ^+ are independent complex variables in Euclidean space, and not constrained by the Majorana condition, there is no need for the gamma matrices to be real, as in eq. (48).

The fermion zero modes satisfy the classical equations of motion**

$$-\mathrm{i}\gamma_\mu\,\partial_\mu\Psi + \frac{2\mathrm{i}}{\chi}\,\gamma_\mu\Psi(\varphi^+\,\partial_\mu\varphi) + \frac{2}{\chi^2}(\Psi^+\Psi)\Psi = 0 \tag{86a}$$

where $\varphi = \varphi_{\mathrm{inst}}$ and we must impose an additional condition

* Instantons in QCD are reviewed in refs. [41, 42].
** The superscript E is hereafter omitted.

$$\Psi^+ \gamma_\mu \Psi = 0 \tag{86b}$$

ensuring the absence of the source term in the classical equation for φ.

In principle, one might try to solve eq. (86) explicitly, but fortunately supersymmetry does the job for us. The solution of the bosonic equation is already known, φ_{inst}. Applying to it a supersymmetry transformation we can generate solutions of the fermionic equation. This fact is general for all supersymmetric theories and was first mentioned in refs. [43, 44].

Let us consider the Euclidean version of the superfield Φ as defined in eq. (72), and put $\Psi = F = 0$, $\varphi = \varphi_{\text{inst}}$. Shifts of θ evidently lead to no changes of Φ because x_{ch} stays the same. On the contrary, under shifts of θ^+, $\theta^+ \to \theta^+ + \zeta^+$

$$(x_{\text{ch}})_\mu \to (x_{\text{ch}})_\mu + 2\mathrm{i}\zeta^+ \gamma_\mu \theta\,.$$

In this way we generate Ψ,

$$\begin{aligned}\Psi^{(1)}_{\text{ss}} &= -2\sqrt{2}\,\mathrm{i}\alpha^+\, y/(z-z_0)^2\,,\\ \Psi^{(2)}_{\text{ss}} &= 0\,, \qquad F = 0\end{aligned} \tag{87}$$

where (1, 2) are the spinor (Lorentz) indices, $\alpha^+ \equiv \zeta^{+(1)}$, and the meaning of the subscript ss (supersymmetry) will become clear in a moment. Now, the reader may easily check that the eqs. (86) are satisfied identically.

This is not the end of the story, however. Just like supersymmetric gluodynamics, the Lagrangian (85) contains no dimensional parameters and its scale invariance at the classical level is obvious. Combining this scale invariance with supersymmetry we arrive at the so-called superconformal group, which includes conformal transformations and their supersymmetric partners (for a discussion of superconformal symmetry in gluodynamics see e.g. ref. [34]).

The fermion sector of this group consists of the ordinary supersymmetry transformations with parameters ζ and ζ^+ – constant Grassmann spinors – plus the same transformations with x-dependent parameters,

$$\xi^{+(\alpha)} = \epsilon^{\alpha\beta}\, x_\mu (\gamma_\mu)_{\beta\delta}\, \zeta^{(\delta)}\,.$$

Both scale invariance and the superconformal invariance are broken at the quantum level by anomalies. They are valid symmetrices, however, if we are interested in the classical solutions. Performing the superconformal transformations we get another fermion solution*, introducing in this way an extra fermion collective coordinate β

$$\begin{aligned}\Psi^{(1)}_{\text{sc}} &= -2\sqrt{2}\,\mathrm{i}\beta\, y/(z-z_0)\,,\\ \Psi^{(2)}_{\text{sc}} &= 0\end{aligned} \tag{88}$$

where β is some Grassmann number.

* Let us notice that a simple examination of eq. (86) shows that any spinor of the type $\Psi^{(1)} = f(z)$, $\Psi^{(2)} = 0$, where $f(z)$ is an arbitrary function of z, satisfies this equation. One may ask: "why is preferance given to the solutions quoted and not to some others?" On reflection one will conclude that eqs. (87), (88) give the only ***normalizable*** solutions with the appropriate analytic behaviour ($f(z)$ must be meromorphic).

Thus, we have two complex fermion collective coordinates which are equivalent to four real zero modes, the number expected from the axial anomaly. It is worth emphasizing that the modes found possess definite chirality

$$\gamma^5 \Psi_{\mathrm{ss(sc)}} = \Psi_{\mathrm{ss(sc)}} . \tag{89}$$

Finally, we may write the super-instanton in the following concise form:

$$\begin{aligned} \Phi_{\mathrm{inst}} &= \frac{y(1+4\mathrm{i}\theta^{(2)}\beta)}{z_{\mathrm{ch}} - z_0 - 4\mathrm{i}\theta^{(2)}\alpha^+} , \\ z_{\mathrm{ch}} &= z + 2\mathrm{i}\theta^{+(1)}\theta^{(2)} . \end{aligned} \tag{90}$$

6.3. Instanton measure

The reader familiar with instanton calculus in QCD certainly remembers that the basic object is the amplitude for the one-instanton transition (another name for the same quantity is the instanton contribution to the vacuum energy). In QCD this amplitude is often called the instanton density, and it plays the key role in instanton-based vacuum models [45–47]. In supersymmetric gluodynamics we have managed to find the one-instanton amplitude exactly [19] and this fact has allowed us to extract the exact β function. The description of the method used in supersymmetric gluodynamics is our main goal in this section and we thus proceed to a discussion of the one-instanton amplitude in the σ model.

The instanton centre may be located at any point in space and, moreover, the instanton size and orientation may be also arbitrary. It is intuitively clear that the amplitude of the one-instanton transition I must be proportional to the integral over x_0 and ρ. (For a more strict argument see refs. [48, 41].) In the supersymmetric model, integration over the bosonic collective coordinates is necessarily accompanied by an integration over their fermionic counter-partners. In such a case after integrating over the instanton orientation, I reduces to a function of g_0, M_0 (ultraviolet cut-off) and ρ times

$$\mathrm{d}x_0\, \mathrm{d}\rho\, \mathrm{d}\alpha\, \mathrm{d}\alpha^+\, \mathrm{d}\beta\, \mathrm{d}\beta^+ .$$

The integrand bears a natural name: the instanton measure.

In ref. [19] it has been shown that, due to supersymmetry, the instanton measure is actually fixed and can be written out almost immediately, up to an overall numerical factor. Let us explain this important point in some detail. First of all, the leading exponential factor

$$I \sim \exp(-4\pi/g_0^2)$$

is determined by the instanton action, $S(\Phi_{\mathrm{inst}})$. The pre-exponential loop corrections account for the fact that the fields fluctuate around Φ_{inst}.

To get the one-loop correction we represent Φ as Φ_{inst} plus a small deviation,

$$\Phi = \Phi_{\mathrm{inst}} + \delta\Phi$$

and expand the action in $\delta\Phi$ keeping only bilinear terms. The resulting functional integral is Gaussian and, in principle, can be done. Symbolically, it can be written as a product of determinants in the background instanton field.

Moreover, the calculation of multi-loop corrections is, generally speaking, a very difficult task. Fortunately, in supersymmetric theories one does not need to calculate anything. All multiloop contributions to the one-instanton amplitude vanish and the same would be valid for the one-loop contribution if it were not for the zero modes. The situation is reminiscent of the well-known cancellation of boson and fermion vacuum loops in the "empty" vacuum (with no instanton background) [49]. The presence of the instanton actually does not affect this cancellation.

6.3.1. One-loop contribution, non-zero modes

Let us forget for a while about the existence of the zero modes. We shall shortly return to their discussion and first integrate over "orthogonal directions" in functional space.

It is convenient to introduce the variations $\delta\varphi$ and $\delta\Psi$ in the following way

$$\varphi = \varphi_{\text{inst}} + \frac{g}{\sqrt{2}}\delta\varphi\,, \qquad \Psi = \Psi_{\text{inst}} + \frac{g}{\sqrt{2}}\delta\Psi \tag{91}$$

where φ_{inst} and Ψ_{inst} stand for the classical solutions described above. Substituting eq. (91) into eq. (85) yields, in the bilinear approximation [38]

$$S^{(2)} = \int \mathrm{d}^2x \left\{\delta\varphi^+\left(-4\frac{\partial}{\partial z}\chi_0^{-2}\frac{\partial}{\partial\bar z}\right)\delta\varphi - 2\mathrm{i}\delta\Psi^+\begin{bmatrix} 0 & \dfrac{\partial}{\partial z}\chi_0^{-2} \\ \chi_0^{-2}\dfrac{\partial}{\partial\bar z} & 0\end{bmatrix}\delta\Psi\right\} \tag{92}$$

where

$$\chi_0 = 1 + \varphi^+_{\text{inst}}\varphi_{\text{inst}} = 1 + |y|^2/|z-z_0|^2\,.$$

Now, we proceed to diagonalize the bilinear form figuring in eq. (92). To this end we must find the eigenfunctions of the corresponding operators. The bosonic modes are defined by the equation

$$-\frac{\partial}{\partial z}\chi_0^{-2}\frac{\partial}{\partial\bar z}\,\delta\varphi_n = E_n^2\chi_0^{-2}\delta\varphi_n \tag{93}$$

where E_n^2 is the nth eigenvalue and $\delta\varphi_n$ is the corresponding eigenfunction, normalized by the condition

$$\int \delta\varphi_n^+\,\delta\varphi_n\,\chi_0^{-2}\,\mathrm{d}^2x = 1\,.$$

Notice that the operator $-(\partial/\partial z)\chi_0^{-2}(\partial/\partial\bar z)$ indeed possesses only non-negative eigenvalues.

We shall not try to solve eq. (93). Instead it will be shown that each bosonic mode (with $E_n \neq 0$) is necessarily accompanied by two "degenerate" fermionic modes. The equation for the fermion eigenfunctions is

$$\begin{bmatrix} 0 & \frac{\partial}{\partial z}\chi_0^{-2} \\ \chi_0^{-2}\frac{\partial}{\partial \bar z} & 0 \end{bmatrix} \begin{bmatrix} \delta\Psi_n^{(1)} \\ \delta\Psi_n^{(2)} \end{bmatrix} = \mathscr{E}_n \chi_0^{-2} \begin{bmatrix} \delta\Psi_n^{(1)} \\ \delta\Psi_n^{(2)} \end{bmatrix}. \tag{94}$$

It has two solutions – this fact immediately stems from eq. (93). Indeed, the first solution

$$\delta\Psi_n^{(1)} = \frac{1}{\sqrt{2}}\delta\varphi_n\,, \qquad \delta\Psi_n^{(2)} = \frac{1}{E_n}\frac{1}{\sqrt{2}}\frac{\partial}{\partial \bar z}\delta\varphi_n$$

corresponds to $\mathscr{E}_n = -E_n$, while the second one

$$\delta\Psi_n^{(1)} = \frac{1}{\sqrt{2}}\delta\varphi_n\,, \qquad \delta\Psi_n^{(2)} = -\frac{1}{E_n}\frac{1}{\sqrt{2}}\frac{\partial}{\partial \bar z}\delta\varphi_n$$

corresponds to $\mathscr{E}_n = E_n$. The two-fold "degeneracy" of the fermion modes is a consequence of the γ_5 invariance of the model. Needless to say that the boson-fermion "degeneracy" reflects the supersymmetry.

Perhaps, many of the readers are already convinced that boson and fermion contributions cancel each other. For those who are not yet convinced we can make a few explanatory remarks. According to the standard rules of functional integration

$$\int \mathrm{D}\delta\varphi^+(x)\,\mathrm{D}\delta\varphi(x) \to \left\{\mathrm{Det}\left[-4\frac{\partial}{\partial z}\chi_0^{-2}\frac{\partial}{\partial \bar z}\right]\right\}^{-1} = \prod_n \frac{1}{4E_n^2}\,. \tag{95}$$

On the other hand,

$$\int \mathrm{D}\delta\Psi^+(x)\,\mathrm{D}\delta\Psi(x) \to \mathrm{Det}\begin{bmatrix} 0 & 2\mathrm{i}\frac{\partial}{\partial z}\chi_0^{-2} \\ 2\mathrm{i}\chi_0^{-2}\frac{\partial}{\partial \bar z} & 0 \end{bmatrix} = \prod_n 2\mathrm{i}\mathscr{E}_n = \prod_n 4E_n^2\,. \tag{96}$$

As a result, all non-zero modes combine to give unity.

These formal manipulations should not overshadow the very simple and general meaning of this phenomenon. In all supersymmetric theories the numbers of boson and fermion degrees of freedom are equal to each other (possibly, with the exception of zero-energy levels). The energy levels are degenerate (supersymmetry!), hence there is no surprise that the fermion loop coincides with the boson one up to a sign (the minus sign appears because of the different statistics). This fact is independent of the presence of the instanton*.

* It took several years to realize that the cancellation avowed for the "empty" vacuum is valid in any self-dual external field as well [50].

6.3.2. One-loop contribution zero modes

The zero-energy levels, the zero modes, are exceptional since they are generally speaking, not paired. All such modes have a geometrical meaning reflecting the symmetries of the classical action. Not all symmetries generate zero modes, though, some are realized trivially. For instance, the Lorentz invariance is not represented by special zero modes. The reason is simple – the instanton solution (79) is a scalar with respect to combined rotations in coordinate space and in isotopic space. Thus, to avoid double counting it is sufficient to account only for isotopic rotations. Likewise, conformal invariance is equivalent to translation and simultaneous dilation.

As for fermion zero modes, one can easily convince oneself that only one of the two possible combinations of the supercurrents generates a non-trivial Ψ_{inst} starting from $(\sigma^a)_{\text{inst}}$, namely, $\frac{1}{2}(S^\mu + \tilde{S}^\mu)$. Another linear combination, $\frac{1}{2}(S^\mu - \tilde{S}^\mu)$, leads to the trivial (vanishing) fermion field.

The integration over the coefficients in front of the boson zero modes is non-Gaussian and, as is well-known, we must switch to integration over the corresponding collective coordinates, but, in passing to collective coordinates, we have to introduce Jacobians. Again, instead of doing a direct computation, we shall try to find a round-about way.

First of all, it is worth recalling that we are interested in the instanton measure averaged over the instanton orientation in isotopic space. Practically, this means that one must integrate over the phase of y. As was explained above there is no need to account for two other possible rotations – this effect is totally cancelled by just the same effect in the "vacuum" amplitude ($\boldsymbol{\sigma}(x) = \text{const.}$).

Secondly, we must take care of the ultraviolet regularization of the theory. It is convenient to use the Pauli–Villars method in the instanton background field (see ref. [48] and a very detailed discussion in the review paper [42]). Introduction of the Pauli–Villars regulators, bosonic and fermionic, results in the following. The determinant figuring, say, in eq. (95) is substituted by

$$\left\{\operatorname{Det}\left[-4\frac{\partial}{\partial z}\chi_0^{-2}\frac{\partial}{\partial\bar{z}}\right]\Big/\operatorname{Det}\left[-4\frac{\partial}{\partial z}\chi_0^{-2}\frac{\partial}{\partial\bar{z}}+M_0^2\right]\right\}^{-1}$$

where M_0 is the regulator mass. In other words, the product of eigenvalues $(4E_n^2)^{-1}$ is substituted by $\Pi_n(M_0^2/4E_n^2)$. The mass term for the fermion regulator is linear in M_0 and, hence, $\Pi 2\mathrm{i}\mathscr{E}_n \to \Pi(2\mathrm{i}\mathscr{E}_n/M_0)$.

The appearance of M_0 obviously does not affect the cancellation of non-zero modes as shown above – each boson mode enters accompanied by a pair of "degenerate" fermion modes, and the corresponding M_0 factors drop out. This is not the case, however, for the zero modes since their number is unbalanced.

As usual [48, 42], each complex boson zero mode yields a pre-exponential factor $S_{\text{inst}}M_0^2 = \text{const}\cdot M_0^2/g_0^2$, where g_0 is the bare charge, $g_0 = g(M_0)$. We have two such modes (see eq. (84)), hence

$$(M_0^2/g_0^2)^2\,. \tag{97}$$

On the other hand, two (complex) fermion zero modes result in the factor

$$(g_0^2/M_0)^2\,. \tag{98}$$

The only point which might deserve a comment is the presence of the g_0^4 factor in eq. (98). In QCD we are used to the fact that the fermion zero modes generate $(S_{\text{inst}})^0$. Actually, our definition of the collective coordinates α and β does not correspond to the standard normalization of the fermion zero modes. Comparing eqs. (87), (88) with eq. (91) we conclude that

$$\int (\delta \Psi^+ \, \delta \Psi)_{\text{zero modes}} \chi_0^{-2} \, d^2x \sim g_0^{-2} \, .$$

This implies in turn that the integral over $d\alpha \, d\alpha^+$, emerging from

$$\operatorname{Det} \begin{bmatrix} 0 & 2i \dfrac{\partial}{\partial z} \chi_0^{-2} \\ 2i\chi_0^{-2} \dfrac{\partial}{\partial \bar{z}} & 0 \end{bmatrix},$$

is accompanied by the factor $(M_0/g_0^2)^{-1}$, coming from the regularizing determinant

$$\left\{ \operatorname{Det} \begin{bmatrix} M_0 & 2i \dfrac{\partial}{\partial z} \chi_0^{-2} \\ 2i\chi_0^{-2} \dfrac{\partial}{\partial \bar{z}} & M_0 \end{bmatrix} \right\}^{-1} .$$

Of course, the same is valid for $d\beta \, d\beta^+$.

Assembling all pieces together we get [64]

$$I = \text{const} \cdot M_0^2 \exp(-4\pi/g_0^2) \, d^2x_0 \frac{d\rho}{\rho} \, d\alpha \, d\alpha^+ \, d\beta \, d\beta^+ \, . \tag{99}$$

Here the power of ρ is reconstructed from dimensional arguments. (Indeed, I must be dimensionless. Moreover, due to the uniformity of superspace, the measure must be independent of x_0, α, β and may depend only on ρ.)

This exhausts the calculation of the instanton measure at the one-loop level.

$I(x_0, \rho, \alpha, \alpha^+, \beta, \beta^+)$ may be considered as an effective interaction with four fermion legs. All vertex corrections to this four-fermion interaction vanish (see subsection 6.4). It is important that renormalization of the Ψ field is also absent; the O(3) sigma model possesses only one Z factor renormalizing the coupling constant (provided that all computations are performed in an O(3) invariant manner; see appendix C for details).

Thus far, our derivation is perfectly analogous to that given previously [19] in supersymmetric gluodynamics. The careful reader might have, perhaps, noticed a peculiarity. The modes $(\varphi_0)_2$ and Ψ_{sc} are actually non-normalizable and call for an *infrared* regularization. We shall not dwell on this issue, referring rather the reader to the very detailed paper [38]. The situation can be summarized as follows. If one introduces an infrared cut-off R it cancels out all the same in I since $(\varphi_0)_2$ yields $\ln R/|y|$ in the numerator of I and Ψ_{sc} produces the same factor in the denominator.

6.4. *Superinvariance of the instanton measure. Multiloop corrections*

In this section we will argue that eq. (99) is actually the exact result for the instanton measure. To prove this let us first show that it is invariant under supersymmetry transformations.

What are the transformation laws of the collective coordinates? The simplest way to answer this question is to examine the instanton superfield (90). Applying supersymmetry transformations to the original solution Φ_{inst} just reshuffles the collective coordinates. Thus, one can readily check that, under the shift $\theta \to \theta + \zeta$, $\theta^+ \to \theta^+ + \zeta^+$, the original solution becomes

$$\begin{aligned}\Phi_{\text{inst}} &\to \frac{y(1 + 4\mathrm{i}\,\theta^{(2)}\beta + 4\mathrm{i}\,\zeta^{(2)}\beta)}{z_{\text{ch}} - z_0 + 4\mathrm{i}\,\zeta^{+(1)}\theta^{(2)} - 4\mathrm{i}\,\theta^{(2)}\alpha^+ - 4\mathrm{i}\,\zeta^{(2)}\alpha^+} \\ &= \frac{y(1 + 4\mathrm{i}\,\zeta^{(2)}\beta)(1 + 4\mathrm{i}\,\theta^{(2)}\beta)}{z_{\text{ch}} - (z_0 + 4\mathrm{i}\,\zeta^{(2)}\alpha^+) - 4\mathrm{i}\,\theta^{(2)}(\alpha^+ + \zeta^{+(1)})},\end{aligned} \tag{100}$$

where we only keep terms linear in the transformation parameters ζ, ζ^+.

In other words, the collective coordinates transform as:

$$y \to y(1 + 4\mathrm{i}\,\zeta^{(2)}\beta), \qquad \bar{y} \to \bar{y}(1 - 4\mathrm{i}\,\beta^+\zeta^{+(2)}); \tag{101a}$$

$$z_0 \to z_0 + 4\mathrm{i}\,\zeta^{(2)}\alpha^+, \qquad \bar{z}_0 \to \bar{z}_0 - 4\mathrm{i}\,\alpha\zeta^{+(2)}; \tag{101b}$$

$$\alpha \to \alpha + \zeta^{(1)}, \qquad \alpha^+ \to \alpha^+ + \zeta^{+(1)}; \tag{101c}$$

$$\beta \to \beta, \qquad \beta^+ \to \beta^+. \tag{101d}$$

It is rather obvious that the instanton measure (99) is invariant under these transformations. Moreover, $\mathrm{d}\ln\rho$ is the only expression compatible with the requirement that the measure should be invariant under the transformation (101a) (we recall the reader that $\rho \equiv |y|$). It follows from this observation alone that higher loops cannot generate corrections of the type $1 + cg_0^2 \ln M_0^2\rho^2 + \cdots$, since this would violate supersymmetry. Shortly we shall show that correction factors of the type $1 + cg_0^2 + c'g_0^4 + \cdots$, which are not ruled out by the above argument, do not appear in I either.

Skipping some subtleties, let us sketch the proof of this assertion. Consider to this end the Feynman graphs in the instanton background field. An example of the three-loop contribution to I is displayed on fig. 9. The graph has two vertices and four superpropagators connecting them. An integration over both supercoordinates (x, θ, θ^+) and (x', θ', θ'^+) is here implied. Furthermore, assume the superpropagators in the instanton background to be known. Fixing the instanton collective coordinates and integrating over (x', θ', θ'^+), we must arrive at an expression of the type

$$\int \mathrm{d}^2x\, \mathrm{d}^2\theta\, \mathrm{d}^2\theta^+\, f(x, \theta, \theta^+, x_0, \rho, \alpha, \beta, \alpha^+, \beta^+) \tag{102}$$

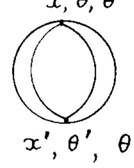


Fig. 9. Multiloop contribution to the vacuum amplitude in the presence of an instanton. The solid lines denote superpropagators in the instanton background field.

where f is some function. The only constraint on f is that it is invariant under supersymmetry transformations. A single glance at eqs. (101) convinces us that, generally speaking, f is a function of the following arguments:

$$z - z_0 - 4\mathrm{i}\theta^{(2)}\alpha^+ , \qquad \bar z - \bar z_0 + 4\mathrm{i}\alpha\theta^{+(2)} ;$$

$$\theta^{(1)} - \alpha , \qquad \theta^{+(1)} - \alpha^+ ;$$

$$y(1 + 4\mathrm{i}\beta\theta^{(2)}) , \qquad \bar y(1 - 4\mathrm{i}\theta^{+(2)}\beta^+) .$$

What is important is that $\theta^{(2)}$ and $\theta^{+(2)}$ do not appear in this list by themselves. As a result, the integration over $\mathrm{d}^2x\, \mathrm{d}^2\theta\, \mathrm{d}^2\theta^+$ yields zero.

Indeed, by shifting x we get rid of the dependence on $\theta^{(2)}\alpha^+$ and $\alpha\theta^{+(2)}$. In order to get a non-vanishing result we must then invoke the dependence on the variables $y(1 + 4\mathrm{i}\beta\theta^{(2)})$ and $\bar y(1 - 4\mathrm{i}\theta^{+(2)}\beta^+)$. This does not help, however; if it did help the answer would be proportional to $\beta\beta^+$. This in turn would imply that integration over two fermionic zero modes (namely, sc-modes) can be performed and that instantons contribute to transition with change of chirality by two units, not four as usual. However, the change of chirality is fixed by the topology of the classical solution and cannot vary in perturbation theory. Hence

$$\int \mathrm{d}^2x\, \mathrm{d}^2\theta\, \mathrm{d}^2\theta^+ f = 0 ,$$

and this completes the proof of our assertion.

It is remarkable that the proof of the non-renormalization theorem for the instanton measure runs very much in parallel to the analogous theorem in supersymmetric gluodynamics [19].

6.5. The exact Gell-Mann–Low function

A last effort, and we shall get as a reward the Gell-Mann–Low function of the model considered, and this to all orders in the coupling constant.

Let us recall that the 2D sigma model is renormalizable. Moreover, all renormalizations reduce to a single factor Z, renormalizing the coupling constant, $Z/g_0^2 = 1/g_R^2$ (see appendix C for some details). All physical amplitudes, being expressed in terms of the renormalized coupling constant, must contain no ultraviolet parameter. In other words, the explicit M_0 dependence, exhibited by eq. (99), must be cancelled out by an implicit dependence entering through $g_0 = g(M_0)$. This requirement immediately yields [64]

$$\begin{aligned} &g^2(M_0)/4\pi = [\ln(M_0^2/m^2)]^{-1} , \\ &\beta = (\mathrm{d}/\mathrm{d}\ln M_0)\, g_0^2/4\pi = -2(g_0^2/4\pi)^2 , \end{aligned} \tag{103}$$

where m is a dimensional constant playing the role of Λ_{QCD}, the scale parameter of quantum chromodynamics.

Note that the exact β function (103) actually coincides with the one-loop result known in the literature [51]. In 4D supersymmetric theories there exists a general theorem [52] stating that, for extended supersymmetries ($n \geq 2$), all the coefficients of the β function (with the possible exception of the first one) vanish. The proof of the theorem relies on some dimensional considerations specific to the 4D case and does not apply directly – as far as we can understand – to the 2D theories considered here. However, the similarity between the final answer for the β function makes us suspect that an analogous general assertion should exist in two dimensions as well. It is all the more probable since the absence of higher loop corrections can be proven – by means of the instanton calculus – for a very broad class of the so-called Käller σ-models [53].

It is worth adding that, unlike in QCD, in the sigma model, one can readily analyse a more general solution with topological charge $q = k$, or k-instanton configurations. These are characterized by $4k$ bosonic and $4k$ fermionic collective coordinates. In particular, each of the k instantons has its own size and orientation. The dependence of the k-instanton transition amplitudes on these parameters is certainly rather complicated and cannot be extracted along the lines sketched above. However, the M_0 dependence of the measure can be found.

As a matter of fact, the corresponding formula

$$I_k \sim \{M_0^2 \exp(-4\pi/g_0^2)\}^k .$$

is a trivial generalization of eq. (53). If we apply the same procedure as that followed above to extract the β function, we arrive at the same result, as it should be the case in any self-consistent theory.

6.6. *Instanton effects*

As an example of more practical applications we shall calculate in this section the two-point function [38]

$$\Pi(x) = \langle 0|\mathrm{T}\{\mathcal{O}(x)\, \mathcal{O}(0)\}|0\rangle \tag{104}$$

where

$$\mathcal{O} = \tfrac{1}{4}(\bar{\psi}^a\psi^a + \mathrm{i}\epsilon^{abc}\bar{\psi}^a\gamma^5\psi^b\sigma^c) = \chi^{-2}\, \bar{\Psi}(1+\gamma^5)\, \Psi ,$$

$$\gamma^5 = \sigma_3.$$

This exercise will allow one to find the vacuum condensate $\langle\bar{\psi}^a\psi^a\rangle$. Moreover it is prompted by similar calculations in supersymmetric gluodynamics [54], which establish the existence of a vacuum condensate of the gluino fields, $\langle\bar{\lambda}^a\lambda^a\rangle \neq 0$.

The strategy for evaluating vacuum condensates by means of the instanton calculus is as follows. Consider an n-point function, vanishing to all orders in perturbation theory by virtue of chirality conservation. Take care, however, that this n-point function includes contributions by instantons. Then, at short distances, the function is saturated by the contribution of small-size instantons and can be reliably calculated. The crucial point is that there always exists some specific n-point function which reduces to a constant in this approximation. Moreover, supersymmetry guarantees that, once the function is a constant at short distances, it remains the same at all distances. In this way the cluster decomposition gets violated at large distances, signalling the presence of a certain vacuum condensate.

All these points can be illustrated using the example of the two-point function (104). First, the operator $\mathcal{O}(x)$ is invariant under O(3) rotations. Second, due to γ^5 invariance, the correlator (104) vanishes to any finite order of perturbation theory. Third, there exists a non-trivial one-instanton contribution to $\Pi(x)$; moreover, only $q = 1$ configurations are relevant since, in this case, the instanton is accompanied by exactly 4 fermion legs, annihilated by the external fermionic sources figuring in eq. (104). Similarly, k-instanton configurations might contribute to the $2k$-point functions of the fermion densities.

Let us consider first the two-point function $\Pi(x)$ in the limit of vanishing x, $x \to 0$, where it is dominated by small-size fluctuations. Moreover, because of asymptotic freedom, the calculation is well under control and is reliable.

To perform an explicit calculation we introduce, following ref. [38], complex fields Ψ, Ψ^+ and make a Wick rotation (note that, in Euclidean space, the γ^5 matrix is just the same, $\gamma^5_{\mathrm{E}} = \sigma_3$). Then we get

$$\Pi(x) = \mathrm{const} \cdot \langle 0| \left\{ \frac{\Psi^+(x)\,(1+\gamma_5)\,\Psi(x)}{\chi^2(x)}, \frac{\Psi^+(0)\,(1+\gamma_5)\,\Psi(0)}{\chi^2(0)} \right\} |0\rangle \tag{105}$$

(see eq. (70)).

The remaining computations are extremely trivial. We simply substitute χ and Ψ by their instanton values:

$$\left(\frac{1+\gamma_5}{2}\right) \Psi(x) \to -2\sqrt{2}\,\mathrm{i}y \left[\frac{1}{(z-z_0)^2}\alpha^+ + \frac{1}{z-z_0}\beta\right],$$

$$\Psi^+\left(\frac{1+\gamma_5}{2}\right) \to \text{h.c.}, \qquad \chi(x) \to 1 + \frac{|y|^2}{|z-z_0|^2},$$

and insert the instanton measure. Simple algebraic manipulation then yields:

$$\begin{aligned}\Pi(x) &= \mathrm{const} \cdot M_0^2 \exp(-4\pi/g_0^2) \int \mathrm{d}^2x_0 \frac{\mathrm{d}\rho}{\rho}\, \mathrm{d}\alpha^+\, \mathrm{d}\alpha\, \mathrm{d}\beta^+\, \mathrm{d}\beta \\ &\quad\times \frac{\rho^4 |z|^2 \alpha^+ \alpha \beta^+ \beta}{|z-z_0|^4\,(1+\rho^2/|z-z_0|^2)^2\,|z_0|^4\,(1+\rho^2/|z_0|^2)^2} \\ &= \mathrm{const} \cdot M_0^2 \exp(-4\pi/g_0^2) \int \mathrm{d}^2x_0 \frac{\mathrm{d}\rho}{\rho} \frac{\rho^4 |z|^2}{(|z-z_0|^2+\rho^2)^2\,(|z_0|^2+\rho^2)^2}\,.\end{aligned}$$

An integration over x_0 and ρ can be performed, for example by means of a Feynman parametrization of the integral. We finally arrive at [38]:

$$\Pi(x) = \mathrm{const} \cdot m^2 \neq 0 \tag{106}$$

where m is the constant entering in the definition of the running coupling constant, see eq. (103).

The result is obviously renormalization-group invariant. Moreover, reiterating the argument presented in subsection 6.4 one can convince oneself that higher order g^2 corrections do not modify eq. (106). In other words, the expression for $\Pi(x)$ quoted in eq. (106) is exact.

Even more surprising is the fact that $\Pi(x)$ is actually x independent. Of course, literally speaking the assertion refers to small x, where the one-instanton approximation used above is justified. However, having established eq. (106) at short distances we can show that, by virtue of supersymmetry, it must be valid for all x.

Just the same situation is realized for the two-point function of gluino densities,

$$\langle 0|\mathrm{T}\left\{\bar{\lambda}^a \frac{1+\gamma^5}{2}\lambda^a(x), \bar{\lambda}^b \frac{1+\gamma^5}{2}\lambda^b(0)\right\}|0\rangle = \text{const.}$$

Moreover, in this case the gluino density $\bar{\lambda}^a(1+\gamma^5)\lambda^a$ (a is the colour index) is the lowest component of a chiral superfield so that the correlator considered can be nothing else but a constant (see ref. [54], or the more recent paper [55]). A similar argument works conceptually (but not technically) in the case of the sigma model.

Let us consider for definiteness the Euclidean version of the model. The operator

$$\mathcal{O} = 2\chi^{-2}\Psi^{+(1)}\Psi^{(1)}$$

is a component of the real superfield $\ln(1+\Phi^+\Phi)$,

$$\ln(1+\Phi^+\Phi) = \cdots + \theta^{+(2)}\theta^{(2)}\, 2\chi^{-2}\Psi^{+(1)}\Psi^{(1)} + \cdots \tag{107}$$

Note that, since $\gamma_{\mathrm{E}}^5 = \sigma_3$, it is perfectly legitimate to trace only the upper (or only the lower) components of the spinors without violating Lorentz invariance.

Among other components of the same superfield there is a fermion operator which enters with $\theta^{(\alpha)}\theta^{(\beta)}\epsilon_{\alpha\beta}\theta^{+(2)}$. We denote it by $K^{+(1)}$. Moreover, it is evident that

$$\langle 0|\, K^{+(1)}(x)\, \mathcal{O}(0)\, |0\rangle = 0\,; \tag{108}$$

since there can be no correlation between the boson and fermion operators in the vacuum. We can perform now a supersymmetry transformation with the parameters

$$\zeta^{(1)} = \zeta^{(2)} = 0\,; \qquad \zeta^{+(1)} = 0\,; \qquad \zeta^{+(2)} \neq 0\,. \tag{109}$$

The key observation is that, under such a transformation

$$\delta K^{+(1)} = \mathrm{i}\zeta^{+(2)}\lambda_\mu\, \partial_\mu \mathcal{O}\ (\lambda_1 = \tfrac{1}{2}, \lambda_2 = \mathrm{i}/2) \text{ and } \delta\mathcal{O} = 0\,,$$

and, applying this transformation to eq. (108) we find

$$\partial_\mu \langle 0|\, \{\mathcal{O}(x)\, \mathcal{O}(0)\}\, |0\rangle = 0$$

or

$$\langle 0|\, \{\mathcal{O}(x)\, \mathcal{O}(0)\}\, |0\rangle = \text{const}\,. \tag{110}$$

Since we have managed to show that, for $x \to 0$, the constant on the right-hand side is non-vanishing, supersymmetry boosts this non-vanishing result to all distances. But, at large distances, eq. (110) would violate the cluster decomposition. It is natural to interpret [55] this fact as due to the occurrence of the corresponding vacuum condensate,

$$\langle 0|\mathcal{O}|0\rangle = \mathrm{const}\cdot m\,. \tag{111}$$

The validity of this interpretation, a biquark condensate, can be checked within CP(N) models [56]. Indeed, in these models, such a biquark condensate is known to exist in the large N limit [57, 58] and its value is fixed theoretically. On the other hand, instanton calculations yield a result analogous to (111). The Green function $G(x_1, \ldots, x_{N+1})$ involving $(N+1)$ pairs of $\Psi^+\Psi$ has been found in ref. [56]. Analogously to eqs. (106), (110), this $(N+1)$-point function turned out to be x_i independent.

In vein with the discussion above, this fact implies that

$$\langle\bar{\Psi}\Psi\rangle_{\mathrm{inst}} = \{\lim_{|x_i-x_j|\to\infty} G(x_1, \ldots x_{N+1})\}^{1/(N+1)} = \mathrm{const}\cdot m\,,$$

what is remarkable since the instanton result coincides exactly [56] with that obtained within the leading $1/N$ approximation [57, 58].

It is amusing to observe once more that the analogy between gluodynamics and the sigma model goes very far. In particular, although the condensates considered are perfectly consistent with supersymmetry as far as the gauge- or O(3)-invariant sectors are concerned, they seemingly do indicate supersymmetry breaking in the gauge- or O(3)-noninvariant sectors.

To begin with, consider $\bar{\lambda}^a(1+\gamma_5)\lambda^a$. On the one hand this operator is proportional to the lowest component of the superfield $W_\alpha W^\alpha$. Moreover, it is well known that if the lowest component develops a non-vanishing vacuum expectation value this does not mean any violation of supersymmetry since the lowest component cannot be represented as an (anti)commutator with the supercharge of some other component.

On the other hand, the same operator is a component of the superfield VW, occupying this time a middle position. More exactly [59]:

$$\bar{\lambda}^a\lambda^a = \frac{1}{4\mathrm{i}}\{Q_\alpha\bar{\lambda}^a_\alpha\gamma_\mu A^a_\mu\}\,. \tag{112}$$

If so, $\langle 0|\bar{\lambda}\lambda|0\rangle \neq 0$ implies that $Q|0\rangle \neq 0$ and supersymmetry is spontaneously broken. The paradox is seemingly solved as follows. Supersymmetry is spontaneously broken but this breakdown manifests itself only in the gauge-noninvariant sector of the theory, while the gauge-invariant sector remains supersymmetric. The situation is reminiscent of the famous U(1) problem. In that case we have a spurious pole generated by instantons and coupled to $K_\mu = \epsilon_{\mu\nu\alpha\beta}(A^a_\nu\,\partial_\alpha A^a_\beta + \frac{1}{3}gf^{abc}A^a_\nu A^b_\alpha A^c_\beta)$. This Goldstone pole, however, decouples from gauge-invariant operators. Here the same thing happens with $\bar{\lambda}^a(x)\,\gamma_\mu\,A^a_\mu(x)$. The corresponding correlation function falls off as x^{-3} at large distances in the presence of an instanton, thus signaling the presence of a pole coupled to $\bar{\lambda}^a\gamma_\mu A^a_\mu$.

By the same token, in the O(3) sigma model the operator $\mathcal{O}$ is not the lowest component of the superfield $\ln(1+\Phi^+\Phi)$. However the lower components of the superfield – their anticommutator with Q_α yields $\mathcal{O}$ – do not possess the O(3) symmetry. Thus, supersymmetry may be broken in the O(3) non-invariant sector and, simultaneously, the O(3)-invariant sector may well be perfectly supersymmetric.

7. Conclusions

This review could be called "Selected topics in non-perturbative effects in sigma models and in QCD". We have concentrated on those aspects which are interesting primarily to QCD investigators, and are rather trivial from the point of view of 2D theorists. Present understanding of two-dimensional models is of course, much deeper. In two dimensions, as opposed to QCD, there exist various subtle and powerful methods allowing one to give exhaustive answers to such dynamical questions as the mass spectrum, the scattering matrix, etc. Suffice it to mention, for instance, the recent complete solution of the O(4) sigma model [60], or older results [61] based on the Zamolodchikov S-matrix approach [62]. Unfortunately, one can hardly hope to extend these methods and results to four-dimensional theories, and we thus leave them aside here.

The present paper touches upon issues for which there exist a close parallel between 4D gauge theories and sigma models. First of all, we investigate Wilson's operator expansion in the presence of non-perturbative effects. A general formulation is given allowing one to construct the expansion to any order. The crucial role of an auxilary mass parameter, the normalization point μ, is emphasized. It is shown that the procedure, which is well-defined theoretically, necessarily requires the introduction of a parameter μ. Once this is done, both the coefficient functions in OPE and the vacuum expectation values of composite operators can be calculated with no ambiguities.

Power (non-perturbative) effects result in the mixing of operators with different normal dimensions. This point seems to be unusual for those who are used to the logarithmic perturbative situation.

Moreover, the operator expansion in the O(N) sigma model ($N \to \infty$) nicely illustrates a property inherent to QCD. Namely that the operators entering the expansion of the correlators, like

$$\langle 0|\mathrm{T}\{f(\partial_\mu \sigma^a(x))^2, f(\partial_\mu \sigma^a(0))^2\}|0\rangle$$

are μ independent in the leading N approximation. As a result, OPE turns out to be especially simple (and there is no difficulty in checking it). Likewise, in QCD, the operators we deal with in applications have a very weak μ dependence, totally negligible from a numerical point of view. For this reason and as a rule for practical problems in QCD (but not in theoretical investigations of OPE!) one can accept a simplified recipe which reads:

(i) draw all relevant Feynman graphs and calculate the coefficients C_i in standard perturbation theory;

(ii) parametrize non-perturbative effects by vacuum expectation values $\langle 0|\mathcal{O}_i|0\rangle$, assuming that $\langle 0|\mathcal{O}_i|0\rangle$ vanish in perturbation theory for all operators except the unit one.

Both prescriptions bear an approximate character – generally speaking, there may exist non-perturbative contributions to $C_i(\mu)$ and, vice-versa, perturbative contributions to $\langle 0|\mathcal{O}_i(\mu)|0\rangle$. We have used the O($N$) sigma model in the next-to-leading approximation to illustrate the latter assertion.

Another interesting issue considered above is the low-energy theorems. They express the existence of symmetries, exact or approximate, and establish strict relations between various amplitudes. Corresponding relations in QCD serve as a unique source of information in connection with confinement dynamics. We have investigated two classes of low-energy theorems: (i) those for the trace of the energy-momentum tensor and (ii) those for the topological charge density. Their validity is confirmed by direct comparison with exact results known in the O(N) and CP($N-1$) models. The analogy between sigma models and QCD turns out to be very close, but not absolute – some details are different. The nature of these distinctions has been explained.

In the second part of the review we have shown that a combination of supersymmetry with instanton calculus results in peculiar consequences. In particular, the exact Gell-Mann–Low function of SUSY models is calculable. The corresponding method, developed previously for Yang–Mills theories, is applicable to 2D sigma models as well. Moreover, in SUSY sigma models, it becomes even simpler, and we couldn't help using this fact for pedagogical purposes. The main lesson is that the coefficients of the β function in SUSY theories have a geometrical meaning and are fixed by the number of zero modes, bosonic and fermionic, in the instanton field.

Another result, discussed in detail, is the instanton calculation of bifermionic condensates. In connection with this problem there is a very interesting question which is left partly unsolved. Namely the fact that the existence of $\Psi^+\Psi$ condensates stands for violation of supersymmetry in the unphysical sectors of the theory (i.e. the gauge non-invariant sector for SUSY Yang–Mills, the O(3) non-invariant sector for the sigma model) may or may not lead to a situation compatible with general theorems, for instance the Witten index theorem [63]?

In a broader context, our consideration confirms that there exists a new way of studying non-perturbative effects in SUSY theories. Namely, a specific example of non-perturbative fluctuations (instanton) used together with some general properties allows one to evaluate unambiguously some quantities which are considered usually to be governed by poorly-understood large-distance dynamics. The supersymmetric theories are very specific since a rigid connection between short and large distance dynamics takes place. Indeed, nothing can prevent appearance of instantons of arbitrary small size in asymptotically free theory where the physical vacuum at short distances is as simple as that of perturbation theory. Proceeding from this fact alone, one is able to fix the full vacuum condensate.

As for the similarity between sigma models and gauge theories it is supported once more. Results first derived in gauge theories do find their analogues in the sigma models. Moreover, in one but important case, the study of sigma models has an advantage – the instanton-based evaluation of the vacuum condensate (111) is confirmed by an alternative (and technically very different) calculation of the same quantity.

In conclusion we would like to thank A.M. Polyakov who a few years ago suggested us to use sigma models to check various elements of our approach to QCD (low-energy theorems, operator product expansion, vacuum condensates). We are also grateful to A. Morozov and V. Fateev for useful discussions.

Appendix A. Vacuum condensates from functional integrals

In this appendix we return to the issue of vacuum condensate and discuss it from the point of view of functional integration.

Let us introduce the generating functional (see section 3.1):

$$Z[\varphi]=\int\prod_x \mathrm{D}\alpha(x)\,\mathrm{D}\sigma(x)\exp\left\{-\frac{1}{2}\int \mathrm{d}^2x\left[(1+\varphi)(\partial_\mu\sigma)^2+\frac{\alpha(x)}{\sqrt{N}}\left(\sigma^2-\frac{N}{f}\right)\right]\right\}. \tag{A.1}$$

If one does not take care of the normalization of the integration measure, the following definition of the vacuum expectation value seems to be natural

$$\left\langle-\frac{1}{2}(\partial_\mu\sigma)^2\right\rangle_{\mathrm{E}}=\frac{1}{Z}\frac{\delta Z}{\delta\varphi}\bigg|_{\varphi=0}. \tag{A.2}$$

This recipe immediately reproduces the result of the naive calculation (12)

$$\langle(\partial_\mu\sigma)^2\rangle_{\rm E} = N \int\limits_{p^2<\mu^2} \frac{d^2p}{(2\pi)^2}\frac{p^2}{p^2+m^2} = \frac{N}{4\pi}\mu^2 - \frac{Nm^2}{f(\mu)}. \tag{A.3}$$

It is easy to show, however, that such a calculation cannot be correct. Indeed, rescaling the integration variables

$$\sigma \to (1+\varphi)^{-1/2}\sigma, \qquad \alpha \to (1+\varphi)\alpha$$

and assuming, for the moment, that φ is x-independent, we get

$$Z[\varphi] = \int \prod_x D\alpha(x)\, D\sigma(x) \exp\left\{-\frac{1}{2}\int d^2x \left[(\partial_\mu\sigma)^2 - \frac{\alpha}{\sqrt{N}}\left(\sigma^2 - \frac{N}{f}(1+\varphi)\right)\right]\right\}. \tag{A.4}$$

Now calculation of the same vacuum expectation value according to eq. (A.2) yields

$$\frac{1}{Z}\frac{\delta Z}{\delta\varphi} = \frac{\sqrt{N}}{2f}\langle\alpha\rangle = \frac{N}{2f}m^2. \tag{A.5}$$

This expression coincides with the result for $(\partial_\mu\sigma)^2$ stemming from the Heisenberg equations of motion. It also coincides with the result obtained within the regularization procedure of section 3.1 involving a point-splitting technique and the Dyson T-product. However, by no means does it reproduce the preceding calculation, eq. (A.3).

The solution of the paradox is as follows. Usually, the measure of the normalized functional integral is independent of the sources. Therefore, to avoid cumbersome expressions associated with normalization, one usually calculates the vacuum expectation values starting from a formula of the type $(1/Z)\delta Z/\delta J$, where J denotes a source, and Z is the generating functional, normalized arbitrarily. In the case at hand the normalization factor for the functional (A.1) does depend on φ, and hence, one should vary over φ not only $\exp(-S_{\rm eff}(\varphi))$ but the integration measure as well. In other words, the use of the formula $Z^{-1}\,\delta Z/\delta J$ is unjustified.

Variation of the measure just cancels out the terms $(N/4\pi)\mu^2$ in eq. (A.3). As for the functional (A.4) its measure is φ-independent, and the standard expression for the vacuum expectation value leads to the correct result. Thus, all definitions of the matrix element $\langle(\partial_\mu\sigma)^2\rangle$ agree with each other.

Let us sketch the basic points of the proof. At first, we shall dwell on the question of the φ dependence of the integration measure. To elucidate the issue it is convenient to start from the functional integral in phase space. For one degree of freedom it has the form

$$\int \prod_t \frac{d\pi(t)\, dq(t)}{2\pi} \exp\left\{i\int [\pi\dot{q} - H(\pi, q)]\, dt\right\}$$

where $H(\pi, q)$ is the Hamilton function, π is the canonical momentum. If we divide the time interval into N small intervals, the correct integration measure is

$$\prod_{i=1}^{N-1}\left[\frac{\mathrm{d}\pi_i\,\mathrm{d}q_i}{2\pi}\right]\frac{\mathrm{d}\pi_N}{2\pi}.$$

Integrating over the canonical momentum π automatically yields the normalized configurational functional. It is rather evident that, if $H(\pi, q)$ is bilinear in π, the norm emerging in this way is entirely determined by the coefficient of π^2 and is independent of interaction.

The canonical momentum for (A.1) is

$$\pi(x) = (1+\varphi)\,\partial_0\sigma\,,$$

and the kinetic energy reduces to

$$T = \frac{1}{2}\int \frac{1}{1+\varphi(x)}\,\pi(x)\,\pi(x)\,\mathrm{d}x\,.$$

Therefore, after integration over π, we arrive at the following expression in the configurational space containing an explicit dependence on φ:

$$\mathrm{D}\sigma(x,t) = \prod_{i=1}^{N}\prod_{x}\mathrm{d}\sigma(x,t_i)\,C\sqrt{1+\varphi(x,t_i)}$$

where C stands for a numerical constant.

The functional (A.4), on the other hand, contains φ only in the interaction terms, and, hence, the corresponding measure in configurational space is φ-independent. This is the reason why the standard definition of the vacuum expectation value (A.2) leads in this case to the correct result for $\langle(\partial_\mu\sigma)^2\rangle$ (see section 3.1).

For completeness we should check that differentiating the measure cancels the quadratically divergent term $(N/4\pi)M_0^2$ in the integral (A.3). From the technical point of view it is rather difficult to check the cancellation literally. Indeed, the M_0^2 divergence of the Feynman integral must be expressed in terms of divergences of the type $(1/\epsilon)^N$ where $\epsilon = t_{i+1} - t_i$ is a short time interval, appearing when the functional integral is substituted by a product of ordinary integrals (see above).

There exists, however, a round-about, discussed with great care in the book by Feynman and Hibbs [24] who devote a whole chapter to the consideration of this problem for a dynamical system with one degree of freedom. In their case, instead of $\langle(\partial_\mu\sigma)^2\rangle$,eq. (A.2), one deals with the vacuum expectation value of the kinetic term,

$$\langle T\rangle = \left\langle\frac{m}{2}\,\dot q(t)\,\dot q(t)\right\rangle = \frac{1}{\int \mathrm{D}q\,\exp(\mathrm{i}S/\hbar)}\int \mathrm{D}q\,\frac{m\dot q^2}{2}\exp(\mathrm{i}S/\hbar)\,. \tag{A.6}$$

It turns out that $\langle T\rangle$, defined in this way, diverges as

$$\langle T\rangle \approx \mathrm{i}\hbar/2\epsilon\,.$$

Moreover, it is easy to show that the divergence disappears provided that one uses another definition,

$$\langle T\rangle = -\mathrm{i}m\,\frac{\delta}{\delta m}\int \mathrm{D}q(t)\exp(\mathrm{i}S/\hbar)\,,$$

and takes into account the fact that the integration measure depends on m (m plays the role of a source analogous to φ). The latter definition of the average kinetic energy is the reasonable one. It is very important that the final result emerging after the cancellation of the divergences, actually reduces to a matrix element of product of velocities at slightly separated points,

$$\langle T\rangle = \lim_{\epsilon\to 0}\frac{m}{2}\langle \dot q(t+\epsilon)\,\dot q(t)\rangle\,. \tag{A.7}$$

The fact that the two definitions, (A.6) and (A.7) are different is due to the equality

$$\langle m\,\dot q(t+\epsilon)\,q(t)\rangle - \langle q(t+\epsilon)\,m\,\dot q(t)\rangle = -\mathrm{i}\hbar\,,$$

or to the commutator term. Thus, on this simple example, we see that the accurate variation over the source, and the Dyson T-product, both yield coinciding results.

It is clear that field theory is not worse in this respect than a system with finite degrees of freedom: the only important point in the Feynman proof sketched above is the bilinearity in the canonical momentum of the Hamiltonian. This property is valid in field theory. Hence, in field theory we also have the equivalence of the two definitions of $\langle(\partial_\mu\sigma)^2\rangle$, obtained with variation over the source and with the Dyson T-product.

Appendix B. Perturbation theory for σ fields

This appendix is devoted to a variant of perturbation theory. When discussing the operator expansion in the main body of the paper, we have used as dynamical variable fields, ϕ^i. They are connected with the original fields σ^i of the sigma model by a non-linear transformation. Here we describe how one can develop perturbation theory directly in terms of the σ^i fields.

Let us first explain why it is difficult to deal with the σ^i fields. Eliminating the Nth component we arrive at the Lagrangian (see eq. (17)):

$$\mathcal{L} = \frac{1}{2}\left\{\partial_\mu\sigma^i\,\partial_\mu\sigma^i + \frac{1}{N/f-\sigma^i\sigma^i}(\sigma^k\,\partial_\mu\sigma^k)(\sigma^e\,\partial_\mu\sigma^e)\right\} \tag{B.1}$$

($i=1,\dots,N-1$; a summation over repeated indices is assumed here). This Lagrangian is adjusted to the description of the situation where the Nth component of the σ field is large,

$$\sigma^N = (N/f-\sigma^i\sigma^i)^{1/2} \sim (N/f)^{1/2}\,,$$

and while all other components are small. The first term in eq. (B.1) then describes small oscillations in a "plane" orthogonal to the direction of σ^N, the second term accounts for curvature. From the exact solution of the problem at large N it is known that the fluctuations of all the components of the σ field are actually large, and that the corresponding matrix elements in the ground state are equal to one another,

$$\langle \sigma^i \sigma^k \rangle = \frac{1}{f} \delta^{ik} ,$$

$$\langle \partial_\mu \sigma^i \, \partial_\nu \sigma^k \rangle = \frac{1}{2f} \delta_{\mu\nu} \delta^{ik} (-m^2) , \qquad (i, k = 1, \ldots, N-1, N) .$$

Now let us try to calculate the vacuum expectation value of the Lagrangian (B.1). There is absolutely no difficulty with the first term,

$$\left\langle \frac{1}{2} \partial_\mu \sigma^i \, \partial_\mu \sigma^i \right\rangle = \frac{N}{2f} (-m^2) [1 + \mathrm{O}(1/N)] .$$

As for the second term, both its numerator

$$\langle \sigma^k \, \partial_\mu \sigma^k \, \sigma^e \, \partial_\mu \sigma^e \rangle = N^{-1} \langle \sigma^2 \rangle \langle (\partial_\mu \sigma)^2 \rangle ,$$

and denominator

$$\left\langle \frac{N}{f} - \sigma^i \sigma^i \right\rangle = \frac{N}{f} - \frac{N}{f} (1 + \mathrm{O}(1/N)) \sim \mathrm{O}(1/f) ,$$

vanish in the leading N approximation. To eliminate this uncertainty of the 0/0 type, one must ensure a higher accuracy. In principle, this is possible, but all computations become unjustifiably complicated. Analogous difficulties arise in the calculation of the σ propagation function and of other correlators. As it is seen from section 3 the $\sigma_i \to \phi_i$ transformation has allowed one to automatically get rid of the uncertainty 0/0 and to deal only with quantities referring to the leading N approximation.

There is an alternative formalism, however, that leads to the same goal. One can calculate all Green functions working with the original σ fields in the leading N approximation, provided that one uses the external field technique.

At first we decompose $\sigma^a(x)$

$$\sigma^a(x) = c^a(x) + q^a(x)$$

where $c^a(x)$ is a classical background field while $q^a(x)$ describe quantum fluctuations. Just the same decomposition is assumed to be performed with the Lagrange multiplier $\alpha(x)$,

$$\alpha(x) = \alpha_{\mathrm{c}}(x) + \alpha_{\mathrm{qu}}(x)$$

(the α field enters only in intermediate formulae). The classical field dynamics is described by the Lagrangian

$$\mathscr{L}^{(0)} = \frac{1}{2} \left\{ \partial_\mu c^a \, \partial_\mu c^a + \frac{\alpha_{\mathrm{c}}}{\sqrt{N}} \left(c^a c^a - \frac{N}{f} \right) \right\} \tag{B.2}$$

while the quantum fluctuations are governed by the Lagrangian

$$\mathcal{L}^{(2)} + \mathcal{L}^{(3)} = \frac{1}{2}\left\{\partial_\mu q^a\, \partial_\mu q^a + \frac{\alpha_c}{\sqrt{N}} q^a q^a + \frac{\alpha_{qu}}{\sqrt{N}} 2c^a q^a + \frac{\alpha_{qu}}{\sqrt{N}} q^a q^a\right\} - J^a q^a . \tag{B.3}$$

The term linear in q^a and α_{qu} reduces to a full derivative and drops off the action provided that $c^a(x)$ and $\alpha_c(x)$ satisfy the classical equations of motion

$$c^a(x)\, c^a(x) = N/f ,$$

$$\partial^2 c^a(x) = \frac{\alpha_c(x)}{\sqrt{N}} c^a(x) ,$$

$$\frac{\alpha_c(x)}{\sqrt{N}} \frac{N}{f} = -(\partial_\mu c^a)^2 .$$

The equations of motion for the quantum fields are

$$2c^a q^a + q^a q^a = 0 , \tag{B.4}$$

$$\partial^2 q^a = \frac{\alpha_c}{\sqrt{N}} q^a + \frac{\alpha_{qu}}{\sqrt{N}} c^a - J^a + \frac{\alpha_{qu}}{\sqrt{N}} q^a . \tag{B.5}$$

Now we are able to find the propagation function of q^a in the background field c^a. This exercise will allow us to build the operator expansion for the exact propagator $\langle \sigma^a(x)\, \sigma^b(0)\rangle$.

The first step is linearization of eqs. (B.4) and (B.5). The relation $c^a q^a = 0$ implies

$$q^a = \Pi^{ab} \varphi^b , \tag{B.6}$$

where φ^b is an arbitrary field and Π^{ab} is the projection operator

$$\Pi^{ab} = \delta^{ab} - c^a c^b / c^2 ,$$

$$\Pi^{ab} c^b = 0 ,$$

$$\Pi^2 = \Pi .$$

Moreover, the linearized eq. (B.5) is

$$\begin{aligned} &[\partial^2 - M^2(c)]\, (\Pi\varphi)^a = -J^a + \frac{\alpha_{qu}}{\sqrt{N}} c^a , \\ &M^2(c) = -(\partial_\mu c)^2 / c^2 . \end{aligned} \tag{B.7}$$

Applying the operator Π to both sides of eq. (B.7), we get rid of the undesirable term proportional to $\alpha_{\rm qu}$,

$$[\Pi\partial^2 - M^2(c)]\,(\Pi\varphi) = -\Pi J$$

and the propagation function of the q field takes the form

$$D^{ab}(x,y) = -\langle x|\Big\{\Pi\,\frac{1}{\Pi\partial^2 + (\partial_\mu c)^2/c^2}\,\Pi\Big\}^{ab}|y\rangle\,. \tag{B.8}$$

Here we have used the equations of motion and expressed $\alpha_c(x)$ in terms of $c(x)$ and $\partial_\mu c(x)$. In the momentum representation the propagator reduces to

$$D^{ab}(q) \equiv \int \mathrm{d}^2x\,\mathrm{e}^{-\mathrm{i}qx}\,D^{ab}(x,0) = \int \mathrm{d}^2x\,\langle x|\Big\{\Pi\,\frac{1}{\Pi(q+\hat p)^2 + M^2(c)}\,\Pi\Big\}^{ab}|y\rangle$$

where $\hat p_\mu = -\mathrm{i}\partial_\mu$. For large q^2 we can expand $D^{ab}(q^2)$ in a series in $1/q^2$:

$$D^{ab}(q) = \Pi^{ab}[c(0)]\,\frac{1}{q^2} - M^2[c(0)]\,\delta^{ab} + \frac{1}{q^4}[\Pi^{ab}[c(x)], \hat p^2]|_{x=0} + \cdots \tag{B.9}$$

(this expression assumes an averaging over the directions of the momentum q).

Identifying the functions of the external field figuring in eq. (B.9) with the vacuum expectation values of the corresponding operators built from σ, we get the operator expansion. Thus,

$$\begin{aligned} D^{ab}(q) &= \delta^{ab}\Big\{\frac{1}{q^2} + \frac{1}{q^4}\langle 0|\,\frac{f}{N}\,(\partial\sigma)^2|0\rangle + \cdots\Big\}(1 + \mathrm{O}(1/N)) \\ &= \delta^{ab}\{1/q^2 - m^2/q^4 + \cdots\}\,, \end{aligned}$$

which reproduces the exact propagator at large q^2.

Let us now turn to the correlator S and begin with tree graphs. Expanding the current j_s and keeping the term linear in the quantum correction we, evidently, get a full derivative,

$$j_s = 2f\,\partial_\mu[\partial_\mu c^a\,\Pi^{ab}q^b]\,. \tag{B.10}$$

The reason is simple: j_s is proportional to the Lagrangian and the field $c^a(x)$ is assumed to satisfy the classical equation of motion. Eq. (B.10) implies that the two-point function S is proportional to the external momentum q. More exactly,

$$\begin{aligned} S(q^2) &= 4f^2\int \mathrm{d}^2x\,\mathrm{e}^{-\mathrm{i}qx}\,\vec\partial_\mu\langle x|\partial_\mu c^a\Big\{\partial\Pi\,\frac{(-1)}{\Pi\partial^2 - M^2}\,\Pi\Big\}^{ab}\,\partial_\nu c^b|0\rangle\overleftarrow{\partial}_\nu \\ &= 4f^2 q_\mu q_\nu\int \mathrm{d}^2x\,\langle x|\partial_\mu c\,\Pi\,\frac{1}{\Pi(q+\hat p)^2 + M^2}\,\Pi\,\partial_\nu c|0\rangle\,. \end{aligned}$$

It is convenient to average $S(q)$ over the directions of q, then

$$S(q^2) = 4f^2 \frac{1}{2} \Big\{ \partial_\mu c^a \, \partial_\mu c^a + \frac{f}{N} (c \, \partial_\mu c)^2 - \frac{1}{q^2} [(\partial c)^2 \, \alpha_c - \partial_\mu c \Pi \partial_\mu \, \partial^2 c + 3\partial^2(\partial_\mu c \, \Pi) \, \partial_\mu c] + \mathrm{O}\Big(\frac{1}{q^4}\Big) \Big\} .$$

The vacuum expectation value of $S(q^2)$ emerges after the following substitution

$$\partial_\mu c^a \, \partial_\nu c^b \to \langle 0 | \partial_\mu \sigma^a \, \partial_\nu \sigma^b | 0 \rangle , \qquad \text{etc.}$$

The resulting expression for $S(q^2)$ at large q^2 is

$$S(q^2) = \frac{4f^2}{2} \Big\{ -\frac{N}{f} m^2 - \frac{2Nm^4}{fq^2} + \cdots \Big\}$$

and coincides with the expansion quoted in the text. Intermediate computations turned out to be even simpler.

The calculation of one-loop diagrams requires a somewhat larger effort than in the case of tree graphs. We shall not dwell on this issue here. It is worth noticing, though, that it is now necessary to solve the constraint

$$2c^a q^a + q^a q^a = 0$$

with higher accuracy

$$q^a = \Pi^{ab} \varphi^b - \frac{c^a}{2c^2} (\Pi \varphi)^2 + \cdots$$

and this changes the vertices for the φ fields. As a result, the form of the current j_s varies. However, the complications are not drastic. The external field technique remains the most economic one.

Appendix C. Uniqueness of the Z factors

In deriving the exact β function in section 6 we have used the fact that there is only one renormalization constant in the O(3) sigma model. More exactly, if the original Lagrangian is written as*

$$\mathcal{L} = \frac{2}{g_0^2} \frac{\partial_\mu \overset{+}{\varphi}_0 \, \partial^\mu \varphi_0}{(1 + \overset{+}{\varphi}_0 \varphi_0)^2} \tag{C.1}$$

* For simplicity we shall consider here the non-supersymmetric variant which, nevertheless, provides the opportunity to answer all questions referring to renormalization.

then the radiative corrections imply the following counterterm:

$$\Delta\mathscr{L} = \frac{2}{g_0^2}(Z-1)\frac{\partial_\mu\varphi_0^+\,\partial^\mu\varphi_0}{(1+\varphi_0^+\varphi_0)^2} \tag{C.2}$$

and its effect reduces to a renormalization of g_0,

$$g_R^2 = g_0^2 Z^{-1}\,. \tag{C.3}$$

However, the reader may find in the literature the opposite assertion according to which there are two distinct Z factors (see e.g. [6b]); one of them renormalizes g and another one renormalizes φ. In other words, $\mathscr{L}+\Delta\mathscr{L}$ is represented in the form

$$\frac{2Z_1}{g_0^2}\frac{\partial_\mu\varphi_0^+\,\partial^\mu\varphi_0}{(1+Z_2\varphi_0^+\varphi_0)^2}\,. \tag{C.4}$$

To avoid confusion we are in a hurry to make a few explanatory remarks.

The appearance of two Z factors is due to the fact that calculations as they are sometimes performed do not respect the full symmetry of the original Lagrangian, the O(3) symmetry. If the symmetry is maintained at each step (including the regularization procedure) the occurrence of Z_2 is certainly impossible. Moreover, there exists a well-developed method, namely, the background field formalism, which automatically respects the full symmetry of the theory.

This method has been already discussed in various contexts in the present review and we shall therefore not dwell on details, but recapitulate the basic points.

The starting point is the introduction of an external field, φ_0

$$\varphi = \varphi_0 + q \tag{C.5}$$

where q represents a quantum piece to be integrated over (q propagates only in loops). The invariance of the Lagrangian

$$\varphi \to \varepsilon + \varepsilon^*\varphi^2\,, \qquad \varphi^* \to \varepsilon^* + \varepsilon(\varphi^*)^2$$

becomes now an invariance with respect to the simultaneous transformations

$$\varphi_0 \to \varepsilon + \varepsilon^*\varphi_0^2\,, \qquad \varphi_0^* \to \varepsilon^* + \varepsilon(\varphi_0^*)^2\,, \tag{C.6a}$$

$$q \to 2\varepsilon^*\varphi_0 q\,, \qquad q^* \to 2\varepsilon\varphi_0^* q^*\,. \tag{C.6b}$$

Notice that the transformation for q is homogeneous; this fact enables one to introduce a mass term for the q field without violating the symmetry, and hence to regularize the theory both in the infrared and ultraviolet limits.

Integrating over q we are left with an effective Lagrangian depending on φ_0 and possessing an invariance with respect to (C.6a). Thus it is intuitively clear that the only counterterm emerging in this way reduces to (C.2). Below we shall illustrate this statement with a one-loop calculation leaving to the reader the opportunity to extend the procedure to multiloop graphs.

In the bilinear approximation, the Lagrangian governing quantum fluctuations, is

$$\mathcal{L}^{(2)} = \frac{2}{g_0^2}\left[\frac{\partial_\mu q^+\,\partial^\mu q - \mu^2 q^+ q}{(1+\varphi_0^+\varphi_0)^2} - 2(\partial_\mu q^+\,\partial^\mu\varphi_0 + \partial_\mu\varphi_0^+\,\partial^\mu q)\frac{\varphi_0^+ q + \varphi_0 q^+}{(1+\varphi_0^+\varphi_0)^3}\right.$$
$$\left. + \partial_\mu\varphi_0^+\,\partial^\mu\varphi_0\left(\frac{3(\varphi_0^+ q + \varphi_0 q^+)^2}{(1+\varphi_0^+\varphi_0)^2} - \frac{2q^+ q}{1+\varphi_0^+\varphi_0}\right)\frac{1}{(1+\varphi_0^+\varphi_0)^2}\right]. \tag{C.7}$$

We have introduced here a small mass term $\mu^2 q^+ q(1+\varphi_0^+\varphi_0)^{-2}$ ensuring infrared regularization. Ultraviolet regularization is achieved by virtue of the Pauli–Villars field R – we add to (C.7) just the same Lagrangian with the substitution

$$q \to R\,, \qquad q^+ \to R^+\,, \qquad \mu \to M_0$$

ascribing to R the opposite metric (with respect to q).

Using the Lagrangian (C.7) we must find the effective one-loop Lagrangian depending on the background field, φ_0. Generally speaking, the calculation can be readily performed for an arbitrary background field. However, since the functional form of the effective Lagrangian is known beforehand (see, e.g. (C.4)) the problem can be essentially simplified if we choose φ_0 in a special way. The most convenient choice is a plane wave,

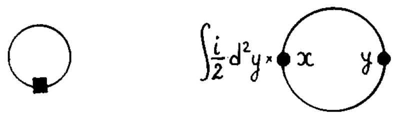
$$\varphi_0 = f\,e^{ikx}\,, \qquad \varphi_0^+ = f^+\,e^{-ikx} \tag{C.8}$$

where f is a dimensionless constant. The value of f is arbitrary while the parameter k is assumed to be small. Moreover, it is clear from eq. (C.4) that we must keep all orders in f, f^+ but can expand in k and keep only bilinear terms in k.

For a plane wave (C.8)

$$\varphi_0^+\varphi_0 = f^+ f = \text{const.}\,,$$

and the first term in $\mathcal{L}^{(2)}$ reduces just to an ordinary boson Lagrangian, quadratic in q. We shall fix the propagator of the q field from this term and shall treat the second ($O(k)$) and the third ($O(k^2)$) terms as an interaction Lagrangian. Both pieces induce vertices with two q lines. Contracting these lines we get the effective Lagrangian $\Delta\mathcal{L}$. The corresponding diagrams are shown in fig. 10 where the solid lines



(a) (b)

Fig. 10. Diagrams determining $\Delta\mathcal{L}$ in the one-loop approximation.

$$\blacksquare = \frac{2}{g_0^2}\,\partial_\mu\varphi_0^+\,\partial^\mu\varphi_0\left(\frac{6\varphi_0^+\varphi_0}{(1+\varphi_0^+\varphi_0)^2} - \frac{2}{1+\varphi_0^+\varphi_0}\right)\frac{q^+ q}{(1+\varphi_0^+\varphi_0)^2}$$

$$\bullet = \frac{2}{g_0^2}(-2)(\partial_\mu q^+\,\partial^\mu\varphi_0 + \partial_\mu\varphi_0^+\,\partial^\mu q)\frac{\varphi_0^+ q + \varphi_0 q^+}{(1+\varphi_0^+\varphi_0)^3}\,.$$

denote the propagation function of q,

$$D_q = \frac{g_0^2}{2}(1+\varphi_0^+\varphi_0)^2 \frac{1}{-\square - \mu^2} \tag{C.9}$$

and the vertices are either bilinear in k (one vertex in fig. 10a) or linear in k (two vertices in fig. 10b).

Computation of the tadpole diagram of fig. 10a is perfectly trivial. The result is

$$\Delta\mathcal{L}\ (\text{fig. 10a}) = \partial_\mu\varphi_0^+\,\partial^\mu\varphi_0\left[\frac{6\varphi_0^+\varphi_0}{(1+\varphi_0^+\varphi_0)^2} - \frac{2}{1+\varphi_0^+\varphi_0}\right]\frac{1}{4\pi}\ln\frac{M_0^2}{\mu^2}. \tag{C.10}$$

As for the other diagram, one should keep in mind that

$$(\partial_\mu\varphi_0(x)\,\partial^\nu\varphi_0^+(x))\int \mathrm{d}^2y\,\langle\partial^\mu(q^+(x)\,q^+(x)),\,\partial_\nu(q(y)\,q(y))\rangle\,\mathrm{e}^{2ik(x-y)} = \mathrm{O}(k^4)$$

and that we are thus left with the correlation function of the type

$$\partial_\mu\varphi_0^+(x)\,\partial^\nu\varphi_0(x)\int \mathrm{d}^2y\,\langle(\partial_\nu q^+(x))\,q(x),\,(\partial^\mu q^+(y))\,q(y)\rangle\,.$$

Integrating over y yields

$$\begin{aligned}\Delta\mathcal{L}\ (\text{fig. 10b}) &= \partial_\mu\varphi_0^+\,\partial^\mu\varphi_0\left[-\frac{4\varphi_0^+\varphi_0}{(1+\varphi_0^+\varphi_0)^2}\right]\int\frac{\mathrm{d}p^2}{4\pi}\left[\frac{p^2}{(p^2+\mu^2)^2} - \frac{p^2}{(p^2+M_0^2)^2}\right] \\ &= \partial_\mu\varphi_0^+\,\partial^\mu\varphi_0\left[-\frac{4\varphi_0^+\varphi_0}{(1+\varphi_0^+\varphi_0)^2}\right]\frac{1}{4\pi}\ln\frac{M_0^2}{\mu^2}.\end{aligned} \tag{C.11}$$

Assembling the two pieces together we get

$$\Delta\mathcal{L} = \frac{\partial_\mu\varphi_0^+\,\partial^\mu\varphi_0}{(1+\varphi_0^+\varphi_0)^2}\left(-\frac{1}{2\pi}\right)\ln\frac{M_0^2}{\mu^2}, \tag{C.12}$$

the counterterm, which is indeed proportional to the original Lagrangian,

$Z_2 = 1\,,\qquad$ q.e.d.

Perhaps, it is worth adding that eq. (C.12) implies that

$$g_\mathrm{R}^2 = \frac{g_0^2}{1-(g_0^2/4\pi)\ln(M_0^2/\mu^2)}\,. \tag{C.13}$$

Thus, we recover in perturbation theory (at the one-loop level) the result derived in the bulk of the review within the framework of instanton calculus.

References

[1] A.M. Polyakov, Phys. Lett. 59B (1975) 79.
[2] A. Gell-Mann and M. Levy, Nuovo Cim. 16 (1960) 705.
[3] A. Belavin and A. Polyakov, Pis'ma ZhETF 22 (1975) 503.
[4] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu.S. Tyupkin, Phys. Lett. 59B (1975) 85.
[5] H. Eichenherr, Nucl. Phys. B146 (1978) 215;
V.L. Golo and A.M. Perelomov, Phys. Lett. 79B (1978) 112;
E. Cremmer and J. Scherk, Phys. Lett. 74B (1978) 341.
[6] a) W. Bardeen, B. Lee and R. Shrock, Phys. Rev. D14 (1976) 985.
b) E. Brezin and J. Zinn-Justin, Phys. Rev. B14 (1976) 3110.
[7] a) D'Adda, M. Lüscher and R. Di Vecchia, Nucl. Phys. B146 (1978) 63.
b) E. Witten, Nucl. Phys. B149 (1979) 285.
[8] M. Shifman, A. Vainshtein and V. Zakharov, Nucl. Phys. B147 (1979) 385.
[9] a) K. Wilson, Phys. Rev. 179 (1969) 1499;
b) R. Brandt, Ann. Phys. (N.Y.) 44 (1967) 221;
W. Zimmermann, in: Lectures on Elementary Particles and Quantum Field Theory, eds. S. Deser, H. Grisaru and H. Pendleton (M.I.T. Press, Cambridge, Mass., 1971) vol. 1;
C.G. Callan, Phys. Rev. D5 (1972) 3202.
[10] F. David, Nucl. Phys. B209 (1982) 433.
[11] F. David, Nucl. Phys. B234 (1984) 237.
[12] S. Gupta and H.R. Quinn, Phys. Rev. D26 (1982) 499; D27 (1983) 980;
C. Taylor and B. McClain, Phys. Rev. D28 (1983) 1364.
[13] a) M. Soldate, preprint SLAC-PUB-3054, Stanford (1983), submitted to Ann. of Phys.
b) A.N. Tavkhelidze and V.F. Tokarev, Preprint P-0314, Nuclear Research Inst., Moscow, 1983.
c) G.S. Danilov and I.T. Dyatlov, Yad. Fiz., to be published.
[14] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Phys. Lett. 78B (1978) 443.
[15] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Nucl. Phys. B191 (1981) 301.
[16] S. Weinberg, Phys. Rev. D11 (1975) 3583.
[17] E. Witten, Phys. Rev. D16 (1977) 2991.
[18] P. Di Vecchia and S. Ferrara, Nucl. Phys. B130 (1977) 93.
[19] V. Novikov, M. Shifman, A. Vainshtein and V. Zakharov, Nucl. Phys. B229 (1983) 381.
[20] S. Coleman, Lecture at 1979 Int. School of Subnuclear Physics, Erice, Italy (SLAC-PUB-2484 (1980)).
[21] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Nucl. Phys. B174 (1980) 378.
[22] A.A. Migdal, unpublished.
[23] A. Di Giacomo and G. Paffuti, Phys. Lett. 108B (1982) 327;
A.Di Giacomo and G.C. Rossi, Phys. Lett. 100B (1981) 481;
T. Banks et al., Nucl. Phys. B190 (1981) 692;
J. Kripfganz, Phys. Lett. 101B (1981) 189.
[24] R. Feynman and A. Hibbs, Quantum Mechanics and Path Integrals (McGraw-Hill Book Co., New York, 1965).
[25] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Nucl. Phys. B, in press.
[26] E.V. Shuryak, Phys. Lett. 79B (1978) 135;
C.G. Callan, R.J. Dashen and D.J. Gross, Phys. Rev. D19 (1979) 1826.
[27] E.V. Shuryak, Phys. Reports 61 (1980) 72.
[28] E.M. Lifshitz and L.P. Pitaevskii, Relativistic Quantum Theory, part 2 (Moscow, Nauka, 1971) sects. 141, 142;
A.I. Akhiezer and V.B. Berestetskii, Quantum-Electrodynamics (Moscow, Nauka, 1969) sect. 40.2.
[29] V. Novikov, M. Shifman, A. Vainshtein and V. Zakharov, Nucl. Phys. B165 (1980) 67.
[30] A.A. Migdal and M.A. Shifman, Phys. Lett. 114B (1982) 445.
[31] R. Rajaraman, Solitons and Instantons (North-Holland, Amsterdam-N.Y.-Oxford, 1982).
[32] E. Witten, Nucl. Phys. B156 (1979) 269;
G. Veneziano, Nucl. Phys. B159 (1979) 213.
[33] Yu.A. Golfand and E.P. Likhtman, in: Problems of Theor. Physics (I.E. Tamm Memorial Volume) (Moscow, Nauka, 1972) p. 37; ZhETF Pisma 13 (1971) 452;
D.V. Volkov and V.P. Akulov, ZhETF Pisma 16 (1972) 621; Phys. Lett. 46B (1973) 109;
J. Wess and B. Zumino, Nucl. Phys. B70 (1974) 39; Phys. Lett. 49B (1974) 52.
[34] P. Fayet and S. Ferrara, Phys. Reports 32 (1977) 249;
V.I. Ogievetski and L. Mezincescu, Usp. Fiz. Nauk 117 (1975) 637 [Sov. Phys. Usp. 18 (1975) 960].
[35] J. Wess and J. Bagger, Supersymmetry and Supergravity (Princeton University Press, Princeton, 1983);
S.J. Gates, M.T. Grisaru, M. Roček and W. Siegel, Superspace (Benjamin-Cummings, Reading, MA, 1983).

[36] B. Zumino, Phys. Lett. 87B (1979) 203.
[37] V. Novikov, M. Shifman, A. Vainshtein and V. Zakharov, preprint ITEP-140, Moscow, 1983 [Fortschritte der Physik, to be published].
[38] H. Bohr, E. Katznelson and K.S. Narain, Nucl. Phys. B238 (1984) 407.
[39] S. Helgason, Differential Geometry, Lie Groups and Symmetric Spaces (Academic Press, N.Y., 1978).
[40] B.A. Dubrovin, S.P. Novikov and A.T. Fomenko, Modern Geometry (Moscow, 1979).
[41] S. Coleman, The Uses of Instantons, Erice Lectures, preprint HUPT-78/004, Harvard 1977.
[42] A. Vainshtein, V. Zakharov, V. Novikov and M. Shifman, Usp. Fiz. Nauk 136 (1982) 553; Sov. Phys. Usp. 25 (1982) 195.
[43] S. Chadha et al., Phys. Lett. 67B (1977) 103.
[44] B. Zumino, Phys. Lett. 69B (1977) 369.
[45] C. Callan, R. Dashen and D. Gross, Phys. Rev. D19 (1979) 1826.
[46] E.-M. Ilgenfritz and M. Mueller-Preussker, Nucl. Phys. B184 (1981) 443.
[47] D.I. Dyakonov and V. Yu. Petrov, preprint LINP-900, Leningrad 1983, Nucl. Phys., submitted.
[48] G't Hooft, Phys. Rev. D14 (1976) 3132.
[49] B. Zumino, Nucl. Phys. B89 (1975) 535;
P. West, Nucl. Phys. B106 (1976) 219.
[50] D'Adda and Di Vecchia, Phys. Lett. 73B (1978) 162.
[51] E. Brezin and J. Zinn-Justin, Phys. Rev. B14 (1976) 3110.
[52] P.S. Howe, K.S. Stelle and P.S. West, Phys. Lett. 124B (1983) 55.
[53] A.Yu. Morozov, A.M. Perelomov and M.A. Shifman, Nucl. Phys. B, in press.
[54] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Nucl. Phys. B229 (1983) 407.
[55] G.C. Rossi and G. Veneziano, Phys. Lett. 138B (1984) 195;
E. Cohen and C. Gomez, Phys. Rev. Lett. 52 (1984) 237.
[56] K.S. Narain, Nucl. Phys. B243 (1984) 131.
[57] E. Witten, Nucl. Phys. B149 (1979) 285.
[58] A. D'Adda, P. Di Vecchia and M. Lüscher, Nucl. Phys. B146 (1978) 63.
[59] M. Dine, W. Fischler and M. Srednicki, Nucl. Phys. B189 (1981) 575.
[60] A.M. Polyakov and P.B. Wiegmann, Phys. Lett. 141B (1984) 223.
[61] R. Shankar and E. Witten, Phys. Rev. D17 (1978) 2134.
[62] A.B. Zamolodchikov, ZhETF Pis'ma 25 (1977) 499;
A.B. Zamolodchikov and Al.B. Zamolodchikov, Ann. Phys. 120 (1979) 253.
[63] E. Witten, Nucl. Phys. B202 (1982) 253.
[64] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Phys. Lett. 139B (1984) 389.