Published for SISSA by Springer

Received: *February 19, 2020*
Accepted: *March 23, 2020*
Published: *April 6, 2020*

UV sensitivity of the axion mass from instantons in partially broken gauge groups

Csaba Csáki,[a] **Maximilian Ruhdorfer**[b] **and Yuri Shirman**[c]

[a] *Laboratory for Elementary Particle Physics, Cornell University, Ithaca, NY 14853, U.S.A.*
[b] *Physik-Department, Technische Universität München, 85748 Garching, Germany*
[c] *Department of Physics and Astronomy, University of California, Irvine, CA 92697, U.S.A.*

E-mail: csaki@cornell.edu, max.ruhdorfer@tum.de, yshirman@uci.edu

Abstract: We examine the contribution of small instantons to the axion mass in various UV completions of QCD. We show that the reason behind the potential dominance of such contributions is the non-trivial embedding of QCD into the UV theory. The effects from instantons in the partially broken gauge group appear as "fractional instanton" corrections in the effective theory. These will exhibit unusual dependences on the various scales in the problem whenever the index of embedding is non-trivial. We present a full one-instanton calculation of the axion mass in the simplest product group models, carefully keeping track of numerical prefactors. Rather than using a 't Hooft operator approximation we directly evaluate the contributions to the vacuum bubble, automatically capturing the effects of closing up external fermion lines with Higgs loops. This approach is manifestly finite and removes the uncertainty associated with introducing a cutoff scale for the Higgs loops. We verify that the small instantons may dominate over the QCD contribution for very high breaking scales and at least three group factors.

Keywords: Anomalies in Field and String Theories, Beyond Standard Model, CP violation, Solitons Monopoles and Instantons

ArXiv ePrint: 1912.02197

https://doi.org/10.1007/JHEP04(2020)031

JHEP04(2020)031

Contents

JHEP04(2020)031

1 Introduction

The past few decades have seen axions become an ever more important ingredient of modern particle physics beyond the standard model (BSM). The QCD axion provides the most plausible solution to the strong CP problem [1–4], and at the same time is also a natural dark matter candidate. Besides the QCD axion, axion-like particles (ALPs) are also ubiquitous in string theory, and can be used for many different purposes in BSM model building. For a pedagogic introduction to the axion and the strong CP problem, see for example [5].

While the coupling of the QCD axion is set by the unknown large Peccei-Quinn (PQ) symmetry breaking scale f_a, its mass is surprisingly well predicted. Even though it is due to uncalculable strongly coupled QCD effects, chiral symmetry relates the uncalculable axion mass to the equally uncalculable pion mass, and one obtains the famous relation (see e.g. [6])

$$m_a^2 = \frac{m_u m_d}{(m_u + m_d)^2} \frac{m_\pi^2 f_\pi^2}{f_a^2} \tag{1.1}$$

This formula depends only on known IR quantities in addition to the axion decay constant f_a (which sets the coupling strength of the axion), and has been the basis of axion

physics ever since the first attempts to directly detect axions. Eq. (1.1) seems like a robust prediction: it is an IR effect where the QCD instantons are strongly coupled, and the expectation is that they will dominate over any additional UV instanton effect, which would be weakly coupled, and hence suppressed. Indeed one can easily check that for the simplest UV completions of QCD the effects of small instantons are strongly suppressed, as long as the theory remains weakly coupled. One possible way to enhance the contributions of small instantons is to change the running of coupling in the UV and make QCD or its UV completion strongly coupled again [7–9]. A particularly elegant realization is to embed QCD into a higher dimensional theory, and it was indeed shown in [10] that small instanton contributions are naturally enhanced in some 5D theories. This observation allowed [11] to construct a 5D model where the axion mass is raised by small instantons.

However, recently Agrawal and Howe (AH) [12, 13] presented the surprising result that for a particular type of UV completion based on product groups small instantons could provide the dominant contribution to the axion mass even if the UV theory remains weakly coupled (and hence fully calculable). This possibility opens up new regions of the parameter space on the m_a^2, f_a plane. Interesting models implementing the mechanism of [12] were proposed in [14–18], applications to models of inflation were studied in [19]. Other approaches to modify the axion mass within QCD were proposed in [20–23] while in [24–30] the axion mass is raised by coupling it to a new confining gauge group.

In this paper we re-examine the models presented in [12, 13] in order to identify the underlying dynamics responsible for an enhancement of small instanton contributions. We identify the non-trivial embedding of QCD into a high-energy gauge group G as the main source behind this enhancement. It is well-known that a spontaneous symmetry breaking can result in unusual matching relations [31, 32] of the form

$$\left(\frac{\Lambda_G}{M}\right)^{kb_G} = \left(\frac{\Lambda_{\rm QCD}}{M}\right)^{b_{\rm QCD}} \tag{1.2}$$

where the integer k is commonly referred to as the index of embedding [31], Λ_G and $\Lambda_{\rm QCD}$ (b_G and $b_{\rm QCD}$) are the dynamical scales (beta functions) of the high and low energy theories respectively, and M is the symmetry breaking scale. Such a scale matching relation implies that the ordinary 1-instanton solution of the low energy theory is identified with a k-instanton solution of the high energy theory [32]. To be more precise, there are certain small instanton configurations that live fully in the broken group, and do not have corresponding instantons in the low energy theory. We will show that the contributions of these configurations to the QCD axion mass scale as

$$\frac{m_k^2}{m_{\rm QCD}^2} \propto \frac{1}{(4\pi)^F}\left(\frac{\Lambda_{\rm QCD}}{v}\right)^F \left(\frac{M}{\Lambda_{\rm QCD}}\right)^{4-\frac{b_{\rm QCD}}{k}} \tag{1.3}$$

where $\Lambda_{\rm QCD}$ is the QCD scale, v is the Higgs VEV, F is the number of flavors and $b_{\rm QCD} = \frac{11}{3}N_c - \frac{2}{3}F$. While for $k = 1$ every factor in (1.3) is smaller than 1 leading to a strong suppression, we find that already for $k = 2$ small instanton contributions are enhanced by powers of $M/\Lambda_{\rm QCD}$ that may overwhelm the other suppression factors for sufficiently large M.

JHEP04(2020)031

The aim of this paper is twofold. First we want to explain how (1.3) is obtained, and the physics leading to it in terms of the effects of the instantons in the partially broken group. Our second aim is to present a detailed estimate of the actual contribution of these instantons to the axion mass. After accounting for all $\mathcal{O}(1)$ factors, including the perturbation of the classical instanton action in the presence of the Higgs VEV,[1] we will be able to identify specific models which successfully implement the Agrawal-Howe enhancement mechanism.

The paper is organized as follows: in section 2 we present a back-of-the-envelope estimate for the scaling of the various small instanton contributions in partially broken gauge theories, and explain why the case with the non-trivial embeddings of the low-energy instanton is the most interesting one. The actual instanton calculation is set up in section 3 where we show how to do the instanton calculation in a completely broken SU(N) theory. Note that in this section we show how to obtain a non-vanishing contribution in the presence of fermions without using the 't Hooft operator approximation, but rather directly performing the integral over the fermionic and bosonic modes, which automatically includes the effects of additional scalar loops closing up the external fermion lines in the 't Hooft operator. We apply these results to the product group theories in section 4 and there we show how much enhancement we can obtain for the axion mass in the various models. We conclude in section 5. We also present two appendices. Appendix A contains a detailed description of how to use the 't Hooft operator approximation and a comparison to the full calculation, while in appendix B we present the conversion from the Pauli-Villars regulator scheme to the commonly used $\overline{\text{MS}}$ scheme.

2 Small instantons in partially broken groups and index of embedding

Before diving into the details of the full instanton calculation we would like to present a back-of-the-envelope estimate for the magnitude of the instanton corrections for various UV completions of QCD. There are two novel aspects of the calculation of [12] both related to the fact that we are considering small instantons of size $\rho \ll \Lambda_{\text{QCD}}^{-1}$.

- At high energies the Higgs boson(s) become propagating particles allowing us to also consider the effects of closing up the fermion legs of the instanton vertex using Higgs loops (rather than Higgs VEV insertions as is usually done)
- There may be non-trivial embeddings of QCD into the UV theory where the small instantons of the UV theory correspond to "fractional instantons" of QCD.

Below we will be estimating the effects of small instantons using both the traditional Higgs VEV insertions as well as the novel loop diagrams. We will see that for the simplest embeddings of QCD into the UV gauge theory all such effects will be negligible. However we will explain that for the cases with non-trivial embeddings there could be an enhancement by some power of the ratio M/Λ_{QCD} which opens up the possibility of these contributions

[1] The importance of this perturbation was also pointed out by the authors of [15] who considered a similar setup.

JHEP04(2020)031

to dominate over the IR contributions of the ordinary QCD instantons. We will show that the examples of small instanton dominance presented in [12] fall in this category of non-trivial embeddings.

Let us assume that the high energy gauge group G is broken to the low energy group H (in phenomenological applications we will, of course, choose H to be $\mathrm{SU}(3)_{\mathrm{QCD}}$) at the scale M by the VEV of some heavy scalars. We will assume that the theory has F flavors of matter fields in the fundamental representation of G. In expectation of our results to the Standard Model we will choose F to be even. In addition, we will introduce gauge singlet scalars H coupled to the matter fields through Yukawa couplings y. These scalars will eventually be identified with the Higgs scalar(s) of the SM. Thus we will assume that in the low energy theory H has both a VEV and a mass of order v. Finally, we will assume that the Yukawa couplings of H are small. This leads us to consider the following hierarchy of scales

$$yv \ll \Lambda_{\mathrm{QCD}} \ll v \ll \Lambda \ll M\,, \tag{2.1}$$

where Λ and Λ_{QCD} denote RG invariant scales of high and low energy theories respectively. When the embedding of the low energy group into G is trivial the matching relation between these scales is given by

$$\left(\frac{\Lambda_{\mathrm{QCD}}}{M}\right)^{b_{\mathrm{QCD}}} = \left(\frac{\Lambda}{M}\right)^{b_G}\,. \tag{2.2}$$

Our choice of the hierarchy of scales leads to several important consequences. First, the contributions of the instantons in the broken group (i.e. instantons of size $\rho \lesssim 1/M$) to the effective action are completely calculable. Furthermore, the contributions of small instantons with size $\rho \ll 1/\Lambda_{\mathrm{QCD}}$ (and, in particular, of size $\rho \lesssim 1/v$) within the low energy theory but still above the QCD scale are also calculabe. Finally, the Higgses H decouple from the low energy physics within the weak coupling regime while the matter fields are effectively massless[2] even at the strong coupling scale Λ_{QCD}.

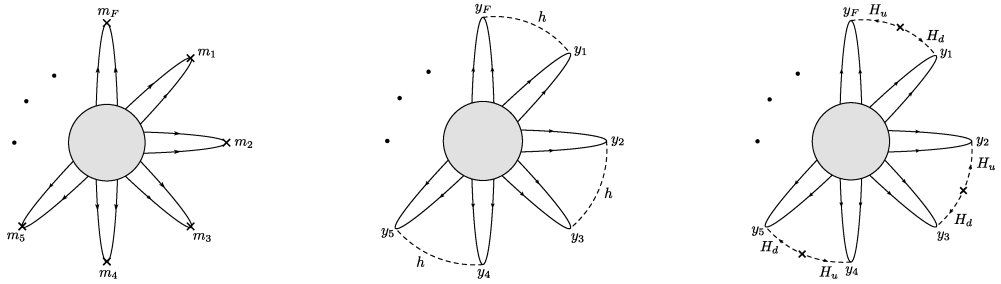
To obtain a simple estimate of the magnitudes of the effects of the small instantons we use an effective Lagrangian below the symmetry breaking scale M. Integrating over the instantons of size $\rho < 1/M$ generates a 't Hooft operator which must be included in the Lagrangian of the effective theory

$$\frac{\Lambda^{b_G}}{M^{b_G+3F-4}} \prod_i^F \psi_i \bar{\psi}_i\,. \tag{2.3}$$

These 't Hooft operators will also contribute to the mass of the axion once the fermion legs are closed up with Higgs VEV insertions or via Higgs loops. Such contributions can be represented by the diagrams in figure 1. One important issue to consider is which of these diagrams can possibly contribute to the axion mass. The axion is the Goldstone boson resulting from the spontaneous breaking of the anomalous $\mathrm{U}(1)_{PQ}$ symmetry at a high scale f_a. However if the classical action possesses additional exact anomalous unbroken

[2]To streamline the analysis we assume here that all the matter fields are light compared to Λ_{QCD}. Accounting for the mass of heavy SM flavors, t, b, c, will not affect the relative importance of contributions from different energy scales.

JHEP04(2020)031



(a) (b) (c)

Figure 1. Contributions to the axion mass obtained from closing up the instanton induced 't Hooft operators. On the left we use Higgs VEV insertions, in the middle we use loops of a dynamical Higgs boson in a single Higgs theory, while on the right we use Higgs loops in a 2HDM. Note that the arrows correspond to chiralities.

symmetries, one can always redefine the broken $\mathrm{U}(1)_{PQ}$ to be anomaly free and the axion remains exactly massless (this is for example the case when one of the SM quarks are exactly massless). As usual the presence of an exact anomalous symmetry will also imply that the QCD θ angle is unphysical. The Yukawa coupling of the SM fermions breaks any additional global symmetries, hence to obtain a contribution to the axion mass one needs to have a diagram proportional to all SM Yukawa couplings. In models with a single Higgs (like the KSVZ-type axion models [35, 36])

$$\sum_{i=1}^{F/2} yH\psi_i\bar{\psi}_i + \sum_{i=F/2+1}^{F} yH^\dagger\psi_i\bar{\psi}_i + \text{h.c.} \tag{2.4}$$

we can obtain a contribution either through Higgs insertions or via closing up the diagrams using Higgs loops as already depicted in figure 1(a) and 1(b). In other common axion models like DFSZ [37, 38] there are two Higgs doublets (2HDM), with Yukawa couplings of the sort

$$\sum_{i=1}^{F/2} yH_u\psi_i\bar{\psi}_i + \sum_{i=F/2+1}^{F} yH_d\psi_i\bar{\psi}_i + \text{h.c.} \tag{2.5}$$

In this case we can still use Higgs insertions, however in order to be able to produce diagrams with Higgs loops one needs an additional B_μ-like term $f_a H_u H_d + h.c.$ Such terms are usually readily present in complete axion models like the DFSZ axion [37, 38], and the actual diagram will be of the sort presented in figure 1(c). The effective theory below f_a will be a one-Higgs doublet model of the sort (2.4). As long as $f_a > M$ we can work in the effective one-Higgs doublet model. However if $f_a < M$ one expects the loop diagrams in 2HDMs to be suppressed by powers of f_a/M.

We can now compare the contributions to the axion mass from the Higgs VEV insertion diagram

$$m_M^2 f_a^2 = \left(\frac{yv}{M}\right)^F \frac{\Lambda^{b_G}}{M^{b_G-4}}\,, \tag{2.6}$$

JHEP04(2020)031

to the contributions from the diagram obtained by closing the Higgs loops:

$$m_M'^2 f_a^2 = \left(\frac{y}{4\pi}\right)^F \frac{\Lambda^{b_G}}{M^{b_G-4}}\,. \tag{2.7}$$

where m_M^2 and $M_a'^2$ represent the contributions of small instantons[3] ($\rho \lesssim 1/M$) to the axion mass obtained from VEV-insertion and loop-induced diagrams respectively. For sufficiently large symmetry breaking scale, the suppression of (2.6) by $(v/M)^F$ may easily overcome the suppression of the loop-induced contribution by loop factors, making (2.7) the dominant contribution from this regime.

There will be similar small instanton contribution to the axion mass from instantons of size $1/M \lesssim \rho \lesssim 1/v$:

$$\begin{aligned} m_v^2 f_a^2 &= y^F \frac{\Lambda_{\rm QCD}^{b_{\rm QCD}}}{v^{b_{\rm QCD}-4}} \\ m_v'^2 f_a^2 &= \left(\frac{y}{4\pi}\right)^F \frac{\Lambda_{\rm QCD}^{b_{\rm QCD}}}{v^{b_{\rm QCD}-4}}\,, \end{aligned} \tag{2.8}$$

where m_v^2 and and $m_v'^2$ denote the VEV insertion and the Higgs loop induced contributions respectively.[4] It is easy to see that loop-induced contributions in (2.7) are small both compared to VEV-insertion and loop-induced contributions in (2.8).

Below the Higgs mass v the Higgs decouples from the theory and loop-induced contributions are absent. However, given our choice of small Yukawas, the fermions remain light and the instanton diagrams with Higgs VEV insertions still contribute both to the 't Hooft operator and the axion potential. These contributions remain calculable in the $1/v < \rho \ll 1/\Lambda_{\rm QCD}$ regime and at the renormalization scale μ satisfying $\Lambda_{\rm QCD} \ll \mu \ll v$ are given by

$$m_\mu^2 f_a^2 = (yv)^F \frac{\Lambda_{\rm QCD}^{b_{\rm QCD}}}{\mu^{b_{\rm QCD}-4}}\,. \tag{2.9}$$

Once again, instanton contributions from lower scales in (2.9) dominate over the instantons of size $1/v$ in (2.8) and instantons of size $1/M$ in (2.6)and (2.7). As the renormalization scale μ approaches the actual strong coupling scale $\Lambda_{\rm QCD}$, the perturbative calculation in the one instanton background becomes unreliable. In this regime the contributios of the non-perturbative dynamics to the axion mass are a priori incalculable, however they can be obtained from chiral perturbation theory by relating the axion mass to the pion mass. Nevertheless, one can estimate the final axion mass by taking a naive $\mu \to \Lambda_{\rm QCD}$ limit:

$$m_{\rm QCD}^2 f_a^2 = (yv)^F \Lambda_{\rm QCD}^{4-F} = m^F \Lambda_{\rm QCD}^{4-F}\,, \tag{2.10}$$

where F is the number of flavors that remain light at $\Lambda_{\rm QCD}$ and $m_{\rm QCD}$ represents the QCD contribution to the axion mass.

We can now estimate the ratio of loop-induced small instanton and QCD contributions to the axion mass:

$$\frac{M_a'^2}{m_a^2} \sim \frac{1}{(4\pi)^F}\left(\frac{\Lambda_{\rm QCD}}{v}\right)^F \left(\frac{\Lambda_{\rm QCD}}{M}\right)^{b_{\rm QCD}-4}\,. \tag{2.11}$$

[3] These effects are dominated by instantons of inverse size M.

[4] These effects are dominated by instantons of inverse size v.

JHEP04(2020)031

As expected the axion mass is dominated by strong coupling QCD contributions while the contributions of small instantons are highly suppressed by powers of Λ/M and otherwise are UV independent. Indeed, every term in (2.11) is smaller than one. As a reminder, $b_{\rm QCD}$ is the QCD beta function just below the matching scale with all flavors assumed to be massless: $b_{\rm QCD} = \frac{11}{3}N_c - \frac{2}{3}F = 7$ for QCD with 6 flavors, but most importantly $b_{\rm QCD} > 4$ implying a strong suppression by powers of $\Lambda_{\rm QCD}/M$.

There is however an important caveat in the above argument, which is what Agrawal and Howe have exploited [12, 13]. The matching relation (2.2) can be modified if the embedding of QCD into the bigger group G is non-trivial. In fact, (2.2) has a very simple and intuitive interpretation: the 1-instanton solution of the low energy H theory is also a 1-instanton solution of the high energy G theory (with additional bosonic zero modes of the high energy theory lifted by spontaneous symmetry breaking). However, other kinds of embedding are possible [32] — for example, the 1-instanton solution of the low energy theory may represent a 2-instanton, or in general a k-instanton configuration in the high energy theory. The first examples of non-trivial effects due to such instantons were identified in the context of exact results in SUSY gauge theories by Intriligator, Seiberg and Leigh [31]. In this case the matching relation (2.2) would be modified to

$$\left(\frac{\Lambda}{M}\right)^{kb_G} = \left(\frac{\Lambda_{\rm QCD}}{M}\right)^{b_{\rm QCD}} \tag{2.12}$$

where the integer k is usually referred to as the index of embedding, first identified in [31, 33, 34] and explained extensively in [32]. Such a non-trivial factor usually appears when there are instantons in the partially broken gauge group [32], meaning that the instantons of the unbroken group do not map one-to-one to the instantons of the high energy theory. Topologically it is the homotopy group $\pi_3(G/H)$ that will be relevant, and when both G and H are simple one can show that $\pi_3(G/H) = Z_k$, where k is the index of embedding. In this paper we are interested in models where one breaks a product group to its diagonal subgroup. For example, when the symmetry breaking pattern is given by $\mathrm{SU}(N) \times \mathrm{SU}(N) \to \mathrm{SU}(N)$ the 1-instanton of the low-energy theory actually corresponds to a $(1,1)$ of the UV theory, while for $\mathrm{SU}(N)^k \to \mathrm{SU}(N)$ the 1-instanton will be a $(1,1,1,\ldots,1)$ instanton. For the product group case the relevant homotopy group will be $\pi_3(\mathrm{SU}(N)^k/\mathrm{SU}(N)) = Z \times Z \times \ldots Z$ with $k-1$ Z-factors. Either way, if dynamical scales and beta function coefficients of all UV gauge group factors are equal, the matching relation will be given by eq. (2.12). More generally the scale matching relation (2.12) is replaced by a relation where factors of dynamical scale on the right-hand side are replaced by a product of 1-instanton weights of UV gauge group factors:

$$\prod_i^k \left(\frac{\Lambda_i}{M}\right)^{b_i} = \left(\frac{\Lambda_D}{M}\right)^{b_D} . \tag{2.13}$$

We can see now how this non-trivial mapping of instantons (and matching of dynamical scales) would possibly lead to an enhancement of the small instanton contributions. When one has a non-trivial index of embedding, some of the broken instantons are actually

JHEP04(2020)031

topologically distinct from those eventually giving rise to the QCD instanton corrections, hence they will scale differently. From the point of view of scaling they will appear as "fractional" $1/k$ instantons, and their contributions may be enhanced compared to the usual QCD instantons. For a case with index of embedding k while the expression of the contribution of the small instantons from the partially broken group are still given by (2.7), the use of the modified matching (2.12) will result in

$$\frac{M_a'^2}{m_a^2} \sim \frac{1}{(4\pi)^F} \left(\frac{\Lambda_{\rm QCD}}{v}\right)^F \left(\frac{\Lambda_{\rm QCD}}{M}\right)^{\frac{b_{\rm QCD}}{k}-4} . \tag{2.14}$$

Already for $k = 2$ the sign of the exponent of $\Lambda_{\rm QCD}/M$ will flip, and lead to the possibility of these terms dominating over the ordinary QCD contributions when M is taken to be large.

In the rest of the paper we will present a detailed calculation of the small instanton effects in the partially broken gauge group to replace (2.14) with a more precise expression, paying careful attention to all the relevant $\mathcal{O}(1)$ factors and perturbation of the classical instanton action in the presence of spontaneous symmetry breaking. This will give us a better understanding of models and parameter regions in which small instanton contributions are dominant.

3 One instanton effects in a broken SU(N)

We now turn to the actual instanton calculation that will verify the validity of our estimates in section 2 and provide us with more precise results. As we have explained, contributions of small instantons that are topologically equivalent to single instanton configurations of the low energy theory are always subleading. Instead we will consider instantons of the high energy theory that are absent from the effective low energy description. These instantons must be carefully integrated out and their effects must be taken into account explicitly when constructing the low energy description. We will be especially interested in product group theories broken to a diagonal subgroup, for example, $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2 \to \mathrm{SU}(N)_D$. The low-energy $\mathrm{SU}(N)_D$ theory contains only a subset of the instanton solutions of the full $\mathrm{SU}(N)_1 \times \mathrm{SU}(N)_2$ theory [32]. For example a 1-instanton configuration in the diagonal subgroup is a $(1,1)$ combination of simultaneous 1-instanton solutions in the individual $\mathrm{SU}(N)$ factors. However, configurations with instantons in only one of the $\mathrm{SU}(N)$ factors (e.g. $(1,0)$ or $(0,1)$) are absent from the effective theory. Since these instantons are embedded in the completely broken factor of the high energy gauge group, it is useful to review the instanton calculus in (spontaneously broken) $\mathrm{SU}(N)$ gauge theories before re-examining the explicit models in [12]. We will loosely follow the instanton calculation in supersymmetric QCD by Cordes [39] with slight modifications due to the non-sypersymmetric nature of the problem at hand. It is common practice to perform instanton calculations using Pauli-Villars (PV) regularization, which we will also use here. However, in perturbative calculations dimensional regularization and the MS or $\overline{\rm MS}$ scheme are more common. We summarize the formulae needed to convert from PV to $\overline{\rm MS}$ scheme in appendix B.

JHEP04(2020)031

In the following we consider an SU(N) gauge theory with a matter sector consisting of $S \geq N-1$ scalars[5] ϕ_n, $n = 1, \ldots, S$ and F (approximately) massless fermions ψ_f, $f = 1, \ldots, F$ in the fundamental representation of SU(N). The euclidean action for this model is given by

$$S_E = S_G + S_\phi + S_\psi \,, \tag{3.1}$$

where

$$S_G = \int d^4x \left[\frac{1}{4} G^A_{\mu\nu} G^A_{\mu\nu} + i\theta \frac{g^2}{32\pi^2} G^A_{\mu\nu} \tilde{G}^A_{\mu\nu} + \mathcal{L}_{\text{ghost}}(\eta, \bar{\eta}) \right], \tag{3.2}$$

$$S_\phi = \int d^4x \left[(D_\mu \phi_n)^\dagger D_\mu \phi_n + V(\phi_n) \right], \tag{3.3}$$

$$S_\psi = \int d^4x \, \bar{\psi}_f (-i\gamma_\mu D_\mu) \psi_f \,, \tag{3.4}$$

with $D_\mu \phi_n = (\partial_\mu - igA^A_\mu T^A)\phi_n$. T^A, $A = 1, \ldots, N^2-1$ are the SU(N) generators. A sum over the scalar and fermion generations is implied. We assume the scalar potential $V(\phi_f)$ to be such that the scalars develop a VEV which breaks the SU(N) gauge symmetry completely.

For vanishing scalar VEVs the euclidean Yang-Mills action S_G possesses exact instanton solutions for the classical equations of motion. The one instanton solution, centered at x_0, with unit topological charge ($Q = 1$) in singular gauge takes the form [40]

$$A^{Q=1}_\mu(x) = \frac{2\rho^2}{g} \bar{\eta}_{a\mu\nu} \frac{(x-x_0)_\nu}{(x-x_0)^2((x-x_0)^2+\rho^2)} J^a \,, \tag{3.5}$$

where $\bar{\eta}_{a\mu\nu}$ are 't Hooft symbols, ρ is the instanton size and J^a, $a = 1, 2, 3$ are the generators of the SU(2) subgroup into which the instanton is embedded. In the following we will work with the so called minimal embedding, where one embeds the SU(2) into the 2×2 upper-left-hand corner of the $N \times N$ matrices which generate the fundamental representation of SU(N).

Once the scalars obtain a VEV $|\langle \phi_n \rangle| > 0$ and break SU(N) completely, no exact instanton solutions exist.[6] However, one expects that for small instantons, $g\rho|\langle \phi_n \rangle| \ll 1$, the solution (3.5) remains a good approximation and the path integral is still dominated by instanton-like configurations. The path integral can be performed by using the constrained instanton formalism of Affleck [41]. In the constrained instanton formalism the scalars satisfy the equation of motion in the classical instanton background, $D^2(A_{cl})\phi = 0$. As a result to leading order in $g\rho|\langle \phi \rangle|$ the scalar profile is given by

$$\phi_{in}(x) = \begin{cases} \left(\frac{x^2}{x^2+\rho^2} \right)^{1/2} \langle \phi_{in} \rangle & \text{for } i = 1, 2 \\ \langle \phi_{in} \rangle & \text{for } i = 3, \ldots, N \end{cases} , \tag{3.6}$$

where i is the SU(N) index of the scalar multiplets.

[5] In order to break SU(N) completely one needs at least $N-1$ scalar fields.

[6] If SU(N) is only partially broken with an unbroken residual SU(2) subgroup, i.e. rank($\langle \phi_{in} \rangle$) $< N-1$, exact instanton solutions still exist in the unbroken SU(2).

JHEP04(2020)031

Evaluating the classical action of the constrained instanton with $Q = 1$ in the presence of the scalar profile, one finds[7]

$$S_0(\rho) = \frac{8\pi^2}{g^2} + 2\pi^2\rho^2 \sum_{i=1}^{2}\sum_{n=1}^{S} |\langle\phi_{in}\rangle|^2 + i\theta\,. \tag{3.7}$$

Thus large instantons ($g\rho|\langle\phi_f\rangle| \gg 1$) are exponentially suppressed. This provides a natural IR cutoff for the instanton size and makes instanton contributions to observables calculable.

In the following we are interested in $W_{\mathrm{SU}(N)}$, the one-instanton semi-classical approximation of the vacuum to vacuum amplitude

$$W_{\mathrm{SU}(N)} \equiv \langle 0|0\rangle_{\Delta Q=1} = \frac{\int_{1-\mathrm{inst}} \mathcal{D}A_\mu \mathcal{D}\eta \mathcal{D}\bar{\eta} \mathcal{D}\phi_f \mathcal{D}\phi_f^\dagger \mathcal{D}\psi \mathcal{D}\bar{\psi}_n\, e^{-S_E}}{\int_{A_\mu^{\mathrm{cl}}=0} \mathcal{D}A_\mu \mathcal{D}\eta \mathcal{D}\bar{\eta} \mathcal{D}\phi_f \mathcal{D}\phi_f^\dagger \mathcal{D}\psi \mathcal{D}\bar{\psi}_n\, e^{-S_E}}\,. \tag{3.8}$$

We can evaluate the functional integral in eq. (3.8) in the semi-classical approximation by expanding the euclidean action to second order in the fields around the classical solutions in eq. (3.6) and (3.5)

$$S_E = S_0(\rho) + \int d^4x \left[\frac{1}{2} A_\mu M_A A_\mu + \bar{\eta} M_{\mathrm{ghost}} \eta + \phi^\dagger M_\phi \phi + \bar{\psi} M_\psi \psi\right], \tag{3.9}$$

where $\phi = (\phi_1, \ldots, \phi_S)^T$ and $\psi = (\psi_1, \ldots, \psi_F)^T$ are vectors containing all scalar and fermion generations, and perform the functional integral.

The various contributions to the generating functional will be discussed thoroughly in the next sections, but we already present the final expression of the general result here. For the above field content the vacuum to vacuum amplitude is given by

$$W_{\mathrm{SU}(N)} = e^{-i\theta} \int \frac{d^4x_0 d\rho}{\rho^5} d_N(\rho) \int d\tilde{\mu}\, e^{-2\pi^2\rho^2 \sum_{i=1}^{2}\sum_{n=1}^{S} |\langle\tilde{\phi}_{in}\rangle(\tilde{\mu})|^2} \int \prod_{f=1}^{2F} \rho^{1/2} d\xi_f^{(0)}\,, \tag{3.10}$$

where $d_N(\rho)$ is the instanton density in vacuum

$$d_N(\rho) = \frac{C_1 e^{-(S-2F)\alpha(1/2)}}{(N-1)!(N-2)!} \left(\frac{8\pi^2}{g^2}\right)^{2N} e^{-\frac{8\pi^2}{g^2(1/\rho)} - C_2 N}\,. \tag{3.11}$$

C_1 and C_2 are defined as

$$C_1 = \frac{2e^{5/6}}{\pi^2} \approx 0.466\,, \tag{3.12}$$

$$C_2 = \frac{5}{3}\ln 2 - \frac{17}{36} + \frac{1}{3}(\ln 2\pi + \gamma) + \frac{2}{\pi^2}\sum_{s=1}^{\infty} \frac{\ln s}{s^2} \approx 1.678\,. \tag{3.13}$$

$\alpha(t)$ is defined in [40] (with $\alpha(0) = 0$, $\alpha(1/2) = 0.145873$, $\alpha(1) = 0.443307$), $\int d\tilde{\mu}$ is the integral over the collective coordinates corresponding to the orientation of the instanton within SU(N) normalized to unity, and $\int d\xi_f^{(0)}$ is the integral over the fermion zero modes. Note that $\langle\tilde{\phi}_{in}\rangle(\tilde{\mu})$ are the scalar VEVs rotated in group space to account for the arbitrary location of the instanton SU(2) inside SU(N).

[7] Note that in the background of an instanton with topological charge Q, $\frac{g^2}{32\pi^2}\int d^4x (G^A_{\mu\nu}\tilde{G}^A_{\mu\nu})_{\mathrm{inst}} = Q$.

JHEP04(2020)031

3.1 Bosonic contributions

Performing the integral in eq. (3.8) over the bosonic sector of the theory, i.e. integrating over the gauge, scalar and ghost fields one obtains

$$W_{\mathrm{SU}(N)} = \int \prod_i d\gamma_i J(\gamma) e^{-S_0(\rho)} I_\psi(\gamma) \frac{(\det' M_A(\gamma))^{-1/2} (\det' M_{\mathrm{ghost}}(\gamma)) (\det' M_\phi(\gamma))^{-2}}{((\det M_A)^{-1/2} (\det M_{\mathrm{ghost}}) (\det M_\phi)^{-2})_{A_\mu^{\mathrm{cl}}=0}}, \tag{3.14}$$

where the contribution from the fermions is encoded in I_ψ, which will be computed later, and the determinants $\det'$ are taken over non-zero modes only.

The zero modes are flat directions in the action and can be parameterised in terms of collective coordinates γ_i

$$\gamma_i = \begin{cases} (x_0)_i & i = 1, \ldots, 4 \\ \rho & i = 5 \\ t^A & i = A + 5 = 6, \ldots, N^2 + 4 \end{cases} \tag{3.15}$$

where x_0 is the instanton position, ρ its size and t^A are the $N^2 - 1$ parameters describing general SU(N) transformations. The group theoretic zero modes depend on the embedding of the instanton into SU(N) and their effects can be found by classifying how the generators T^A of the full group transform under the J^a generators of the SU(2) subgroup in which the instanton is embedded. For SU(N) one finds i) one triplet (J^a themselves) ii) $2(N-2)$ doublets and iii) N^2-4N+4 singlets. There are no normalizable zero modes corresponding to singlet generators, which means there are $4N$ normalizable zero modes altogether.

Replacing the integration over the zero modes in the functional integral by an integration over the collective coordinates introduces the Jacobian $J(\gamma)$ in eq. (3.14). Using the normalization of the zero modes one finds (see e.g. [39] and [42])[8]

$$\int \prod_i^{4N} d\gamma_i \, J(\gamma) = \int d^4x_0 \, d\rho \, d\mu \, \frac{2^7}{\rho^5} \left(\frac{\pi \rho^2}{g^2} \right)^{2N}, \tag{3.16}$$

where $d\mu$ is the Haar measure of the quotient group SU(N)/T_N, with T_N being the stability group of the instanton, i.e. the subgroup of SU(N) that leaves the instanton invariant. In [39] it is shown that for integrands invariant under T_N, the group integration can be expressed as

$$\begin{aligned} \int_{\mathrm{SU}(N)/T_N} d\mu &= \frac{V(\mathrm{SU}(N-1))}{V(T_N)} \int_{\mathrm{SU}(N)/\mathrm{SU}(N-1)} d\mu \\ &= \frac{2^{4N-6} \pi^{N-2}}{(N-2)!} \int_{S^{2N-1}} \delta(\sqrt{\textstyle\sum |y_i|^2} - 1) \, d^2y_1 \ldots d^2y_N \, . \end{aligned} \tag{3.17}$$

We will denote a general element of the coset SU(N)/SU($N-1$) by Ω. It is possible to parameterise Ω in terms of the y_i [39], but the explicit form of Ω will not be needed in

[8] Cordes [39] and Bernard [42] use different normalizations for the SU(N) generators, which is reflected in their different results for the zero-mode normalization (apart from the missing factor of ρ in $||A_\mu^{(\mathrm{isodoub})}||$ in eq. (5.7) of [39], which is clearly a typo). In the following we will follow Cordes' conventions.

JHEP04(2020)031

the following. Using the fact that the surface of the S^{2N-1} sphere is given by $S(S^{2N-1}) = 2\pi^N/(N-1)!$, we can define a normalized integration measure

$$\int d\tilde{\mu} = \frac{(N-1)!}{2\pi^N} \int_{S^{2N-1}} \delta\left(\sqrt{\sum |y_i|^2} - 1\right) d^2y_1 \dots d^2y_N \,. \tag{3.18}$$

As a last step we need to evaluate the functional determinants over the non-zero modes. This calculation has been done in 't Hooft's original paper [40] for an SU(2) gauge theory. The generalization to SU(N) is straightforward (see e.g. [42]) and yields in Pauli-Villars regularization

$$\begin{aligned} &\frac{(\det' M_A(\gamma))^{-1/2} (\det' M_{\text{ghost}}(\gamma)) (\det' M_\phi(\gamma))^{-2}}{((\det M_A)^{-1/2} (\det M_{\text{ghost}}) (\det M_\phi)^{-2})_{A_\mu^{\text{cl}}=0}} \\ &\quad = \exp\left[-\left(\frac{1}{3}N + \frac{1}{6}\sum_t S(t)C(t)\right) \ln(\mu_0\rho) - \alpha(1) - 2(N-2)\alpha(1/2) - \sum_t S(t)\alpha(t) \right], \end{aligned} \tag{3.19}$$

where t denotes the isospin representation under the instanton SU(2). $S(t)$ is the number of scalar multiplets with isospin t, where each complex multiplet counts as 1 and each real multiplet as $1/2$ and $C(t) = \frac{2}{3}t(t+1)(2t+1)$. Each scalar fundamental contributes one multiplet in the isospin $1/2$ representation and $(N-2)$ singlets.

Substituting eqs. (3.16), (3.17), (3.18) and (3.19) into eq. (3.14) and recalling that in Pauli-Villars regularization each zero-mode yields a factor μ_0 of the regulator field, we obtain

$$\begin{aligned} W_{\text{SU}(N)} = &\frac{C_1\, e^{-N_S\alpha(1/2)}}{(N-1)!(N-2)!} e^{-i\theta} \left(\frac{8\pi^2}{g^2}\right)^{2N} \int \frac{d^4x_0\, d\rho}{\rho^5} (\mu_0\rho)^{b_0} e^{-8\pi^2/g^2 - C_2 N} \\ &\times \int d\tilde{\mu}\, I_\psi(\gamma)\, e^{-2\pi^2\rho^2 \sum_{i=1}^{2}\sum_{n=1}^{S} |\langle\phi_{in}\rangle|^2} \,, \end{aligned} \tag{3.20}$$

where $b_0 = \frac{11}{3}N - \frac{1}{6}S$ is the bosonic contribution to the β-function and

$$C_1 = \frac{4\, e^{-\alpha(1)+4\alpha(1/2)}}{\pi^2} \,, \qquad C_2 = 2\,\ln 2 + 2\,\alpha(1/2) \,. \tag{3.21}$$

Note that when using the explicit expression for $\alpha(t)$, the above definition of C_1 and C_2 agrees with eqs. (3.12) and (3.13).

The group integration $\int d\tilde{\mu}$ in eq. (3.20) corresponds to rotating the instanton embedding in SU(N). This is equivalent to keeping the instanton fixed and instead rotating all other fields, in particular the scalar fields and their VEVs, by a general SU(N)/SU($N-1$) group element Ω, i.e. in eq. (3.20) we should make the replacement

$$\langle\phi_{in}\rangle \to \langle\tilde{\phi}_{in}\rangle(\tilde{\mu}) = \sum_{j=1}^{N} \Omega_{ij} \langle\phi_{jn}\rangle \,. \tag{3.22}$$

JHEP04(2020)031

3.2 Fermionic contributions

Analogously to the bosonic contributions to the vacuum to vacuum amplitude, one can isolate the zero modes in the integration over the fermionic fields, i.e.

$$\mathcal{D}\psi\mathcal{D}\bar{\psi} = \prod_f ||\psi_f^{(0)}||^{-1}\, d\xi_f^{(0)} \prod_{f'} ||\bar{\psi}_{f'}^{(0)}||^{-1}\, d\bar{\xi}_{f'}^{(0)}\, \mathcal{D}\psi'\, \mathcal{D}\bar{\psi}'\,, \tag{3.23}$$

where $\psi_f^{(0)}$ and $\bar{\psi}_f^{(0)}$ are the zero mode wave functions of the Dirac operator $M_\psi^{mn} = -i\delta^{mn}\gamma_\mu D_\mu$ and $d\xi_f^{(0)}, d\bar{\xi}_f^{(0)}$ are Grassmann integration measures with mass dimension $[d\xi_f^{(0)}] = [d\bar{\xi}_f^{(0)}] = \frac{1}{2}$. The explicit form of the normalized zero-modes in singular gauge, for an instanton centered at x_0, is given by [43]

$$\psi_f^{(0)}(x)_{\alpha i} = \frac{\rho}{\pi}\frac{(x-x_0)_\mu}{((x-x_0)^2)^{1/2}((x-x_0)^2+\rho^2)^{3/2}} \begin{pmatrix} 0 \\ i(\tau_\mu^+)_i^j \varphi_{\alpha j} \end{pmatrix} \epsilon_{\alpha k}\,, \tag{3.24}$$

where $\alpha, i, j = 1, 2$ are the spinor and SU(N) indices (restricted to the instanton SU(2) with $\psi_f^{(0)}(x)_{\alpha i} = 0$ for $i = 3, \ldots, N$), respectively.[9] τ_μ^+ is defined as $\tau_\mu^+ = (\vec{\tau}, -i)$ with $\vec{\tau}$ being the Pauli matrices. $\varphi_{\alpha j}$ is a two component Weyl spinor with $\varphi_{\alpha j} = \epsilon_{\alpha j}$. Note that for small instantons, far from the instanton center, the zero mode wavefunction is proportional to the free fermion propagator $S_F(x) = \frac{\gamma_\mu x_\mu}{2\pi^2(x^2)^2}$. Each massless Dirac fermion in the fundamental representation possesses two zero modes, one for each chirality, in the one instanton background. This implies that in the model with F fermion flavors we have $2F$ fermionic zero modes.

The integral over the non-zero modes can be directly performed, which yields

$$I_\psi = \int \prod_{f=1}^{2F} d\xi_f^{(0)} \mu_0^{-F} \left(\frac{\det' M_\psi^\dagger M_\psi}{(\det M_\psi^\dagger M_\psi)_{A_\mu^{\rm cl}=0}} \right)^{1/2}, \tag{3.25}$$

where we assumed normalized zero modes and collectively denoted the zero mode integration measure as $d\xi_f^{(0)}$. Additionally we inserted a factor $\mu_0^{-1/2}$ of the regulator field for each of the $2F$ zero modes, since we work in Pauli-Villars regularization scheme.

The non-zero mode determinant was computed by 't Hooft in his original paper [40]

$$\left(\frac{\det' M_\psi^\dagger M_\psi}{(\det M_\psi^\dagger M_\psi)_{A_\mu^{\rm cl}=0}} \right)^{1/2} = \exp\left[\frac{1}{3} F \ln(\rho\mu_0) + 2F\alpha(1/2) \right]. \tag{3.26}$$

Combining eqs. (3.25) and (3.26), we obtain the full fermionic contribution to $W_{\mathrm{SU}(N)}$

$$I_\psi = \rho^F e^{-\frac{2}{3}F\ln(\rho\mu_0)+2F\alpha(1/2)} \int \prod_{f=1}^{2F} d\xi_f^{(0)}\,. \tag{3.27}$$

Plugging this result into eq. (3.20), one obtains the vacuum to vacuum amplitude for a broken SU(N) gauge theory in a one instanton background, which we already previewed in eq. (3.10).

[9] Note that the zero modes naively seem to have the wrong dimension (mass dimension 2 instead of 3/2), but the combination with the corresponding Grassmann variable $\xi_f^{(0)}$ in the expansion $\psi_f(x) = \sum_k \psi_f^{(k)}(x)\,\xi_f^{(k)}$ has the right dimension ($[\xi_f^{(k)}] = -1/2$, s.t. $\int d\xi_f^{(k)}\, \xi_f^{(k)} = 1$).

JHEP04(2020)031

3.3 Vacuum energy/axion potential

Instanton configurations in the vacuum to vacuum amplitude generate a contribution to the vacuum energy which depends on the θ angle. This can be encoded in terms of an effective Lagrangian that captures the 1-(anti)instanton effects in terms of a potential for the θ angle, which in the presence of an axion will be interpreted as an effective potential/mass term for the axion itself. In a theory without massless fermions this potential is simply given by

$$-\delta\mathcal{L}^{F=0} = 2\int\frac{d\rho}{\rho^5}\int d\tilde{\mu}\, C_N(\rho)\cos(\theta)\,, \tag{3.28}$$

where $C_N(\rho)$ contains the instanton density and the action of the Higgs scalars

$$C_N(\rho) = d_N(\rho)\, e^{-2\pi^2\rho^2\sum_{i=1}^{2}\sum_{n=1}^{S}|\langle\tilde{\phi}_{in}\rangle|^2}\,. \tag{3.29}$$

If the theory contains massless fermions, eq. (3.27) implies that due to the $\xi_f^{(0)}$ integration any correlation function, including the vacuum to vacuum amplitude, which does not include the full set of $2F$ chiral fermions vanishes. Effectively the integration projects out the zero mode wave functions, i.e. for a fermion field expanded in eigenmodes $\psi_f = \psi_f^{(0)}\xi_f^{(0)} + \ldots$, the integration yields $\int d\xi_f^{(0)}\psi_f = \psi_f^{(0)}$. Thus the effect of massless fermions in the instanton background is captured by an effective $2F$-fermion operator, the so called 't Hooft operator.

However, even in the presence of massless fermions instantons can still generate a potential for the θ angle if the theory contains further interactions. The easiest way to see that is by working in the effective theory with a 't Hooft operator and closing up the external legs using the additional interaction terms forming a vacuum bubble (see figure 1), which contributes to the vacuum energy. Alternatively one can calculate the non-vanishing contribution to the vacuum to vacuum amplitude directly from the path integral by including higher orders in the interaction that includes all massless fermions. In the following we will pursue the second approach, which corresponds to the full calculation. We do expect the effective 't Hooft operator approach to be a good approximation to the full calculation, which we will indeed verify in appendix A where we present the 't Hooft operator method and also compare the results of the two approaches.

Let us assume the theory contains an additional scalar H (which we will later identify with the SM Higgs), which couples to the massless fermions via Yukawa interactions, i.e. let us add the following term to the Euclidean action

$$\Delta S = S_0[H] - i\int d^4x\sum_{f=1}^{F}\frac{y_f}{\sqrt{2}}H(x)\bar{\psi}_f(x)\psi_f(x)\,, \tag{3.30}$$

where $S_0[H]$ is the free action for the scalar H. With this addition the vacuum to vacuum amplitude now takes the form

$$\begin{aligned} W_{\mathrm{SU}(N)} = e^{-i\theta}\int d^4x_0\int d\tilde{\mu}\int\frac{d\rho}{\rho^5}C_N(\rho)\int\mathcal{D}H\, e^{-S_0[H]} \\ \times\int\prod_{f=1}^{F}\rho\, d\xi_f^{(0)}d\bar{\xi}_f^{(0)}e^{i\int d^4x\sum_{f=1}^{F}\frac{y_f}{\sqrt{2}}H(x)\bar{\psi}_f(x)\psi_f(x)}\,. \end{aligned} \tag{3.31}$$

JHEP04(2020)031

At order F in the Yukawa couplings, the expansion of the exponential contains a term with all $2F$ massless fermions. The integration over $\xi_f^{(0)}$ and $\bar{\xi}_f^{(0)}$ projects out the corresponding zero mode wave functions and all lower order terms vanish due to this integration. The leading contribution to $W_{\mathrm{SU}(N)}$, assuming F is even so that the path integral of the Higgs field does not vanish (ie. the Higgs loops can be closed up), is therefore[10]

$$\begin{aligned} W_{\mathrm{SU}(N)} = & e^{-i\theta} \int d^4x_0 \int d\tilde{\mu} \int \frac{d\rho}{\rho^5} C_N(\rho) \int \mathcal{D}H \, e^{-S_0[H]} \\ & \times \prod_{f=1}^{F} \left(\frac{i y_f \rho}{\sqrt{2}} \int d^4x H(x) \bar{\psi}_f^{(0)}(x) \psi_f^{(0)}(x) \right) . \end{aligned} \tag{3.32}$$

Performing the path integral for H, only fully contracted Higgs fields survive, each contraction giving a Feynman propagator

$$W_{\mathrm{SU}(N)} = e^{-i\theta} \int d^4x_0 \int d\tilde{\mu} \int \frac{d\rho}{\rho^5} C_N(\rho) \, \kappa_F \prod_{f=1}^{F} \left(\frac{y_f \rho}{\sqrt{2}} \right) \mathcal{I}^{F/2} , \tag{3.33}$$

where $\kappa_F = (F-1) \cdot (F-3) \cdots 1$ counts the number of equivalent contractions and $\mathcal{I}$ is the integral over the fermion zero modes and scalar Feynman propagators $\Delta_F(x_1 - x_2)$

$$\mathcal{I} = - \int d^4x_1 \int d^4x_2 \, \bar{\psi}_f^{(0)}(x_1) \psi_f^{(0)}(x_1) \bar{\psi}_{f'}^{(0)}(x_2) \psi_{f'}^{(0)}(x_2) \Delta_F(x_1 - x_2) . \tag{3.34}$$

Using the explicit form for the fermion zero modes[11] $\psi_f^{(0)}$ in eq. (3.24) $\mathcal{I}$ simplifies to

$$\mathcal{I} = \frac{\rho^4}{4\pi^8} \int d^4x_1 \int d^4x_2 \int d^4p \frac{1}{p^2 + m_H^2} \frac{e^{-ipx_1}}{(x_1^2 + \rho^2)^3} \frac{e^{ipx_2}}{(x_2^2 + \rho^2)^3} . \tag{3.35}$$

Using the identity

$$\int d^4x \frac{e^{-ipx}}{(x^2 + \rho^2)^3} = \frac{\pi^2}{2\rho^2} (p\rho) K_1(p\rho) , \tag{3.36}$$

where K_1 is a modified Bessel function of the second kind, we can evaluate $\mathcal{I}$ explicitly in the limit $\rho \ll 1/m_H$

$$\mathcal{I} \simeq \frac{1}{12\pi^2\rho^2} . \tag{3.37}$$

Plugging this into eq. (3.33) we can immediately write down the leading contribution to the potential for the θ angle, generated by 1-(anti)instanton configurations, for theories with massless fermions and a Yukawa interaction

$$-\delta\mathcal{L}^F = 2 \int \frac{d\rho}{\rho^5} \int d\tilde{\mu} \, C_N(\rho) \, \kappa_F \prod_{f=1}^{F} \left(\frac{y_f}{\sqrt{24}\pi} \right) \cos(\theta) . \tag{3.38}$$

[10] Note that the $1/F!$ from the expansion of the exponential is compensated by $F!$ terms which are identical after renaming the integration variables.

[11] Note that similarly to the scalars ϕ_{in} one should rotate $\psi_f^{(0)}$ with the general $\mathrm{SU}(N)/\mathrm{SU}(N-1)$ coset element Ω. However, due to the $\mathrm{SU}(N)$ invariant Yukawa interaction, the Ω dependence cancels out and $\mathcal{I}$ is independent of $\tilde{\mu}$.

JHEP04(2020)031

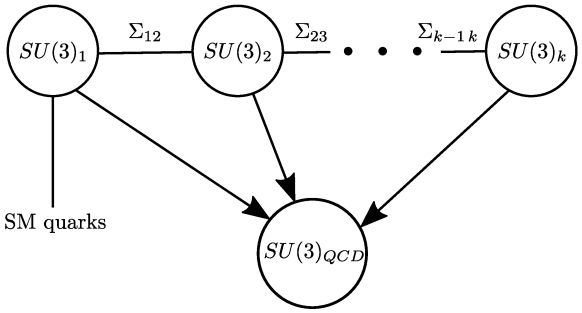


Figure 2. Illustration of the product gauge group model introduced in [12] to enhance small instanton contributions to the axion mass. The k SU(3) factors are broken at a scale M by scalar link fields $\Sigma_{i\,i+1}$ in the bifundamental representation of $\mathrm{SU}(3)_i \times \mathrm{SU}(3)_{i+1}$ to their diagonal combination which is identified with $\mathrm{SU}(3)_{\mathrm{QCD}}$. We further assume that the SM quarks are only charged under $\mathrm{SU}(3)_1$.

It is worth emphasizing that $\mathcal{I}$ could be estimated in the effective field theory by soaking up fermion legs of the 't Hooft operator with the Higgs propagators. However, the EFT result would be cutoff dependent while the above computation is completely convergent and calculable. For more on the correct value of the cutoff scale see appendix A.

4 Small instantons in product group models

Small instanton contributions to the axion mass can dominate over the non-perturbative large QCD instantons in partially broken gauge theories with a non-trivial embedding of $\mathrm{SU}(3)_{\mathrm{QCD}}$ (see section 2). An example of such a setup are the models proposed by Agrawal and Howe [12, 13], in which a product gauge group consisting of k SU(3) factors is spontaneously broken at a scale M to its diagonal subgroup by $k-1$ link fields $\Sigma_{i\,i+1}$

$$\mathrm{SU}(3)_1 \times \mathrm{SU}(3)_2 \times \ldots \times \mathrm{SU}(3)_k \to \mathrm{SU}(3)_{\mathrm{QCD}}\,. \tag{4.1}$$

The diagonal subgroup can then be identified with $\mathrm{SU}(3)_{\mathrm{QCD}}$. In the following we will assume that all SM quarks are only charged under $\mathrm{SU}(3)_1$. For a diagrammatic depiction of the model see figure 2. The individual SU(3) factors by themselves are completely broken and therefore the 1-instanton effects are calculable and finite. The 1-instanton configuration in low energy QCD corresponds to k-instantons of the UV theory with one instanton in each SU(3) factor. In the following we will first discuss some details of the model before we explicitly compute the small instanton contributions to the axion potential in the two simplest realizations with $k = 2, 3$ and compare the results to [12]. Note that in this section we work in Minkowski space.

4.1 Axions in product group models

Each of the SU(3) gauge factors comes with its own CP violating θ angle. Therefore we assume that there is also one anomalous $\mathrm{U}(1)_{PQ}$ for each factor, which is spontaneously

JHEP04(2020)031

broken at $f_{a_i} > M$. This yields one axion for each SU(3)

$$\mathcal{L} = \sum_{i=1}^{k} \mathcal{L}_i \,, \qquad \mathcal{L}_i = -\frac{1}{4} G^a_{i\,\mu\nu} G_i^{a\,\mu\nu} + \frac{g_i^2}{32\pi^2} \left(\frac{a_i}{f_{a_i}} - \theta_i \right) G^a_{i\,\mu\nu} \tilde{G}_i^{a\,\mu\nu} \,. \tag{4.2}$$

As depicted in figure 2, the gauge group is broken to $\mathrm{SU}(3)_{\mathrm{QCD}}$ by higgsing it with $k-1$ scalar link fields $\Sigma_{i\,i+1}$, which transform as a bifundamental $(3, \bar{3})$ under $\mathrm{SU}(3)_i \times \mathrm{SU}(3)_{i+1}$. A potential[12] of the form [12, 44]

$$V(\Sigma) = -m_\Sigma^2 \mathrm{Tr}(\Sigma\Sigma^\dagger) + \frac{\lambda}{2} [\mathrm{Tr}(\Sigma\Sigma^\dagger)]^2 + \frac{\kappa}{2} \mathrm{Tr}(\Sigma\Sigma^\dagger\Sigma\Sigma^\dagger) \tag{4.3}$$

for each of the link fields induces a VEV

$$\langle \Sigma \rangle = \frac{m_\Sigma}{\sqrt{\kappa + 3\lambda}} \mathbb{1}_3 \equiv v_\Sigma \mathbb{1}_3 \,, \tag{4.4}$$

which for simplicity we take to be the same for all link fields. Each symmetry breaking VEV results in one massive gauge and one massive scalar multiplet in the adjoint representation of the unbroken diagonal group. The masses of gauge and scalar multiplets are of the order[13] $g_i v_\Sigma$ and κv_Σ and they can be integrated out. The dynamical scale of the low energy effective field theory is given by

$$\Lambda_{\mathrm{QCD}}^{b_{\mathrm{QCD}}} = \frac{\prod_i^k \Lambda_i^{b_i}}{M^{\sum_i b_i - b_{\mathrm{QCD}}}} \,, \tag{4.5}$$

where the matching scale M is the geometric mean of the eigenvalues of the mass matrix for the heavy states. In terms of the QCD coupling constant g_s this implies the usual matching relation at M

$$\frac{1}{g_s^2(M)} = \sum_{i=1}^{k} \frac{1}{g_i^2(M)} \,, \tag{4.6}$$

The QCD θ angle is simply the sum of the individual $\mathrm{SU}(3)_i$ θ angles

$$\bar{\theta}_{\mathrm{QCD}} = \sum_{i=1}^{k} \bar{\theta}_i \,, \tag{4.7}$$

where $\bar{\theta} = \theta + \arg\det M_f$ is the physical theta angle, which contains a possible CP violating phase from the fermion mass matrix. At the same time one also has to integrate out the small instantons in the UV theory, which generate a potential for the axions. Thus the effective Lagrangian for the axion fields takes the form[14]

$$\mathcal{L}_a = \sum_{i=1}^{k} m_{a_i}^2 f_{a_i}^2 \cos\left(\frac{a_i}{f_{a_i}} - \bar{\theta}_i \right) + \frac{g_s^2}{32\pi^2} \sum_{i=1}^{k} \left(\frac{a_i}{f_{a_i}} - \bar{\theta}_i \right) G^a_{\mu\nu} \tilde{G}_i^{\mu\nu} \,. \tag{4.8}$$

[12] One can add U(1) factors to forbid terms like $\mu \det \Sigma$ [12].

[13] For simplicity we will assume that $g^2/\kappa \sim 1$ and will not distinguish between the gauge boson and scalar thresholds.

[14] In [12] the mass scale of the potential $m_{a_i}^2 f_{a_i}^2$ was denoted Λ_i^4.

JHEP04(2020)031

One can see $\bar{\theta}_{\rm QCD}$ is relaxed to zero due to two independent effects. First, small instanton contributions in broken gauge factors relax each individual $\bar{\theta}_i$ to zero. In addition, once QCD confines, the potential is generated for the linear combination $a/f_a = \sum_i a_i/f_{a_i}$ which relaxes $\bar{\theta}_{\rm QCD}$ to zero just like the usual axion would. In contrast to standard axion models there is not just one but k axions in the IR spectrum and it is the lightest mass eigenstate which plays the role of the QCD axion. When small instanton contributions are dominant the mass of this lightest state can be significantly higher than the standard QCD prediction in eq. (1.1).

4.2 Small instanton contributions

When working in the EFT one has to take into account the instanton configurations which are not mapped to the low energy theory, i.e. QCD. These are the independent 1-instanton contributions from $\mathrm{SU}(3)_1, \ldots, \mathrm{SU}(3)_k$. Since they are broken to their diagonal combination each SU(3) factor considered separately is completely broken and therefore we can use the formalism of section 3 with three Higgs scalars ϕ_n, $n = 1, 2, 3$ for each link field, which develop a VEV[15]

$$\langle \phi_{in} \rangle = v_\Sigma \, \delta_{in} \,, \tag{4.9}$$

where $i = 1, 2, 3$ are the SU(3) indices. This allows us to evaluate the classical action for the Higgs scalars from one of the link fields in the instanton background explicitly

$$S_0^\phi(\rho) = 2\pi^2\rho^2 \sum_{i=1}^{2}\sum_{n=1}^{3} |\langle \tilde{\phi}_{in} \rangle|^2 = 2\pi^2\rho^2 v_\Sigma^2 \sum_{i=1}^{2}\sum_{n=1}^{3} |\Omega_{in}|^2 = 4\pi^2\rho^2 v_\Sigma^2 \,, \tag{4.10}$$

where we considered the rotated VEVs (see eq. (3.22)) to account for arbitrary instanton locations inside SU(3).[16] The result is independent of $\tilde{\mu}$ and we can therefore do the now trivial group integration in the results of section 3.3. Note that the scalar action for $\mathrm{SU}(3)_2, \ldots, \mathrm{SU}(3)_{k-1}$ is twice as large, since each of them couples to two link fields.

We begin by considering the SU(3) sectors without fermions. The last of these sectors, i.e. $\mathrm{SU}(3)_k$, has only one scalar link, i.e. $S = 3$ scalars in the fundamental representation, and the beta function coefficient $b_k = 21/2$. For this sector the vacuum-vacuum amplitude contributes directly to the axion potential (see eq. (3.28)) with a mass scale m_{a_k} of

$$m_{a_k}^2 f_{a_k}^2 = \left(\frac{\Lambda_k}{M}\right)^{b_k} \left(\frac{M}{2\pi v_\Sigma}\right)^{b_k - 4} M^4 \,, \tag{4.11}$$

where the factor $(M/2\pi v_\Sigma)^{b_k-4}$ converts between the physical mass threshold at M and the effective cutoff of the instanton size integral at $1/\rho \sim 2\pi v_\Sigma$, while the RG invariant scale of $\mathrm{SU}(3)_k$ sector is defined by

$$\Lambda_k^{b_k} = d_3(M)|_{S_k, F=0} \, \Gamma[b_k/2 - 2] M^{b_k} \,, \tag{4.12}$$

[15] From the point of view of one of the $\mathrm{SU}(3)_i$ factors the bifundamental $\Sigma_{i\,i+1}$ looks like three scalars in the fundamental representation.

[16] Note that the explicit form of Ω in SU(3) is not needed to obtain the factor of 2. $\sum_{i=1}^{2}\sum_{f=1}^{3} |\Omega_{if}|^2$ sums the norms of the first two row vectors in Ω and since $\Omega \in \mathrm{SU}(3)$ each row vector is normalized to unity.

JHEP04(2020)031

and the instanton weight $d_3(M)|_{S_k,F=0}$ is given in (3.11). The remaining sectors $i = 2,\ldots,k-1$ have two link fields, i.e. $S = 6$ scalars in the fundamental representation and the beta function coefficient $b_i = 10$. The vacuum-vacuum amplitude contributes to the axion potential in these sectors with a mass scale m_{a_i} which is given by eq. (4.11) after the replacement $k \to i$ and $v_\Sigma \to \sqrt{2}\, v_\Sigma$. The additional suppression by $2^{2-b_i/2}$ originates from the scalar action which is twice as large, since all of these sectors couple to two link fields.

All the SM quarks are charged under the $\mathrm{SU}(3)_1$ sector. Thus its particle content is characterized by $F = 6$ approximately massless fermions[17] and $S = 3$ scalars in the fundamental representation, corresponding to a beta function coefficient of $b_1 = 13/2$. Taking the result for the vacuum energy in the instanton background with massless quarks and a Yukawa interaction from eq. (3.38) for $N = 3$ and $\theta = \bar{\theta}_1 - \frac{a_1}{f_{a_1}}$ and matching it to the axion potential in eq. (4.8) we obtain the scale $m_{a_1}^2 f_{a_1}^2$

$$m_{a_1}^2 f_{a_1}^2 = K \int \frac{d\rho}{\rho^5}\, 2C_3(\rho)\,. \tag{4.13}$$

where K is given by

$$K = \frac{40}{9} \frac{y_u y_d y_s y_c y_b y_t}{(16\pi^2)^3}\,. \tag{4.14}$$

Note that K reproduces a loop factor expected from an EFT diagram in figure 1(b) and included in the results of [12]. However, the full calculation of correlation functions in the instanton background performed in section 3.3 allows us to extract the exact numerical coefficient multiplying this loop factor. Performing the ρ integral in eq. (4.13) we find

$$m_{a_1}^2 f_{a_1}^2 = K \left(\frac{\Lambda_1}{M}\right)^{b_1} \left(\frac{M}{2\pi v_\Sigma}\right)^{b_1-4} M^4\,, \tag{4.15}$$

where the dynamical scale of $\mathrm{SU}(3)_1$ is defined by

$$\Lambda_1^{b_1} = d_3(M)|_{S=3,F=6}\, \Gamma[b_1/2-2] M^{b_1}\,, \tag{4.16}$$

and once again the instanton weight $d_3(M)|_{S=3,F=6}$ is given in (3.11). Note that these results are in agreement with the qualitative discussion of section 2.

The unusual scaling of the axion mass with the physical QCD scale can be seen from the fact that $d_3(M) \sim \exp(-\frac{8\pi^2}{g_i^2(M)})$ where g_i^2 is the coupling of the ith SU(3) factor rather than the actual QCD coupling, implying that $\Lambda_i^{b_i}$ will be a fractional power of $\Lambda_{\mathrm{QCD}}^{b_{\mathrm{QCD}}}$, where the actual fraction depends on the ratios of coupling strengths and the distribution of the matter fields among the different group factors.

However the full expression of the corrections to the axion mass eqs. (4.11)–(4.15) also includes an additional suppression factors, for example the conversion factor $(M/2\pi v_\Sigma)^{b_i-4}$. Indeed the presence of this factor implies that, up to an order one coefficient, our results for $m_{a_i}^2 f_{a_i}^2$ are smaller than the previous estimates $(\tilde{m}_{a_i}^2 f_{a_i}^2)$ in [12] by a factor of

$$\frac{m_{a_i}^2}{\tilde{m}_{a_i}^2} \simeq 2^{-6} \cdot \left(\frac{M}{2\pi v_\Sigma}\right)^{b_i-4}\,. \tag{4.17}$$

[17] To a good approximation all SM quarks are massless at scales $M \gg$ TeV.

JHEP04(2020)031

This suppression is due to two independent reasons:

- Our vacuum instanton density $d_N(\rho)|_{S=F=0}$ is smaller by a factor of 2^{-2N} than the one used in [12]. This discrepancy originates from a small error in 't Hooft's original calculation [40], which was later corrected in an Erratum. However, the source for the instanton density [45] cited in [12] still contains this error.

- In [12] the ρ integration was cut off at $\rho = 1/M$ by hand. However, when working in the constrained instanton framework the ρ integral is convergent and we find that the actual cutoff is roughly $\rho \sim 1/(2\pi v_\Sigma)$ (see also [15]).

The actual size of the suppression depends on the relation between the matching scale M and the VEV v_Σ. Since M corresponds to the mass scale of the massive gauge bosons, it scales like $M = g_{\text{eff}}\, v_\Sigma$, where g_{eff} is some combination of $g_1, \ldots, g_k$. For couplings of $\mathcal{O}(1)$ this leads to a suppression of $(2\pi)^{4-b_i}$, which is strongest for the SU(3) group factors that do not couple to fermions. As we will show momentarily, this suppression is significant in the minimal model with only two group factors, but is less important once more SU(3) factors are included and the matching relation in eq. (4.6) allows larger couplings in the individual SU(3) factors.

4.3 Example $\mathbf{SU(3)^2}$, $\mathbf{SU(3)^3 \to SU(3)_{QCD}}$

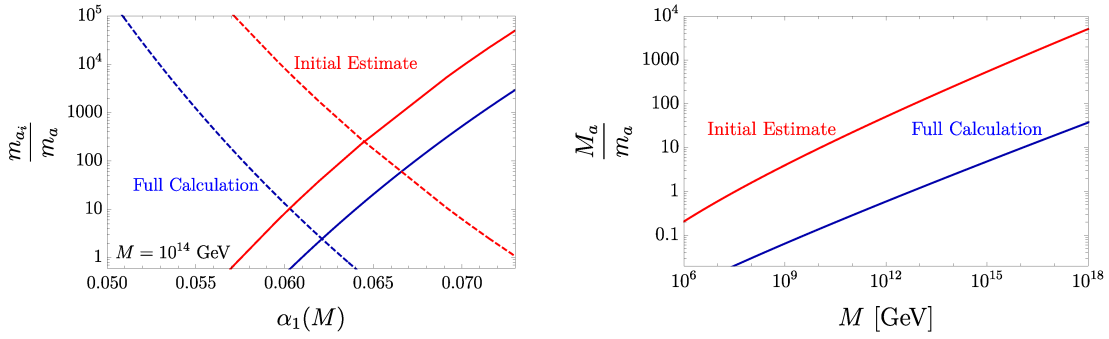
Let us now have a look at the minimal model with $k = 2$. In this case the matching scale is directly set by masses of the heavy gauge bosons (and scalars)

$$M^2 = (g_1^2 + g_2^2) v_\Sigma^2 \,. \tag{4.18}$$

In order to do the matching we use the RG equation to run the $\overline{\text{MS}}$ QCD coupling from the top mass at $\alpha_s(m_t) = 0.10$ to the matching scale M. The small instanton contribution to the axion mass relative to the QCD contribution can now be computed using the mass scales $m_{a_2}^2 f_{a_2}^2$ and $m_{a_1}^2 f_{a_1}^2$ from eqs. (4.11) and (4.15) respectively. For simplicity we will assume that $f_{a_1} = f_{a_2} = f_a$ and use eq. (1.1) to obtain a numerical value for $f_a m_a = (75.5\text{ MeV})^2$.

This ratio is shown for both axions (solid for m_{a_1}/m_a and dashed for m_{a_2}/m_a) for the choice of $M = 10^{14}\,\text{GeV}$ for the symmetry breaking scale in figure 3(a). In contrast to previous estimates [12] (shown in red), the full calculation shows that there is no region in parameter space where both axion masses are enhanced by more than an $\mathcal{O}(1)$ factor compared to the pure QCD prediction at the same time. One of the axions might be heavy, but then the other will be dominated by the QCD contribution to its potential and will therefore be like the standard QCD axion. The largest effect of small instanton contributions to both axion masses is found at the intersection of the two curves where both axions have the same mass which is about $M_a/m_a = 2.4$ times heavier than the standard QCD axion. Figure 3(b) shows the maximal enhancement of the axion mass due to small instantons as a function of the symmetry breaking scale M. This shows that even taking M to be at the Planck scale the axion mass cannot deviate by more than a factor of ~ 100 from the QCD prediction. Due to the suppression factor in eq. (4.17) the enhancement is lower by about two orders of magnitude than the initial prediction in [12].

JHEP04(2020)031



(a) (b)

Figure 3. Small instanton contribution to the axion mass relative to the IR QCD contributions in the model based on the symmetry breaking structure $\mathrm{SU}(3) \times \mathrm{SU}(3) \to \mathrm{SU}(3)_{\mathrm{QCD}}$. On the left we show our results for the full calculation in blue compared to the initial estimates in previous work [12] in red at a breaking scale of $M = 10^{14}\,\mathrm{GeV}$ as a function of $\alpha_1 = \frac{g_1^2}{4\pi}$. The solid (dashed) curves show m_{a_1}/m_a (m_{a_2}/m_a), which intersect at $M_a/m_a = 2.4$ in the full calculation and at $M_a/m_a = 251$ in previous estimates. On the right we show the values for M_a/m_a at the intersection point of m_{a_1}/m_a and m_{a_2}/m_a for a wide range of breaking scales M. In both plots we took $f_{a_1} = f_{a_2} = f_a$.

We can therefore conclude that it is hard to get significant contributions from small instantons to the axion mass in the minimal model. However, according to our parametric estimate in section 2, we expect a larger mass enhancement in models with more SU(3) factors. In the following we demonstrate that this conclusion is indeed correct by considering the next to minimal model with $k = 3$ factors.

In the model with $k = 3$ group factors $\mathrm{SU}(3)^3$ is broken by the VEVs of two link fields, which we both take to be $\langle \Sigma \rangle = v_\Sigma \mathbb{1}_3$. Note that since $\mathrm{SU}(3)_2$ couples to both link fields, not all gauge bosons get the same masses. One linear combination, corresponding to the QCD gluons, stays massless as before, whereas the masses of the other two linear combinations are given by

$$M^2_{V_{1/2}} = \frac{v_\Sigma^2}{2}\left(g_1^2 + 2g_2^2 + g_3^2 \pm \sqrt{4g_2^4 + (g_1^2 - g_3^2)^2} \right). \tag{4.19}$$

The matching threshold is given by the geometric mean of these two mass eigenvalues

$$M = \left(g_1^2 g_2^2 + g_2^2 g_3^2 + g_1^2 g_3^2\right)^{1/4} v_\Sigma\,. \tag{4.20}$$

As in the minimal model we take $f_{a_1} = f_{a_2} = f_{a_3} = f_a$ and show our result (blue) in figure 4 for the small instanton contributions to the axion mass compared to the estimates in [12] (red), fixing in both cases $g_2 = g_3$. In figure 4(a) we again show m_{a_1}/m_a (solid) and m_{a_2}/m_a (dashed) at a breaking scale of $M = 10^{14}\,\mathrm{GeV}$. Note that m_{a_3} is always larger than m_{a_2} for identical couplings, since m_{a_2} is suppressed by an additional factor of $2^{2-b_2/2}$. As can be seen, even though the mass enhancement is again smaller in the full calculation than in the initial estimate, small instantons can still enhance the mass of all three axions

JHEP04(2020)031

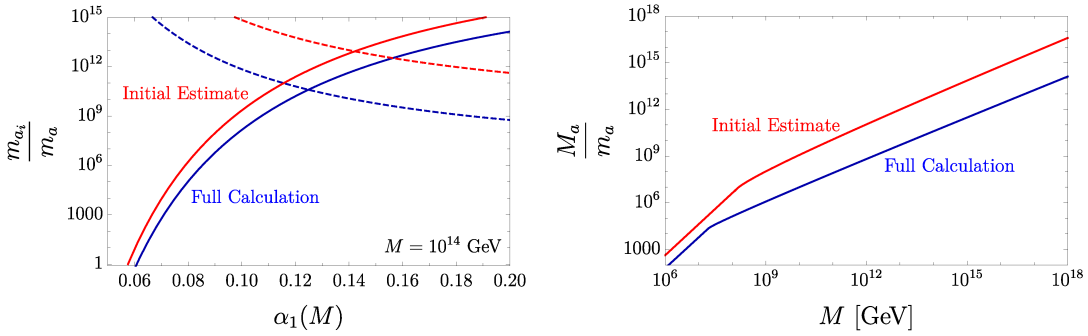


Figure 4. Small instanton contribution to the axion mass relative to the IR QCD contributions in the model based on the symmetry breaking structure $\mathrm{SU}(3) \times \mathrm{SU}(3) \times \mathrm{SU}(3) \to \mathrm{SU}(3)_{\mathrm{QCD}}$. On the left we show the results for the full calculation in blue compared to the initial estimates in previous work [12] in red at a breaking scale of $M = 10^{14}\,\mathrm{GeV}$ as a function of $\alpha_1 = \frac{g_1^2}{4\pi}$. The solid (dashed) curves show m_{a_1}/m_a (m_{a_2}/m_a), which intersect at $M_a/m_a \simeq 4 \cdot 10^{10}$ in the full calculation and at $M_a/m_a = 9 \cdot 10^{12}$ in previous estimates. On the right we show the values for M_a/m_a at the intersection point of m_{a_1}/m_a and m_{a_2}/m_a or at the maximum of m_{a_1}/m_a if they do not intersect for a wide range of breaking scales M. In both plots we took $f_{a_1} = f_{a_2} = f_{a_3} = f_a$ and fixed $g_2 = g_3$, which implies that m_{a_3} is always slightly larger than m_{a_2}.

simultaneously by up to a factor of $4 \cdot 10^{10}$ compared to the QCD contribution at the intersection point. Figure 4(b) shows that small instantons give dominant contributions to the axion mass also at smaller breaking scales M, making the axion considerably heavier than in the standard QCD axion scenario. Note that at small M, m_{a_1}/m_a and m_{a_2}/m_a do not intersect anymore. When this is the case $m_{a_1}/m_a < m_{a_2}/m_a$ due to its suppression by the Yukawa couplings and therefore we take the maximum of m_{a_1}/m_a as an estimate for the maximal simultaneous enhancement of all axion masses. This is the reason for the kink in the curves in the figure 4(b).

Adding additional SU(3) factors increases the possible enhancement of the axion mass even further. It was already noted in [12] that for $k \gg 1$ the axion masses scale as $m_{a_i} \sim M^2/f_{a_i}$ for $i = 2, \ldots, k$ and $m_{a_1} \sim \sqrt{K} M^2/f_{a_1}$, where the first axion mass is parametrically suppressed relative to the others by the Yukawa couplings and loop factors $\sqrt{K} \approx 10^{-12}$. With the help of eq. (2.14) we can now understand the scaling of the axion mass with M^2 as the limit $b_{\mathrm{QCD}}/k \xrightarrow{k\to\infty} 0$ in eq. (2.14).

5 Conclusions

We have presented a full calculation of the effects of small instantons on the axion mass in product group extensions of QCD. We found that a non-trivial embedding of the QCD instanton into the UV group will lead to an unusual scaling of small instanton contributions, which will appear as fractional instantons from the low-energy point of view. This opens up the possibility for small instantons in partially broken gauge groups to dominate over

JHEP04(2020)031

the large QCD instantons and significantly raise the axion mass away from the usual $m_a^2 - f_a$ relation.

We have carried out the full 1-instanton calculation of the vacuum-to-vacuum transition amplitude of the broken product gauge group theories. As a first step we calculated the 1-instanton contribution to the vacuum bubble for a fully broken bosonic SU(N) theory by performing the integral over the bosonic zero modes and non-zero modes. Since the gauge group is broken the effects of large instantons are exponentially cut off, and the calculation can be reliably carried out. While the inclusion of fermions and their interactions is usually handled using a 't Hooft operator approximation, we were able to evaluate the effects of the fermionic modes along with the Higgs loops needed to close up the zero modes directly without resorting to the 't Hooft operator approximation. This has the advantage that the result is manifestly finite and does not require the introduction of a regulator via a cutoff (whose exact definition in simple estimates usually leads to some uncertainty on the exact numerical value of the corrections).

Using this method we were able to perform the calculation in the full theory with product groups broken to the diagonal and verify the scalings expected from our simple estimates. While the numerical value of the enhancement is not significant for the simplest 2 product group extension, already for 3 group factors we can obtain a large enhancement of the axion mass.

Acknowledgments

We thank Paddy Fox, Tony Gherghetta, Valya Khoze, Alex Pomarol and Stefan Stelzl for useful discussions. C.C., M.R. and Y.S. thank the MIAPP at TU Munich for its hospitality while this project was initiated. C.C. also thanks the Aspen Center for Physics and the KITP at Santa Barbara for their hospitality while this work was in progress. M.R. thanks the Cornell particle theory group for its hospitality while this research was in progress. Y.S. thanks the Aspen Center for Physics for the hospitality while this work was in progress. The research of C.C. was supported in part by the NSF grant PHY-1719877 and in part by the US-Israeli BSF grant 2016153, as well as a Humboldt Foundation research prize. The work of M.R. was partially supported by the Collaborative Research Center SFB1258, by the DFG under Germany's Excellence Strategy — EXC-2094 — 390783311 and by the Studienstiftung des deutschen Volkes. Y.S. was supported in part by the NSF grant PHY-1915005.

A 't Hooft operator approach

In this appendix we compute the small instanton contribution to the vacuum energy or axion potential in the presence of massless fermions using the 't Hooft operator approximation and compare it to the full calculation in section 3.3.

In a gauge theory with F massless fermion flavors in the fundamental representation of SU(N) the pure vacuum-vacuum amplitude in the instanton background vanishes and the instanton configuration only contributes to correlation functions in which each fermion

JHEP04(2020)031

flavor and chirality appears at least once, i.e. for example $\langle 0|\prod_f(\bar{\psi}_f\psi_f)|0\rangle_{\Delta Q=1}\neq 0$. The effect of the instanton can thus be captured by the 't Hooft operator, which is an effective $2F$ fermion operator of the form (see e.g. [40])

$$-\delta\mathcal{L}^F = e^{-i\theta}\int\frac{d\rho}{\rho^5}C_N(\rho)\rho^{3F}\left(\kappa_N^{(N_f)}\right)^{i_1\cdots i_{2F}}\det_{f,f'}\left(\bar{\psi}_{R\,f}(x_0)\psi_{L\,f'}(x_0)\right)_{i_1\cdots i_{2F}}+\text{h.c.}\,, \tag{A.1}$$

where the determinant goes over flavor indices and the hermitian conjugate results from the anti-instanton configuration. $C_N(\rho)$ is defined in eq. (3.29) and $\left(\kappa_N^{(N_f)}\right)^{i_1\cdots i_{2F}}$ is obtained by computing the $2F$ fermion correlation function in the instanton background and matching the result to the above effective operator. Note that the integration over the instanton location inside SU(N), for which we assumed that $\sum_{i=1}^{2}\sum_{n=1}^{3}|\langle\tilde{\phi}_{in}\rangle|^2$ inside $C_N(\rho)$ is independent of the instanton position, projects out all invariant contractions of the fermion SU(N) indices $i_1,\ldots,i_F$. For one fermion flavor the matching is straightforward (see [46] for an example in SU(2) and SU(3)) and gives

$$-\delta\mathcal{L}^{F=1} = e^{-i\theta}\int\frac{d\rho}{\rho^5}C_N(\rho)\rho^3\kappa_N^{(1)}\bar{\psi}_{R\,1}(x_0)\psi_{L\,1}(x_0)+\text{h.c.}\,, \tag{A.2}$$

where we used that $\left(\kappa_N^{(1)}\right)^{i_1 i_2}=\kappa_N^{(1)}\delta^{i_1 i_2}$ (for example for SU(3): $\kappa_3^{(1)}=\frac{4\pi^2}{3}$).

Since we want to close the 't Hooft operator with Higgs loops, we are only interested in flavor diagonal SU(N) contractions of the form $(\bar{\psi}_R\psi_L)^F$. Therefore we will consider the effective Lagrangian

$$-\delta\mathcal{L}^F \simeq e^{-i\theta}\int\frac{d\rho}{\rho^5}C_N(\rho)(\rho^3\kappa_N^{(1)})^F\prod_{f=1}^{F}\bar{\psi}_{R\,f}(x_0)\psi_{L\,f}(x_0)+\text{h.c.}\,. \tag{A.3}$$

Note that due to Fierz relations among SU(N) invariants, the prefactor $(\kappa_N^{(1)})^F$ is not exact, but will deviate from the full prefactor by an $\mathcal{O}(1)$ factor.

Such a 't Hooft operator contributes to the axion potential if one closes the fermion legs with loops. The leading contribution arises from closing the operator with Higgs loops via Yukawa couplings to the fermions as shown in figure 1(b). This is the case since the diagram only includes marginal couplings and therefore scales as M_{cut}^{3F} where M_{cut} is the cutoff for the divergent loop integrals.

Focusing on SU(3) and identifying $\theta=\bar{\theta}-a_1/f_{a_1}$, we can match the resulting operator to the effective Lagrangian in eq. (4.8) to obtain $m_{a_1}f_{a_1}$

$$m_{a_1}^2 f_{a_1}^2 = 2K\int\frac{d\rho}{\rho^5}C_3(\rho)(4\pi^2M^3\rho^3)^F\,, \tag{A.4}$$

where K contains the Yukawa couplings and loop factors

$$K=\prod_{f=1}^{F}\frac{y_f}{4\pi}\,. \tag{A.5}$$

Note that we canceled a factor $N=3$ from the sum over colors in the loop for each fermion flavor with the 3^{-F} from $(\kappa_3^{(1)})^F$. Computing the ρ integral one obtains

$$m_{a_1}^2 f_{a_1}^2 = K\,d_3(M)|_{S=3,F}\,(4\pi^2)^F\Gamma\left[\frac{3F+b_0^{(1)}-4}{2}\right]\left(\frac{M}{2\pi v_\Sigma}\right)^{b_0^{(1)}-4}\left(\frac{M_{\text{cut}}}{2\pi v_\Sigma}\right)^{3F}M^4\,, \tag{A.6}$$

where M is the matching scale for the couplings.

JHEP04(2020)031

Comparing this result to eq. (4.15), which was obtained by including the SM Higgs and Yukawa couplings directly in the path integral evaluation of the vacuum-vacuum amplitude, one finds that with $M_{\rm cut}$, the cutoff of the loop integrals, an additional scale appears. However, the exact definition of M_{cut} is ambiguous and always introduces an uncertainty. Since $M_{\rm cut}$ enters $m_{a_1}^2 f_{a_1}^2$ with a large power, even $\mathcal{O}(1)$ changes in the definition of $M_{\rm cut}$ can have a significant impact on $m_{a_1}^2 f_{a_1}^2$. This ambiguity is removed in the calculation in section 3.3, since the result is manifestly finite.

Both methods are equivalent and therefore we can use the result from section 3.3 to infer the appropriate definition of $M_{\rm cut}$ for this process. We find that both approaches yield the same result, up to an $\mathcal{O}(1)$ factor, if one identifies $M_{\rm cut} \simeq v_\Sigma$, in nice agreement with our intuitive expectations.

B Converting results to $\overline{\rm MS}$ scheme

All results in section 3 were derived in the Pauli-Villars regularization scheme. However, in perturbative calculations dimensional regularization and the MS or $\overline{\rm MS}$ scheme are more common. In this appendix we briefly summarize how to convert the results to these schemes.

Already in [40] 't Hooft showed that in order to convert the results to dimensional regularization one has to do the substitutions

$$\ln\mu_0 \to \frac{1}{4-n} - \frac{1}{2}\gamma + \frac{1}{2}\ln 4\pi \qquad \text{(zero-modes)}\,, \tag{B.1}$$

$$\ln\mu_0 \to \frac{1}{4-n} - \frac{1}{2}\gamma + \frac{1}{2}\ln 4\pi - \frac{1}{2} \qquad \text{(kinetic terms)}\,, \tag{B.2}$$

where the first substitution has to be made for the μ_0 originating from gauge and fermion zero-modes and the second for μ_0 from kinetic terms, i.e. from the non-zero modes and scalar fields.[18] This substitution only affects the running coupling in the exponential

$$-\frac{8\pi^2}{g^2(1/\rho)} = -\frac{8\pi^2}{g_B^2(\mu_0)} + \ln(\mu_0\rho)\left[(4N-F) + \left(\frac{1}{3}F - \frac{1}{3}N - \frac{1}{6}S\right)\right] \tag{B.3}$$

$$\to \frac{8\pi^2}{g_B^2(n)} + \left(\ln\rho + \frac{1}{4-n} + \frac{1}{2}(\ln 4\pi - \gamma)\right)b_0 - \frac{1}{2}\left(\frac{1}{3}F - \frac{1}{3}N - \frac{1}{6}S\right), \tag{B.4}$$

where we separated in eq. (B.3) the contributions to b_0 originating from zero modes (first bracket) from the ones from non-zero modes (second bracket). The renormalized coupling now depends on the renormalization scheme. Here we will consider MS and $\overline{\rm MS}$ scheme

[18] In [40] 't Hooft found $-\frac{5}{12}$ instead of the $-\frac{1}{2}$ in eq. (B.2). This mistake was noted by Hasenfratz and Hasenfratz [47] and reconciled the disagreement with Shore, who did the instanton calculation using dimensional regularization [48]. 't Hooft corrected the $-\frac{5}{12}$ in eq. (B12) of [49] to -1. However, this was later again corrected by Shifman [43] to the $-\frac{1}{2}$ we use in eq. (B.2).

JHEP04(2020)031

which are defined by

$$\frac{8\pi^2}{g^2_{\rm MS}(1/\rho)} = \frac{8\pi^2}{g^2_B(n)} + \left(\ln\rho + \frac{1}{4-n}\right) b_0\,, \tag{B.5}$$

$$\frac{8\pi^2}{g^2_{\overline{\rm MS}}(1/\rho)} = \frac{8\pi^2}{g^2_B(n)} + \left(\ln\rho + \frac{1}{4-n} + \frac{1}{2}(\ln 4\pi - \gamma)\right) b_0\,. \tag{B.6}$$

Note that in the above we have to identify [40]

$$g_B(n) \to g_{\rm MS}(\mu) \qquad \text{and} \qquad \ln\rho + \frac{1}{4-n} \to \ln(\rho\mu)\,, \tag{B.7}$$

$$g_B(n) \to g_{\overline{\rm MS}}(\mu) \qquad \text{and} \qquad \ln\rho + \frac{1}{4-n} + \frac{1}{2}(\ln 4\pi - \gamma) \to \ln(\rho\mu)\,, \tag{B.8}$$

where μ is the renormalization scale in dimensional regularization. Thus to convert our results to $\overline{\rm MS}$ scheme we have to replace

$$e^{-8\pi^2/g^2(1/\rho) - C_2 N} \to e^{-\frac{1}{12}(2F-S)} e^{-8\pi^2/g^2_{\overline{\rm MS}}(1/\rho) - C_2^{\overline{\rm MS}} N}\,, \tag{B.9}$$

with $C_2^{\overline{\rm MS}}$ given by

$$C_2^{\overline{\rm MS}} = C_2 - \frac{1}{6}\,. \tag{B.10}$$

Using this the instanton density in $\overline{\rm MS}$ scheme is given by

$$d_N^{\overline{\rm MS}}(\rho)\Big|_{F,S} = e^{-\frac{1}{12}(2F-S)+\frac{1}{6}N}\, d_N(\rho)|_{F,S}\,. \tag{B.11}$$

References

[1] R.D. Peccei and H.R. Quinn, *CP conservation in the presence of instantons*, *Phys. Rev. Lett.* **38** (1977) 1440 [INSPIRE].

[2] R.D. Peccei and H.R. Quinn, *Constraints Imposed by CP conservation in the presence of instantons*, *Phys. Rev.* **D 16** (1977) 1791 [INSPIRE].

[3] S. Weinberg, *A new light boson?*, *Phys. Rev. Lett.* **40** (1978) 223 [INSPIRE].

[4] F. Wilczek, *Problem of strong P and T invariance in the presence of instantons*, *Phys. Rev. Lett.* **40** (1978) 279 [INSPIRE].

[5] A. Hook, *TASI lectures on the strong CP problem and axions*, PoS(TASI2018)004 [arXiv:1812.02669] [INSPIRE].

[6] G. Grilli di Cortona, E. Hardy, J. Pardo Vega and G. Villadoro, *The QCD axion, precisely*, *JHEP* **01** (2016) 034 [arXiv:1511.02867] [INSPIRE].

[7] B. Holdom and M.E. Peskin, *Raising the axion mass*, *Nucl. Phys.* **B 208** (1982) 397 [INSPIRE].

JHEP04(2020)031

[8] B. Holdom, *Strong QCD at high-energies and a heavy axion*, *Phys. Lett.* **B 154** (1985) 316 [*Erratum ibid.* **B 156** (1985) 452].

[9] J.M. Flynn and L. Randall, *A computation of the small instanton contribution to the axion potential*, *Nucl. Phys.* **B 293** (1987) 731 [INSPIRE].

[10] E. Poppitz and Y. Shirman, *The strength of small instanton amplitudes in gauge theories with compact extra dimensions*, *JHEP* **07** (2002) 041 [hep-th/0204075] [INSPIRE].

[11] T. Gherghetta, V.V. Khoze, A. Pomarol and Y. Shirman, *The axion mass from 5D small instantons*, *JHEP* **03** (2020) 063 [arXiv:2001.05610] [INSPIRE].

[12] P. Agrawal and K. Howe, *Factoring the strong CP problem*, *JHEP* **12** (2018) 029 [arXiv:1710.04213] [INSPIRE].

[13] P. Agrawal and K. Howe, *A flavorful factoring of the strong CP problem*, *JHEP* **12** (2018) 035 [arXiv:1712.05803] [INSPIRE].

[14] M.K. Gaillard et al., *Color unified dynamical axion*, *Eur. Phys. J.* **C 78** (2018) 972 [arXiv:1805.06465] [INSPIRE].

[15] J. Fuentes-Martín, M. Reig and A. Vicente, *Strong CP problem with low-energy emergent QCD: the* 4321 *case*, *Phys. Rev.* **D 100** (2019) 115028 [arXiv:1907.02550] [INSPIRE].

[16] D. Croon, R. Houtz and V. Sanz, *Dynamical axions and gravitational waves*, *JHEP* **07** (2019) 146 [arXiv:1904.10967] [INSPIRE].

[17] M. Reig, *On the high-scale instanton interference effect: axion models without domain wall problem*, *JHEP* **08** (2019) 167 [arXiv:1901.00203] [INSPIRE].

[18] M.B. Gavela, M. Ibe, P. Quilez and T.T. Yanagida, *Automatic Peccei–Quinn symmetry*, *Eur. Phys. J.* **C 79** (2019) 542 [arXiv:1812.08174] [INSPIRE].

[19] M.A. Buen-Abad and J. Fan, *Dynamical axion misalignment with small instantons*, *JHEP* **12** (2019) 161 [arXiv:1911.05737] [INSPIRE].

[20] M. Dine and N. Seiberg, *String theory and the strong CP problem*, *Nucl. Phys.* **B 273** (1986) 109 [INSPIRE].

[21] K. Choi, C.W. Kim and W.K. Sze, *Mass renormalization by instantons and the strong CP problem*, *Phys. Rev. Lett.* **61** (1988) 794 [INSPIRE].

[22] K. Choi and H.D. Kim, *Small instanton contribution to the axion potential in supersymmetric models*, *Phys. Rev.* **D 59** (1999) 072001 [hep-ph/9809286] [INSPIRE].

[23] A. Hook, *Solving the hierarchy problem discretely*, *Phys. Rev. Lett.* **120** (2018) 261802 [arXiv:1802.10093] [INSPIRE].

[24] V.A. Rubakov, *Grand unification and heavy axion*, *JETP Lett.* **65** (1997) 621 [hep-ph/9703409] [INSPIRE].

[25] H. Fukuda, K. Harigaya, M. Ibe and T.T. Yanagida, *Model of visible QCD axion*, *Phys. Rev.* **D 92** (2015) 015021 [arXiv:1504.06084] [INSPIRE].

[26] Z. Berezhiani, L. Gianfagna and M. Giannotti, *Strong CP problem and mirror world: the Weinberg-Wilczek axion revisited*, *Phys. Lett.* **B 500** (2001) 286 [hep-ph/0009290] [INSPIRE].

[27] A. Hook, *Anomalous solutions to the strong CP problem*, *Phys. Rev. Lett.* **114** (2015) 141801 [arXiv:1411.3325] [INSPIRE].

[28] N. Blinov and A. Hook, *Solving the wrong hierarchy problem*, *JHEP* **06** (2016) 176 [arXiv:1605.03178] [INSPIRE].

JHEP04(2020)031

[29] S. Dimopoulos, A. Hook, J. Huang and G. Marques-Tavares, *A collider observable QCD axion*, *JHEP* **11** (2016) 052 [arXiv:1606.03097] [INSPIRE].

[30] T. Gherghetta, N. Nagata and M. Shifman, *A visible QCD axion from an enlarged color group*, *Phys. Rev.* **D 93** (2016) 115010 [arXiv:1604.01127] [INSPIRE].

[31] K.A. Intriligator and N. Seiberg, *Duality, monopoles, dyons, confinement and oblique confinement in supersymmetric* $\mathrm{SO}(N_c)$ *gauge theories*, *Nucl. Phys.* **B 444** (1995) 125 [hep-th/9503179] [INSPIRE].

[32] C. Csáki and H. Murayama, *Instantons in partially broken gauge groups*, *Nucl. Phys.* **B 532** (1998) 498 [hep-th/9804061] [INSPIRE].

[33] K.A. Intriligator, R.G. Leigh and N. Seiberg, *Exact superpotentials in four-dimensions*, *Phys. Rev.* **D 50** (1994) 1092 [hep-th/9403198] [INSPIRE].

[34] K.A. Intriligator and N. Seiberg, *Phases of* $N = 1$ *supersymmetric gauge theories in four-dimensions*, *Nucl. Phys.* **B 431** (1994) 551 [hep-th/9408155] [INSPIRE].

[35] J.E. Kim, *Weak interaction singlet and strong CP invariance*, *Phys. Rev. Lett.* **43** (1979) 103 [INSPIRE].

[36] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Can confinement ensure natural CP invariance of strong interactions?*, *Nucl. Phys.* **B 166** (1980) 493 [INSPIRE].

[37] M. Dine, W. Fischler and M. Srednicki, *A simple solution to the strong CP problem with a harmless axion*, *Phys. Lett.* **B 104** (1981) 199.

[38] A.R. Zhitnitsky, *On Possible Suppression of the Axion Hadron Interactions. (In Russian)*, *Sov. J. Nucl. Phys.* **31** (1980) 260 [INSPIRE].

[39] S.F. Cordes, *The instanton induced superpotential in supersymmetric QCD*, *Nucl. Phys.* **B 273** (1986) 629 [INSPIRE].

[40] 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].

[41] I. Affleck, *On constrained instantons*, *Nucl. Phys.* **B 191** (1981) 429 [INSPIRE].

[42] C.W. Bernard, *Gauge zero modes, instanton determinants and QCD calculations*, *Phys. Rev.* **D 19** (1979) 3013 [INSPIRE].

[43] M. Shifman, *Advanced topics in quantum field theory*, Cambridge University Press, Cambridge U.K. (2012).

[44] Y. Bai and B.A. Dobrescu, *Minimal* $\mathrm{SU}(3) \times \mathrm{SU}(3)$ *symmetry breaking patterns*, *Phys. Rev.* **D 97** (2018) 055024 [arXiv:1710.01456] [INSPIRE].

[45] C.G. Callan Jr., R.F. Dashen and D.J. Gross, *Toward a theory of the strong interactions*, *Phys. Rev.* **D 17** (1978) 2717 [INSPIRE].

[46] M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, *Instanton density in a theory with massless quarks*, *Nucl. Phys.* **B 163** (1980) 46 [INSPIRE].

[47] A. Hasenfratz and P. Hasenfratz, *The scales of Euclidean and Hamiltonian lattice QCD*, *Nucl. Phys.* **B 193** (1981) 210 [INSPIRE].

[48] G.M. Shore, *Dimensional regularization and instantons*, *Annals Phys.* **122** (1979) 321 [INSPIRE].

[49] G. 't Hooft, *How instantons solve the* U(1) *problem*, *Phys. Rept.* **142** (1986) 357 [INSPIRE].

JHEP04(2020)031