

# Spontaneous $CP$ breaking in QCD and the axion potential: an effective Lagrangian approach

Paolo Di Vecchia,<sup>a,b</sup> Giancarlo Rossi,<sup>c,d</sup> Gabriele Veneziano<sup>e,f</sup>  
and Shimon Yankielowicz<sup>g</sup>

<sup>a</sup>The Niels Bohr Institute,

Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

<sup>b</sup>Nordita, KTH Royal Institute of Technology and Stockholm University,

Roslagstullsbacken 23, SE-10691 Stockholm, Sweden

<sup>c</sup>Dipartimento di Fisica, Università di Roma “Tor Vergata” and INFN Sezione Roma 2,

Via della Ricerca Scientifica, 00133 Roma, Italy

<sup>d</sup>Centro Fermi, Museo Storico della Fisica e Centro Studi e Ricerche “E. Fermi”,

Piazza del Viminale 1, 00184 Roma, Italy

<sup>e</sup>Collège de France,

11 place M. Berthelot, 75005 Paris, France

<sup>f</sup>Theory Division, CERN,

CH-1211 Geneva 23, Switzerland

<sup>g</sup>School of Physics and Astronomy, Tel-Aviv University,

Ramat-Aviv 69978, Israel

E-mail: [divecchi@nbi.dk](mailto:divecchi@nbi.dk), [rossig@roma2.infn.it](mailto:rossig@roma2.infn.it),

[Gabriele.Veneziano@cern.ch](mailto:Gabriele.Veneziano@cern.ch), [Shimonya@post.tau.ac.il](mailto:Shimonya@post.tau.ac.il)

**ABSTRACT:** Using the well-known low-energy effective Lagrangian of QCD — valid for small (non-vanishing) quark masses and a large number of colors — we study in detail the regions of parameter space where  $CP$  is spontaneously broken/unbroken for a vacuum angle  $\theta = \pi$ . In the  $CP$  broken region there are first order phase transitions as one crosses  $\theta = \pi$ , while on the (hyper)surface separating the two regions, there are second order phase transitions signalled by the vanishing of the mass of a pseudo Nambu-Goldstone boson and by a divergent QCD topological susceptibility. The second order point sits at the end of a first order line associated with the  $CP$  spontaneous breaking, in the appropriate complex parameter plane. When the effective Lagrangian is extended by the inclusion of an axion these features of QCD imply that standard calculations of the axion potential have to be revised if the QCD parameters fall in the above mentioned  $CP$  broken region, in spite of the fact that the axion solves the strong- $CP$  problem. These last results could be of interest for axionic dark matter calculations if the topological susceptibility of pure Yang-Mills theory falls off sufficiently fast when temperature is increased towards the QCD deconfining transition.

**KEYWORDS:** 1/N Expansion, Anomalies in Field and String Theories

**ARXIV EPRINT:** [1709.00731](https://arxiv.org/abs/1709.00731)

---

**Contents**

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Chiral, large-<math>N</math> QCD at arbitrary <math>\theta</math>: a reminder</b>	<b>4</b>
<b>3</b>	<b>QCD phase diagrams</b>	<b>9</b>
3.1	$N_f = 1$	10
3.2	$N_f = 2$	13
3.3	$N_f \geq 3$	15
<b>4</b>	<b>Spontaneous <math>CP</math> violation and the axion potential</b>	<b>17</b>
4.1	Including the axion in the QCD effective Lagrangian	17
4.2	The axion potential	21
4.2.1	$N_f = 1$	21
4.2.2	$N_f \geq 2$ and discussion	23
<b>5</b>	<b>Conclusions</b>	<b>26</b>
<b>A</b>	<b>Ward-Takahashi identities</b>	<b>28</b>

---

**1 Introduction**

Already in the early seventies Dashen recognized [1] that phases in the quark mass matrix could spontaneously break  $CP$  and the possibility that such a phenomenon could explain the observed  $CP$  violation in kaon physics was explored [2]. It turned out that these violations were too large to explain the experiments with  $K$  mesons and would give a much too high value for the electric dipole moment of the neutron and for the  $\eta \rightarrow 2\pi$  decay amplitude [3]. At about the same time Weinberg pointed out [4, 5] that possible  $CP$  violating phases can be eliminated through chiral rotations of the quark fields. These rotations included an anomalous  $U_A(1)$  transformation and therefore generated a  $CP$  violating term proportional to  $F\tilde{F}$ . However, at the time such a term was considered innocuous since it amounts to adding to the Lagrangian a total derivative (and, indeed, it is irrelevant at all orders in perturbation theory). It looked therefore as if QCD did automatically conserve  $CP$ .

The phenomenological problem with that naive conclusion is that the same triviality of  $F\tilde{F}$  implies the famous  $U(1)$  problem, expressed for instance by the anomalously large  $\eta'$  mass. After the discovery of the instanton solutions and the presence of different topological sectors in pure Yang-Mills (YM) theory, it was soon realized [6, 7] that the  $U(1)$  problem might be solved although this remained somewhat controversial for a while [8]. The observation [9, 10] that, in the framework of large- $N$  QCD, the mass matrix of the

mesons contains, besides the terms related to the masses of the quark, an extra parameter connected to the topological susceptibility of pure YM theory, opened the way to a quantitative resolution of the U(1) problem [11–14].<sup>1</sup>

Unfortunately, the resolution of the U(1) problem brought back the question of  $CP$  conservation in strong interactions. Indeed,  $CP$  violating phases of the quark mass matrix could no longer be rotated away so that QCD would not automatically preserve  $CP$ . The YM Lagrangian could be supplemented with an extra term, given by the topological charge density and containing a parameter, the so-called vacuum angle  $\theta$ , that also breaks  $CP$ . By performing an anomalous  $U_A(1)$  transformation of the quark fields, it turns out that the relevant observable quantity is a combination of the  $\theta$  parameter and the phases present in the quark mass matrix  $M$ , given by  $\bar{\theta} \equiv \theta + \arg \det m$ . The  $CP$  violation induced by a non-vanishing  $\bar{\theta}$  was first used to estimate the resulting electric dipole moment of the neutron in [16]. It was later refined in [17] by identifying a leading logarithmic contribution thus establishing a limit on  $\bar{\theta}$  of order  $10^{-9}$ – $10^{-10}$  for the smallness of which QCD, on its own, has no explanation. In section 4 we will come back to this problem and to its resolution with the help of an axion.

The next step was the construction and study of an extension [18–21] of the effective Lagrangian of the light pseudo Nambu-Goldstone bosons (the non-linear  $\sigma$ -model) to include a term linear in the topological charge density and reproducing both the  $U_A(1)$  anomaly and the  $\theta$  term of the microscopic theory, as well as a quadratic term whose coefficient is associated with the topological susceptibility of pure YM theory.<sup>2</sup>

The  $\theta$  dependence of physical quantities, in the framework of the effective Lagrangian for mesons, was studied in detail in refs. [19, 21] where it was found that for a generic non-zero value of  $\theta$   $CP$  is broken but, for  $\theta = \pi$  (where  $CP$  is a symmetry of the theory) could be either spontaneously broken or independent of the values of the quark masses and the topological susceptibility.

The possibility of spontaneously breaking of  $CP$  from the introduction of phases in the quark mass matrix was taken up again in [27–30] in the framework of low-energy effective Lagrangian for the pseudoscalar mesons, where it was shown that at  $\theta = \pi$  there are indeed two regions in parameter space, one where  $CP$  is conserved and the other where  $CP$  is broken, separated by a surface whose shape depends on the quark mass ratios. An important result of the analysis of ref. [28] is that, on the separating surface, one of the mesons becomes massless.

Recently, the discussion of the case  $\theta = \pi$  has been taken up again in a very interesting paper [31] where it was proven, under a few very plausible assumptions, that, even for finite  $N$ ,  $CP$  must be spontaneously broken at  $\theta = \pi$  in  $SU(N)$  YM theory. The main ingredient in the derivation of this result is the use of 't Hooft's anomaly constraint for the mixed anomaly of the discrete  $CP$  and center symmetries. This first order transition nicely fits with the spontaneous  $CP$  breaking in QCD at  $\theta = \pi$  in the decoupling (heavy quark mass) limit.

<sup>1</sup>A big role in the solution of the U(1) problem was played by the analogy of QCD with the  $CP^{n-1}$  model in two dimensions [15].

<sup>2</sup>Together with refs. [18–21] see also refs. [22–24] and refs. [25, 26] for an old and a more recent review.

In the first part of this paper we discuss again the  $\theta$  dependence of chiral, large- $N$  QCD in its low-energy approximation, using the above mentioned effective Lagrangian and concentrating our attention on what happens in the neighborhood of  $\theta = \pi$ . Besides the quark masses, parametrized in terms of the  $N_f$  parameters  $-2m_i\langle\bar{\psi}\psi\rangle \equiv \mu_i^2 F_\pi^2$ , there is an additional parameter, the topological susceptibility of YM theory,  $\chi_{\text{YM}}$ , which, as already mentioned, plays a crucial role in the large- $N$  resolution of the U(1) problem. In this enlarged parameter space (w.r.t. the one considered in [28]) there is an hypersurface separating the region where  $CP$  is conserved from the one where  $CP$  is spontaneously broken. On the hypersurface itself the theory exhibits a second order phase transition where one of the pseudo Nambu-Goldstone bosons (PNGBs) becomes exactly massless and the topological susceptibility of QCD diverges. Inside the  $CP$  broken region the ground state makes a sudden, finite jump as  $\bar{\theta}$  goes from  $\pi - \epsilon$  to  $\pi + \epsilon$  corresponding to a first order phase transition. In an appropriate complex parameter space (discussed in section 3) the second order point resides at the endpoint of a first order line associated with  $CP$  breaking and starting at  $-\infty$  where the decoupling to YM occurs. The position of the second order endpoint depends on all the other parameters (mass ratios, topological susceptibility).

These results can be seen as a rather straightforward generalization of those of [28–30] to the case of a generic value of  $\chi_{\text{YM}}$  and of [32, 33] to the case of a generic quark mass matrix (the equal mass case is indeed quite special since it is always in the  $CP$  broken phase except in the case of a single light flavor). In [33] the issue of  $CP$  breaking in QCD was also addressed for finite  $N$ , and the theories residing on the resulting domain walls were studied.

In the second part of this paper we turn our attention to the case in which QCD has been augmented by the addition of an axion field, the best known way to solve, in a natural way, the strong- $CP$  problem. The axion can be easily incorporated in the effective Lagrangian (see e.g. [26]). We then find that the QCD results of the previous sections have an interesting bearing on the properties of the axion potential near the boundary of its periodicity interval. Depending again on where one is in the QCD parameter space the axion potential can differ significantly from the one commonly used in the literature (see e.g. [34]). Furthermore, in the immediate vicinity of the critical hypersurface the very concept of an axion potential ceases to be physically meaningful since the dynamics is described by *two* very light pseudoscalars whose mass is of the order of the geometric mean between the PNGB mass and the conventional axion mass. Quite naturally, in that region the mass eigenstates are strongly mixed combinations of the two. Although at zero temperature real QCD is quite deeply inside the  $CP$  conserving region, one cannot exclude a-priori the possibility that, as one moves towards the deconfining, chiral-symmetry-restoring temperature, QCD may move (in parameter space) towards the critical hypersurface or even inside the  $CP$  breaking region. If true, this could have interesting physical effects, e.g. on the standard computation of axionic dark matter abundance. As we will discuss, some precise lattice calculations in quenched QCD at finite temperature would be highly desirable in order to settle this point.

The paper is organized as follows. In section 2 we review the main properties and consequences of the low-energy effective Lagrangian at generic values of the  $\theta$  angle and

quark masses. In section 3.1 we study in detail the behavior at  $\theta = \pi$  in the case of a single flavor, while in sections 3.2 and 3.3 we discuss the case of two or more flavors respectively. Non-trivial checks that the results derived from the effective Lagrangian exactly satisfy general Ward-Takahashi identities (WTIs) are presented in appendix A. In section 4 we consider QCD with a very generic additional axionic degree of freedom and discuss the axion potential in the different situations described above. In particular we examine the “realistic” case of two or three unequal mass light flavors. Some final remarks are presented in section 5.

## 2 Chiral, large- $N$ QCD at arbitrary $\theta$ : a reminder

For the sake of being self-contained we summarize in this section some already known facts. We will refer, where appropriate, to the original literature for further details.

Assuming confinement and spontaneous chiral symmetry breaking by a quark-antiquark condensate at a generic value of  $\theta$ , QCD, for three light quarks ( $m_i \ll \Lambda_{\text{QCD}}$ ) and a large number of colors ( $N \gg 1$ ),<sup>3</sup> is described at low-energy by the following effective Lagrangian [18–21]

$$\begin{aligned}
 L = & \frac{1}{2} \text{Tr} \left( \partial_\mu U \partial^\mu U^\dagger \right) + \frac{F_\pi}{2\sqrt{2}} \text{Tr} \left[ \mu^2 (U + U^\dagger) \right] \\
 & + \frac{Q^2}{2\chi_{\text{YM}}} + \frac{i}{2} Q \text{Tr} \left[ \log U - \log U^\dagger \right] - \theta Q.
 \end{aligned}
 \tag{2.1}$$

Here  $F_\pi$  is the pion decay constant ( $F_\pi \sim 95 \text{ MeV}$  in the real world with  $N = 3$ )<sup>4</sup> and the  $3 \times 3$  matrix  $U$  describes, non-linearly, the spontaneous breaking of the approximate  $U(3)_L \otimes U(3)_R$  chiral symmetry in terms of nine light PNBs so that

$$U = \frac{F_\pi}{\sqrt{2}} e^{i\sqrt{2}\Phi/F_\pi} \quad ; \quad \Phi = \Pi^a T_{ij}^a,
 \tag{2.2}$$

where  $T_{ij}^a$  are the matrices satisfying the algebra of  $U(3)$  normalized as  $\text{Tr}(T^a T^b) = \delta^{ab}$ . Furthermore,  $\mu^2$  is proportional to the quark mass matrix<sup>5</sup> which, without loss of generality, can be taken to be real, diagonal and non negative (provided a  $\theta$ -term is added). More precisely, in terms of the quark masses  $m_i$  and condensate at  $\theta = 0$ ,  $\langle \bar{\psi}\psi \rangle$ ,  $\mu^2$  is defined by

$$\mu_{ij}^2 = \mu_i^2 \delta_{ij} = -2m_i \langle \bar{\psi}\psi \rangle F_\pi^{-2} \delta_{ij}.
 \tag{2.3}$$

Although the physically relevant case is the one with two or three light flavors, for the sake of generality, we will consider hereafter the case of  $N_f$  light flavors (hence now

---

<sup>3</sup>We also assume to be below the so-called conformal window whose beginning is expected to occur at a value of  $N_f$  proportional to  $N$ . Using the two-loop beta function it is found to occur at  $N_f = 34N^3/(13N^2 - 3) \sim \frac{34}{13}N$ .

<sup>4</sup>Remember that  $F_\pi$  grows like  $\sqrt{N}$  for large  $N$ .

<sup>5</sup>In the literature  $\mu^2$  is often denoted by  $M$ . In this paper we prefer this different notation in order to avoid confusion with a different use of the symbol  $M$ .

$i, j = 1, \dots, N_f$ ).  $Q$  is the QCD topological charge density that appears in the divergence of the  $U_A(1)$  current

$$\begin{aligned} \partial_\mu J_5^\mu &= 2N_f Q + 2 \sum_{i=1}^{N_f} m_i P_i ; & Q &= \frac{g^2}{32\pi^2} F_{\mu\nu}^a (\tilde{F}^a)^{\mu\nu} ; & (\tilde{F}^{\mu\nu})^a &= \frac{1}{2} \epsilon^{\mu\nu\rho\sigma} F_{\rho\sigma}^a \\ J_5^\mu &= \sum_{i=1}^{N_f} \bar{\psi}_i \gamma^\mu \gamma_5 \psi_i ; & P_i &= i \bar{\psi}_i \gamma_5 \psi_i . \end{aligned} \quad (2.4)$$

Modulo the mass term, the Lagrangian (2.1) is invariant under  $SU(N_f)_L \otimes SU(N_f)_R \otimes U(1)_V$  transformations while, under the  $U_A(1)$  transformation  $U \rightarrow U e^{-2i\alpha}$ , one has

$$\frac{i}{2} \text{Tr} \left( \log U - \log U^\dagger \right) \rightarrow \frac{i}{2} \text{Tr} \left( \log U - \log U^\dagger \right) + 2\alpha N_f Q , \quad (2.5)$$

as needed. The quadratic term in  $Q$  contains a coefficient,  $\chi_{\text{YM}}$ , which turns out to be nothing but the topological susceptibility of pure YM theory in the large- $N$  limit. Finally, the last term takes into account of the presence of a non-zero  $\theta$  parameter.

The  $2\pi$  periodicity in  $\theta$  (which in the underlying QCD theory is related to the quantization of  $\nu \equiv \int d^4x Q(x)$ ) can be easily checked at the level of (2.1). Indeed, a shift in  $\theta$  by  $2\pi$  can be reabsorbed, thanks to the anomaly term in (2.1), by a chiral rotation by  $2\pi$  of a component (say  $U_{11}$ ) of  $U$  under which even the mass term in (2.1) is invariant. We also note that, under  $CP$ ,  $Q \rightarrow -Q$  and  $U \rightarrow U^\dagger$ . Thus naively, in our convention of real positive quark masses, only the last term in (2.1) breaks  $CP$  unless  $\theta = 0$ .<sup>6</sup> However, even if  $\theta = \pm\pi$ ,  $CP$  is not *explicitly* broken since  $2\pi$  periodicity implies that  $\theta = +\pi$  and  $\theta = -\pi$  are equivalent. Nonetheless, as discussed below,  $CP$  can be *spontaneously* broken at  $\theta = \pm\pi$ .

In the infinite- $N$  limit the anomaly effectively turns off and the physical PNCB spectrum consists of  $N_f^2$  unmixed states of mass

$$M_{ij}^2 = \frac{1}{2} (\mu_i^2 + \mu_j^2) , \quad i, j = 1, 2, \dots, N_f . \quad (2.6)$$

In general, one could add to the previous Lagrangian a  $U(N_f)_L \otimes U(N_f)_R$  invariant function of  $Q$ ,  $U$  and  $U^\dagger$ . However, it can be shown [18–21] that the only surviving terms at large  $N$  are those appearing in (2.1).

Before we proceed further let us notice that the Lagrangian (2.1) for a single flavor is exactly the Lagrangian one gets by using the two-dimensional bosonization rules in the massive Schwinger model, where the kinetic term of the gauge field corresponds to the first term in the second line of (2.1) with  $a \equiv \frac{e^2}{\pi}$ ,  $F_\pi = \frac{1}{\sqrt{2\pi}}$ , while the term coupling the fermions to the gauge field corresponds to the anomaly term with the logarithm. The other terms are also reproduced as also noticed in refs. [29, 30]. A similar structure appears also in other two-dimensional models as the one discussed in ref. [35]. In those models, as also in the massive Schwinger model, the bosonized Lagrangian is equivalent to the original

---

<sup>6</sup>It is believed, and supported by lattice calculations and the chiral Lagrangian approach, that at  $\theta = 0$  the vacuum is non-degenerate and the theory is gapped with no spontaneous  $CP$  breaking.

microscopic Lagrangian, while, in our case, the effective Lagrangian (2.1) is only valid at low energy, for small quark masses, and for large  $N$ . However, the fact that in all these cases one gets the same Lagrangian indicates that our results may not necessarily be valid only at large  $N$ .

Since the equation of motion of  $Q(x)$  is algebraic, we could integrate out  $Q(x)$  from the start. However, as later on we will want to compute the  $\langle QQ \rangle$  correlator, we prefer to rewrite eq. (2.1) as follows:

$$L = \frac{1}{2} \text{Tr} \left( \partial_\mu U \partial^\mu U^\dagger \right) + \frac{F_\pi}{2\sqrt{2}} \text{Tr} \left( \mu^2 (U + U^\dagger) \right) - \frac{\chi_{\text{YM}}}{2} \left[ \theta - \frac{i}{2} \text{Tr} \left( \log U - \log U^\dagger \right) \right]^2 + \frac{1}{2\chi_{\text{YM}}} \left[ Q - \chi_{\text{YM}} \left( \theta - \frac{i}{2} \text{Tr} \left( \log U - \log U^\dagger \right) \right) \right]^2. \quad (2.7)$$

The presence of the  $\theta$  term implies that, for unequal masses, the vacuum does not correspond anymore to  $\langle U \rangle$  being proportional to the unit matrix.<sup>7</sup> We are obliged to introduce a separate VEV for each flavor by writing

$$\langle \Phi_{ij} \rangle = -\frac{F_\pi}{\sqrt{2}} \phi_i \delta_{ij}. \quad (2.8)$$

Inserting eq. (2.8) in the previous Lagrangian the vacua of the theory correspond to the minima of the following potential

$$V(\phi_i) = -\frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \cos \phi_i + \frac{\chi_{\text{YM}}}{2} \left( \theta - \sum_{i=1}^{N_f} \phi_i \right)^2, \quad (2.9)$$

and are therefore obtained by looking for the stable solutions of the equations

$$\mu_i^2 \sin \phi_i - a \left( \theta - \sum_{j=1}^{N_f} \phi_j \right) = 0 \quad ; \quad i = 1, \dots, N_f, \quad (2.10)$$

where we have defined

$$a = \frac{2\chi_{\text{YM}}}{F_\pi^2}. \quad (2.11)$$

The eqs. (2.10) determine  $\phi_i$  and all physical quantities in terms of  $\mu_i^2$ ,  $a$  and  $\theta$ . Denoting this solution by  $\phi_i = \hat{\phi}_i(\mu_i^2, a, \theta)$ , and computing  $\langle Q \rangle$  from the quadratic part of the Lagrangian in (2.14), we finally identify  $\langle Q \rangle$  with  $\chi_{\text{YM}} \left( \theta - \sum \hat{\phi}_i \right)$ .

Defining a new  $\hat{U}$  matrix in terms of the shifted fields

$$\hat{U} \equiv \frac{F_\pi}{\sqrt{2}} e^{i\frac{\sqrt{2}}{F_\pi} \hat{\Phi}} \quad ; \quad \hat{\Phi} = \Phi - \langle \Phi \rangle, \quad (2.12)$$

---

<sup>7</sup>In spite of appearance, this does not correspond to a spontaneous breaking of  $\text{SU}(N_f)_V$  since phases can always be rotated away into the quark mass matrix.

as well a shifted  $Q$  field

$$\hat{Q} = Q - \chi_{\text{YM}} \left( \theta - \sum_{i=1}^{N_f} \hat{\phi}_i \right), \quad (2.13)$$

we get a Lagrangian that depends on  $\hat{U}$  and  $\hat{Q}$  as follows

$$\begin{aligned} L = & -V(\hat{\phi}_i) + \frac{1}{2} \text{Tr} \left( \partial_\mu \hat{U} \partial^\mu \hat{U}^\dagger \right) + \frac{F_\pi^2}{2} \text{Tr} \left[ \mu^2(\theta) \left( \cos \left( \frac{\sqrt{2}}{F_\pi} \hat{\Phi} \right) - 1 \right) \right] - \frac{a}{2} \left[ \text{Tr} \left( \hat{\Phi} \right) \right]^2 \\ & + \chi_{\text{YM}} \left( \theta - \sum_{i=1}^{N_f} \hat{\phi}_i \right) \text{Tr} \left[ \sin \left( \frac{\sqrt{2}}{F_\pi} \hat{\Phi} \right) - \frac{\sqrt{2}}{F_\pi} \hat{\Phi} \right] \\ & + \frac{1}{2\chi_{\text{YM}}} \left[ \hat{Q} - \chi_{\text{YM}} \frac{\sqrt{2}}{F_\pi} \text{Tr} \hat{\Phi} \right]^2, \end{aligned} \quad (2.14)$$

where we have defined

$$\mu_{ij}^2(\theta) \equiv \mu_i^2 \cos \hat{\phi}_i \delta_{ij}. \quad (2.15)$$

The first line of eq. (2.14) (apart from the first term which is a constant) describes the spectrum and the interaction of the PNGBs, the second, being odd under  $\hat{\Phi} \rightarrow -\hat{\Phi}$ , gives the  $CP$  violating contributions (controlled by its coefficient  $\chi_{\text{YM}}(\theta - \sum_{j=1}^{N_f} \hat{\phi}_j)$ ), and the third line will be useful to determine the topological susceptibility in QCD. As we shall see below, while for  $\theta = 0$  the  $CP$  violating coefficient is zero, for  $\theta = \pm\pi$  it can be non-zero. The latter case has to be attributed to the spontaneous breaking of  $CP$  by some non- $CP$ -invariant VEVs.

The spectrum of the PNGBs is obtained by restricting our attention to the terms quadratic in  $\hat{\Phi}$ , coming from the first line of (2.14), for which we get

$$L_2 = \frac{1}{2} \text{Tr} \left( \partial_\mu \hat{\Phi} \partial^\mu \hat{\Phi} \right) - \frac{1}{2} \text{Tr} \left( \mu^2(\theta) \hat{\Phi}^2 \right) - \frac{a}{2} \left[ \text{Tr} \left( \hat{\Phi} \right) \right]^2. \quad (2.16)$$

Separating in  $\hat{\Phi}$  the generators in the Cartan sub-algebra from the others

$$\hat{\Phi} = \tilde{T}_{ij}^{\alpha\beta} \tilde{\Pi}^{\alpha\beta} + v_i \delta_{ij}, \quad (2.17)$$

we have from  $L_2$  the following two-point correlation functions in momentum space

$$\langle \tilde{\Pi}^{\alpha\beta}(x) \tilde{\Pi}^{\gamma\delta}(y) \rangle^{\text{F.T.}} = \frac{i\delta^{\alpha\gamma} \delta^{\beta\delta}}{p^2 - M_{\alpha\beta}^2}; \quad M_{\alpha\beta}^2 = \frac{1}{2} (\mu_\alpha^2(\theta) + \mu_\beta^2(\theta)) \quad (2.18)$$

and

$$\langle v_i(x) v_j(y) \rangle^{\text{F.T.}} = iA_{ij}^{-1}(p^2), \quad (2.19)$$

where

$$A_{ij}(p^2) = (p^2 - \mu_i^2) \delta_{ij} - aH_{ij} \equiv p^2 \delta_{ij} - M_{ij}^2 \quad (2.20)$$



and  $H_{ij}$  is a matrix with 1 in all entries. The masses  $M_i$  of the physical states in the Cartan sub-algebra are obtained by diagonalizing the matrix  $M_{ij}^2$  and satisfy the equation

$$\det M^2 = \prod_{i=1}^{N_f} M_i^2(\theta) = \prod_{i=1}^{N_f} \mu_i^2(\theta) \left[ 1 + a \sum_{i=1}^{N_f} \frac{1}{\mu_i^2(\theta)} \right]. \quad (2.21)$$

For  $p^2 \neq 0$  one gets

$$\det A = \prod_{i=1}^{N_f} (p^2 - M_i^2(\theta)) = \prod_{i=1}^{N_f} (p^2 - \mu_i^2(\theta)) \left[ 1 - a \sum_{i=1}^{N_f} \frac{1}{p^2 - \mu_i^2(\theta)} \right]. \quad (2.22)$$

In the last part of this section we use the Lagrangian (2.14) to compute the two-point correlator of  $\hat{Q}$  (note that, by definition  $\langle \hat{Q} \rangle = \langle v_i \rangle = 0$ ) and relate the topological susceptibilities of YM and QCD. Since there is no quadratic term involving  $v_i$  with the combination of  $\hat{Q}$  and  $v_j$  appearing in the last line of eq. (2.14), we get immediately the following two-point correlation function

$$\left\langle \left( \hat{Q}(x) - \chi_{\text{YM}} \frac{\sqrt{2}}{F_\pi} \sum_{k=1}^{N_f} v_k(x) \right) v_j(y) \right\rangle = 0, \quad (2.23)$$

which implies

$$\langle \hat{Q}(x) v_j(y) \rangle^{\text{F.T.}} = \frac{\chi_{\text{YM}} \sqrt{2}}{F_\pi} \sum_{k=1}^{N_f} \langle v_k(x) v_j(y) \rangle^{\text{F.T.}} = i \frac{\chi_{\text{YM}} \sqrt{2}}{F_\pi} \sum_{k=1}^{N_f} A_{kj}^{-1}(p^2), \quad (2.24)$$

where in the last step we have used eq. (2.19). From the relation

$$\sum_{k=1}^{N_f} A_{kj}^{-1}(p^2) = \frac{1}{p^2 - \mu_j^2(\theta)} \frac{\prod_{i=1}^{N_f} (p^2 - \mu_i^2(\theta))}{\prod_{i=1}^{N_f} (p^2 - M_i^2(\theta))} \quad (2.25)$$

the correlator (2.24) becomes

$$\langle \hat{Q}(x) v_j(y) \rangle^{\text{F.T.}} = i \frac{\chi_{\text{YM}} \sqrt{2}}{F_\pi} \frac{1}{p^2 - \mu_j^2(\theta)} \frac{\prod_{i=1}^{N_f} (p^2 - \mu_i^2(\theta))}{\prod_{i=1}^{N_f} (p^2 - M_i^2(\theta))}. \quad (2.26)$$

Finally, from the last line of eq. (2.14) we get

$$\left\langle \left( \hat{Q} - \chi_{\text{YM}} \frac{\sqrt{2}}{F_\pi} \sum_{j=1}^{N_f} v_j \right) (x) \times \left( \hat{Q} - \chi_{\text{YM}} \frac{\sqrt{2}}{F_\pi} \sum_{j=1}^{N_f} v_j \right) (y) \right\rangle = i \chi_{\text{YM}} \delta^{(4)}(x - y). \quad (2.27)$$

Using eq. (2.26) and

$$\sum_{h,k=1}^{N_f} A_{hk}^{-1}(p^2) = \frac{\sum_{k=1}^{N_f} \frac{1}{p^2 - \mu_k^2(\theta)}}{1 - a \sum_{k=1}^{N_f} \frac{1}{p^2 - \mu_k^2(\theta)}} = \sum_{k=1}^{N_f} \frac{1}{p^2 - \mu_k^2(\theta)} \frac{\prod_{i=1}^{N_f} (p^2 - \mu_i^2(\theta))}{\prod_{i=1}^{N_f} (p^2 - M_i^2(\theta))}, \quad (2.28)$$

we get

$$\langle Q(x)Q(y) \rangle_{\text{conn.}}^{\text{F.T.}} = \langle \hat{Q}(x)\hat{Q}(y) \rangle^{\text{F.T.}} = i \frac{\chi_{\text{YM}}}{1 - a \sum_{k=1}^{N_f} \frac{1}{p^2 - \mu_k^2(\theta)}}. \quad (2.29)$$

In particular, for  $p^2 = 0$  one gets the topological susceptibility in QCD with  $N_f$  flavors

$$\chi_{\text{QCD}} = \frac{\chi_{\text{YM}}}{1 + a \sum_{i=1}^{N_f} \frac{1}{\mu_i^2(\theta)}} = \chi_{\text{YM}} \left( 1 - \frac{\chi_{\text{YM}}}{\sum_{k=1}^{N_f} (m_i \langle \bar{\psi}\psi \rangle)} \right)^{-1}. \quad (2.30)$$

Since our effective Lagrangian is, strictly speaking, valid for  $N \rightarrow \infty$  (where the  $\eta'$  is a PNCB), the quark condensate in the previous equation should be evaluated in the leading planar order proportional to  $N$ . The next to the leading terms should not be included. In particular, it means that the next to the leading contributions which are affected by logarithmic divergencies [36–38] and make the quenched quark condensate ill-defined, are avoided.<sup>8</sup>

Finally as a last remark we wish to stress an important property of both eqs. (2.21) and (2.30), namely that they both reduce to the case of a theory with  $N_f - 1$  flavors when one of the quark masses becomes very large. If all quarks become much heavier than  $a$  (which can still be the case in the chiral regime since  $a$  scales like  $1/N$  at large  $N$ ) then  $\chi_{\text{QCD}} \rightarrow \chi_{\text{YM}}$ . Finally, when any quark flavor becomes massless the QCD topological susceptibility goes to zero as it should on general grounds.

In appendix A we provide the form of various two-point functions at small (but not necessarily vanishing) momenta and show that they satisfy exactly (i.e. without  $O(1/N)$  corrections) all the expected anomalous and non-anomalous Ward-Takahashi identities (WTIs).

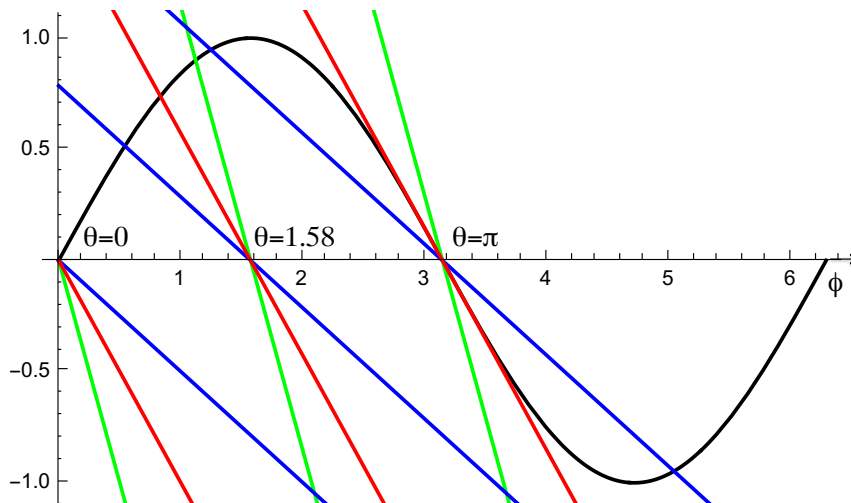
### 3 QCD phase diagrams

In this section we discuss the phase diagrams of QCD at zero temperature and chemical potential for different numbers of quark flavors  $N_f$ . The parameter space in which we consider possible phase transitions is spanned by the  $(N_f + 1)$  parameters  $\mu_i^2 \geq 0$  and  $\theta$  (with  $0 \leq \theta < 2\pi$ ) while considering  $\chi_{\text{YM}}$  and  $F_\pi$  (and thus  $a$ ) as given. In section 4 we will see how those phase diagrams acquire a different meaning in the presence of a QCD axion and also briefly mention possible non-zero temperature effects.

Just to make our terminology clear. We will be talking about  $CP$  conservation or violation referring, respectively, to the vanishing or non-vanishing of the quantity  $\chi_{\text{YM}} \left( \theta - \sum_{j=1}^{N_f} \phi_j \right)$  in eq. (2.14). Sometimes the breaking of  $CP$  is explicit (e.g. for generic values of  $\theta$ ) while in some other cases it is spontaneous (like for  $\theta = \pi$ ). We will try to make the distinction when needed in order to avoid confusion.

---

<sup>8</sup>Even though our analysis is valid for large  $N$ , in the spirit of the large  $N$  expansion, we will use it for the physical  $N = 3$  with the hope that even for this value the leading term (in the large  $N$  expansion), dominates.



**Figure 1.** Solutions of  $V'(\phi) = 0$  are given by the intersections of the curve  $\sin \phi$  (black) with the straight lines  $(\theta - \phi)/\epsilon$  for  $\theta = 0$ ,  $\theta = \pi$  and a generic value taken to be  $\theta = 1.58$ . Code color is as follows:  $\epsilon < 1$  green lines,  $\epsilon = 1$  red lines,  $\epsilon > 1$  blue lines.

### 3.1 $N_f = 1$

In the case of a single flavor the potential in eq. (2.9) becomes, up to an irrelevant factor

$$\frac{V(\phi)}{a} = -\epsilon \cos \phi + \frac{1}{2}(\theta - \phi)^2 ; \quad \epsilon \equiv \frac{\mu^2}{a}, \quad (3.1)$$

from which we can compute its derivatives with respect to  $\phi$

$$\begin{aligned} \frac{V'}{a} &= \epsilon \sin \phi + \phi - \theta ; & \frac{V''}{a} &= \epsilon \cos \phi + 1 \\ \frac{V'''}{a} &= -\epsilon \sin \phi ; & \frac{V''''}{a} &= -\epsilon \cos \phi. \end{aligned} \quad (3.2)$$

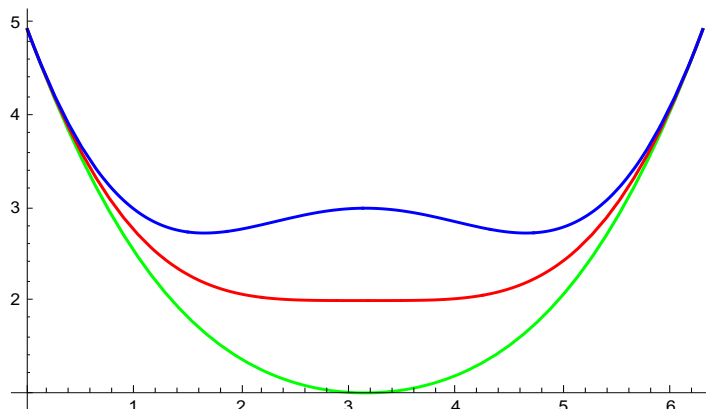
Let us distinguish two cases:

- $\epsilon < 1$

In this case  $V'' > 0$  so that there can only be a single stable minimum with positive mass. This is confirmed by solving graphically the equation  $V' = 0$ , as illustrated in figure 1. At  $\theta = 0$  the minimum is at  $\phi = 0$  while at  $\theta = \pi$  it is at  $\phi = \pi$ . In both cases  $CP$  is unbroken. At  $0 < \theta < \pi$  ( $\pi < \theta < 2\pi$ ) the minimum is at some  $0 < \phi < \theta$  ( $\theta < \phi < 2\pi$ ) and  $CP$  is explicitly broken.

- $\epsilon \geq 1$

This case is much richer. Since now  $V''$  can be negative, some stationary points can correspond to maxima rather than minima of  $V$ . For a zero mass ground state we should require  $V' = V'' = 0$ . But for it to be the absolute minimum we should also have  $V''' = 0$  and  $V'''' > 0$ . However, from (3.2) we see that  $V''' = 0$  is only possible if  $\phi = \pi \text{ mod}(\pi)$  and therefore (from the first and last of eqs. (3.2)) if  $\theta = \pi$ . Let us then consider this case in more detail.



**Figure 2.**  $V(\phi)$  of eq. (3.1) at  $\theta = \pi$ , and  $\epsilon = 0.5$  (green curve),  $\epsilon = 1.0$  (red) and  $\epsilon = 2.0$  (blue).

For  $\theta = \pi$  there is always a stationary point at  $\phi = \pi$  which, however, for the case  $\epsilon > 1$ , corresponds to a maximum ( $V'' < 0$ ). Since  $V$  is bounded from below there should be minima elsewhere. Indeed, for  $\epsilon = 1 + \delta, \delta \ll 1$ , one easily finds two (degenerate) minima. For  $\epsilon = 1$  the three stationary points degenerate at  $\phi = \pi$  and the stable minimum corresponds to a massless  $CP$  conserving ground state.

To make the discussion more quantitative let us assume that  $\theta = \pi$  and that  $\phi = \pi - \delta$  where  $\delta$  is a small quantity. We can determine  $\delta$  by plugging it into the first equation in (3.2) getting

$$\delta \left( \frac{\delta^2 \epsilon}{6} + 1 - \epsilon \right) = 0. \quad (3.3)$$

In this way we find again the solution  $\delta = 0$ , which corresponds to a maximum, together with two stable minima related by  $CP$  (see below) at

$$\delta_{\pm} = \pm \sqrt{\frac{6(\epsilon - 1)}{\epsilon}}. \quad (3.4)$$

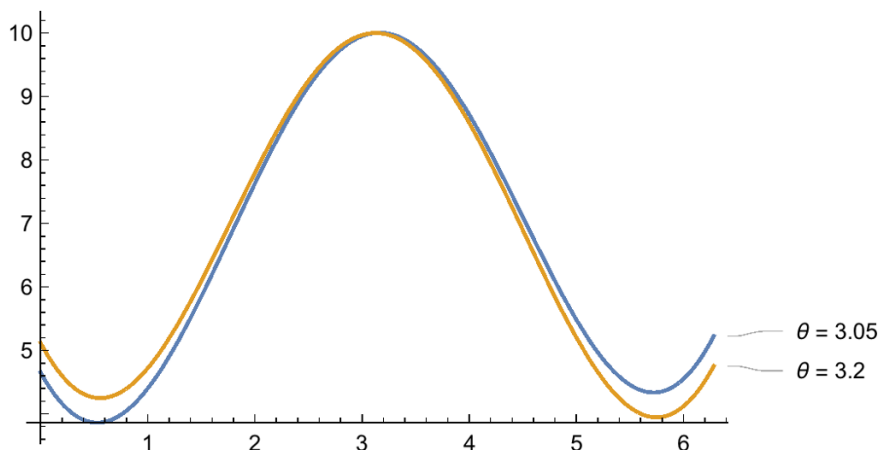
This can be seen by plugging (3.4) in the second of the equations (3.2) obtaining respectively

$$\frac{V''}{a} \Big|_{\delta=0} = 1 - \epsilon \quad ; \quad \frac{V''}{a} \Big|_{\delta_{\pm}} = 2(\epsilon - 1). \quad (3.5)$$

This implies that the solution with  $\delta = 0$  is a stable one for  $\epsilon \leq 1$ , while the two other solutions are stable for  $\epsilon > 1$  (see figure 2). At  $\epsilon = 1$  there is a second order phase transition where the PNCB becomes massless. Indeed the mass square is given by the second derivative of the potential computed at the minimum, yielding

$$M^2 = \mu^2(\theta) + a = \mu^2 \cos \phi + a, \quad (3.6)$$

as follows from (2.21) with  $N_f = 1$ . Notice that  $M^2$  goes to zero for  $\epsilon = 1, \theta = \phi = \pi$ .



**Figure 3.**  $V(\phi)$  for two values of  $\theta$  on opposite sides of  $\pi$  and  $\epsilon = 5$ . The true minimum swaps abruptly as one goes through  $\theta = \pi$ . For the apparent lack of  $2\pi$  periodicity in this figure see the discussion in the text.

If we move away from  $\theta = \pi$  while  $\epsilon > 1$  we can have different situations. Below a critical  $\epsilon(\theta)$  there is only one minimum while above it an extra couple of stationary points pops out. One of them is a local maximum, the other a local minimum. Which is the absolute minimum depends on  $\theta$ . For  $\theta < \pi$  the true minimum is at  $\phi < \theta$  while for  $\theta > \pi$  it is at  $\phi > \theta$  as illustrated in figure 3. Precisely at  $\theta = \pi$  there is a two-fold degeneracy easily understood as due to the spontaneous breaking of  $CP$ .<sup>9</sup> This abrupt change in the minimum of the potential around  $\theta = \pi$  signals a first order phase transition all along the line  $\mu^2 e^{i\theta} = [-\infty, -a^2]$  ending at the second order phase transition point  $\theta = \pi, \mu^2 = a$  as first observed in [29, 30] and more recently discussed in [32, 33].

The second order phase transition is not only signalled by the mass gap going to zero, but also from the divergence of the topological susceptibility (generally defined as the  $\langle Q Q \rangle$  correlator at zero momentum) at  $\epsilon = 1, \theta = \pi$ . This follows from eq. (2.30) for  $N_f = 1$

$$\chi_{\text{QCD}} = \frac{\chi_{\text{YM}}}{1 + \frac{a}{\mu^2(\theta)}} = \frac{\chi_{\text{YM}} \epsilon \cos \phi}{1 + \epsilon \cos \phi}, \quad (3.7)$$

which diverges for  $\epsilon = 1$  at  $\theta = \phi = \pi$ .

Figures 2 and 3 illustrate the shape of the potential for different values of  $\epsilon$  and for  $\theta = \pi$  or  $\theta \neq \pi$ , respectively. Note that the potentials shown in figures 2 and 3 do not look periodic in  $\phi$  while they should. Indeed the potential is multi valued because of the log term in the effective Lagrangian (2.7) and the correct branch has to be chosen as we vary  $\phi$ . Periodicity is thus restored at the expense of non-analyticity points (cusps) in  $V$  at particular values of  $\phi$ . For instance, for  $\theta = \pi$  (figure 2) the cusp are at  $\phi = 0 \text{ mod}(2\pi)$ , while for a generic  $\theta$  they are at  $\theta + \pi \text{ mod}(2\pi)$ .

<sup>9</sup>Indeed the two minima appear to be symmetric with respect to  $\phi = \pi$  and become equal and opposite after a trivial  $2\pi$  shift of one of them.

### 3.2 $N_f = 2$

In the case  $N_f = 2$  with unequal masses (say,  $\mu_1^2 < \mu_2^2$ ) the equations to be solved are

$$\epsilon_1 \sin \phi_1 = \epsilon_2 \sin \phi_2 = \theta - \phi_1 - \phi_2 \quad ; \quad \epsilon_i \equiv \frac{\mu_i^2}{a}. \quad (3.8)$$

For  $\theta = \pi$  the solutions are simply

$$\phi_1 = \pi \quad ; \quad \phi_2 = 0 \quad \text{or} \quad \phi_1 = 0 \quad ; \quad \phi_2 = \pi. \quad (3.9)$$

The masses of the two pseudoscalar mesons can be read from eq. (2.22) and are given by

$$M_{1,2}^2 = a + \frac{\mu_1^2(\theta) + \mu_2^2(\theta)}{2} \pm \sqrt{a^2 + \left( \frac{\mu_1^2(\theta) - \mu_2^2(\theta)}{2} \right)^2}, \quad (3.10)$$

valid for arbitrary  $\theta$ . It is easy to check that the mass squared with the minus sign is massless if the following condition is satisfied

$$a(\mu_2^2(\theta) + \mu_1^2(\theta)) = \left( \frac{\mu_1^2(\theta) - \mu_2^2(\theta)}{2} \right)^2 - \left( \frac{\mu_1^2(\theta) + \mu_2^2(\theta)}{2} \right)^2. \quad (3.11)$$

Notice that, if both  $\mu_{1,2}^2(\theta)$  are positive, the previous condition cannot be satisfied because the r.h.s. is always negative, while the l.h.s. is always positive. In particular, it cannot be satisfied at  $\theta = 0$ . But at  $\theta = \phi_1 = \pi$ , the previous condition becomes

$$a(\mu_2^2 - \mu_1^2) = \mu_1^2 \mu_2^2 \implies \frac{1}{a} + \frac{1}{\mu_2^2} = \frac{1}{\mu_1^2}. \quad (3.12)$$

This means that, if the condition

$$\frac{1}{\mu_1^2} - \frac{1}{\mu_2^2} \geq \frac{1}{a} \quad (3.13)$$

is fulfilled,  $CP$  is unbroken because  $\theta - \phi_1 - \phi_2 = 0$ . Although the second solution in (3.9) conserves  $CP$ , it does not correspond to the absolute minimum and does not satisfy (3.11).

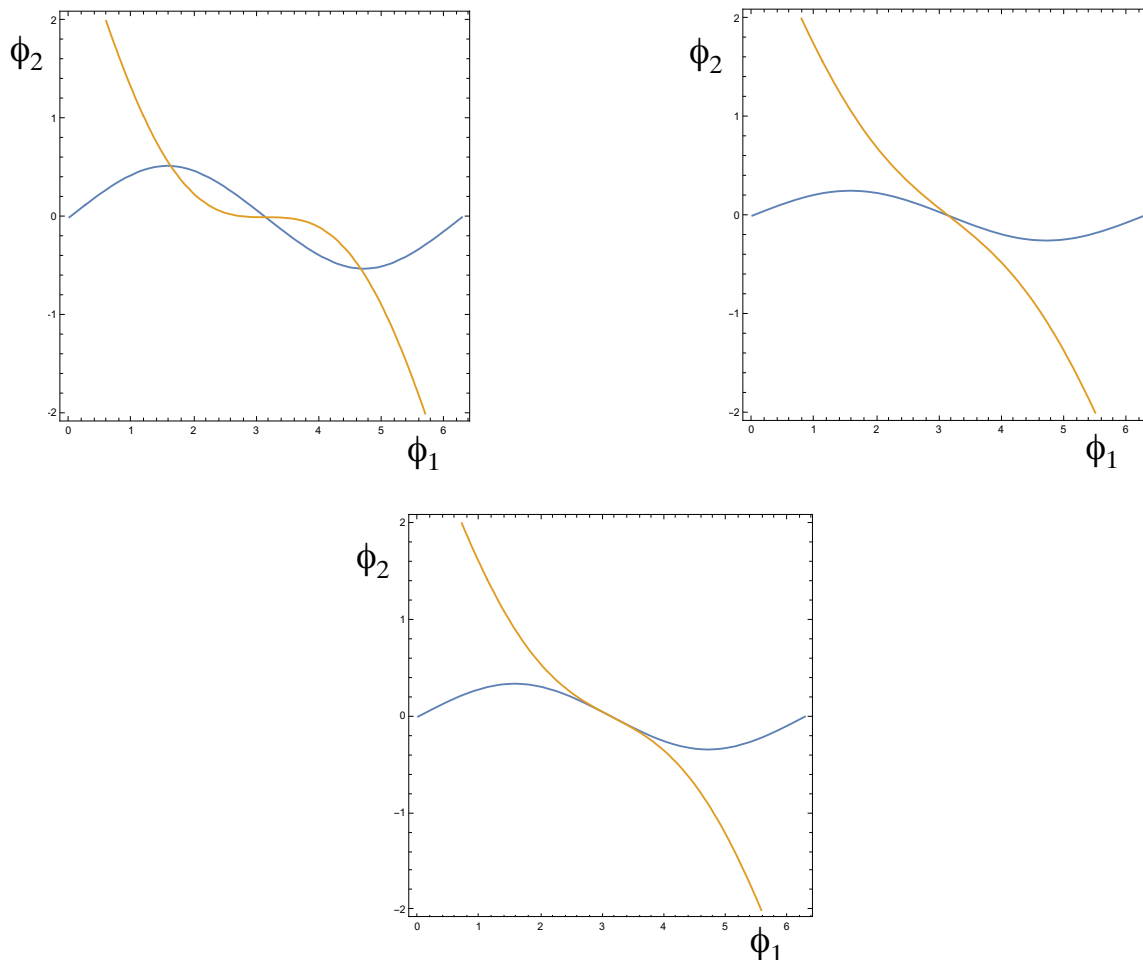
On the other hand, if  $\mu_1^{-2} < \mu_2^{-2} + a^{-1}$  not even the first solution in eq. (3.9) corresponds to a minimum and other solutions takes over. As in the case  $N_f = 1$ , let us consider the following example. Defining

$$\epsilon_i = \mu_i^2/a \quad ; \quad \rho = \epsilon_1/\epsilon_2 \quad ; \quad \sigma = \epsilon_1 + \rho - 1, \quad (3.14)$$

one finds, to leading order in  $\sigma \ll 1$ , the two further solutions

$$\phi_1 = \pi - \delta_1 \quad ; \quad \phi_2 = \delta_2 \quad ; \quad \delta_1 = \pm \sqrt{\frac{6\sigma}{1 - \rho^3}} \quad ; \quad \delta_2 = \rho \delta_1. \quad (3.15)$$

In the general case the solutions can be found numerically. Figure 4 illustrates again the three distinct cases for  $\theta = \pi$ , while figure 5 does the same for  $\theta \neq \pi$ . We see clearly that, as in the  $N_f = 1$  case, the critical surface  $\mu_1^{-2} = \mu_2^{-2} + a^{-1}$  separates the situation with a single solution from the one with several solutions. In the latter case  $CP$  is spontaneously

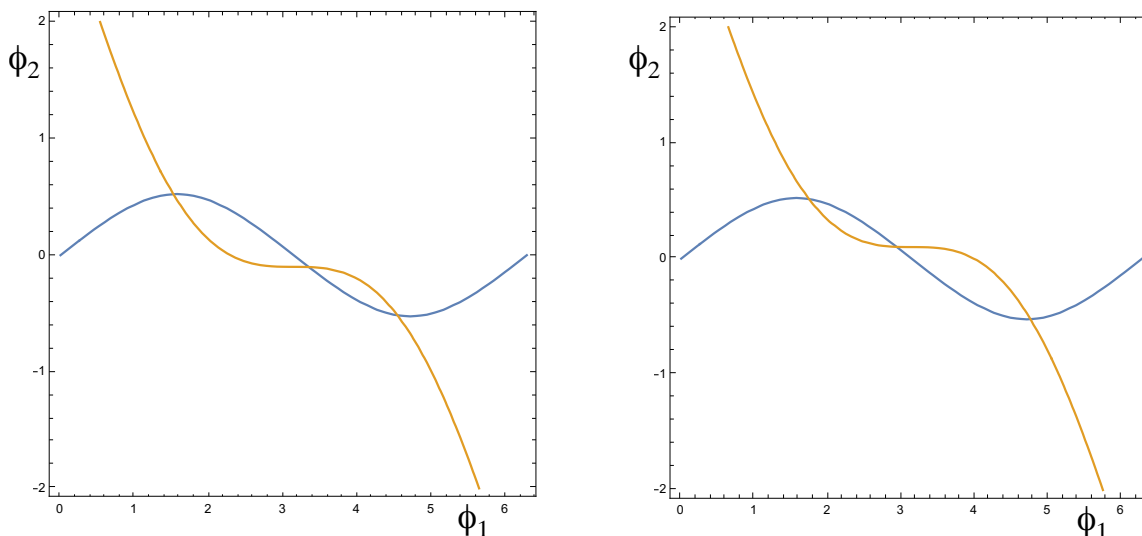


**Figure 4.** Solutions of the stationarity conditions for  $N_f = 2$ ,  $\mu_d^2 = 2\mu_u^2$  and  $\theta = \pi$  are given by the intersections of the curves shown in different color. The two situations with one or three solutions are shown together with the limiting case corresponding to a second order phase transition.

broken and the ground state jumps as we go from  $\theta < \pi$  to  $\theta > \pi$ . On the critical surface there is a massless excitation and the QCD topological susceptibility blows up.

In this generic case the phase structure resembles the  $N_f = 1$  case. In the complex  $\mu_1^2 e^{i\theta}$  plane ( $\mu_1^2$  is the smallest mass parameter) we find a line of first order transitions along the negative axis ending on a second order transition point where one mass goes to zero. The position of the second order point depends on the other parameters (mass ratios,  $a$ ). We can also see this structure in the complex  $\det \mu^2$  plane, as discussed in the next subsection.

Let us close with a short discussion of the peculiarities of the equal mass case,  $\mu_1^2 = \mu_2^2 = \mu^2$ . In this case the condition (3.13) cannot be satisfied except, asymptotically, if we send  $\mu^2/a$  to zero. In other words, as discussed in [33], the first order phase transition line now extends over the whole negative real axis terminating at the origin. However, before jumping too quickly to this conclusion we should observe that the potential becomes very flat for small  $\mu^2/a$ , so much that it develops a flat direction at  $O(\mu^2/a)$ . This continuous vacuum degeneracy is lifted at  $O((\mu^2/a)^2)$  so that the  $CP$  violating minimum



**Figure 5.** Same as figure 4 in the  $CP$  broken situation, but for two values of  $\theta$  on opposite sides of  $\theta = \pi$ : (a):  $\theta < \pi$ , (b):  $\theta > \pi$ . The true minimum (corresponding to the intersection which is farther away from the middle one) swaps abruptly as one goes through  $\theta = \pi$ .

is found to lie  $O((\mu^2/a)^2)$  below the  $CP$  conserving one. The existence of this quasi-flat direction and its lifting to  $O(m^2)$  was first pointed out in [27] and further discussed in [33]. In general,  $O(m^2)$  corrections are not included in effective Lagrangians like (2.1) but, in the context of our double limit  $m/\Lambda \rightarrow 0, N \rightarrow \infty$  with  $mN/\Lambda$  fixed (recall  $a \sim \Lambda^2/N$ ), the split in the potential between the two vacua is of order  $\Lambda^4(mN/\Lambda)^2$  while the  $O(m^2)$  corrections we are ignoring are at least a factor  $1/N$  lower. We can thus conclude that, above a sufficiently large  $N$ ,  $CP$  is broken for two equal mass flavors.<sup>10</sup>

### 3.3 $N_f \geq 3$

For a generic number of flavors we have to solve eqs. (2.10). It can be immediately seen that for  $\theta = \pi$  we have the following solution that generalizes to  $N_f$  flavors what we found for two flavors, namely<sup>11</sup>

$$\phi_1 = \pi \ ; \ \phi_2 = \phi_3 = \dots = \phi_{N_f} = 0 \ ; \ \mu_1^2 \leq \mu_i^2 \ \text{for } i \neq 1. \quad (3.16)$$

It can be immediately checked that the determinant in eq. (2.21) is positive if the condition

$$\Delta \equiv \frac{1}{\mu_1^2} - \frac{1}{a} - \sum_{i=2}^{N_f} \frac{1}{\mu_i^2} > 0 \quad (3.17)$$

<sup>10</sup>We thank Z. Komargodski for having raised with us the issue of flat directions and for useful correspondence about it.

<sup>11</sup>We can find many other stationary solution that preserve  $CP$  by choosing an arbitrary number of  $\phi_i$  to be  $\pm\pi$  with their sum adding up to  $\theta = \pi$ . However, it is trivial to show that the solution in eq. (3.16) is, among those, the one with the lowest energy and thus the one to be compared with other (in general  $CP$  breaking) solutions.



is satisfied. In the corresponding region of parameter space we have a  $CP$  conserving stable solution since  $\theta - \sum_{i=1}^{N_f} \phi_i = 0$ . On the surface where (3.17) is replaced by an equality, the topological susceptibility diverges, as follows from eq. (2.30), and there is a massless state, signalling a second order phase transition. In the region where, instead,  $\Delta < 0$ , the solution in eq. (3.16) ceases to be a minimum and we have to look for new solutions corresponding to minima where we will find that  $CP$  is spontaneously broken.

In terms of the dimensionless quantities

$$\epsilon_i \equiv \frac{\mu_i^2}{a}, \quad \rho_i \equiv \mu_1^2/\mu_i^2 = \epsilon_1/\epsilon_i, \quad (i = 2, \dots, N_f); \quad (0 \leq \rho_i \leq 1), \quad (3.18)$$

the criticality condition can be written as

$$\epsilon_1 = 1 - \Sigma; \quad \Sigma \equiv \sum_{i=2}^{N_f} \rho_i \quad (3.19)$$

and the zero-mass eigenvector is simply given by

$$\mathcal{V}(M = 0) \propto (1, -\rho_2, \dots, -\rho_{N_f}). \quad (3.20)$$

Clearly the above expression is consistent with decoupling when one of the  $\rho$ 's goes to zero.  $CP$  is broken (unbroken) when the l.h.s. of (3.19) is larger (smaller) than the r.h.s. It is always broken if  $\Sigma > 1$ . If instead we look at the equal mass case,  $\rho_i = 1$ , we see that  $\Delta < 0$  except in the case  $N_f = 1$  and  $\mu^2/a < 1$  and in the case  $N_f = 2$  and  $\mu = 0$  [33].

As before, in the generic mass case we have a line of first order transition in the complex  $\mu_1^2 e^{i\theta}$  plane ending on a second order point where one physical mass goes to zero. The position of the second order point resides at the intersection of the negative  $\mu_1^2$  line with the critical hyper surface and therefore depends on the other parameters (mass ratios,  $a$ ).

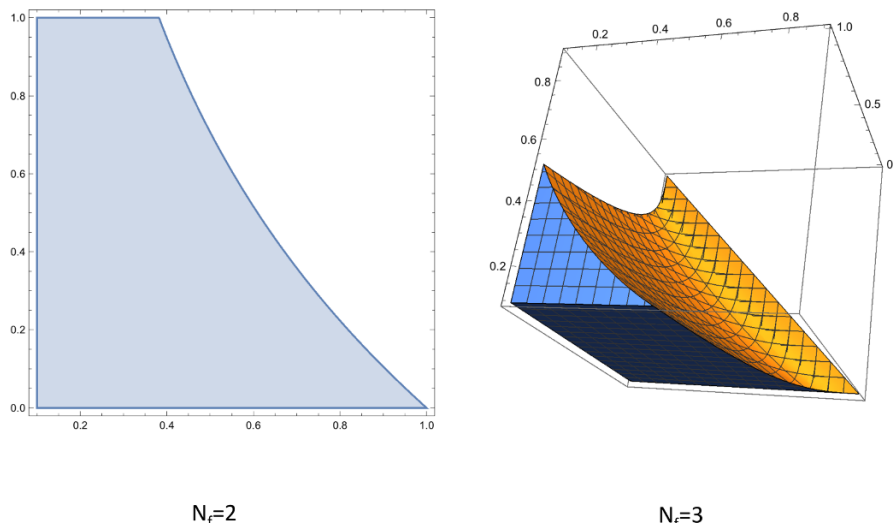
We end this section giving a definition of the critical hypersurface in terms of the quantity  $D \equiv \det(\mu^2/a^2) = \det(\epsilon)$ , where, however,  $\mu^2$  is now the matrix introduced in (2.1) after having absorbed the  $\theta$  angle by a chiral rotation.<sup>12</sup> The critical value of  $D$ ,  $D_c$ , is negative (corresponding to  $\bar{\theta} = -\arg D = \pm\pi$ ) and its absolute value depends only on the ratios  $\rho_i$  introduced earlier. Indeed the condition for  $CP$  violation can be expressed as follows

$$|D| > |D_c|; \quad |D_c|^{1/N_f} = (1 - \Sigma)\Pi^{-1/N_f}; \quad \Pi = \prod_{i=2}^{N_f} \rho_i. \quad (3.21)$$

It can be checked that also the above expression satisfies decoupling when one of the  $\rho_i$ 's goes to zero. Plots of the  $CP$  conserving regions and of the critical lines (surfaces) for  $N_f = 2$  ( $N_f = 3$ ) are shown in figures 6.

---

<sup>12</sup>We recall that only  $\bar{\theta} = \theta + \arg D$  is physically relevant. In the rest of the paper we adopted the convention of having  $\mu^2$  diagonal, real and positive and  $\theta$  arbitrary. For the rest of this section, instead,  $\theta = 0$  and  $\arg D$  is arbitrary.



**Figure 6.**  $CP$  conserving (filled) and  $CP$  breaking (empty) regions for  $N_f = 2$  and  $N_f = 3$ . The vertical axis is  $D$ , the horizontal is (are) the mass ratio(s).

#### 4 Spontaneous $CP$ violation and the axion potential

We shall now discuss some consequences of the considerations made in the previous sections when an extra dynamical low-energy degree of freedom, the axion, is added to those of chiral QCD. As pointed out independently by Weinberg [39] and Wilczek [40], the existence of an axion is a necessary consequence of the Peccei-Quinn (PQ) resolution [41, 42] of the strong- $CP$  problem. The latter consists in the observation that present bounds on the electric dipole moment of the neutron force the  $\theta$  angle (actually  $\bar{\theta}$ ) to be less than  $10^{-9}$  [17].

Of course, if one of the quarks is massless, the strong- $CP$  problem would be automatically solved since  $\theta$  could be rotated away (equivalently  $\bar{\theta} = 0$ ). Unfortunately, the low-energy spectrum of QCD is inconsistent with the data if one of the quark flavors is massless. A generic way to introduce the PQ resolution of the problem, and the axion, parallels the massless quark solution while avoiding its unwanted consequences. One assumes the existence a new axial  $U(1)$  global symmetry, only broken by the QCD anomaly (in QCD that symmetry would be the chiral rotation of the massless quark field). Then the existence of the axion follows from Goldstone's theorem associated with the spontaneous breaking of this symmetry. The axion is only a PGB because there is an no anomaly-free spontaneously broken exact symmetry. The only additional free-parameters with respect to QCD are the so-called axion decay constant  $F_\alpha$ , the analog of  $F_\pi$ , and  $\alpha_{PQ}$ , denoting the strength of the contribution of the new sector to the  $U_A(1)$  anomaly. Instead, the  $\theta$  parameter can be rotated away as we shall now discuss in detail.

##### 4.1 Including the axion in the QCD effective Lagrangian

In view of the above considerations, the axion can be easily incorporated in the QCD effective Lagrangian discussed in section 2 as if there were an extra zero-mass fermion, condensing at the scale  $F_\alpha$ , and contributing to the anomaly with a coefficient  $\alpha_{PQ}$  (relative

to the weight of a QCD fermion). This can be simply implemented by introducing, together with  $U$  and  $\Phi$ , similarly related axionic fields  $\alpha$  and  $N$

$$N = \frac{F_\alpha}{\sqrt{2}} e^{i\frac{\sqrt{2}}{F_\alpha}\alpha}. \quad (4.1)$$

The generalization of the Lagrangian (2.1) then reads<sup>13</sup>

$$L = \frac{1}{2} \text{Tr} \left( \partial_\mu U \partial^\mu U^\dagger \right) + \frac{1}{2} \partial_\mu N \partial^\mu N^\dagger + \frac{F_\pi}{2\sqrt{2}} \text{Tr} \left( \mu^2 (U + U^\dagger) \right) + \frac{Q^2}{2\chi_{\text{YM}}} + \frac{i}{2} Q \left[ \log U - \log U^\dagger + \alpha_{PQ} \left( \log N - \log N^\dagger \right) \right] - \theta Q. \quad (4.2)$$

Restricting, for the sake of simplicity, our analysis to the fields in the Cartan sub-algebra of the QCD pseudoscalar mesons, the previous Lagrangian becomes

$$L = \frac{1}{2} \sum_{i=1}^{N_f} \partial_\mu v_i \partial^\mu v_i + \frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \cos \left( -\phi_i + \frac{\sqrt{2}}{F_\pi} v_i \right) + \frac{Q^2}{2\chi_{\text{YM}}} + \frac{1}{2} (\partial_\mu \alpha)^2 - Q \left( \theta - \sum_{i=1}^{N_f} \phi_i - \beta + \frac{\sqrt{2}}{F_\pi} \sum_{i=1}^{N_f} v_i + \frac{\alpha_{PQ} \sqrt{2}}{F_\alpha} \sigma \right), \quad (4.3)$$

where again we have allowed for a non-trivial expectation  $\langle U \rangle$  as in eq. (2.8) and we have also introduced an expectation value for  $\alpha(x)$  and a shifted axion field  $\sigma$  as  $\alpha(x) = -\frac{\alpha_{PQ} \sqrt{2}}{F_\alpha} \beta + \sigma(x)$ .

Proceeding now as in section 2, we determine the phases  $\phi_i$  and  $\beta$  by minimizing

$$V(\phi_i, \beta) = -\frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \cos \phi_i + \frac{\chi_{\text{YM}}}{2} \left( \theta - \sum_{i=1}^{N_f} \phi_i - \beta \right)^2. \quad (4.4)$$

The stationary points of this potential are solutions of the equations

$$-\frac{F_\pi^2}{2} \mu_i^2 \sin \phi_i + \chi_{\text{YM}} \left( \theta - \sum_i \phi_i - \beta \right) = 0 \quad ; \quad i = 1, 2, \dots, N_f$$

$$\theta - \sum_i \phi_i - \beta = 0, \quad (4.5)$$

and are given by

$$\hat{\phi}_i = 0 \quad \text{mod}(\pi) \quad ; \quad \hat{\beta} = \theta - \sum_{i=1}^{N_f} \phi_i. \quad (4.6)$$

We notice that the choice

$$\hat{\phi}_i = 0 \quad ; \quad i = 1, 2, \dots, N_f \quad ; \quad \hat{\beta} = \theta \quad (4.7)$$

---

<sup>13</sup>See ref. [26].

corresponds to the minimum of the potential, while the other choices correspond to maxima or to saddle points. Setting the expectation values to (4.6), eq. (4.3) takes the form

$$L = -V(\hat{\phi}_i, \hat{\beta}) + \frac{1}{2} \sum_{i=1}^{N_f} \partial_\mu v_i \partial^\mu v_i + \frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \cos\left(\frac{\sqrt{2}}{F_\pi} v_i\right) + \frac{Q^2}{2\chi_{\text{YM}}} + \frac{1}{2} (\partial_\mu \sigma)^2 - Q \left( \frac{\sqrt{2}}{F_\pi} \sum_{i=1}^{N_f} v_i + \frac{\alpha_{PQ} \sqrt{2}}{F_\alpha} \sigma \right), \quad (4.8)$$

where  $V(\hat{\phi}_i, \hat{\beta})$  is a constant. Thus unlike the QCD case, physics has become  $\theta$ -independent and  $CP$  conserving. As we shall see in the following subsection, the full richness of the QCD case reappears once we consider the axion potential.

For the moment, in analogy with eq. (2.14), we rewrite (4.8) in the form

$$L = -V(\hat{\phi}_i, \hat{\beta}) + \frac{1}{2} \sum_{i=1}^{N_f} \partial_\mu v_i \partial^\mu v_i + \frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \left( \cos\left(\frac{\sqrt{2}}{F_\pi} v_i\right) - 1 \right) + \frac{1}{2} (\partial_\mu \sigma)^2 - \frac{\chi_{\text{YM}}}{2} \left( \frac{\sqrt{2}}{F_\pi} \sum_{i=1}^{N_f} v_i + \frac{\alpha_{PQ} \sqrt{2}}{F_\alpha} \sigma \right)^2 + \frac{1}{2\chi_{\text{YM}}} \left( Q - \chi_{\text{YM}} \left( \frac{\sqrt{2}}{F_\pi} \sum_{i=1}^{N_f} v_i + \frac{\alpha_{PQ} \sqrt{2}}{F_\alpha} \sigma \right) \right)^2. \quad (4.9)$$

The mass spectrum of the system can be found by diagonalizing the quadratic part of eq. (4.9) which reads

$$L_2 = \frac{1}{2} \sum_{i=1}^{N_f} \partial_\mu v_i \partial^\mu v_i - \frac{1}{2} \sum_{i=1}^{N_f} \mu_i^2 v_i^2 - \frac{\chi_{\text{YM}}}{2} \left( \frac{\sqrt{2}}{F_\pi} \sum_{i=1}^{N_f} v_i + \frac{\sqrt{2} \alpha_{PQ}}{F_\alpha} \sigma \right)^2 + \frac{1}{2} (\partial_\mu \sigma)^2 = \frac{1}{2} \sum_{a=1}^{N_f+1} \partial_\mu H_a \partial^\mu H_a - \frac{1}{2} H^T A H, \quad (4.10)$$

where  $H$  is an  $N_f + 1$ -column vector and  $A$  is the squared-mass matrix

$$H = \begin{pmatrix} \sigma \\ v_1 \\ v_2 \\ \vdots \\ v_{N_f} \end{pmatrix}; \quad A = \begin{pmatrix} b^2 a & ba & ba & ba & \dots & ba \\ ba & \mu_1^2 + a & a & a & \dots & a \\ ba & a & \mu_2^2 + a & a & \dots & a \\ \dots & \dots & \dots & \dots & \dots & \dots \\ ba & a & a & a & \dots & \mu_{N_f}^2 + a \end{pmatrix}. \quad (4.11)$$

The mass spectrum is the result of the diagonalization of  $A$  and can be read off from

$$\begin{aligned} \det(p^2\delta_{ij} - A_{ij}) &= p^2 \prod_{i=1}^{N_f} (p^2 - \mu_i^2) \left[ 1 - a \left( \sum_{i=1}^{N_f} \frac{1}{p^2 - \mu_i^2} + \frac{b^2}{p^2} \right) \right] \\ &= \prod_{i=1}^{N_f+1} (p^2 - M_i^2), \end{aligned} \quad (4.12)$$

where  $a = \frac{2\chi_{\text{YM}}}{F_\pi^2}$  (as in eq. (2.11)) and  $b = \frac{F_\pi \alpha_{PQ}}{F_\alpha}$ . The  $M_i$  are the masses of the physical states that diagonalize the mass matrix. By going to  $p^2 = 0$ , eq. (4.12) implies

$$\det A = ab^2 \prod_{i=1}^{N_f} \mu_i^2 = \prod_{j=1}^{N_f+1} M_j^2, \quad (4.13)$$

where the product on the r.h.s. includes the axion as well as the Cartan PNGB masses. Note that, unlike the non-axionic case, for non-vanishing  $m_i$ ,  $a$  and  $b$ , this determinant is always positive implying no massless state (and indeed a non-tachyonic spectrum). This would have also been the case had we considered QCD with one massless flavor (in that case  $b = 1$ ). In particular, for small  $b$ , the mass of the axion is given by looking for a zero at small  $p^2$  of the term in square brackets in eq. (4.12). Neglecting  $p^2$  with respect to  $\mu_i^2$  one obtains

$$M_{\text{axion}}^2 = \frac{b^2}{\frac{1}{a} + \sum_{i=1}^{N_f} \frac{1}{\mu_i^2}}. \quad (4.14)$$

This reduces to the usual expression for the axion mass [39, 43] in the limit  $a, \mu_s^2 \gg \mu_{u,d}^2$ . Alternatively, using eq. (2.30) and the definition of  $b$ , we can write

$$M_{\text{axion}}^2 = \frac{2\alpha_{PQ}^2}{F_\alpha^2} \chi_{\text{QCD}}, \quad (4.15)$$

another formula often used in the literature (see e.g. ref. [44]).

Finally, from the term in the last line of eq. (4.9) and the matrix definition in eq. (4.10) we get (having  $\langle Q \rangle = 0$ ) the following two-point correlation function

$$\langle Q(x)Q(y) \rangle^{\text{F.T.}} = i\chi_{\text{YM}} \frac{p^2 \prod_{i=1}^{N_f} (p^2 - \mu_i^2)}{\prod_{i=1}^{N_f+1} (p^2 - M_i^2)} = \frac{i\chi_{\text{YM}}}{\left[ 1 - a \left( \sum_{i=1}^{N_f} \frac{1}{p^2 - \mu_i^2} + \frac{b^2}{p^2} \right) \right]}, \quad (4.16)$$

that vanishes at  $p^2 = 0$  signalling that the topological susceptibility in a theory where QCD is “augmented” by another sector that includes the axion, is zero consistently with the fact that the dependence on the  $\theta$  parameter disappears.

For the physically interesting case we have to take  $b \ll 1$  so that the spectrum should contain a very light pseudo-scalar, the physical axion, which is the original field  $\sigma$  up to an  $O(b)$  admixture of PNGBs. This is all well known. We will now discuss how things take an interesting turn when we go from properties of the spectrum (i.e. of small fluctuations around the minimum of  $V$ ) to those of the full potential at a finite distance from its minimum.

## 4.2 The axion potential

From eq. (4.9) we can immediately read the axion-PNGB potential

$$V(v_i, \sigma) = -\frac{F_\pi^2}{2} \sum_{i=1}^{N_f} \mu_i^2 \cos\left(\frac{\sqrt{2}}{F_\pi} v_i\right) + \frac{a}{2} \left(\sum_{i=1}^{N_f} v_i + b\sigma\right)^2. \quad (4.17)$$

In the literature one introduces the concept of an axion potential after integrating out the remaining  $N_f$  degrees of freedom in the assumption that they are much heavier than the axion. In principle this requires diagonalizing the mass matrix so as to be in position of identifying the lowest lying state, the physical axion that will be a mixture of  $\sigma$  and the  $v_i$ . In the limit of very small  $b$ , which is where physics lies, one can neglect these mixings and identify  $\sigma$  with the axion modulo some exceptional cases to be discussed below.

For the physically interesting case of two light flavors the axion potential was first derived in [19] under the assumption  $\mu_1^2, \mu_2^2 \ll a$  with the result [34]

$$V_{\text{axion}}(\sigma) = -\frac{F_\pi^2}{2} \sqrt{(\mu_1^2 + \mu_2^2)^2 - 4\mu_1^2\mu_2^2 \sin^2\left(\frac{\alpha_{PQ}\sigma}{\sqrt{2}F_\alpha}\right)} + \mathcal{O}(\mu_i^2/a), \quad (4.18)$$

which for  $N_f = 1$  simply becomes

$$V_{\text{axion}}(\sigma) = -\frac{F_\pi^2}{2} \mu^2 \cos\left(\frac{\sqrt{2}\alpha_{PQ}\sigma}{F_\alpha}\right) + \mathcal{O}(\mu^2/a). \quad (4.19)$$

We see, however, that by having considered the axion potential at a generic value of  $\sigma$  we have effectively recovered, *mutatis mutandis*, the situation discussed in QCD at fixed  $\theta$ . This is why the discussion of section 3 becomes very relevant here. Indeed, the previous analysis shows that, precisely around  $\sigma = \frac{\pi F_\alpha}{\sqrt{2}\alpha_{PQ}}$ , some PNGB mass can become arbitrarily small. In this case integrating out the PNGB fields is no longer justified and a more careful analysis is needed. In other cases the naive solution for the  $v_i$  corresponds to a maximum and it has to be replaced with the right solution. The rest of this section is devoted to such an analysis for different numbers of quark flavors.

In the following, for simplicity of notation, we shall denote by  $\varphi_i$  and  $\zeta$  the dimensionless quantities  $-\frac{\sqrt{2}}{F_\pi} v_i$  and  $\frac{\sqrt{2}\alpha_{PQ}}{F_\alpha} \sigma$ , respectively. In this notation the potential (4.17) simply reads

$$2F_\pi^{-2} V(\zeta, \varphi_i) = -\sum_{i=1}^{N_f} \mu_i^2 \cos \varphi_i + \frac{a}{2} \left(\sum_{i=1}^{N_f} \varphi_i - \zeta\right)^2. \quad (4.20)$$

### 4.2.1 $N_f = 1$

The potential  $V(\zeta, \varphi)$  has two distinct stationary points, one at  $\zeta = \varphi = 0$  and one at  $\zeta = \varphi = \pi$ . The first is a true minimum, the second a saddle point. Let us now consider the stationary points in  $\varphi$  at fixed  $\zeta$  in order to compute  $V_{\text{axion}}(\zeta)$ , distinguishing three cases (looking at figure 1 can help following the discussion).

- $\mu^2/a < 1$ . In this case there is a single stationary point at  $\hat{\varphi}(\zeta) \leq \zeta$  which grows monotonically with  $\zeta$  interpolating between the two stationary points of  $V$ . In this case the potential (4.19) is easily recovered. At  $\zeta = \pi$  the potential is smooth and reaches a maximum lying  $\mu^2 F_\pi^2$  above the absolute minimum. One can easily check that, for  $\mu^2/a$  not too close to 1, the mass of the PNGB is always much larger than the scale of variation of the axion potential so that integrating out that degree of freedom is justified. We shall discuss separately the case  $|1 - \mu^2/a| \ll 1$ .
- $\mu^2/a > 1$ . In this case, as one varies  $\zeta$  from 0 to  $\pi$ ,  $\hat{\varphi}(\zeta)$  remains always smaller than  $\zeta$ . Actually, above a value of  $\zeta$  that depends on  $\mu^2/a$ , new stationary points in  $\varphi$  (lying above  $\varphi = \pi$ ) appear but they have higher energy. This is nothing but the situation we have described and discussed around figure 3. In particular, as we approach  $\zeta = \pi$ ,  $\hat{\varphi}$  approaches a finite value smaller than  $\pi$  and behaving as  $\pi a/\mu^2$  for  $\mu^2/a \gg 1$ . Precisely at  $\zeta = \pi$  this minimum becomes degenerate with one at  $\varphi > \pi$  which, upon a shift by  $2\pi$  is just its  $CP$  transformed. Again, for  $\mu^2/a$  not too close to 1, integrating out the PNGB appears fully justified but, instead of (4.19), we get

$$V_{\text{axion}}(\sigma) = \frac{1}{2} \chi_{\text{YM}} \left( \frac{\sqrt{2} \alpha_{PQ} \sigma}{F_\alpha} \right)^2 + O(a/\mu^2), \quad (4.21)$$

where for a moment we have reintroduced the canonical  $\sigma$  field. In particular, the axion mass is now controlled by  $a$  rather than by  $\mu^2$ . At the boundary of its periodicity interval  $V_{\text{axion}}$  now reaches its maximal value  $\frac{1}{2} \chi_{\text{YM}} \pi^2 \ll \mu^2 F_\pi^2$  (in the small- $a$  limit). Furthermore, at that point its first derivative is non-vanishing (and positive) and, since the potential is periodic, its first derivative will be discontinuous, giving a spike at  $\zeta = \pi$ . This, of course, is related to the fact that the solution for  $\hat{\varphi}$  jumps abruptly as we go through  $\theta = \pi$  (see again figure 3).

- $|1 - \mu^2/a| \ll 1$ . This third regime is perhaps the most interesting one, at least theoretically. Let us consider the mass matrix (better the matrix of second derivatives) around  $\zeta = \varphi = \pi$ . It takes the form

$$A = \begin{pmatrix} b^2 a & ba \\ ba & -\mu^2 + a \end{pmatrix}. \quad (4.22)$$

We see that, if  $|\mu^2 - a| = O(ba)$ , the off-diagonal entries become of the same order as the difference between the two diagonal ones (remember that  $b \ll 1$ ). This is precisely the situation in which the two eigenvectors are strongly mixed w.r.t. the original (axion-PNGB) basis. Indeed the maximal mixing occurs at  $\mu^2 = a(1 - b^2)$  since then the matrix  $A$  becomes

$$A = \begin{pmatrix} b^2 a & ba \\ ba & b^2 a \end{pmatrix}, \quad (4.23)$$

whose eigenvectors are  $(1, \pm 1)$ , with eigenvalues  $b^2 a \pm ba$ . In fact, as we go through the point  $\mu^2 = a$ , the two eigenvectors evolve very quickly (i.e. as  $\mu^2$  goes from  $a - O(ab)$

to  $a + O(ab)$ ) from almost pure axion to almost pure PNGB or vice versa. This is clearly shown by the numerical calculation presented in figure 7. Since  $\det A < 0$  the spectrum always consists of a normal and a tachyonic state, but the latter is mainly in the PNGB direction at large  $\mu^2$  while it becomes mainly axion-like at small  $\mu^2$ . That means that, had we started the evolution of the PNGB plus axion system at  $\zeta = \varphi = \pi$  the evolution would go immediately towards smaller  $\zeta$ 's if  $\mu^2 < a$  while, for  $\mu^2 > a$ , it would first roll down to the true minimum in  $\varphi$  and only then will roll down towards  $\zeta = 0, \varphi = 0$ .

It is also quite clear that in this particular range of  $\mu^2/a$  and  $\zeta$  it is not possible to describe the system only in terms of a  $V_{\text{axion}}(\zeta)$  since the other degree of freedom is as light as the axion itself. Only a description in terms of a  $V(\zeta, \varphi)$  is fully adequate.

### 4.2.2 $N_f \geq 2$ and discussion

The real world has two very light quarks,  $u$  and  $d$ , a light one,  $s$ , and three heavy quarks. The latter play no role in our discussion. Thus the case of physical interest is  $N_f = 2$  or 3. Also, at zero temperature, the quantitative solution of the U(1) problem requires [9, 10]  $\mu_u^2 < \mu_d^2 \ll \mu_s^2 < a$ . The ratios  $\mu_u^2 : \mu_d^2 : \mu_s^2 : a$  are about 1 : 2 : 40 : 18. In what follows we shall use these numbers together with the results we obtained from the large- $N$  effective action approach, even though in the real world  $N = 3$ . The success of the large- $N$  solution to the U(1) problem suggests that, at least in this sector, the large- $N$  expansion converges quite fast.

We should keep in mind, however, that, while quark mass ratios are expected to be constant below the QCD deconfining temperature (they depend on phenomena occurring at the electroweak-breaking scale), the temperature dependence of  $\chi_{\text{YM}}$  could possibly differ from that of the quark condensate meaning a possible (strong?)  $T$ -dependence of  $\mu^2/a$ . An increase of that ratio by an order of magnitude would bring us inside the  $CP$  broken region. The available lattice measurements [45–47] do not seem to favor this possibility. We defer further comments on this issue to the conclusion section.

In the following we will consider therefore the case of two or three quark flavors of different masses and allow for arbitrary ratios  $\mu_i^2/a$ . The situation is now more involved than in the  $N_f = 1$  case, but qualitatively similar. The stationary points of the potential (4.20) are

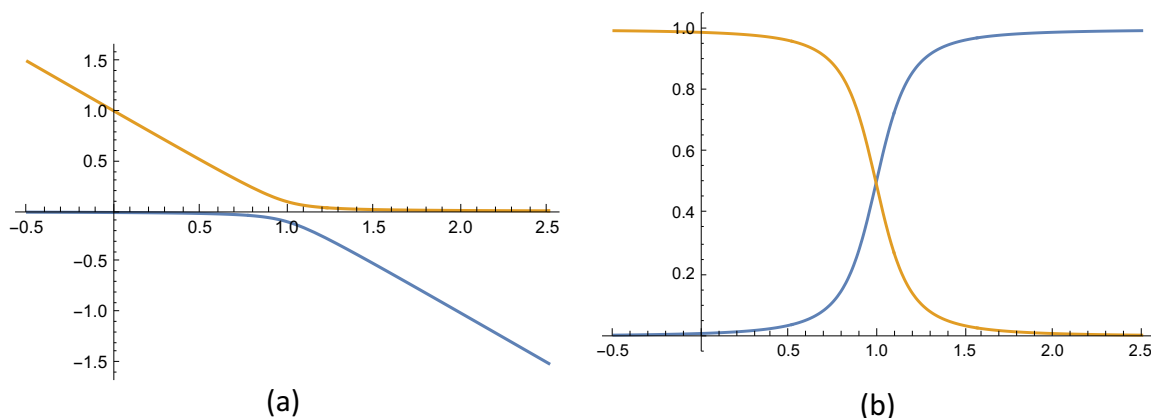
$$\zeta = 0, \pi \pmod{2\pi} \ ; \ \varphi_i = 0, \pi \pmod{2\pi} \ ; \ \sum \varphi_i = \zeta. \quad (4.24)$$

The absolute minimum is as usual the trivial one  $\zeta = \varphi_i = 0$ . In general it is legitimate to integrate out the PNGB degrees of freedom by minimizing their potential at fixed  $\zeta$  and then insert the solution  $\hat{\varphi}_i(\zeta)$  in  $V(\zeta, \varphi_i)$ . If  $\mu_i^2 \ll a$  this can be easily done. In the two-flavor case this gives the result (4.18). In the three-flavor case recalling that

$$\sin \phi_s = \mu_u^2 / \mu_s^2 \sin \phi_u \ll \sin \phi_u, \quad (4.25)$$

we see that the result (4.18) still holds up to corrections  $O(\mu_{u,d}^2/\mu_s^2)$ . This is indeed the result used in the literature.





**Figure 7.**  $N_f = 1$ . (a) Evolution of the two eigenvalues of (4.22) for  $b = 0.1$  as one varies  $\mu^2/a$ . The lower eigenvalue is tachyonic. (b) Projections of the two corresponding eigenvectors along the PNCB direction. Maximal mixing occurs in the vicinity of the critical point  $\mu^2/a = 1$ .

What happens if, for some physical reason,  $\chi_{\text{YM}}$  drops so fast with  $T$  that  $a$  becomes of order  $\mu_{u,d}^2$  or even smaller? We can understand the situation by considering what happens at the saddle point corresponding to

$$\zeta = \varphi_u = \pi, \quad \varphi_d = \varphi_s = 0. \tag{4.26}$$

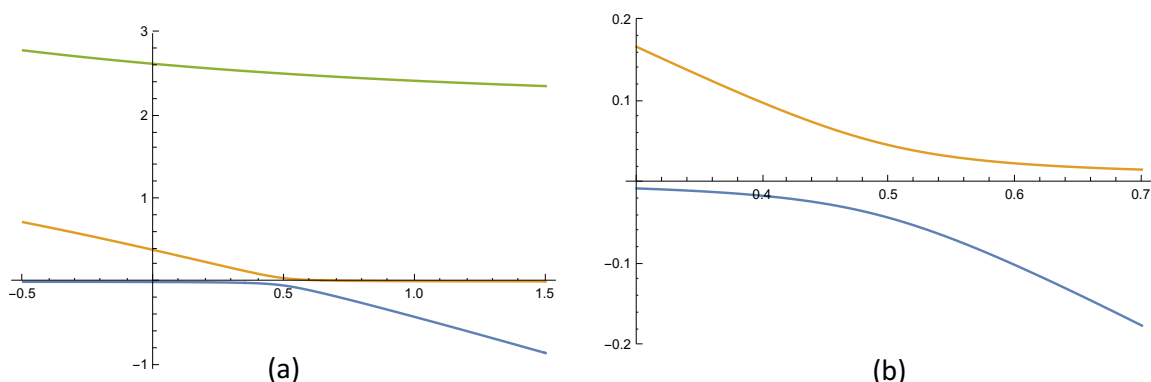
We have seen in sections 3.2 and 3.3 that the condition for having a massless boson (in the absence of the axion) is

$$\frac{1}{\mu_u^2} = \frac{1}{a} + \frac{1}{\mu_d^2} + \frac{1}{\mu_s^2} \sim \frac{1}{a} + \frac{1}{\mu_d^2} \Rightarrow a(\mu_d^2 - \mu_u^2) = \mu_u^2 \mu_d^2. \tag{4.27}$$

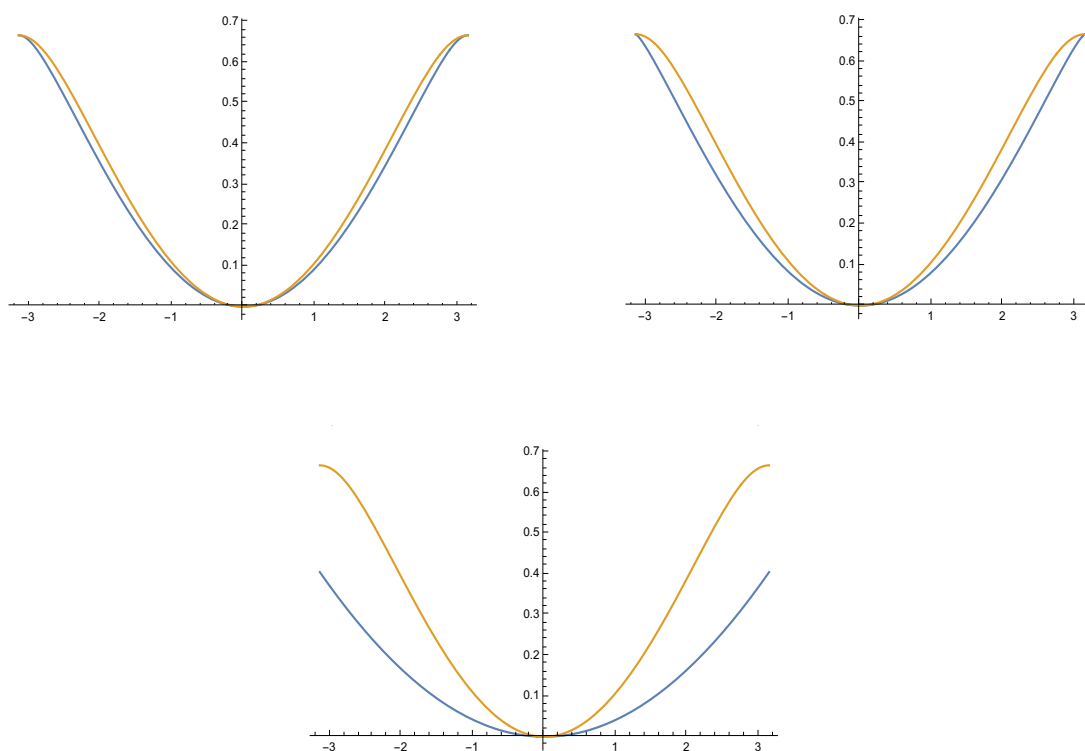
Precisely around this point we expect a large mixing to occur between the would-be massless PNCB and the axion and, as one goes through that region, we expect the tachyonic boson to change its dominant component from axionic to mesonic.

This is indeed fully supported by the numerical results shown in figures 7 and 8 for  $N_f = 1$  and  $N_f = 2$ , respectively. We have solved, using Mathematica, the minimization conditions at fixed  $\zeta$  and reconstructed this way the axion potential (see figure 9). We then clearly see that, while at small  $\mu_{u,d}^2/a$  the potential has a regular maximum around  $\zeta = \pi$  which coincides with the one of (4.18) and agrees well with it elsewhere, as we increase  $\mu_{u,d}^2/a$  above the critical value  $1 - \mu_u^2/\mu_d^2$  (see eq. (4.27)), the potential is lower than the one given by (4.18) even at  $\zeta = \pi$  and, by periodicity must develop a spike at that point. As we finally go much beyond the critical point, the true potential has nothing to do with the conventional one.

As in the  $N_f = 1$  case also here, the description of physics in terms of a single axion field is no longer appropriate when we are the vicinity of the condition (4.27). In that case only one “heavy” field can be integrated out and a description in terms of two light fields is more appropriate.



**Figure 8.**  $N_f = 2$ . (a) Evolution of the three eigenvalues as one varies  $\mu^2/a$  for  $b = 0.1$  and  $\mu_2^2 = 2\mu_1^2$ . One of the three eigenvalue always lies much higher than the other two and is not much affected by the axion. (b) Blow up of the lower part of the figure showing the repulsion (and mixing) of the two lower eigenvalues.



**Figure 9.** Comparing the conventional axion potential (yellow curves) with the “exact” one (blue curves) for  $N_f = 2$ ,  $\mu_d^2 = 2\mu_u^2$  and at three values of  $\mu_u^2/a$ : 0.25, 0.5 (critical value), 2.5. In the first two cases the two potentials (but not necessarily their derivatives) agree at  $\zeta = \pm\pi$  while in the third (overcritical) case even the values of the potentials disagree at the boundary of the periodicity interval.

## 5 Conclusions

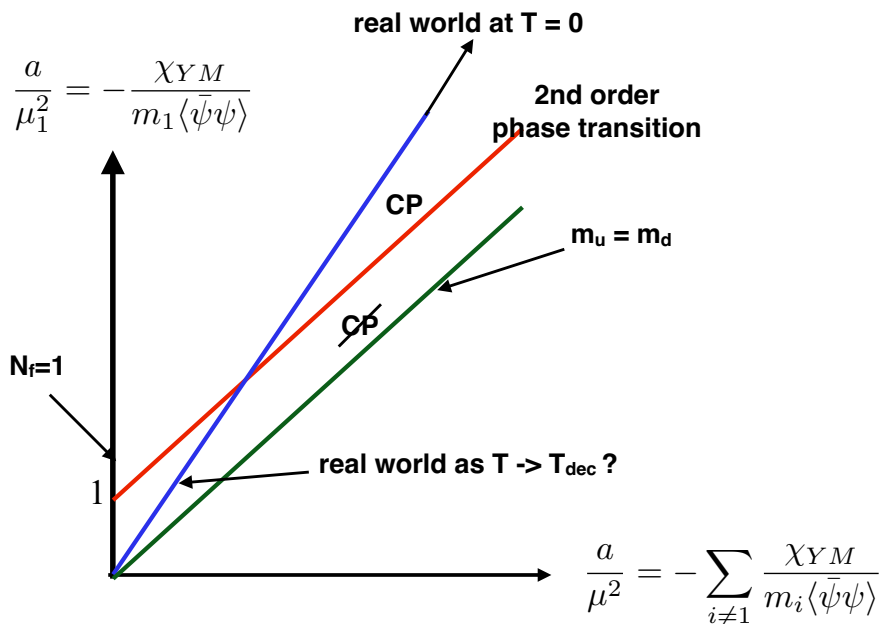
The phase structure of QCD associated with spontaneous  $CP$  breaking at  $\theta = \pi$  may, potentially (depending on parameters like quark masses and topological susceptibility, their ratios and temperature dependence), have important implications on the axion potential and its cosmological “phenomenology”.

In the present work we employed the effective chiral Lagrangian approach to investigate the inter-relation between spontaneous  $CP$  breaking in QCD at  $\theta = \pi$  and the axion potential near the boundary of its periodicity interval. Formally, the effective Lagrangian approach is applicable at low energies and, in particular, when all mass parameters (notably quark masses) are small with respect to the QCD scale,  $\Lambda$ . We also look at the large- $N$  limit in which we can have ratios of quark masses to  $\Lambda$  small but still much larger than  $1/N$ . This allows us to identify and reliably investigate the existence, at  $\theta = \pi$ , of a second order phase transition point on the hypersurface dividing the region in parameters space where  $CP$  is spontaneously broken from the one where it is not. The second order point is characterized by one of the PNCB mass going to zero and by the topological susceptibility (which can be seen as the order parameter) to diverge.

For generic masses the phase structure of QCD reveals a line of first order transitions, associated with spontaneous  $CP$  breaking at  $\theta = \pi$ , along the negative real axis in the complex  $\mu_1^2 e^{i\theta}$  mass plane ( $\mu_1$  being the lowest quark mass). The first order line extends all the way from  $-\infty$  to the second order point without reaching the chiral point at the origin. The position of the second order transition depends on all other parameters (mass ratios and the susceptibility related parameter we called  $a$ ). A similar phase structure is obtained by working in the complex quark-mass-determinant plane.

It is the existence of this second order point which has the most dramatic effect on the axion potential. Clearly, upon introducing the axionic field into the effective Lagrangian there is no more a  $\theta$  dependence and no strong- $CP$  breaking. However, precisely around the point in parameter space (quark masses and topological susceptibility) where, in the absence of the axion, the condition for having a zero mass boson is met, we find large mixing between the would be massless particle and the axion. In this region one cannot integrate out all the PNCB since one of them becomes very light with a mass of the order of the axion mass. Hence, in this region, the notion of an axionic potential which depends on just the axion field (obtained upon integration out all the PNCB) is not viable and should be replaced by a potential which depends on the two above mentioned light degrees of freedom as discussed in section 4. This potential is obtained upon integrating out all the other much heavier PNCBs.

Given the actual physical numerical values of the parameters (for  $N_f = 2$  and  $N_f = 3$ ) we see that, at zero temperature, we are not in the region of the parameter space where the concept of an axion potential and the derived result for the axion mass should be modified. However, if, as we raise the temperature while staying below the deconfinement transition (which for QCD is not a sharp transition), the corresponding YM topological susceptibility (and hence the parameter  $a$ ) drops faster with the temperature than the quark condensate so as to allow  $\mu^2/a$  to increase by about an order of magnitude, we will enter into this intriguing region (see figure 10).



**Figure 10.** Phase diagram for  $N_f = 2$ . The red line separates the phases with broken and unbroken  $CP$  at  $\theta = \pi$  and corresponds to a second-order transition with a massless particle. For  $m_d \gg m_u$  we recover the  $N_f = 1$  case represented by the vertical axis. Also shown is the  $m_u = m_d$  case lying entirely in the  $CP$  broken phase. The real world at  $T = 0$  is far up on the blue line (representing  $m_d \sim 2m_u$ ). As  $T$  is increased towards  $T_{\text{dec}}$  the real world will stay on the blue line (since  $m_d/m_u$  is  $T$ -independent) but may move down and cross the red line as indicated in the picture. Present lattice data seem to disfavour this possibility.

It seems, however, that lattice calculations (see e.g. [45–47] as well as [48, 49]) show a rather mild  $T$ -dependence of both  $\chi_{\text{YM}}$  and the quenched chiral condensate with a sharp drop (but not necessarily vanishing) of both above a similar value of  $T$ . There does not seem to be a clean window in which  $\mu^2/a$  increases by the above-mentioned order of magnitude. It would be desirable to have detailed lattice data on both  $\chi_{\text{YM}}$  and the planar chiral condensate by a single group using the same Monte Carlo configurations. It would be particularly interesting to study the pure number  $\chi_{\text{YM}}/\langle m\bar{\psi}\psi \rangle$  in the vicinity of the above-mentioned drop and also check its  $N$ -dependence (expected to be  $1/N$ ).

An obviously related issue is whether there is a critical temperature  $T_{\text{top}}$  above which  $\chi_{\text{YM}}$  vanishes, at least in the large- $N$  limit (dilute instantons [50–52], for instance, predict  $\chi_{\text{YM}} \sim e^{-cN}$ ) and, in that case, whether  $T_{\text{top}}$  can be higher than  $T_{\text{ch}}$ , the temperature above which chiral symmetry is restored. Under reasonable assumptions, claims that  $\chi_{\text{YM}}$  should vanish above  $T_{\text{ch}}$  were made in the past [53, 54] leaving open the possibility that  $\chi_{\text{YM}}$  goes to zero either together or before  $\langle \bar{\psi}\psi \rangle$  does it.

Although some old lattice calculations [55] appear to point in the opposite direction (and such a possibility has its own effective Lagrangian formulation [56, 57]), more recent simulations of the pure gauge theory [58, 59] suggest the existence of a similar (or even identical) value for the temperatures of deconfinement, chiral restoration and  $U_A(1)$  restoration. Above the transition temperature the dilute instanton gas approximation seems to set in.

Actually there is lattice evidence [60] that  $\chi_{\text{YM}}$  drops rather fast above  $T_c$  for large  $N$  (and even at  $N = 3$  a substantial decrease of  $\chi_{\text{YM}}$  is visible [61–63]) and may actually go to zero above it for  $N \rightarrow \infty$ . However, it is not clear what the ratio  $\langle \bar{\psi}\psi \rangle_{\text{planar}}/\chi_{\text{YM}}$  does around  $T_c$ . It would thus be very interesting to plan new lattice projects dedicated to the calculation of  $\chi_{\text{YM}}$  and  $\langle \bar{\psi}\psi \rangle$  in the planar limit across the phase transition.

Recently, using the mixed  $CP$ /Center discrete anomaly matching (together with some other plausible assumptions), it was shown [31] that in YM theory the  $CP$  symmetry is spontaneously broken at  $\theta = \pi$  and zero temperature and that the temperature  $T_{\text{res}}$  at which  $CP$  is restored is higher than the deconfinement temperature, i.e.  $T_{\text{res}} \geq T_{\text{dec}}$ . This result seems to be going in favor of the scenario advocated in [53, 54]. Breaking of  $CP$  in YM connects smoothly with  $CP$ -breaking in, say,  $N_f = 1$  QCD at  $\mu^2/a > 1$ . As we increase the temperature, if  $CP$  were restored before reaching  $T_{\text{dec}}$ , it would suggest that, in its QCD analog,  $\mu^2/a$  would go down till, at  $T_{\text{res}}$ , it reaches 1, which is precisely the opposite of what we were advocating, i.e. a ratio  $\mu^2/a$  increasing with temperature. Hence the statement  $T_{\text{res}} \geq T_{\text{dec}}$  is an (admittedly very mild) indication in favor of the scenario in which the finite temperature axion potential has to be revised in a certain range of temperature. Even if such a revision would be necessary, it remains to be seen whether it would make any substantial difference with respect to the standard calculations [44] (see also [64, 65]) of axionic dark matter abundance.

## Acknowledgments

We thank D. Gaiotto, Z. Komargodski and N. Seiberg for informing us of their work [33] prior to posting it. We also thank Z. Komargodski for useful comments on a preliminary version of this manuscript as well as M. D’Elia, L. Giusti and E. Vicari for discussions about lattice results on Yang-Mills and quenched QCD at finite temperature. S.Y. would like to thank O. Aharony and M. Peskin for discussions. G.V. wishes to acknowledge an illuminating discussion with M. Shifman. The work of S.Y. is supported in part by the I-CORE program of the Planning and Budgeting Committee (grant number 1937/12), the US-Israel Binational Science Foundation (BSF), the Israel-Germany Foundation (GIF) and the ISF Center of Excellence.

## A Ward-Takahashi identities

In this appendix we derive the WTIs for the anomalous  $U_A(1)$  currents in QCD and check that the two-point amplitudes derived from the effective Lagrangian in section 2 exactly satisfy them. We start from the anomaly equation in (2.4), but written for a single flavor

$$\partial_\mu J_{5i}^\mu = 2Q + 2m_i P_i \quad ; \quad J_{5i}^\mu = \bar{\psi}_i \gamma^\mu \gamma_5 \psi_i \quad ; \quad P_i = i\bar{\psi}_i \gamma_5 \psi_i. \quad (\text{A.1})$$

Inserting the previous anomaly equation in a two-point amplitudes with another operator  $O(y)$  we get

$$\partial_\mu \langle J_{5i}^\mu O(y) \rangle = \langle 2Q(x) O(y) \rangle + \delta(x^0 - y^0) \langle [J_{5i}^0, O(y)] \rangle + \langle 2m_i P_i(x) O(y) \rangle, \quad (\text{A.2})$$

that in Fourier space, after a partial integration, becomes

$$\begin{aligned} & \int d^4x e^{ipx} \langle 2Q(x)O(y) \rangle + \langle [Q_{5i}, O(y)] \rangle + \int d^4x e^{ipx} \langle 2m_i P_i(x)O(y) \rangle \\ &= -i \int d^4x e^{ipx} \langle p_\mu J_{5i}^\mu(x)O(y) \rangle \quad ; \quad i = 1, \dots, N_f, \end{aligned} \quad (\text{A.3})$$

where  $Q_{5i} = \int d^3x J_{5i}^0(x)$ . For  $O(y) = Q(y)$  the second term does not contribute and we get

$$\int d^4x e^{ipx} \langle 2Q(x)Q(y) \rangle + \int d^4x e^{ipx} \langle 2m_i P_i(x)Q(y) \rangle = -i \int d^4x e^{ipx} \langle p_\mu J_{5i}^\mu(x)Q(y) \rangle, \quad (\text{A.4})$$

while, for  $O(y) = 2m_j P_j(y)$ , the commutator gives  $[Q_{5i}, P_j] = -2i\bar{\psi}_i\psi_i\delta_{ij}$  and we get

$$\begin{aligned} & \int d^4x e^{ipx} \langle 2Q(x)2m_j P_j(y) \rangle + 2i\mu_i^2 F_\pi^2 + \int d^4x e^{ipx} \langle 2m_i P_i(x)2m_j P_j(y) \rangle \\ &= -i \int d^4x e^{ipx} \langle p_\mu J_{5i}^\mu(x)2m_j P_j(y) \rangle, \end{aligned} \quad (\text{A.5})$$

having made use of the Gell-Mann-Oakes-Renner relation  $-2\delta_{ij}m_i\langle\bar{\psi}_i\psi_i\rangle = \delta_{ij}\mu_i^2 F_\pi^2$ .

One checks that the following two-point amplitudes satisfy the previous anomalous WTIs and we get

$$\int d^4x e^{ipx} \langle Q(x)Q(y) \rangle = i \frac{aF_\pi^2}{2} \prod_{i=1}^{N_f} \frac{p^2 - \mu_i^2}{p^2 - M_i^2} = i \frac{aF_\pi^2}{2} \left[ 1 - a \sum_{i=1}^{N_f} \frac{1}{p^2 - \mu_i^2} \right]^{-1}, \quad (\text{A.6})$$

$$\int d^4x e^{ipx} \langle Q(x)2m_i P_i \rangle = i \frac{2\mu_i^2}{p^2 - \mu_i^2} \frac{aF_\pi^2}{2} \prod_{j=1}^{N_f} \frac{p^2 - \mu_j^2}{p^2 - M_j^2}, \quad (\text{A.7})$$

$$\int d^4x e^{ipx} \langle J_{5\mu}^{(i)}(x)Q(y) \rangle = -\frac{2p_\mu}{p^2 - \mu_i^2} \frac{aF_\pi^2}{2} \prod_{j=1}^{N_f} \frac{p^2 - \mu_j^2}{p^2 - M_j^2}, \quad (\text{A.8})$$

$$\int d^4x e^{ipx} \langle 2m_i P_i(x)2m_j P_j \rangle = i \frac{2F_\pi^2 \mu_i^4}{p^2 - \mu_i^2} \delta_{ij} + i \frac{4\mu_i^2 \mu_j^2}{(p^2 - \mu_i^2)(p^2 - \mu_j^2)} \frac{aF_\pi^2}{2} \prod_{k=1}^{N_f} \frac{p^2 - \mu_k^2}{p^2 - M_k^2}, \quad (\text{A.9})$$

$$\int d^4x e^{ipx} \langle J_{5\mu}^{(i)}(x)2m_j P_j \rangle = -\frac{2F_\pi^2 \mu_i^2 p_\mu}{p^2 - \mu_i^2} \delta_{ij} - \frac{4p_\mu \mu_j^2}{(p^2 - \mu_i^2)(p^2 - \mu_j^2)} \frac{aF_\pi^2}{2} \prod_{k=1}^{N_f} \frac{p^2 - \mu_k^2}{p^2 - M_k^2}. \quad (\text{A.10})$$

Furthermore, the poles at  $p^2 = \mu_i^2$  apparently present in (A.9) and (A.10) can be shown to be absent. The only poles present are at  $p^2 = M_i^2$  and correspond to the masses of the physical mesons. The previous two-point amplitudes reproduce those in section 2 with the identification

$$m_i P_i \implies \frac{F_\pi}{\sqrt{2}} \mu_i^2 v_i. \quad (\text{A.11})$$

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] R.F. Dashen, *Some features of chiral symmetry breaking*, *Phys. Rev. D* **3** (1971) 1879 [[INSPIRE](#)].
- [2] J. Nuyts, *Is CP-invariance violation caused by an SU(3) singlet?*, *Phys. Rev. Lett.* **26** (1971) 1604 [*Erratum ibid.* **27** (1971) 361] [[INSPIRE](#)].
- [3] M.A.B. Bég, *Question of spontaneous genesis of CP-violation*, *Phys. Rev. D* **4** (1971) 3810 [[INSPIRE](#)].
- [4] S. Weinberg, *Nonabelian Gauge Theories of the Strong Interactions*, *Phys. Rev. Lett.* **31** (1973) 494 [[INSPIRE](#)].
- [5] S. Weinberg, *Current algebra and gauge theories. 2. NonAbelian gluons*, *Phys. Rev. D* **8** (1973) 4482 [[INSPIRE](#)].
- [6] G. 't Hooft, *Symmetry Breaking Through Bell-Jackiw Anomalies*, *Phys. Rev. Lett.* **37** (1976) 8 [[INSPIRE](#)].
- [7] G. 't Hooft, *Computation of the Quantum Effects Due to a Four-Dimensional Pseudoparticle*, *Phys. Rev. D* **14** (1976) 3432 [*Erratum ibid.* **D 18** (1978) 2199] [[INSPIRE](#)].
- [8] R. Crewther, *Chiral properties of quantum chromodynamics*, *NATO Sci. Ser. B* **55** (1980) 529 [[INSPIRE](#)].
- [9] E. Witten, *Current Algebra Theorems for the U(1) Goldstone Boson*, *Nucl. Phys. B* **156** (1979) 269 [[INSPIRE](#)].
- [10] G. Veneziano, *U(1) Without Instantons*, *Nucl. Phys. B* **159** (1979) 213 [[INSPIRE](#)].
- [11] P. Di Vecchia, K. Fabricius, G.C. Rossi and G. Veneziano, *Preliminary Evidence for  $U_A(1)$  Breaking in QCD from Lattice Calculations*, *Nucl. Phys. B* **192** (1981) 392 [[INSPIRE](#)].
- [12] L. Giusti, G.C. Rossi, M. Testa and G. Veneziano, *The  $U_A(1)$  problem on the lattice with Ginsparg-Wilson fermions*, *Nucl. Phys. B* **628** (2002) 234 [[hep-lat/0108009](#)] [[INSPIRE](#)].
- [13] L. Del Debbio, L. Giusti and C. Pica, *Topological susceptibility in the SU(3) gauge theory*, *Phys. Rev. Lett.* **94** (2005) 032003 [[hep-th/0407052](#)] [[INSPIRE](#)].
- [14] M. Cé, C. Consonni, G.P. Engel and L. Giusti, *Non-Gaussianities in the topological charge distribution of the SU(3) Yang-Mills theory*, *Phys. Rev. D* **92** (2015) 074502 [[arXiv:1506.06052](#)] [[INSPIRE](#)].
- [15] P. Di Vecchia, *An Effective Lagrangian With No U(1) Problem in  $CP^{(n-1)}$  Models and QCD*, *Phys. Lett. B* **85** (1979) 357 [[INSPIRE](#)].
- [16] V. Baluni, *CP Violating Effects in QCD*, *Phys. Rev. D* **19** (1979) 2227 [[INSPIRE](#)].
- [17] R.J. Crewther, P. Di Vecchia, G. Veneziano and E. Witten, *Chiral Estimate of the Electric Dipole Moment of the Neutron in Quantum Chromodynamics*, *Phys. Lett. B* **88** (1979) 123 [*Erratum ibid.* **B 91** (1980) 487] [[INSPIRE](#)].
- [18] C. Rosenzweig, J. Schechter and C.G. Trahern, *Is the Effective Lagrangian for QCD a  $\sigma$ -model?*, *Phys. Rev. D* **21** (1980) 3388 [[INSPIRE](#)].
- [19] P. Di Vecchia and G. Veneziano, *Chiral Dynamics in the Large-N Limit*, *Nucl. Phys. B* **171** (1980) 253 [[INSPIRE](#)].
- [20] P. Nath and R.L. Arnowitt, *The U(1) Problem: Current Algebra and the Theta Vacuum*, *Phys. Rev. D* **23** (1981) 473 [[INSPIRE](#)].

- [21] E. Witten, *Large- $N$  Chiral Dynamics*, *Annals Phys.* **128** (1980) 363 [INSPIRE].
- [22] K. Kawarabayashi and N. Ohta, *The Problem of  $\eta$  in the Large- $N$  Limit: Effective Lagrangian Approach*, *Nucl. Phys. B* **175** (1980) 477 [INSPIRE].
- [23] K. Kawarabayashi and N. Ohta, *On the Partial Conservation of the U(1) Current*, *Prog. Theor. Phys.* **66** (1981) 1789 [INSPIRE].
- [24] N. Ohta, *Vacuum Structure and Chiral Charge Quantization in the Large- $N$  Limit*, *Prog. Theor. Phys.* **66** (1981) 1408 [Erratum *ibid.* **67** (1982) 993] [INSPIRE].
- [25] P. Di Vecchia, *The Dynamics of the Pseudoscalar Mesons at Arbitrary  $\Theta$  in Large- $N$  Quantum Chromodynamics*, *Acta Phys. Austriaca Suppl.* **22** (1980) 341 [INSPIRE].
- [26] P. Di Vecchia and F. Sannino, *The Physics of the  $\theta$ -angle for Composite Extensions of the Standard Model*, *Eur. Phys. J. Plus* **129** (2014) 262 [arXiv:1310.0954] [INSPIRE].
- [27] A.V. Smilga, *QCD at theta similar to pi*, *Phys. Rev. D* **59** (1999) 114021 [hep-ph/9805214] [INSPIRE].
- [28] M. Creutz, *Spontaneous violation of CP symmetry in the strong interactions*, *Phys. Rev. Lett.* **92** (2004) 201601 [hep-lat/0312018] [INSPIRE].
- [29] M. Creutz, *One flavor QCD*, *Annals Phys.* **322** (2007) 1518 [hep-th/0609187] [INSPIRE].
- [30] T.J. Hollowood and T. Kingaby, *A comment on the  $\chi_y$  genus and supersymmetric quantum mechanics*, *Phys. Lett. B* **566** (2003) 258 [hep-th/0303018] [INSPIRE].
- [31] D. Gaiotto, A. Kapustin, Z. Komargodski and N. Seiberg, *Theta, Time Reversal and Temperature*, *JHEP* **05** (2017) 091 [arXiv:1703.00501] [INSPIRE].
- [32] N. Seiberg, *New phases of QCD3 and QCD4*, talk given at *Strings 2017*, Tel Aviv, Israel, 26–30 June 2017.
- [33] D. Gaiotto, Z. Komargodski and N. Seiberg, *Time-Reversal Breaking in QCD4, Walls and Dualities in 2+1 Dimensions*, arXiv:1708.06806 [INSPIRE].
- [34] G. Grilli di Cortona, E. Hardy, J. Pardo Vega and G. Villadoro, *The QCD axion, precisely*, *JHEP* **01** (2016) 034 [arXiv:1511.02867] [INSPIRE].
- [35] Z. Komargodski, A. Sharon, R. Thorngren and X. Zhou, *Comments on Abelian Higgs Models and Persistent Order*, arXiv:1705.04786 [INSPIRE].
- [36] APE collaboration, M. Crisafulli et al., *Quenched  $B_K$  parameter with the Wilson and Clover actions at  $\beta = 6.0$* , *Nucl. Phys. Proc. Suppl.* **42** (1995) 397 [hep-lat/9505020] [INSPIRE].
- [37] C.W. Bernard and M.F.L. Golterman, *Chiral perturbation theory for the quenched approximation of QCD*, *Phys. Rev. D* **46** (1992) 853 [hep-lat/9204007] [INSPIRE].
- [38] L. Giusti, *Exact chiral symmetry on the lattice: QCD applications*, *Nucl. Phys. Proc. Suppl.* **119** (2003) 149 [hep-lat/0211009] [INSPIRE].
- [39] S. Weinberg, *A New Light Boson?*, *Phys. Rev. Lett.* **40** (1978) 223 [INSPIRE].
- [40] F. Wilczek, *Problem of Strong  $p$  and  $t$  Invariance in the Presence of Instantons*, *Phys. Rev. Lett.* **40** (1978) 279 [INSPIRE].
- [41] R.D. Peccei and H.R. Quinn, *CP Conservation in the Presence of Instantons*, *Phys. Rev. Lett.* **38** (1977) 1440 [INSPIRE].



- [42] R.D. Peccei and H.R. Quinn, *Constraints Imposed by CP Conservation in the Presence of Instantons*, *Phys. Rev. D* **16** (1977) 1791 [INSPIRE].
- [43] W.A. Bardeen and S.H.H. Tye, *Current Algebra Applied to Properties of the Light Higgs Boson*, *Phys. Lett. B* **74** (1978) 229 [INSPIRE].
- [44] J. Preskill, M.B. Wise and F. Wilczek, *Cosmology of the Invisible Axion*, *Phys. Lett. B* **120** (1983) 127 [INSPIRE].
- [45] E. Berkowitz, M.I. Buchoff and E. Rinaldi, *Lattice QCD input for axion cosmology*, *Phys. Rev. D* **92** (2015) 034507 [arXiv:1505.07455] [INSPIRE].
- [46] S. Borsányi et al., *Axion cosmology, lattice QCD and the dilute instanton gas*, *Phys. Lett. B* **752** (2016) 175 [arXiv:1508.06917] [INSPIRE].
- [47] C. Bonati, M. D’Elia, H. Panagopoulos and E. Vicari, *Change of  $\theta$  Dependence in 4D SU(N) Gauge Theories Across the Deconfinement Transition*, *Phys. Rev. Lett.* **110** (2013) 252003 [arXiv:1301.7640] [INSPIRE].
- [48] P. Chen et al., *Toward the chiral limit of QCD: Quenched and dynamical domain wall fermions*, [hep-lat/9812011](#) [INSPIRE].
- [49] R.G. Edwards, U.M. Heller, J.E. Kiskis and R. Narayanan, *Chiral condensate in the deconfined phase of quenched gauge theories*, *Phys. Rev. D* **61** (2000) 074504 [[hep-lat/9910041](#)] [INSPIRE].
- [50] D.J. Gross, R.D. Pisarski and L.G. Yaffe, *QCD and Instantons at Finite Temperature*, *Rev. Mod. Phys.* **53** (1981) 43 [INSPIRE].
- [51] T. Schäfer and E.V. Shuryak, *Instantons in QCD*, *Rev. Mod. Phys.* **70** (1998) 323 [[hep-ph/9610451](#)] [INSPIRE].
- [52] A. Ringwald and F. Schrenpp, *Confronting instanton perturbation theory with QCD lattice results*, *Phys. Lett. B* **459** (1999) 249 [[hep-lat/9903039](#)] [INSPIRE].
- [53] G. Veneziano, *Goldstone Mechanism From Gluon Dynamics*, *Phys. Lett. B* **95** (1980) 90 [INSPIRE].
- [54] T.D. Cohen, *The High temperature phase of QCD and U(1)<sub>A</sub> symmetry*, *Phys. Rev. D* **54** (1996) R1867 [[hep-ph/9601216](#)] [INSPIRE].
- [55] A. Di Giacomo, E. Meggiolaro and H. Panagopoulos, *Behavior of the topological susceptibility across the deconfining phase transition*, *Phys. Lett. B* **277** (1992) 491 [INSPIRE].
- [56] E. Meggiolaro, *Is the U(1) axial symmetry restored above the SU(L) × SU(L) chiral transition?*, *Z. Phys. C* **62** (1994) 669 [INSPIRE].
- [57] E. Meggiolaro, *Saturation of QCD Ward identities above the SU(L) × SU(L) chiral transition*, *Z. Phys. C* **62** (1994) 679 [INSPIRE].
- [58] B. Alles, M. D’Elia and A. Di Giacomo, *Topological susceptibility at zero and finite T in SU(3) Yang-Mills theory*, *Nucl. Phys. B* **494** (1997) 281 [Erratum *ibid.* **B 679** (2004) 397] [[hep-lat/9605013](#)] [INSPIRE].
- [59] L. Del Debbio, H. Panagopoulos and E. Vicari, *Topological susceptibility of SU(N) gauge theories at finite temperature*, *JHEP* **09** (2004) 028 [[hep-th/0407068](#)] [INSPIRE].
- [60] B. Lucini, M. Teper and U. Wenger, *Topology of SU(N) gauge theories at  $T \simeq 0$  and  $T \simeq T_c$* , *Nucl. Phys. B* **715** (2005) 461 [[hep-lat/0401028](#)] [INSPIRE].

- [61] C. Gattringer, R. Hoffmann and S. Schaefer, *The Topological susceptibility of SU(3) gauge theory near  $T_c$* , *Phys. Lett. B* **535** (2002) 358 [[hep-lat/0203013](#)] [[INSPIRE](#)].
- [62] V.G. Bornyakov, E.M. Ilgenfritz, B.V. Martemyanov, V.K. Mitrjushkin and M. Müller-Preussker, *Topology across the finite temperature transition studied by overimproved cooling in gluodynamics and QCD*, *Phys. Rev. D* **87** (2013) 114508 [[arXiv:1304.0935](#)] [[INSPIRE](#)].
- [63] G.-Y. Xiong, J.-B. Zhang, Y. Chen, C. Liu, Y.-B. Liu and J.-P. Ma, *Topological susceptibility near  $T_c$  in SU(3) gauge theory*, *Phys. Lett. B* **752** (2016) 34 [[arXiv:1508.07704](#)] [[INSPIRE](#)].
- [64] L.F. Abbott and P. Sikivie, *A Cosmological Bound on the Invisible Axion*, *Phys. Lett. B* **120** (1983) 133 [[INSPIRE](#)].
- [65] M. Dine and W. Fischler, *The Not So Harmless Axion*, *Phys. Lett. B* **120** (1983) 137 [[INSPIRE](#)].