Eur. Phys. J. C (2019) 79:822
https://doi.org/10.1140/epjc/s10052-019-7304-4

THE EUROPEAN PHYSICAL JOURNAL C

Check for updates

Regular Article - Theoretical Physics

Axions are blind to anomalies

Jérémie Quevillon[a], **Christopher Smith**[b]

Laboratoire de Physique Subatomique et de Cosmologie, Université Grenoble-Alpes, CNRS/IN2P3, Grenoble INP, 38000 Grenoble, France

Received: 17 June 2019 / Accepted: 13 September 2019 / Published online: 5 October 2019
© The Author(s) 2019

Abstract The axion couplings to SM gauge bosons are derived in various models, and shown to always arise entirely from non-anomalous fermion loops. They are thus independent of the anomaly structure of the model. This fact is without consequence for vector gauge interactions like QCD and QED, but has a major impact for chiral gauge theories. For example, in the DFSZ axion model, the couplings of axions to electroweak gauge bosons do not follow the pattern expected from chiral anomalies, as we prove by an explicit calculation. The reason for this mismatch is traced back to triangle Feynman diagrams sensitive to the anomalous breaking of the vector Ward identity, and is ultimately related to the conservation of baryon and lepton numbers. Though our analyses are entirely done for true axion models, this observation could have important consequences for axion-like particle searches.

1 Introduction

Axions have been around for more than 40 years, and even if the simplest models have been ruled out, they still remain probably the best solution for the strong CP problem of the Standard Model. They come under many guises, but the basic principle is always the same: design a spontaneously broken global $U(1)_{PQ}$ symmetry and assign chiral charges to some colored fermions [1,2]. This ensures the associated Goldstone boson a^0, the axion [3,4], has a direct coupling to gluons, and possibly also to photons, of the form

$$\mathcal{L}_{eff} = \frac{a^0}{16\pi^2 v}\left(g_s^2 \mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + e^2 \mathcal{N}_{em} F_{\mu\nu}\tilde{F}^{\mu\nu}\right), \quad (1)$$

where v is the vacuum expectation value (VEV) of the scalar breaking the $U(1)_{PQ}$ symmetry, and $\mathcal{N}_{C,em}$ some constants related to the fermion $U(1)_{PQ}$ and gauge charges. Naively, the CP-violating QCD coupling $\theta_{QCD} G_{\mu\nu}\tilde{G}^{\mu\nu}$ can then be rotated away no matter the value of θ_{QCD} since the Goldstone boson is shifted as $a^0 \to a^0 + v\theta$ under a $U(1)_{PQ}$ transformation of parameter θ. In reality, what happens is that non-perturbative QCD effects dressing $G_{\mu\nu}\tilde{G}^{\mu\nu}$ create an effective potential for the axion field [1,2] (see also e.g. Ref. [6]), whose minimum is attained precisely when $\langle a^0\rangle + v\theta_{QCD} = 0$. In the process, the axion acquires a small QCD-induced mass, typically well below the eV scale [7,8]. For a recent review, see for example Ref. [9].

Though this picture is correct, the nature of the couplings in Eq. (1) is often wrongly ascribed to the anomaly of the $U(1)_{PQ}$ fermionic current. This idea comes from Noether's theorem: the Goldstone boson a^0 is coupled to its symmetry current, $vp^\mu = \langle 0|J^\mu_{PQ}|a^0(p)\rangle$, and the current is anomalous, $\partial_\mu J^\mu_{PQ} \sim g_s^2 \mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + e^2 \mathcal{N}_{em} F_{\mu\nu}\tilde{F}^{\mu\nu}$, hence it seems Eq. (1) is immediately recovered. It has been known for a long time that this derivation of the axion couplings is not correct but sufficient for practical purpose [10,11]. It permits to identify the couplings of axions to gluons and photons as induced by heavy fermions. So, by a common abuse of language, the $a \to \gamma\gamma$ and $a \to gg$ processes are said to be induced by the anomaly in the $U(1)_{PQ}$ current.

However, as we will explicitly demonstrate in this paper, using this same procedure to derive the couplings of the axion to electroweak gauge bosons fails whenever chiral fermions are charged under $U(1)_{PQ}$. In that case, the true gauge couplings do not follow the pattern expected from the anomalies in the $U(1)_{PQ}$ current. More precisely, for the $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ gauge group, one expects from chiral anomalies [10] (see also Ref. [12])

$$\mathcal{L}_{eff} = \frac{a^0}{16\pi^2 v}\left(g_s^2 \mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + g^2 \mathcal{N}_L W^i_{\mu\nu}\tilde{W}^{i,\mu\nu} + g'^2 \mathcal{N}_Y B_{\mu\nu}\tilde{B}^{\mu\nu}\right), \quad (2)$$

where $\mathcal{N}_{C,L,Y}$ are derived from the $U(1)_{PQ}$ charges and gauge quantum numbers of the SM fermions. While for QCD

[a] e-mail: jeremie.quevillon@lpsc.in2p3.fr
[b] e-mail: chsmith@lpsc.in2p3.fr

and QED, which are vector theories, the coefficients $\mathcal{N}_C$ and $\mathcal{N}_{em} = \mathcal{N}_L + \mathcal{N}_Y$ indeed tune the $a^0 \to gg$ and $\gamma\gamma$ decays, we will prove that the $a^0 \to \gamma Z$, ZZ, and W^+W^- decay amplitudes are independent of $\mathcal{N}_L$ and $\mathcal{N}_Y$! The true processes driving these decays are not anomalous and their relative strengths do not follow the pattern expected from Eq. (2). The reason for this mismatch between the anomalous couplings and the true axion couplings will be traced to the existence of additional anomalies, occurring only for chiral theories. These additional anomalies cancel exactly the contributions in Eq. (2), leaving as remainder only non-local, non-anomalous processes.

The paper is organized as follow. To set the stage, we start in the next section by presenting in details an axion toy model [1,2]. Though simple, this model illustrates many important physical features of more realistic axion models. In particular, it will be clear that the axion couplings to gauge bosons are not anomalous, but that this has no quantitative consequence for photons or gluons. Then, before turning our attention to full-fledged axion models, we derive a number of important results for anomalies in Sect. 3. Specifically, to be able to treat chiral gauge theories, the Ward identities applied to AVV or AAA triangle graphs have to be properly calculated, where $V_\mu = \bar{\psi}\gamma_\mu\psi$ and $A_\mu = \bar{\psi}\gamma_\mu\gamma_5\psi$ denote the vector and axial currents. Indeed, it is necessary to go beyond simple regularization procedures to be able to locate the anomaly of the AVV triangle in one of the vector Ward identities or to break explicitly the Bose symmetry of the AAA triangle. Equipped with these results, we turn to the Peccei-Quinn axion model in Sect. 4, derive the correct axion couplings to gauge bosons, and identify precisely where the naive procedure leading to Eq. (2) fails. This analysis is then trivially extended to invisible axion models. Finally, our results are summarized in Sect. 5, along with their implications for axion-like particle searches.

2 An axionic toy model

Let us consider the following simple extension of QED. Starting with a massless Dirac fermion ψ, charged under a local symmetry $U(1)_{em}$, we add a scalar field ϕ charged under a $U(1)_{PQ}$ global symmetry. The complete Lagrangian is

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}_L(i\not{D})\psi_L + \bar{\psi}_R(i\not{D})\psi_R \\ &+(y\phi\bar{\psi}_L\psi_R + h.c.) + \partial_\mu\phi^\dagger\partial^\mu\phi - V(\phi),\end{aligned} \tag{3}$$

where $D^\mu = \partial^\mu + ieA^\mu$ and e is the electric charge of ψ, and $\psi_{L,R} = 1/2(1 \mp \gamma_5)\psi$. When the Yukawa coupling is non-zero, $y \neq 0$, the fermion fields have to be charged under $U(1)_{PQ}$ also. If we normalize the scalar charge to -1, then under a $U(1)_{PQ}$ transformation of parameter θ,

$$\begin{aligned}&\phi \to \exp(-i\theta)\phi, \quad \psi_L \to \exp(i\alpha\theta)\psi_L, \\ &\psi_R \to \exp(i(\alpha+1)\theta)\psi_R.\end{aligned} \tag{4}$$

At the classical level, this clearly leaves the Lagrangian invariant whatever the free parameter α. Actually, this parameter corresponds to the conservation of the global fermion number, which is aligned with the electric charge in this simple one-fermion model.

When the global $U(1)_{PQ}$ symmetry is broken spontaneously, a Goldstone boson is left behind and will be identified with the axion [3,4]. So, we choose the potential as $V(\phi^\dagger\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ with $\mu^2 < 0$. Let us analyze the resulting theory using two different representations for the scalar field.

2.1 Linear representation

Shifted to the true vacuum, the usual linear representation for the scalar field is

$$\phi(x) = \sigma^0(x) + ia^0(x) + v, \tag{5}$$

with $v = -\mu^2/\lambda$. The Goldstone boson a^0 remains massless and is the axion, while the σ^0 field acquires a mass $m_\sigma^2 = 2\lambda v^2$. Assuming y is real and denoting the fermion mass as $m \equiv vy$, the whole Lagrangian when ϕ is expended around the true vacuum becomes

$$\begin{aligned}\mathcal{L}_{\text{Linear}} = &-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \bar{\psi}(i\not{D})\psi + m\bar{\psi}\psi\left(1+\frac{\sigma^0}{v}\right) \\ &+\frac{1}{2}\partial_\mu a^0\partial^\mu a^0 + \frac{m}{v}a^0\bar{\psi}i\gamma_5\psi \\ &+\frac{1}{2}(\partial_\mu\sigma^0\partial^\mu\sigma^0 - 2\lambda v^2(\sigma^0)^2) - \lambda v\sigma^0((\sigma^0)^2 + (a^0)^2) \\ &-\frac{\lambda}{4}((\sigma^0)^2 + (a^0)^2)^2 - \frac{\lambda v^4}{4}.\end{aligned} \tag{6}$$

Since ϕ is not charged under $U(1)_{em}$, a^0 does not directly couple to photons. However, this coupling arises through one-loop triangle graphs, see Fig. 1. Adopting the Pauli-Villars procedure to regulate the loop amplitude in intermediate steps, we compute

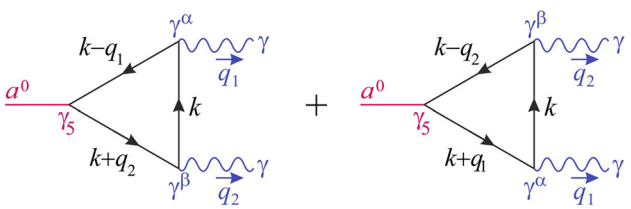

Fig. 1 Triangle graphs inducing the $a^0 \to \gamma\gamma$ decay

$$\begin{aligned}\mathcal{T}^{\alpha\beta}_{PVV} &= \int \frac{d^4k}{(2\pi)^4}(-1)\mathrm{Tr}\\ &\times\left[\frac{i}{\not{k}-\not{q}_1-m}\gamma^\alpha\frac{i}{\not{k}-m}\gamma^\beta\frac{i}{\not{k}+\not{q}_2-m}\gamma_5\right]\\ &+(1,\alpha\leftrightarrow 2,\beta)+(m\to M)\\ &=-i\frac{1}{2\pi^2}\varepsilon^{\alpha\beta\rho\sigma}q_{1,\rho}q_{2,\sigma}(mC_0(m^2)-MC_0(M^2)).\end{aligned} \tag{7}$$

The three-point scalar function $C_0(m^2)\equiv C_0(q_1^2,q_2^2,(q_1+q_2)^2;m^2,m^2,m^2)$ obeys[1]

$$\lim_{m\to\infty}C_0(m^2)=\frac{-1}{2m^2}\,,\quad \lim_{m\to 0}m^2C_0(m^2)=0. \tag{8}$$

The regulator M can thus safely be sent to infinity, and the decay amplitude in the linear representation for the scalar field is

$$\begin{aligned}&\mathcal{M}(a^0\to\gamma\gamma)_{\text{Linear}}\\ &=-i\frac{m}{v}e^2\mathcal{T}^{\alpha\beta}_{PVV}\varepsilon(q_1)^*_\alpha\varepsilon(q_2)^*_\beta\\ &=-\frac{e^2}{2\pi^2 v}m^2C_0(m^2)\varepsilon^{\alpha\beta\rho\sigma}\varepsilon(q_1)^*_\alpha\varepsilon(q_2)^*_\beta q_{1,\rho}q_{2,\sigma}\,.\end{aligned} \tag{9}$$

In the $m\to\infty$ limit, this amplitude corresponds to the local interaction

$$\mathcal{L}^{\text{eff}}_{\text{Linear}}=-\frac{e^2}{16\pi^2 v}a^0F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{10}$$

It is interesting to remark that this term is not decoupling,[2] even though $\mathcal{T}^{\alpha\beta}_{PVV}\to 0$ as $m\to\infty$, thanks to the m factor coming from the a^0 coupling to the fermion. Still, at the level of the linearly-realized theory, anomalies do not show up and the loop amplitude can be safely computed using a naive regularization procedure, even if we know from the fermion charges in Eq. (4) that $U(1)_{PQ}$ is anomalous.

2.2 Polar representation

Instead of the linear representation for the scalar field, there is another representation more suited to the circular geometry of the vacuum.[3] The polar or exponential representation is

$$\phi(x)=\frac{1}{\sqrt{2}}(v+\sigma^0(x))e^{-ia^0(x)/v}. \tag{11}$$

[1] We use the notations and conventions of Ref. [5]

[2] This non-decoupling is to be understood in the formal sense, since strictly speaking the $m\to\infty$ limit is not compatible with our perturbative treatment of the Yukawa couplings.

[3] Note that the polar representation still allows the pseudoscalar of having a small mass which could be treated as a perturbation as this is usually done in axion phenomenology.

Then, under the action of a $U(1)_{PQ}$ rotation of parameter θ, the component fields have simple (though non-linear) transformation properties: the σ field is constant, $\sigma^0\to\sigma^0$, while the axion is shifted by a constant, $a^0\to a^0+v\theta$. Importantly, and even if we are using the same notations a^0 and σ^0, these fields are not the same ones as in Eq. (5). In particular, the a^0 couples differently to fermions. Clearly, expanding the exponential, couplings of arbitrarily high dimensions are present, and the theory is no longer manifestly renormalizable.

As a next step, after plugging the polar representation of ϕ in the Lagrangian, we perform a reparametrization of the fermion fields to make them invariant under $U(1)_{PQ}$:

$$\begin{aligned}&\psi_L(x)\to\exp(i\alpha a^0(x)/v)\psi_L(x),\\ &\psi_R(x)\to\exp(i(\alpha+1)a^0(x)/v)\psi_R(x)\,.\end{aligned} \tag{12}$$

This has three effects. First, the Goldstone field disappears from the Yukawa coupling. Second, the fermion kinetic terms induce derivative interactions,

$$\begin{aligned}\delta\mathcal{L}_{\text{Der}}&=-\frac{\partial_\mu a^0}{v}(\alpha\bar{\psi}_L\gamma^\mu\psi_L+(\alpha+1)\bar{\psi}_R\gamma^\mu\psi_R)\\ &=-\frac{\partial_\mu a^0}{2v}((2\alpha+1)\bar{\psi}\gamma^\mu\psi+\bar{\psi}\gamma^\mu\gamma_5\psi),\end{aligned} \tag{13}$$

where one can recognize the fermionic current associated to the $U(1)_{PQ}$ symmetry. Third, the fermionic path integral measure is not invariant, and the Jacobian of the transformation has to be included as a new local interaction:

$$\delta\mathcal{L}_{\text{Jac}}=\frac{e^2}{16\pi^2 v}a^0(\alpha-(\alpha+1))F_{\mu\nu}\tilde{F}^{\mu\nu}=-\frac{e^2}{16\pi^2 v}a^0F_{\mu\nu}\tilde{F}^{\mu\nu}. \tag{14}$$

The whole non-linear Lagrangian is thus

$$\begin{aligned}\mathcal{L}_{\text{Polar}}&=-\frac{1}{4}F_{\mu\nu}F^{\mu\nu}+\bar{\psi}(i\not{D})\psi+m\bar{\psi}\psi\left(1+\frac{\sigma^0}{v}\right)\\ &+\frac{1}{2}\partial_\mu a^0\partial^\mu a^0\left(1+\frac{\sigma^0}{v}\right)^2\\ &-\frac{\partial_\mu a^0}{2v}((2\alpha+1)\bar{\psi}\gamma^\mu\psi+\bar{\psi}\gamma^\mu\gamma_5\psi)\\ &-\frac{e^2}{16\pi^2 v}a^0F_{\mu\nu}\tilde{F}^{\mu\nu}\\ &+\frac{1}{2}(\partial_\mu\sigma^0\partial^\mu\sigma^0-2\lambda v^2(\sigma^0)^2)\\ &-\lambda v(\sigma^0)^3-\frac{\lambda}{4}(\sigma^0)^4-\frac{\lambda v^4}{4}.\end{aligned} \tag{15}$$

Under this form, the Lagrangian is manifestly $U(1)_{PQ}$-symmetric but for the local Jacobian term $a^0F_{\mu\nu}\tilde{F}^{\mu\nu}$ since σ^0, $\psi_{L,R}$, and A^μ are invariant but a^0 is shifted by a constant. This permits to get rid of any pre-existing $\theta_{em}F_{\mu\nu}\tilde{F}^{\mu\nu}$ term in the Lagrangian, by shifting $a^0\to a^0+v\theta_{em}$. Of

course, for an abelian theory, $F_{\mu\nu}\tilde{F}^{\mu\nu}$ is a total derivative that can be safely discarded, but this toy model can easily be adapted to the non-abelian case, and closely resembles the KSVZ invisible axion model [13,14] when applied to QCD.

Let us now compute $a^0 \to \gamma\gamma$ within the non-linearly realized theory. Together with the local amplitude from $\delta\mathcal{L}_{\text{Jac}}$, we must include the triangle graphs arising from $\delta\mathcal{L}_{\text{Der}}$, which have both a vector and axial component. The vector current $V^\mu = \bar{\psi}\gamma^\mu\psi$ does not contribute since the photon couplings are also vectorial, and the corresponding triangle graph vanishes thanks to Furry's theorem. The axial current $A^\mu = i\bar{\psi}\gamma^\mu\gamma_5\psi$, on the other hand, gives a non-zero contribution. We recognize the well-known AVV triangle graph:

$$\begin{aligned}\mathcal{T}^{\gamma\alpha\beta}_{AVV} = &\int \frac{d^4k}{(2\pi)^4}(-1)\\ &\times \text{Tr}\left[\frac{i}{\not{k}-\not{q}_1-m}\gamma^\alpha\frac{i}{\not{k}-m}\gamma^\beta\frac{i}{\not{k}+\not{q}_2-m}\gamma^\gamma\gamma_5\right]\\ &+(1,\alpha\leftrightarrow 2,\beta).\end{aligned} \tag{16}$$

Adopting a Pauli–Villars regularization, we add to this a second term with $m \to M$, and proceed with the calculation in $d = 4$ dimensions. With the derivative from $\partial_\mu a^0$ amounting to taking the divergence, we find

$$\begin{aligned}\mathcal{M}(a^0\to\gamma\gamma)_{\text{Der}} &= \frac{-e^2}{2v}i(q_1+q_2)_\gamma\mathcal{T}^{\gamma\alpha\beta}_{AVV}\varepsilon(q_1)^*_\alpha\varepsilon(q_2)^*_\beta\\ &= \frac{-e^2}{4\pi^2 v}(2m^2C_0(m^2)-2M^2C_0(M^2))\varepsilon^{\alpha\beta\rho\sigma}\varepsilon(q_1)^*_\alpha\\ &\quad\times\varepsilon(q_2)^*_\beta q_{1,\rho}q_{2,\sigma}.\end{aligned} \tag{17}$$

The regulator term does not decouple in the $M \to \infty$ limit but gives a finite local contribution. Interestingly, this contribution precisely cancel with the local term from $\delta\mathcal{L}_{\text{Jac}}$, so altogether, we recover precisely the result obtained in the linear case, see Eq. (9):

$$\begin{aligned}&\mathcal{M}(a^0\to\gamma\gamma)_{\text{Polar}}\\ &\quad= \mathcal{M}(a^0\to\gamma\gamma)_{\text{Der}}+\mathcal{M}(a^0\to\gamma\gamma)_{\text{Jac}} &(18)\\ &\quad= -\frac{e^2}{2\pi^2 v}m^2C_0(m^2)\varepsilon^{\alpha\beta\rho\sigma}\varepsilon(q_1)^*_\alpha\varepsilon(q_2)^*_\beta q_{1,\rho}q_{2,\sigma}\\ &\quad= \mathcal{M}(a^0\to\gamma\gamma)_{\text{Linear}}. &(19)\end{aligned}$$

This exercise illustrates a number of important features:

- In the polar representation, the appearance of the anomalous local term $a^0F_{\mu\nu}\tilde{F}^{\mu\nu}$ is spurious. When computing specific amplitudes, it only serves to cancel out the anomalous term arising from the derivative interaction $\partial_\mu a^0\bar{\psi}\gamma^\mu\gamma_5\psi$. At the end, this is nothing but an application of the well-known axial current Ward identity:

$$\partial_\mu A^\mu - \frac{e^2}{8\pi^2}F_{\mu\nu}\tilde{F}^{\mu\nu} = 2imP, \tag{20}$$

where $P = \bar{\psi}\gamma_5\psi$, $A^\mu = \bar{\psi}\gamma^\mu\gamma_5\psi$. The right-hand side corresponds to the amplitude in the linear representation, and the left-hand side to the two contributions arising in the polar representation. It is thus clear that it would be wrong to understand the $a^0 \to \gamma\gamma$ coupling as induced by the anomaly. The anomalous interaction of Eq. (14) only arises because of the reparametrization in Eq. (12), and necessarily comes together with the appropriate derivative interactions.
- In the $m \to 0$ limit, the $a^0 \to \gamma\gamma$ amplitude vanishes exactly,

$$\mathcal{M}(a^0\to\gamma\gamma)_{\text{Linear}} = \mathcal{M}(a^0\to\gamma\gamma)_{\text{Polar}} \overset{m\to 0}{=} 0. \tag{21}$$

In the linear representation, this trivially follows from the vanishing of the coupling of a^0 to fermions, see Eq. (6). In the polar representation, it requires an exact cancellation of the local anomalous contribution with the non-local triangle amplitudes. Again, this can be understood in terms of the axial current Ward identity. In this limit, the spurious nature of the contact interaction is manifest.
- The reason why $a^0 \to \gamma\gamma$ is often misinterpreted as induced by the anomaly can be understood looking at the $m \to \infty$ limit. Indeed, the contribution of the derivative term vanishes,

$$\mathcal{M}(a^0\to\gamma\gamma)_{\text{Der}} \overset{m\to\infty}{=} 0, \tag{22}$$

as can be trivially seen in Eq. (17). Since then all that remains in the non-linear theory is the local term from $\delta\mathcal{L}_{\text{Jac}}$, it necessarily corresponds to the contribution surviving in the $m \to \infty$ limit in the linear theory, Eq. (10). Yet, this is only a parametric correspondence, and certainly not a physical identification. The anomaly still cancels in $\mathcal{M}(a^0 \to \gamma\gamma)_{\text{Der}} + \mathcal{M}(a^0 \to \gamma\gamma)_{\text{Jac}}$, and the only surviving contribution is actually the first term of Eq. (17). Ultimately, the local anomalous term $\delta\mathcal{L}_{\text{Jac}}$ in Eq. (14) is no more than a convenient book-keeping device tracking all the fields that have been integrated out [10,11].
- Furry's theorem together with the vector current Ward identity ensures the absence of any dependence on the free parameter α in the non-linear theory, in agreement with the absence of this parameter in the corresponding linear representation.
- As an aside, it should be clear that though the anomalous coupling in Eq. (14) is not affected by radiative corrections [15], neither is the anomalous part of the triangle graph in Eq. (17). So, the two exactly cancel at all orders, but the $a^0 \to \gamma\gamma$ amplitude does get corrected

by higher order effects since it is actually not induced by the anomaly. Further, note that the theory in Eq. (3) is obviously renormalizable, so radiative corrections can be calculated perturbatively using standard techniques. In this respect, if the photons are also coupled to some other fermions χ, the two-loop process $a^0 \to \gamma\gamma \to \bar{\chi}\chi$ is even UV finite in the linear representation [16]. This fact would clearly be difficult to guess using the polar representation, where the local $a^0 \to \gamma\gamma$ vertex from $\delta\mathcal{L}_{\rm Jac}$ leads to a UV divergent diagram.

- As said earlier, the present toy model can be generalized to more complicated, and more realistic KSVZ-like axion models [13,14]. Consider for instance the SM, to which a gauge-singlet scalar ϕ and a set of vector-like fermions ψ are added:

$$\begin{aligned}\mathcal{L}_{\rm KSVZ} = \mathcal{L}_{\rm SM} + \bar{\psi}_L(i\not{D})\psi_L + \bar{\psi}_R(i\not{D})\psi_R \\ +(y\phi\bar{\psi}_L\psi_R + h.c.) + \partial_\mu\phi^\dagger\partial^\mu\phi - V(\phi),\end{aligned} \tag{23}$$

with the same scalar potential as in Eq. (3). A priori, the covariant derivative can include all three SM interactions,

$$D^\mu = \partial^\mu - ig_s G_a^\mu T^a - igW_i^\mu T^i - ig'\frac{Y}{2}B^\mu\,, \tag{24}$$

where T^a and T^i are the $SU(3)_C$ and $SU(2)_L$ generators in the representation carried by ψ, and Y its hypercharge. Then, if $yv \gg v_{EW}$ and $v \gg v_{EW}$, with v_{EW} the electroweak vacuum expectation value, the effect of the heavy fermion is to induce (see Eq. (10))

$$\begin{aligned}\mathcal{L}^{\rm eff,KSVZ}_{\rm Linear} = -\frac{1}{16\pi^2 v}a^0(g_s^2 d_L^\psi C_C^\psi G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} \\ +g^2 d_C^\psi C_L^\psi W^i_{\mu\nu}\tilde{W}^{i,\mu\nu} \\ +g'^2 d_L^\psi d_C^\psi C_Y^\psi B_{\mu\nu}\tilde{B}^{\mu\nu}),\end{aligned} \tag{25}$$

with the quadratic invariants $C_C^\psi\delta^{ab} = {\rm Tr}[T^aT^b]$, $C_L^\psi\delta^{ij} = {\rm Tr}[T^iT^j]$, and $C_Y^\psi = Y^2/4$, and $d^\psi_{C,L}$ the corresponding dimensions of the $SU(3)_C$ and $SU(2)_L$ representations. Yet, several points must be clear: (1) These couplings are not anomalous, but result from the appropriately generalized Eq. (9) in the $y \to \infty$ limit. (2) The integrated fermion must have vector couplings to gauge fields, otherwise the free parameter α corresponding to fermion number does not cancel in Eq. (14). We will see later on how to deal with chiral theories. (3) Even if vector-like, no mass term is allowed for ψ because it would explicitly break the $U(1)_{PQ}$ symmetry. (4) The couplings in $\mathcal{L}^{\rm eff,KSVZ}_{\rm Linear}$ are not protected from radiative corrections since they are not anomalous. (5) Finally, the nature of the divergence arising when using this effective Lagrangian to compute e.g. the couplings of a^0 to SM fermions is clear. In the UV complete theory, Eq. (23), the $a^0 \to \gamma\gamma$, WW, and gg vertices are never local. Instead, the pseudoscalar triangle of Eq. (9) acts as a form factor and is sufficient to regulate the UV behavior of the vector boson loops, making them finite.

3 On the consistent use of anomalies

In the course of the previous section, the anomaly appeared several times, and there is a number of peculiarities that are worth detailing. First, consider the calculation of the local term $\delta\mathcal{L}_{\rm Jac}$, see Eq. (14). The simplest way to compute the Jacobian of the transformation is to compute a triangle graph, this time with three left-(or right-)handed currents (the momentum flow is defined as in Fig. 1):

$$\begin{aligned}\mathcal{T}^{\alpha\beta\gamma}_{LLL} = \int\frac{d^4k}{(2\pi)^4}(-1) \\ \times{\rm Tr}\left[\frac{i}{\not{k}-\not{q}_1}\gamma^\beta P_L\frac{i}{\not{k}}\gamma^\gamma P_L\frac{i}{\not{k}+\not{q}_2}\gamma^\alpha P_L\right] \\ +(1,\beta\leftrightarrow 2,\gamma),\end{aligned} \tag{26}$$

with $P_L = (1-\gamma_5)/2$. Following Ref. [17], the divergences of this amplitude are calculated by first keeping track of the ambiguities in the loop momentum routing, and then from the surviving surface terms, and one finds:

$$i(q_1+q_2)_\alpha\mathcal{T}^{\alpha\beta\gamma}_{LLL} = \frac{1}{8\pi^2}(a-b)\,\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, \tag{27a}$$

$$-i(q_1)_\beta\mathcal{T}^{\alpha\beta\gamma}_{LLL} = \frac{1}{8\pi^2}(1+b)\,\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, \tag{27b}$$

$$-i(q_2)_\gamma\mathcal{T}^{\alpha\beta\gamma}_{LLL} = \frac{1}{8\pi^2}(1-a)\,\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \tag{27c}$$

where a and b are free parameters, such that no choice can force all three divergences to vanish simultaneously. For the purpose of deriving the $\delta\mathcal{L}_{\rm Jac}$ term (Eq. (14)), since two currents have to be conserved to preserve the $U(1)_{em}$ invariance (see Eq. (3)), we can choose $a = -b = 1$. Equation (14) is then recovered including both ψ_L and ψ_R, accounting for the fact that the anomalous terms are opposite for $\mathcal{T}^{\alpha\beta\gamma}_{LLL}$ and $\mathcal{T}^{\alpha\beta\gamma}_{RRR}$.

Yet, it is worth stressing that this choice $a = -b = 1$ is not the one usually adopted, as it breaks the Bose symmetry of the amplitude under the exchange of the three currents. The consistent anomaly is that with $a = -b = 1/3$, which restores the Bose symmetry as

$$\begin{aligned}i(q_1+q_2)_\alpha\mathcal{T}^{\alpha\beta\gamma}_{LLL}|_{Bose} = -i(q_1)_\alpha\mathcal{T}^{\gamma\alpha\beta}_{LLL}|_{Bose} \\ = -i(q_2)_\alpha\mathcal{T}^{\beta\gamma\alpha}_{LLL}|_{Bose} = \frac{1}{12\pi^2}\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}.\end{aligned} \tag{28}$$

Importantly, this choice is not always explicit. For example, using the Pauli-Villars regularization, one immediately arrives at the Bose symmetric result (see Ref. [18]).

In terms of vector and axial currents, the VVV and VAA triangle amplitudes cancel out because of Furry's theorem. This leaves the AVV and AAA diagrams, for which we obtain using the same method as in Ref. [17]:

$$\begin{aligned} i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m=0} &= i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m=0} \\ &= \frac{1}{4\pi^2}(a-b)\,\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, -i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m=0} \quad (29a) \\ &= -i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m=0} \\ &= \frac{1}{4\pi^2}(1+b)\,\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, -i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m=0} \quad (29b) \\ &= -i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m=0} \\ &= \frac{1}{4\pi^2}(1-a)\,\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}. \quad (29c) \end{aligned}$$

These results are valid in the massless limit for the fermions. As for the consistent anomaly, the usual conventions are $a = -b = 1$ for the AVV triangle to keep the vector currents conserved, and $a = -b = 1/3$ for the AAA triangle to enforce the Bose symmetry. The former case is the basis for Eq. (17). These conventions are automatically enforced when using simple regularization procedures like Pauli-Villars or dimensional regularization. However, for the following, the freedom to assign the anomaly in one of the vector currents or to explicitly break the Bose symmetry will be essential, so these regularization procedures are inadequate.

At the level of the AVV and AAA triangles, there is no obstruction to consider massive fermions. These contributions can be separated from the anomalous pieces, and do not depend on the chosen momentum routing. Actually, a simple calculation leads to

$$i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AVV} = 2im\mathcal{T}^{\beta\gamma}_{PVV} + \frac{1}{4\pi^2}(a-b)\,\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, \quad (30a)$$

$$-i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AVV} = \frac{1}{4\pi^2}(1+b)\,\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, \quad (30b)$$

$$-i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AVV} = \frac{1}{4\pi^2}(1-a)\,\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \quad (30c)$$

and

$$i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AAA} = 2im\mathcal{T}^{\beta\gamma}_{PAA} + \frac{1}{4\pi^2}(a-b)\,\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, \quad (31a)$$

$$-i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AAA} = 2im\mathcal{T}^{\alpha\gamma}_{PAA} + \frac{1}{4\pi^2}(1+b)\,\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, \quad (31b)$$

$$-i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AAA} = 2im\mathcal{T}^{\alpha\beta}_{PAA} + \frac{1}{4\pi^2}(1-a)\,\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \quad (31c)$$

with the pseudoscalar triangle amplitudes (see Eq. (7))

$$\mathcal{T}^{\alpha\beta}_{PVV} = -i\frac{1}{2\pi^2}mC_0(m^2)\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \quad (32a)$$

$$\mathcal{T}^{\alpha\beta}_{PAA} = -i\frac{1}{2\pi^2}m(C_0(m^2)+2C_1(m^2))\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \quad (32b)$$

where $C_i(m^2) = C_i(q_1^2, q_2^2, (q_1+q_2)^2, m^2, m^2, m^2)$.

In some sense, the usual conventions are automatic for the mass terms: the vector currents are always conserved and Bose symmetry must be manifest for the AAA triangle. Indeed, once the anomalous terms are separated, the rest obeys the naive Ward identities $\partial_\mu V^\mu = 0$ and $\partial_\mu A^\mu = 2imP$, with P the pseudoscalar current. This means that the mass-dependent term must be absent from $-i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AVV}$ and $-i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AVV}$, and it must be equal for all three divergences of $\mathcal{T}^{\alpha\beta\gamma}_{AAA}$. This picture is trivially confirmed using dispersion relations [19], see Fig. 2: cutting out a vector current necessarily gives zero since the intermediate fermions are on their mass shell and $\partial_\alpha \bar\psi\gamma^\alpha\psi = 0$, while cutting out any one of the axial currents, the identity $\partial_\alpha \bar\psi\gamma^\alpha\gamma_5\psi = 2im\bar\psi\gamma_5\psi$ must hold.

Finally, in the $m \to \infty$ limit, using $C_0 \to -1/(2m^2)$ and $C_1(m^2) \to 1/(6m^2)$, the various divergences are

$$i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m\to\infty} = \frac{1}{4\pi^2}(a-b-2)\,\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, \quad (33a)$$

$$-i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m\to\infty} = \frac{1}{4\pi^2}(1+b)\,\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, \quad (33b)$$

$$-i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AVV}|_{m\to\infty} = \frac{1}{4\pi^2}(1-a)\,\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}, \quad (33c)$$

and

$$i(q_1+q_2)_\alpha \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m\to\infty} = \frac{1}{4\pi^2}\left(a-b-\frac{2}{3}\right)\varepsilon^{\beta\gamma\mu\nu}q_{1\mu}q_{2\nu}, \quad (34a)$$

$$-i(q_1)_\beta \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m\to\infty} = \frac{1}{4\pi^2}\left(1+b-\frac{2}{3}\right)\varepsilon^{\gamma\alpha\mu\nu}q_{1\mu}q_{2\nu}, \quad (34b)$$

$$-i(q_2)_\gamma \mathcal{T}^{\alpha\beta\gamma}_{AAA}|_{m\to\infty} = \frac{1}{4\pi^2}\left(1-a-\frac{2}{3}\right)\varepsilon^{\alpha\beta\mu\nu}q_{1\mu}q_{2\nu}. \quad (34c)$$

Crucially for the following, the vanishing of all the divergences in Eq. (33) in the $m \to \infty$ limit is only true when the anomaly is carried by the axial current in the AVV triangle, which implies the parametric choice $a = -b = 1$. Similarly, the vanishing of all the divergences in Eq. (34) in the $m \to \infty$ limit is only true when the anomaly is carried equally by the

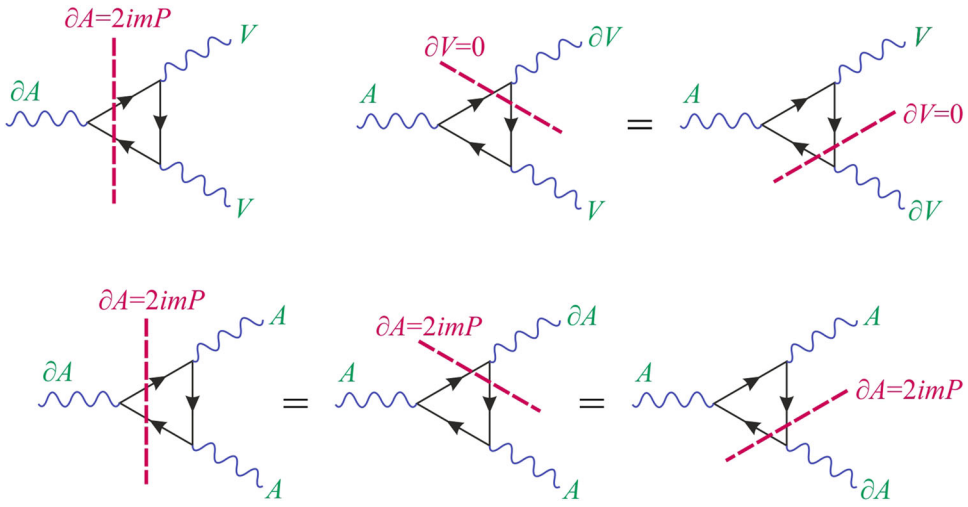

Fig. 2 Non–anomalous Ward identities valid for the mass-dependent terms of the triangle amplitudes. The dashed lines denote the cuts along which the fermionic equation of motion is valid

three axial currents in the AAA triangle, which implies the choice $a = -b = 1/3$. But evidently, there is no reason for this to hold in general and we will see in the next section how this plays a role in the derivation of the axion couplings to electroweak gauge bosons.

4 The Peccei-Quinn axion and its couplings

Historically, the first axion model was proposed by Peccei and Quinn [1,2]. It is based on a Two Higgs Doublet Model (THDM), but with an enlarged symmetry. Denoting the Higgs doublets as

$$\Phi_1 = \begin{pmatrix} H_1^+ \\ H_1^0 \end{pmatrix} \ , \ \Phi_2 = \begin{pmatrix} H_2^+ \\ H_2^0 \end{pmatrix} \ , \tag{35}$$

the scalar potential is assumed to allow for the independent rephasing

$$U(1)_1 \otimes U(1)_2 : \Phi_j \to \exp(i\theta_j)\Phi_j \quad \text{with } j = 1, 2. \tag{36}$$

In the standard notation (see Ref. [20]), this forces the parameters m_{12}, $\lambda_{5,6,7}$ to vanish. One combination of these $U(1)$s can be identified with the hypercharge, since both doublets have the same gauge quantum numbers, while another combination is denoted $U(1)_{PQ}$. The vacuum expectation values

$$\langle \Phi_1 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_1 \end{pmatrix} \ , \ \langle \Phi_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v_2 \end{pmatrix} \ , \tag{37}$$

break $SU(2)_L \otimes U(1)_Y \otimes U(1)_{PQ} \to U(1)_{em}$ spontaneously, where $v_1^2 + v_2^2 \equiv v^2 \approx (246\,\text{GeV})^2$ and $\tan\beta \equiv v_1/v_2$. The $U(1)_Y$ symmetry is local and its Goldstone boson is eaten up by the Z boson, but the $U(1)_{PQ}$ is only global and the massless left over state is the axion. Indeed, assuming Yukawa couplings of Type II,

$$\mathcal{L}_{\text{Yukawa}} = -\bar{u}_R \mathbf{Y}_u q_L \Phi_1 - \bar{d}_R \mathbf{Y}_d q_L \Phi_2^\dagger - \bar{e}_R \mathbf{Y}_e \ell_L \Phi_2^\dagger + h.c., \tag{38}$$

the $U(1)_{PQ}$ symmetry is necessarily chiral and anomalous. Note that at this level, there is a lot of freedom in how to assign $U(1)_1 \otimes U(1)_2$ charges for the fermions, and we will see later on how to (partially) fix them. Remark also that we did not include a phase difference between v_1 and v_2 in Eq. (37) because it can be removed by a $U(1)_{PQ}$ transformation. Actually, the $U(1)_{PQ}$ symmetry being anomalous, this rather moves the phase difference between v_1 and v_2 into the strong and electroweak θ terms. This is precisely the interplay that permits to solve the strong CP puzzle once the $U(1)_{PQ}$ symmetry is spontaneously broken, since it entangles the phase difference with the axion field.

To further explore this simple model, we follow the same strategy as for the toy model of Sect. 2, and perform the analysis using either a linear or a polar representation for the scalar fields $\Phi_{1,2}$.

4.1 Linear representation

The most convenient basis for the scalar states is obtained by performing a rotation of angle β,

$$\begin{pmatrix} \Phi_h \\ \Phi_H \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} \Phi_2 \\ \Phi_1 \end{pmatrix}, \tag{39}$$

so that the VEV is carried by one doublet only

$$\Phi_h = \frac{1}{\sqrt{2}} \exp\{i\boldsymbol{\sigma}\cdot\boldsymbol{G}(x)/v\} \begin{pmatrix} 0 \\ v+\phi_h \end{pmatrix}$$
$$\Phi_H = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}H^+ \\ \phi_H + iA^0 \end{pmatrix}, \tag{40}$$

where only the SM Goldstone bosons have been exponentiated. The pseudoscalar A^0 is the massless Goldstone boson corresponding to the $U(1)_{PQ}$ breaking, and it is parametrized linearly. The scalar states (ϕ_h, ϕ_H) are related to the mass eigenstates (h^0, H^0) through a rotation of angle α which depends on the scalar potential parameters, and H^+ is the charged scalar boson. We note in passing that the charge assignments corresponding to $U(1)_1 \otimes U(1)_2 \leftrightarrow U(1)_Y \otimes U(1)_{PQ}$ thus require orthogonality of the Goldstone states G^0 and A^0, and cannot be guessed a priori from the $U(1)$ charges of the doublets in Eq. (36). Actually, since A^0 is a β-dependent combination of $\mathrm{Im}H_1^0$ and $\mathrm{Im}H_2^0$, the $U(1)_{PQ}$ charges of the doublets are also function of β. We will come back to this point later, and for now, we simply proceed by plugging Eq. (40) in the THDM Lagrangian.

Because Φ_H has no vacuum expectation value, the A^0 boson does not couple to pairs of gauge bosons at tree level. On the other hand, it does couple to fermions. From Eq. (38), we derive

$$\mathcal{L}_{A^0 f\bar{f}} = -i \sum_{f=u,d,e} \frac{m_f}{v} \chi_P^f A^0 \bar{\psi}_f \gamma_5 \psi_f,$$
$$\chi_P^u = \frac{1}{\tan\beta}, \quad \chi_P^d = \chi_P^e = \tan\beta. \tag{41}$$

To reach this form, the mass terms are identified as $\sin\beta v \mathbf{Y}_u \equiv \sqrt{2}\mathbf{m}_u$ and $\cos\beta v \mathbf{Y}_{d,e} \equiv \sqrt{2}\mathbf{m}_{d,e}$ and the fermions are rotated to their mass basis. At the one-loop level, the A^0 then couples to pairs of gauge bosons through the PVV and PAA triangle graphs. Denoting the fermion currents coupled to gauge bosons as $\bar{\psi}_{f'}(g_V^f \gamma^\mu - g_A^f \gamma^\mu \gamma_5)\psi_f$ with, using the short-hand $c_W = \cos\theta_W$, $s_W = \sin\theta_W$:

	g	W	γ	Z
$g_V^f =$	$g_s T_a^f$,	$\frac{g}{\sqrt{2}} T_3^f$,	eQ^f,	$\frac{g}{2c_W}(T_3^f - 2s_W^2 Q^f)$,
$g_A^f =$	0,	$\frac{g}{\sqrt{2}} T_3^f$,	0,	$\frac{g}{2c_W} T_3^f$.

(42)

where $SU(3)$ generators T_a^f are normalized such that $\mathrm{Tr}(T_a^f T_b^f) = 1/2\delta^{ab}$, it is a simple exercise to find (see Eq. (32)):

$$\mathcal{M}(A^0 \to V_1 V_2)_{\mathrm{Linear}} = -i\sigma \sum_{f=u,d,e} \frac{m_f}{v} \chi_P^f \left(g_{V,V_1}^f g_{V,V_2}^{f'} \mathcal{T}_{PVV}^{\alpha\beta}(m_f) + g_{A,V_1}^f g_{A,V_2}^{f'} \mathcal{T}_{PAA}^{\alpha\beta}(m_f) \right) \varepsilon(q_1)_\alpha^* \varepsilon(q_2)_\beta^*, \tag{43}$$

where $\varepsilon(q_1)_\alpha^*$ and $\varepsilon(q_2)_\beta^*$ are the polarizations of V_1 and V_2, carrying respectively the momentum q_1 and q_2. To account for the absence of crossed diagrams for W^+W^-, we also define $\sigma = 1/2$ for $V_{1,2} = W$, $\sigma = 1$ for $V_{1,2} = g, \gamma, Z$. Also, the trace over color is understood for quarks, bringing a factor $\delta^{ab}/2$ for $V_{1,2} = g$, and N_C for $V_{1,2} = W, Z, \gamma$. Finally, the loop functions in $\mathcal{T}_{PVV}^{\alpha\beta}(m_f)$ and $\mathcal{T}_{AVV}^{\alpha\beta}(m_f)$ are the same as in Eq. (32) for $V_{1,2} = g, \gamma, Z$, and generalized as $C_i(m_f^2) = C_i(q_1^2, q_2^2, (q_1+q_2)^2, m_f^2, m_{f'}^2, m_f^2)$ with f' the $SU(2)$ partner of f for W^+W^-. In that case, the CKM entries sum up to one when the three generations are included. The expression Eq. (43) for the decay amplitudes agrees with the standard result of Ref. [21].

A few limiting cases are interesting. First, remark that

$$\mathcal{M}(A^0 \to V_1 V_2)_{\mathrm{Linear}}\Big|_{m_f \to 0} = 0, \tag{44}$$

since the A^0 coupling to fermions is proportional to their masses. Second, in the opposite limit of infinite fermion masses, the loop functions behave as $C_0 \to -1/(2m^2)$ and $C_1(m^2) \to 1/(6m^2)$, so

$$m_f \mathcal{T}_{PVV}^{\alpha\beta}(m_f) = i\frac{1}{4\pi^2} \varepsilon^{\alpha\beta\mu\nu} q_{1\mu} q_{2\nu},$$
$$m_f \mathcal{T}_{PAA}^{\alpha\beta} = i\frac{1}{12\pi^2} \varepsilon^{\alpha\beta\mu\nu} q_{1\mu} q_{2\nu}. \tag{45}$$

The only special case is that of the lepton contribution to $A^0 \to W^+W^-$, since neutrinos remain massless. In that case, $C_0(m_e^2) \to -1/m_e^2$ and $C_1(m_e^2) \to 1/(2m_e^2)$ in the $m_e \to \infty$ limit, and $\mathcal{T}_{PAA}^{\alpha\beta}$ cancels out. Plugging all these limits in the amplitudes, they take the simple form

$$\mathcal{M}(A^0 \to V_1 V_2)_{\mathrm{Linear}}\Big|_{m_{u,d,e}\to\infty} = \frac{1}{4\pi^2 v} \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2} \mathcal{N}_{V_1,V_2},$$
$$\varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2} \equiv \varepsilon^{\alpha\beta\mu\nu} \varepsilon(q_1)_\alpha^* \varepsilon(q_2)_\beta^* q_{1\mu} q_{2\nu}, \tag{46}$$

for some mass-independent coefficients $\mathcal{N}_{V_1,V_2}$. Working out their values, to these amplitudes correspond the effective interactions

$$\mathcal{L}_{\mathrm{Linear}}^{\mathrm{eff}} = \frac{A^0}{16\pi^2 v} \Big(g_s^2 \mathcal{N}_C G_{\mu\nu}^a \tilde{G}^{a,\mu\nu} + e^2 \mathcal{N}_{em} F_{\mu\nu} \tilde{F}^{\mu\nu} + \frac{2e^2}{c_W s_W} \left(\mathcal{N}_0 - s_W^2 \mathcal{N}_{em} \right) Z_{\mu\nu} \tilde{F}^{\mu\nu} + \frac{e^2}{c_W^2 s_W^2} \left(\mathcal{N}_1 - 2s_W^2 \mathcal{N}_0 + s_W^4 \mathcal{N}_{em} \right) Z_{\mu\nu} \tilde{Z}^{\mu\nu} + 2\mathcal{N}_2 g^2 W_{\mu\nu}^+ \tilde{W}^{-,\mu\nu} \Big), \tag{47}$$

with the coefficients (to simplify the comparison, we adopt the standard convention $x \equiv 1/\tan\beta$):

$$\mathcal{N}_C = \frac{1}{2}\left(x + \frac{1}{x}\right), \tag{48a}$$
$$\mathcal{N}_{em} = N_C \left(\frac{4x}{9} + \frac{1}{9x}\right) + \frac{1}{x}, \tag{48b}$$

$$\mathcal{N}_0 = \frac{1}{4}\left(N_C\left(\frac{2x}{3}+\frac{1}{3x}\right)+\frac{1}{x}\right), \tag{48c}$$

$$\mathcal{N}_1 = \frac{1}{12}\left(N_C\left(x+\frac{1}{x}\right)+\frac{1}{x}\right), \tag{48d}$$

$$\mathcal{N}_2 = \frac{1}{12}\left(N_C\left(x+\frac{1}{x}\right)+\frac{3}{2x}\right). \tag{48e}$$

We stress that these effective couplings are directly derived from the well-established $A^0 \to V_1 V_2$ amplitudes computed in the standard THDM,[4] as obtained in Ref. [21].

Often, one interprets the non-decoupling of these amplitudes when $m_{u,d,e} \to \infty$ as an indirect manifestation of the underlying anomalies, but it is important to understand this statement correctly. The pseudoscalar triangles do decouple in that limit, $\mathcal{T}^{\alpha\beta}_{PVV}(m_f) \to 0$ and $\mathcal{T}^{\alpha\beta}_{PAA}(m_f) \to 0$ when $m_f \to \infty$, but this is compensated by the mass-dependent couplings of the A^0 to fermions. The fact that this compensation works, leaving a constant remainder in the $m_f \to \infty$, is guaranteed by the anomalous axial Ward identity of Eq. (20). Yet, there is no anomalous contribution of any kind to the $A^0 \to V_1 V_2$ amplitudes, and those are driven entirely by non-local and well-behaved fermion loops. The situation is quite similar for the loop-induced Higgs decay $h \to \gamma\gamma$ in the SM, whose non-decoupling behavior in the $m_{u,d,e} \to \infty$ limit is interpreted as a manifestation of the trace anomaly [22]. Yet, there is of course no direct contribution of that anomaly to $h \to \gamma\gamma$, which can be safely computed perturbatively.

4.2 Polar representation

The Higgs boson fields are written in the polar representation[5] as (see Eq. (11)):

$$\Phi_1 = \frac{1}{\sqrt{2}}\exp(i\eta_1/v_1)\begin{pmatrix} H_1^+ \\ v_1 + \mathrm{Re}H_1 \end{pmatrix},$$
$$\Phi_2 = \frac{1}{\sqrt{2}}\exp(i\eta_2/v_2)\begin{pmatrix} H_2^+ \\ v_2 + \mathrm{Re}H_2 \end{pmatrix}, \tag{49}$$

where $v_1 = v\sin\beta, v_2 = v\cos\beta$. The η_1 and η_2 states are the Goldstone bosons associated to the $U(1)_1$ and $U(1)_2$ symmetries. They are related to the physical Goldstone bosons a^0 and G^0 by the same rotation as in Eq. (39):

$$\begin{pmatrix} G^0 \\ a^0 \end{pmatrix} = \begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix}\begin{pmatrix} \eta_2 \\ \eta_1 \end{pmatrix}. \tag{50}$$

Note that the G^0 components of Φ_1 and Φ_2 are the same since they have the same hypercharge, and the a^0 and G^0 states are orthogonal to each other. Now, the G^0 can be gauged away via a $U(1)_Y$ rotation, leaving

$$\Phi_1 = \frac{1}{\sqrt{2}}\exp(ia^0 x/v)\begin{pmatrix} H_1^+ \\ v_1 + \mathrm{Re}H_1 \end{pmatrix},$$
$$\Phi_2 = \frac{1}{\sqrt{2}}\exp(-ia^0/xv)\begin{pmatrix} H_2^+ \\ v_2 + \mathrm{Re}H_2 \end{pmatrix}, \tag{51}$$

where we introduce $x = v_2/v_1 = 1/\tan\beta$ to stick to common conventions. The remaining neutral Goldstone field, here denoted a^0 instead of A^0, is the axion. Under this form, the true $U(1)_{PQ}$ charges of the Higgs doublets are finally apparent, since under such a transformation, $a^0 \to a^0 + v\theta$. Then, the Yukawa Lagrangian can also be made invariant under $U(1)_{PQ}$ provided the fermion charges are set as

$$\Phi_1 \to \exp(i\theta x)\Phi_1, \quad \Phi_2 \to \exp(-i\theta/x)\Phi_2, \quad \psi \to \exp(i\chi_\psi\theta)\psi, \tag{52}$$

with

$$\chi_{q_L} = \alpha, \quad \chi_{u_R} = \alpha + x, \quad \chi_{d_R} = \alpha + \frac{1}{x},$$
$$\chi_{\ell_L} = \beta, \quad \chi_{e_R} = \beta + \frac{1}{x}, \tag{53}$$

where α and β are free parameters, corresponding to the conservation of baryon and lepton numbers $\mathcal{B}$ and $\mathcal{L}$.

Following the same steps as in Sect. 2, we now perform a reparametrization of the fermion fields to force them to be invariant under $U(1)_{PQ}$. This is achieved by the field-dependent rotations $\psi \to \exp(i\chi_\psi a^0(x)/v)\psi$. The axion field disappears from the Yukawa Lagrangian, but reappears in derivative interactions from the fermion kinetic terms

$$\delta\mathcal{L}_{\mathrm{Der}} = -\frac{\partial_\mu a^0}{v}\sum_{\psi=q_L,u_R,d_R,\ell_L,e_R}\chi_\psi\,\bar\psi\gamma^\mu\psi$$
$$= -\frac{\partial_\mu a^0}{2v}\sum_{f=u,d,e,\nu}\left(\chi_V^f\,\bar\psi_f\gamma^\mu\psi_f + \chi_A^f\,\bar\psi_f\gamma^\mu\gamma_5\psi_f\right), \tag{54}$$

with

$$\chi_V^u = 2\alpha + x,\ \chi_V^d = 2\alpha + \frac{1}{x},\ \chi_V^e = 2\beta + \frac{1}{x},\ \chi_V^\nu = \beta,$$
$$\chi_A^u = x, \qquad \chi_A^d = \frac{1}{x}, \qquad \chi_A^e = \frac{1}{x}, \qquad \chi_A^\nu = -\beta. \tag{55}$$

The vector current couplings are α and β-dependent, since those parameters reflect the conservation of $\mathcal{B}$ and $\mathcal{L}$, which are vectorial [10]. The β dependence of χ_A^ν is related to the peculiar nature of neutrinos, which are kept massless. As we will see in Sect. 4.3.4, it disappears if right handed neutrinos ν_R together with the $\mathcal{L}$-invariant Yukawa interaction $\bar\nu_R Y_\nu \ell_L \Phi_1$ are added.

[4] These results apply to axion like particles but also to heavier pseudoscalars.

[5] Note that one cannot start from the rotated basis Eq. (40) to exponentiate the Goldstone field A^0, because such a representation would not be canonical for Φ_H since it has no vacuum expectation value [23].

In addition, given the $U(1)_{PQ}$ charges in Eq. (52), the fermion reparametrization is anomalous and its Jacobian has to be included

$$\delta\mathcal{L}_{\rm Jac} = \frac{a^0}{16\pi^2 v}\left(g_s^2\mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + g^2\mathcal{N}_L W^i_{\mu\nu}\tilde{W}^{i,\mu\nu} + g'^2\mathcal{N}_Y B_{\mu\nu}\tilde{B}^{\mu\nu}\right). \quad (56)$$

Given the SM fermion quantum numbers, the coefficients are

$$\begin{aligned}\mathcal{N}_C &= \sum_{\psi=q_L,u_R^\dagger,d_R^\dagger} d_L^\psi C_C^\psi \chi_\psi \\ &= \frac{1}{2}(-2\chi_{q_L} + \chi_{u_R} + \chi_{d_R}) = \frac{1}{2}\left(x + \frac{1}{x}\right), \quad (57)\\ \mathcal{N}_L &= \sum_{\psi=q_L,\ell_L} d_C^\psi C_L^\psi \chi_\psi \\ &= \frac{1}{2}(-N_C\chi_{q_L} - \chi_{\ell_L}) = -\frac{1}{2}(N_C\alpha + \beta), \quad (58)\\ \mathcal{N}_Y &= \sum_{\psi=q_L,u_R^\dagger,d_R^\dagger,\ell_L,e_R^\dagger} d_L^\psi d_C^\psi C_Y^\psi \chi_\psi \\ &= \frac{1}{2}(N_C\alpha + \beta) + N_C\left(\frac{4}{9}x + \frac{1}{9x}\right) + \frac{1}{x}, \quad (59)\end{aligned}$$

where $d^\psi_{C,L}$, $C^\psi_{C,L}$ are the $SU(3)_C$ and $SU(2)_L$ dimensions and quadratic Casimir invariant of the representation carried by the field ψ, respectively, and by extension, $C_Y^\psi = Y(\psi)^2/4$ with $Y(q_L, u_R, d_R, \ell_L, e_R) = (1/3, 4/3, -2/3, -1, -2)$. These expressions correspond to Eq. (27), choosing $a = -b = 1$ to reflect the absence of anomalies for the SM gauge symmetries (so no Bose symmetry here).

The parameters α and β enter in a combination reflecting the anomaly in the $\mathcal{B}+\mathcal{L}$ current. Interestingly, neither the QED nor the QCD coefficients depend on them, but those to electroweak gauge bosons do. To see this, it sufficies to plug $W^3_\mu = c_W Z_\mu + s_W A_\mu$, $B_\mu = -s_W Z_\mu + c_W A_\mu$ and $g' c_W = e = g s_W$ in Eq. (56):

$$\begin{aligned}\delta\mathcal{L}_{\rm Jac} = \frac{a^0}{16\pi^2 v}\Big(&g_s^2\mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + e^2\mathcal{N}_{em}F_{\mu\nu}\tilde{F}^{\mu\nu} \\ &+\frac{2e^2}{c_W s_W}\left(\mathcal{N}_L - s_W^2\mathcal{N}_{em}\right)Z_{\mu\nu}\tilde{F}^{\mu\nu} \\ &+\frac{e^2}{c_W^2 s_W^2}\left((1-2s_W^2)\mathcal{N}_L + s_W^4\mathcal{N}_{em}\right)Z_{\mu\nu}\tilde{Z}^{\mu\nu} \\ &+2\mathcal{N}_L g^2 W^+_{\mu\nu}\tilde{W}^{-,\mu\nu}\Big), \quad (60)\end{aligned}$$

and the free parameters cancel out in

$$\mathcal{N}_{em} \equiv \mathcal{N}_L + \mathcal{N}_Y = N_C\left(\frac{4}{9}x + \frac{1}{9x}\right) + \frac{1}{x}. \quad (61)$$

Concerning the strong CP puzzle, the polar representation makes the shift symmetry explicit. Under a $U(1)_{PQ}$ transformation of parameter θ, all the fields are invariant but for $a^0 \to a^0 + v\theta$. So, if there is a $\theta_{QCD}G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}$ term in the gauge sector, it can be eliminated by an adequate shift of the a^0 field. As said in the introduction, what happens in reality is that one is forced by non-perturbative QCD effects to shift the a^0 field by precisely that amount [1,2]. Once the a^0 shift is set by QCD, the electroweak $\theta_L W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}$ can still be removed by an appropriate choice of $N_C\alpha + \beta$. All that remains then is the abelian $\theta_Y B_{\mu\nu}\tilde{B}^{\mu\nu}$, which is an inoccuous total derivative. We thus recover the well-known fact that the electroweak θ term can be rotated away thanks to the $\mathcal{B}$ and $\mathcal{L}$ invariance of the model.

4.3 Matching the polar and linear representations

The physics must not depend on the representation chosen for the scalar fields or on the parametrization of the fermion fields. So, the one-loop $A^0 \to V_1V_2$ amplitudes of Eq. (43) must match the $a^0 \to V_1V_2$ amplitudes computed using the local terms $\delta\mathcal{L}_{\rm Jac}$ together with the triangle graphs arising from the derivative interactions in $\delta\mathcal{L}_{\rm Der}$ (these two contributions are of the same order in the coupling constants). At this stage, it is immediately clear that the triangle graphs will play a crucial role. Indeed, taken alone, the contributions from the anomalous local terms of $\delta\mathcal{L}_{\rm Jac}$ match those found in the linear case in the $m_{u,d,e} \to \infty$ limit, Eq. (47), only for the $\gamma\gamma$ and gg final states. For the others, there is no dependence on α and β in Eq. (47), but there is one through $\mathcal{N}_L$ in $\delta\mathcal{L}_{\rm Jac}$. Even worse, no choice of these parameters could make the two results compatible because the relative strengths of the $\gamma\gamma$, γZ, ZZ, and W^+W^- decays are irremediably different. As we will now detail, it is only once the anomalous contributions from $\delta\mathcal{L}_{\rm Jac}$ precisely cancel out with those hidden in the triangle graphs that the $a^0 \to V_1V_2$ amplitudes match the $A^0 \to V_1V_2$ amplitudes.

Throughout this section, the momentum flow is defined as $a^0(q_1+q_2) \to V_1(q_1,\alpha)V_2(q_2,\beta)$, with $V_{1,2} = g, \gamma, Z, W$, and neither the final gauge bosons nor the initial axion are necessarily on-shell. The final amplitudes thus depend on q_1^2, q_2^2, and $(q_1+q_2)^2$, and are to be compared to those obtained in the linear case in Eq. (43).

4.3.1 The $a^0 \to \gamma\gamma$ and $a^0 \to gg$ decays

For the decays into photons or gluons, only the AVV triangle contributes since $SU(3)$ and $U(1)_{em}$ are vector symmetries.

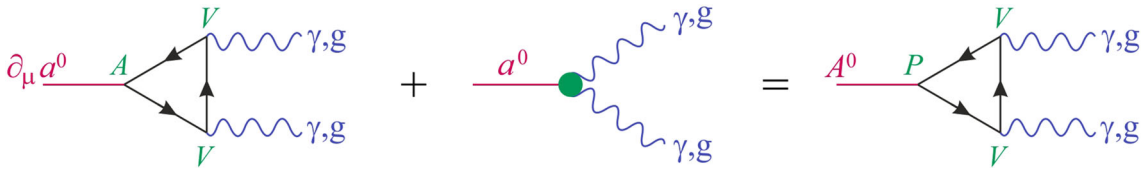

Fig. 3 Representation of the contributions to $a^0 \to \gamma\gamma, gg$ in the polar representation, and their matching with the $A^0 \to \gamma\gamma, gg$ amplitude of the linearly realized theory. The notation P, V, A denotes pseudoscalar, vector, and axial vertices, that is, γ_5, γ^μ, and $\gamma^\mu\gamma_5$ Dirac structures. The green disk depicts the local anomalous vertex derived from Eq. (56). All the SM fermions but the neutrinos circulate in the $\gamma\gamma$ loops, while only quarks occur for the gg loops

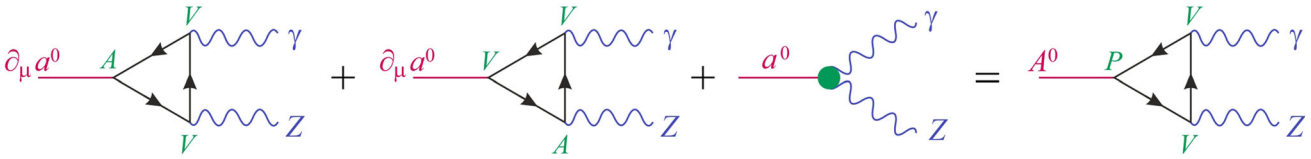

Fig. 4 Contributions to the axion decay into γZ in the polar (left) and linear (right) representations, using the same notations as in Fig. 3. All the SM fermions but the neutrinos circulate in the loops

Then, the situation is exactly the same as in the toy model of Sect. 2. Setting $a = -b = 1$ in Eq. (30), we find

$$\begin{aligned}\mathcal{M}(a^0 \to \gamma\gamma)^{AVV}_{\text{Der}} &= -\frac{e^2}{2v}\sum_{f=u,d,e} d_C^f \times \chi_A^f Q_f^2 \times i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{AVV}(f) \\ &= -\mathcal{M}(a^0 \to \gamma\gamma)_{\text{Jac}} + \mathcal{M}(A^0 \to \gamma\gamma)_{\text{Linear}}\,, \end{aligned} \tag{62}$$

and

$$\begin{aligned}\mathcal{M}(a^0 \to gg)^{AVV}_{\text{Der}} &= -\frac{g_s^2}{2v}\sum_{f=u,d} \chi_A^f C_C^f \times i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{AVV}(f) \\ &= -\mathcal{M}(a^0 \to gg)_{\text{Jac}} + \mathcal{M}(A^0 \to gg)_{\text{Linear}}, \end{aligned} \tag{63}$$

where $d_C^{u,d} = N_C$, $d_C^{e,\nu} = 1$, Q^f is the electric charge of f, and $C_C^{u,d} = 1/2$, $C_C^{e,\nu} = 0$. The amplitudes in the polar representation thus trivially match that in the linear case, see Fig. 3,

$$\begin{aligned}\mathcal{M}(a^0 \to \gamma\gamma, gg)_{\text{Polar}} &= \mathcal{M}(a^0 \to \gamma\gamma, gg)^{AVV}_{\text{Der}} \\ &+\mathcal{M}(a^0 \to \gamma\gamma, gg)_{\text{Jac}} = \mathcal{M}(A^0 \to \gamma\gamma, gg)_{\text{Linear}}. \end{aligned} \tag{64}$$

In other words, the anomalous contact interactions do cancel out systematically with the anomalous part of the triangle graphs.

As for the toy model, the polar representation is thus interesting only to make the shift symmetry manifest, and because the contact $a^0\gamma\gamma$ and a^0gg interactions read off $\delta\mathcal{L}_{\text{Jac}}$ are reliable book-keeping of the effects of heavy fermions. Specifically, $\mathcal{M}(a^0 \to \gamma\gamma, gg)^{AVV}_{\text{Der}} \overset{m\to\infty}{=} 0$ implies that $\mathcal{M}(A^0 \to \gamma\gamma, gg)_{\text{Linear}} \overset{m\to\infty}{=} \mathcal{M}(a^0 \to \gamma\gamma, gg)_{\text{Jac}}$. Finally, remark that the cancellation of the local anomalous terms ensures $\mathcal{M}(a^0 \to \gamma\gamma, gg)_{\text{Polar}} = 0$ in the $m_{u,d,e} \to 0$ limit. So, though interpreting the axion coupling to photons or gluons as induced by the anomaly is incorrect, this misidentification does not lead to serious consequences for those final states. For heavy fermions, the coupling to gluons is tuned by $\mathcal{N}_C$, and that to photons by $\mathcal{N}_{em}$, and their ratio, when restricted to quarks, give back the usual $\mathcal{N}^q_{em}/\mathcal{N}_C = 8/3$. However, as we will see in the next subsection, interpreting the axion coupling involving at least one electroweak gauge boson as induced by the anomaly is not only wrong in principle but also leads to incorrect couplings.

4.3.2 The $a^0 \to \gamma Z$ decay

For electroweak gauge bosons in the final state, the situation is less simple. Consider first the γZ final state. The derivative interactions induce again only the AVV triangle graphs since the photon coupling is vectorial. However, this time we have two possible contributions, depending on which current is carrying the anomaly, see Fig. 4. First, there are the usual $A(\partial_\mu a^0) - V(\gamma) - V(Z)$ triangles, for which the anomaly is in the axial current. Using Eq. (30) with $a = -b = 1$, we find

$$\begin{aligned}\mathcal{M}(a^0 \to \gamma Z)^{AVV}_{\text{Der}} &= -\frac{1}{2v}\sum_{f=u,d,e} d_C^f \times \chi_A^f g^f_{V,\gamma} g^f_{V,Z} \times i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{AVV}(f) \\ &= -\frac{1}{4\pi^2 v}\frac{e^2}{c_W s_W}\left(\mathcal{N}_0 - s_W^2 \mathcal{N}_{em}\right)\varepsilon^{\varepsilon_1^*\varepsilon_2^* q_1 q_2} \\ &+ \mathcal{M}(A^0 \to \gamma Z)_{\text{Linear}}, \end{aligned} \tag{65}$$

where $\varepsilon^{\varepsilon_1^*\varepsilon_2^* q_1 q_2} \equiv \varepsilon^{\alpha\beta\mu\nu}\varepsilon(q_1)^*_\alpha \varepsilon(q_2)^*_\beta q_{1\mu} q_{2\nu}$. But contrary to the $\gamma\gamma$ and gg final state, there are now new contributions

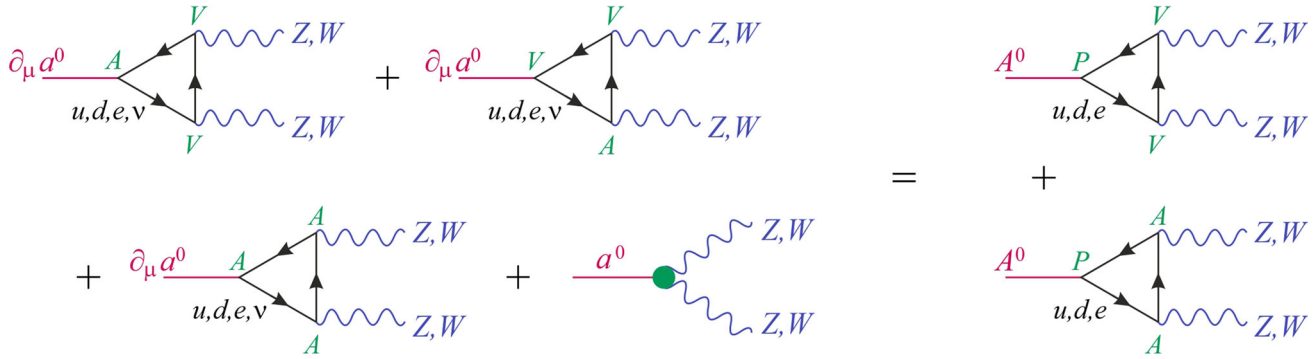

Fig. 5 Contributions to the axion decay into ZZ and W^+W^- in the polar (left) and linear (right) representations, using the same notations as in Fig. 3. For ZZ, all the SM fermions circulate in the loops in the non-linear theory, but the neutrino is absent in the linear one since we take it as massless. For W^+W^-, it is understood that in between the two Ws, the $SU(2)_L$ partner of the indicated fermion propagates

from the $V(\partial_\mu a^0) - V(\gamma) - A(Z)$ triangles, with the axion vector couplings of Eq. (54). Using again Eq. (30) but this time with $a = b = 1$ to preserve $SU(2)_L \otimes U(1)_Y$, and noting that $-i(q_1)_\gamma \mathcal{T}^{\alpha\gamma\beta}_{AVV} = -i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{VAV}$,

$$\begin{aligned}\mathcal{M}(a^0 \to \gamma Z)^{VAV}_{\text{Der}} &= -\frac{1}{2v}\sum_{f=u,d,e} d^f_C \times \chi^f_V g^f_{V,\gamma} g^f_{A,Z} \times (-i)(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{VAV} \\ &= \frac{1}{4\pi^2 v}\frac{e^2}{c_W s_W}(\mathcal{N}_0 - \mathcal{N}_L)\,\varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2}. \end{aligned} \tag{66}$$

As explained in Sect. 3, this contribution is free of any mass-dependent term because the naive vector current conservation must hold, up to the anomaly.

These two contributions from the derivative interactions combine with the Jacobian term from Eq. (60),

$$\mathcal{M}(a^0 \to \gamma Z)_{\text{Jac}} = \frac{1}{4\pi^2 v}\frac{e^2}{c_W s_W}\left(\mathcal{N}_L - s_W^2 \mathcal{N}_{em}\right)\varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2}, \tag{67}$$

to give the total decay amplitude in the polar representation:

$$\begin{aligned}\mathcal{M}(a^0 \to \gamma Z)_{\text{Polar}} &= \mathcal{M}(a^0 \to \gamma Z)^{AVV}_{\text{Der}} \\ &\quad + \mathcal{M}(a^0 \to \gamma Z)^{VAV}_{\text{Der}} + \mathcal{M}(a^0 \to \gamma Z)_{\text{Jac}} \\ &= \mathcal{M}(A^0 \to \gamma Z)_{\text{Linear}}, \end{aligned} \tag{68}$$

as it should. We can now understand why $\mathcal{M}(a^0 \to \gamma Z)_{\text{Jac}}$ does not match $\mathcal{M}(A^0 \to \gamma Z)_{\text{Linear}}$ in the $m_{u,d,e} \to \infty$ limit. Indeed, while $\mathcal{M}(a^0 \to \gamma Z)^{AVV}_{\text{Der}}$ vanishes in the $m_{u,d,e} \to \infty$ limit, $\mathcal{M}(a^0 \to \gamma Z)^{VAV}_{\text{Der}}$ obviously does not since it is independent of $m_{u,d,e}$. In that limit, we should write

$$\begin{aligned}&\mathcal{M}(A^0 \to \gamma Z)_{\text{Linear}}\Big|_{m_{u,d,e}\to\infty} \\ &\quad = \mathcal{M}(a^0 \to \gamma Z)^{VAV}_{\text{Der}} + \mathcal{M}(a^0 \to \gamma Z)_{\text{Jac}}, \end{aligned} \tag{69}$$

in which the parameters α and β cancel out. So, for a chiral gauge theory, the local terms coming from $\delta\mathcal{L}_{\text{Jac}}$ are no longer reliable book-keeping of the effect of heavy fermions, because part of the anomaly is hidden in the VAV triangle.

Actually, the reason why the VAV triangle plays a role can be understood directly from the fermion reparametrization, given the charges in Eq. (52). When α and/or β are different from zero, the fermionic current associated to $U(1)_{PQ}$ has a component aligned with baryon and/or lepton number, respectively. The fermionic reparametrization thus generates the anomalous $\mathcal{B}+\mathcal{L}$ interactions: those correspond to the $N_C\alpha + \beta$ terms of $\mathcal{N}_L$ and $\mathcal{N}_Y$ in Eq. (56). But as for the toy model, these interactions are spurious and must cancel with the anomalies present in the triangle graphs induced by the derivative interactions $\delta\mathcal{L}_{\text{Der}}$. This must necessarily come from a breakdown of the vector Ward identity for the VAV triangle since $\mathcal{B}$ and $\mathcal{L}$ are purely vectorial symmetries [24].

4.3.3 The $a^0 \to ZZ$ and $a^0 \to W^+W^-$ decays

Turning first to the $a^0 \to ZZ$ amplitude, the derivative interactions induce new types of diagrams: the AAA triangles and graphs with neutrinos circulating in the loop, see Fig. 5.

Proceeding with the calculation, the three contributions from the derivative terms are

$$\begin{aligned}
&\mathcal{M}(a^0 \to ZZ)^{AVV}_{\text{Der}} \\
&= -\frac{1}{2v} \sum_{f=u,d,e,\nu} d_C^f \times \chi_A^f g_{V,Z}^f g_{V,Z}^f \times i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{AVV}(f) \\
&= \frac{-1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \left(\frac{3}{4}\mathcal{N}_1 - 2s_W^2 \mathcal{N}_0 + s_W^4 \mathcal{N}_{em} - \frac{\beta}{16} \right) \times \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2} + \mathcal{M}(A^0 \to ZZ)^{PVV}_{\text{Linear}},
\end{aligned} \tag{70}$$

where $\mathcal{M}(A^0 \to ZZ)^{PVV}_{\text{Linear}}$ is the part of Eq. (43) proportional to $\mathcal{T}^{\alpha\beta}_{PVV}$. The neutrino contribution is absent from that term since the loop function in Eq. (32) vanishes for massless fermions. The next piece comes from the divergence of the axionic vector current,

Though the anomalous terms again sum up to zero, they are now completely entangled in the three types of triangle graphs, which all depend on the free parameters α and β. As a result, none of the derivative amplitudes vanish in the $m_{u,d,e} \to \infty$ limit. Specifically, because neutrinos remain massless,

$$\begin{aligned}
&\mathcal{M}(a^0 \to ZZ)^{VAV}_{\text{Der}} \\
&= -\frac{1}{2v} \sum_{f=u,d,e,\nu} d_C^f \times \chi_V^f (g_{V,Z}^f g_{A,Z}^f + g_{A,Z}^f g_{V,Z}^f) \times (-i)(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{VAV} \\
&= \frac{1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \left(\frac{3}{2}\mathcal{N}_1 - 2s_W^2 \mathcal{N}_0 - (1-2s_W^2)\mathcal{N}_L - \frac{\beta}{8} \right) \times \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2}.
\end{aligned} \tag{71}$$

There is no mass-dependent terms here because of the vector Ward identity. Finally, the AAA triangle gives, using Eq. (31) with $a = -b = 1$

$$\begin{aligned}
&\mathcal{M}(a^0 \to ZZ)^{AAA}_{\text{Der}} \\
&= -\frac{1}{2v} \sum_{f=u,d,e,\nu} d_C^f \times \chi_A^f g_{A,Z}^f g_{A,Z}^f \times i(q_1+q_2)_\gamma \mathcal{T}^{\gamma\alpha\beta}_{AAA} \\
&= \frac{-1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \left(\frac{3}{4}\mathcal{N}_1 - \frac{\beta}{16} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2} + \mathcal{M}(A^0 \to ZZ)^{PAA}_{\text{Linear}},
\end{aligned} \tag{72}$$

where $\mathcal{M}(A^0 \to ZZ)^{PAA}_{\text{Linear}}$ is the part of Eq. (43) proportional to $\mathcal{T}^{\alpha\beta}_{PAA}$. Again, the neutrino contribution to that part vanishes. Combining the three contributions of the derivative interactions with that of the Jacobian

$$\begin{aligned}
\mathcal{M}(a^0 \to ZZ)_{\text{Jac}} &= \frac{1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \\
&\times \left((1-2s_W^2)\mathcal{N}_L + s_W^4 \mathcal{N}_{em} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2},
\end{aligned} \tag{73}$$

all the local anomalous contributions cancel exactly, leaving:

$$\begin{aligned}
\mathcal{M}(a^0 \to ZZ)_{\text{Polar}} &= \mathcal{M}(a^0 \to ZZ)^{AVV}_{\text{Der}} \\
&+ \mathcal{M}(a^0 \to ZZ)^{VAV}_{\text{Der}} \\
&+ \mathcal{M}(a^0 \to ZZ)^{AAA}_{\text{Der}} \\
&+ \mathcal{M}(a^0 \to ZZ)_{\text{Jac}} = \mathcal{M}(A^0 \to ZZ)_{\text{Linear}}.
\end{aligned} \tag{74}$$

$$\mathcal{M}(a^0 \to ZZ)^{AVV}_{\text{Der}}\Big|_{m_{u,d,e}\to\infty} = -\frac{1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \left(-\frac{\beta}{16} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2}, \tag{75}$$

where β comes from the neutrino coupling χ_A^ν in Eq. (55). Also,

$$\begin{aligned}
&\mathcal{M}(a^0 \to ZZ)^{AAA}_{\text{Der}}\Big|_{m_{u,d,e}\to\infty} \\
&= -\frac{1}{4\pi^2 v} \frac{e^2}{c_W^2 s_W^2} \left(\frac{1}{2}\mathcal{N}_1 - \frac{\beta}{16} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2},
\end{aligned} \tag{76}$$

because in addition to massless neutrinos, the mass term respects Bose symmetry but not the anomaly which resides only in the axion axial current (there is a relative factor of 1/3, see Eq. (34) with $a = -b = 1$). Finally, $\mathcal{M}(a^0 \to ZZ)^{VAV}_{\text{Der}}$ does not vanish since it is mass-independent. Because of all this, $\mathcal{M}(a^0 \to ZZ)_{\text{Jac}}$ is clearly not identical to $\mathcal{M}(A^0 \to ZZ)_{\text{Linear}}$ in the $m_{u,d,e} \to \infty$ limit, and cannot be used as an estimate of the electroweak couplings of the axion.

The $a^0 \to W^+W^-$ amplitude shares many similarities with that for $a^0 \to ZZ$, so we will be very brief. Following the same steps, the various contributions to the decay in the polar representation are

$$\begin{aligned}
\mathcal{M}(a^0 \to W^+W^-)^{AVV}_{\text{Der}} &= -\frac{g^2}{4\pi^2 v} \left(\frac{3}{4}\mathcal{N}_1 - \frac{\beta}{16} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2} \\
&+ \mathcal{M}(A^0 \to W^+W^-)^{PVV}_{\text{Linear}},
\end{aligned} \tag{77}$$

$$\mathcal{M}(a^0 \to W^+W^-)^{VAV}_{\text{Der}} = +\frac{g^2}{4\pi^2 v} \left(\frac{3}{2}\mathcal{N}_1 - \mathcal{N}_L - \frac{\beta}{8} \right) \varepsilon^{\varepsilon_1^* \varepsilon_2^* q_1 q_2}, \tag{78}$$

$$\mathcal{M}(a^0 \to W^+W^-)^{AAA}_{\rm Der} = -\frac{g^2}{4\pi^2 v}\left(\frac{3}{4}\mathcal{N}_1 - \frac{\beta}{16}\right)\varepsilon^{\varepsilon_1^*\varepsilon_2^* q_1 q_2} + \mathcal{M}(A^0 \to W^+W^-)^{PAA}_{\rm Linear}, \quad (79)$$

$$\mathcal{M}(a^0 \to W^+W^-)_{\rm Jac} = +\frac{g^2}{4\pi^2 v}\left(\mathcal{N}_L\right)\varepsilon^{\varepsilon_1^*\varepsilon_2^* q_1 q_2}. \quad (80)$$

Note well that it is $\mathcal{N}_1$ that occurs in these amplitudes, and not the $\mathcal{N}_2$ coefficients to which $\mathcal{M}(A^0 \to W^+W^-)_{\rm Linear}$ tends to in the $m_{u,d,e} \to \infty$ limit. Still, the sum of these four contributions reproduces again the linear result, $\mathcal{M}(a^0 \to W^+W^-)_{\rm Polar} = \mathcal{M}(A^0 \to W^+W^-)_{\rm Linear}$. As none of the triangle amplitudes vanish in the $m_{u,d,e} \to \infty$ limit, $\mathcal{M}(a^0 \to W^+W^-)_{\rm Jac}$ is clearly not identical to $\mathcal{M}(A^0 \to W^+W^-)_{\rm Linear}$ in the $m_{u,d,e} \to \infty$ limit.

4.3.4 Impact of heavy neutrinos

The effective interactions in Eq. (47) involve five $\mathcal{N}_i$ coefficients, but this is partly due to the non-decoupling of neutrinos. To see this, let us add right handed neutrinos together with the Yukawa interaction $\bar{\nu}_R Y_\nu \ell_L \Phi_1$. Then, the effective amplitude in the linear representation is trivially obtained by extending the sum in Eq. (43) to $f = u, d, e, \nu$. If, for the sake of the argument, we now take the infinite mass limit for all fermions, $m_{u,d,e,\nu} \to \infty$, the effective interactions in Eq. (47) are recovered but with

$$\mathcal{N}_1 = \mathcal{N}_2 = \frac{1}{12}(N_C + 1)\left(x + \frac{1}{x}\right) \equiv \mathcal{N}_\nu. \quad (81)$$

By contrast, it is amusing to remark that in the polar representation, the anomalous interactions of Eq. (56) are not modified by the presence of right-handed neutrinos since they are neutral under the whole SM gauge group. Yet, right-handed neutrinos do modify the derivative interactions in Eq. (55) to $\chi_V^\nu = 2\beta + x$ and $\chi_A^\nu = x$. The various triangle amplitudes are then obtained from those quoted in the previous section by replacing $\mathcal{N}_1 \to \mathcal{N}_\nu$ and discarding the β terms. While $\mathcal{M}(a^0 \to ZZ, W^+W^-)^{AVV}_{\rm Der}$ now vanish in the $m_{u,d,e,\nu} \to \infty$ limit, this does not change our conclusions and $\mathcal{M}(a^0 \to ZZ, W^+W^-)_{\rm Jac}$ remain different from $\mathcal{M}(A^0 \to ZZ, W^+W^-)_{\rm Linear}$ in that limit.

As a final note, it is interesting to remark that once certain that α and β disappear from the physical amplitudes, one can chose to fix them as one wishes. In particular, as said before, they can be set to eliminate the electroweak θ term. Alternatively, one can set $\alpha = \beta = 0$, thereby getting rid of the $a^0 W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}$ coupling altogether. As we have seen, this does not forbids $a^0 \to W^+W^-$ since the amplitude is independent of α and β. Finally, this freedom can also be used to allow for the presence of a lepton number violating Majorana mass term, as well as an effective dimension-five $(\bar{\ell}_L^C H_1)(\ell_L H_1)/\Lambda$ operator. We will not explore such settings here, see Ref. [25] for studies involving axions together with neutrino Majorana mass terms.

4.4 Application to the DFSZ axion model

The Peccei-Quinn model is ruled out experimentally because the coupling of the axion to fermions, tuned by the electroweak vacuum expectation value, is too large. One way to extend the model and render the axion invisible was proposed not long after the original idea and is called the DFSZ axion model [26,27]. Our goal here is to derive the electroweak couplings for that model.

The starting point is to extend the THDM by adding a complex scalar field ϕ, singlet under the SM gauge symmetries, and constrain the scalar potential to

$$\begin{aligned} V_{\rm Scalar} = {} & m_1^2\Phi_1^\dagger\Phi_1 + m_2^2\Phi_2^\dagger\Phi_2 + \frac{\lambda_1}{2}(\Phi_1^\dagger\Phi_1)^2 \\ & + \frac{\lambda_2}{2}(\Phi_2^\dagger\Phi_2)^2 + \lambda_3(\Phi_1^\dagger\Phi_1)(\Phi_2^\dagger\Phi_2) \\ & + \lambda_4(\Phi_2^\dagger\Phi_1)(\Phi_1^\dagger\Phi_2) + \mu^2\phi^\dagger\phi + \frac{\lambda}{2}(\phi^\dagger\phi)^2 \\ & + a_1\phi^\dagger\phi\Phi_1^\dagger\Phi_1 + a_2\phi^\dagger\phi\Phi_2^\dagger\Phi_2 \\ & + \left[-\lambda_{12}^2\phi^2\Phi_1^\dagger\Phi_2 + h.c.\right]. \end{aligned} \quad (82)$$

In this way, the whole Lagrangian remains invariant under independent rephasing of the doublet provided the singlet has the charge:

$$\begin{aligned} & \Phi_1 \to \exp(i\alpha_1\theta)\Phi_1, \quad \Phi_2 \to \exp(i\alpha_2\theta)\Phi_2, \\ & \phi \to \exp(i\alpha\theta)\phi, \quad \alpha_1 - \alpha_2 = 2\alpha. \end{aligned} \quad (83)$$

The Type II Yukawa couplings of Eq. (38) then imply chiral charges for the fermions when $\alpha_1 \neq \alpha_2$, and part of the $U(1)_1 \otimes U(1)_2$ symmetry is anomalous.

Let us take the linear representation (compare with Eq. (40))

$$\begin{aligned} & \Phi_h = \frac{1}{\sqrt{2}}\begin{pmatrix} G^+ \\ v + \phi_h + iG^0 \end{pmatrix}, \quad \Phi_H = \frac{1}{\sqrt{2}}\begin{pmatrix} \sqrt{2}H^+ \\ \phi_H + i\phi_A \end{pmatrix} \\ & \phi = \frac{1}{\sqrt{2}}(v_s + \sigma_S + i\pi_S), \end{aligned} \quad (84)$$

where $\Phi_{h,H}$ are given in terms of $\Phi_{1,2}$ in Eq. (39). Plugging these representations in the potential, and extracting the quadratic terms, the G^0 field is immediately massless and corresponds to the Would-be Goldstone of the Z boson. On the other hand, the ϕ_A and π_S fields mix,

$$\begin{pmatrix} A^0 \\ \pi^0 \end{pmatrix} = \frac{1}{\sqrt{v_s^2 + v^2\sin^2 2\beta}}\begin{pmatrix} v\sin 2\beta & v_s \\ -v_s & v\sin 2\beta \end{pmatrix}\begin{pmatrix} \phi_A \\ \pi_S \end{pmatrix}, \quad (85)$$

with $M^2_{A^0} = 0$ and $M^2_{\pi^0} = 2\lambda_{12}(v^2 \sin 2\beta + v_s^2/\sin 2\beta)$. The A^0 field is the axion, and the π^0 field is a heavy pseudoscalar Higgs boson. While $A^0 \to \phi_A$ when $v_s \to 0$, reproducing the THDM discussed previously, the opposite holds when $v_s \gg v$ [26,27]:

$$A^0 = \pi_S + \frac{v}{v_s}\phi_A \sin 2\beta + \mathcal{O}(v^2/v_s^2). \tag{86}$$

Then, since π_S is not coupled to fermions, it is easy to see that the $A^0 \to V_1 V_2$ amplitudes are simply rescaled by $v \sin 2\beta/v_s$, i.e.,

$$\begin{aligned}\mathcal{L}^{\text{eff,DFSZ}}_{\text{Linear}} = \frac{A^0 \sin 2\beta}{16\pi^2 v_s}\Big(& g_s^2 \mathcal{N}_C G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} + e^2 \mathcal{N}_{em} F_{\mu\nu}\tilde{F}^{\mu\nu} \\ &+ \frac{2e^2}{c_W s_W}\left(\mathcal{N}_0 - s_W^2 \mathcal{N}_{em}\right) Z_{\mu\nu}\tilde{F}^{\mu\nu} \\ &+ \frac{e^2}{c_W^2 s_W^2}\left(\mathcal{N}_1 - 2s_W^2 \mathcal{N}_0 + s_W^4 \mathcal{N}_{em}\right) Z_{\mu\nu}\tilde{Z}^{\mu\nu} \\ &+ 2\mathcal{N}_2 g^2 W^+_{\mu\nu}\tilde{W}^{-,\mu\nu}\Big),\end{aligned} \tag{87}$$

with the same coefficients as in Eq. (48). In particular, the axion couplings to γZ, ZZ, and W^+W^- are not aligned with the purely anomalous interactions one derives looking only at the PQ charges of the fermion fields.

4.5 Anomaly cancellation in generic axion models

In the previous sections, we showed how the anomalies present in the non-linearly realized axion models necessarily cancel out, since they are spuriously generated by reparametrizing an anomaly-free model. Here, we want to follow the opposite route. Starting from a Non Renormalizable (NR) effective Lagrangian valid above the electroweak scale:

$$\begin{aligned}\mathcal{L}^{\text{eff}}_{NR} =& \frac{1}{2}\partial_\mu a^0 \partial^\mu a^0 - \frac{\partial_\mu a^0}{2v_a} \\ &\times \sum_{f=u,d,e,\nu}\left(\chi_V^f\, \bar{\psi}_f \gamma^\mu \psi_f + \chi_A^f\, \bar{\psi}_f \gamma^\mu \gamma_5 \psi_f\right) \\ &+ \frac{a^0}{16\pi^2 v_a}\Big(g_s^2 \mathcal{N}_C^{eff} G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} \\ &+ g^2 \mathcal{N}_L^{eff} W_{\mu\nu}\tilde{W}^{\mu\nu} + g'^2 \mathcal{N}_Y^{eff} B_{\mu\nu}\tilde{B}^{\mu\nu}\Big),\end{aligned} \tag{88}$$

where v_a is some high scale, $\mathcal{N}^{eff}_{C,L,Y}$ as well as $\chi^f_{V,A}$ some a priori free Wilson coefficients, we want to show that whenever the anomalies cancel, this model collapses to a simple renormalizable theory.

To do this, starting from $\mathcal{L}^{\text{eff}}_{NR}$, we require the anomalous interactions $\mathcal{N}^{eff}_{C,L,Y}$ to cancel with the local anomalous terms present in the AVV, VAV, and AAA triangle graphs. Imposing this for the five channels $a^0 \to V_1 V_2$, $V_{1,2} = g, \gamma, Z, W$, four independent constraints are derived:

$$\mathcal{N}_C^{eff} = \frac{1}{2}(\chi_A^u + \chi_A^d), \tag{89}$$

$$\mathcal{N}_{em}^{eff} \equiv \mathcal{N}_L^{eff} + \mathcal{N}_Y^{eff} = \frac{4}{9}N_C \chi_A^u + \frac{1}{9}N_C \chi_A^d + \chi_A^e, \tag{90}$$

$$\begin{aligned}\mathcal{N}_L^{eff} =& \frac{1}{6}N_C(\chi_A^u - \chi_V^u) \\ &+ \frac{1}{12}N_C(\chi_A^d - \chi_V^d) + \frac{1}{4}(\chi_A^e - \chi_V^e),\end{aligned} \tag{91}$$

together with

$$\chi_A^\nu - \chi_V^\nu = \frac{1}{3}N_C(\chi_A^u - \chi_V^u) - \frac{1}{3}N_C(\chi_A^d - \chi_V^d) + \chi_A^e - \chi_V^e\,. \tag{92}$$

Once these conditions are imposed, all the $a^0 \to V_1 V_2$ processes are entirely driven by the non-anomalous parts of the AVV and AAA triangle graphs (see Figs. 3, 4, 5). The VAV triangles disappear since they have no mass-dependent terms. But then, these non-anomalous amplitudes satisfy the naive Ward identities, and are equivalently obtained from PVV and PAA triangles, provided

$$\mathcal{L}^{\text{eff}}_R = \frac{1}{2}\partial_\mu a^0 \partial^\mu a^0 - i\sum_{f=u,d,e}\frac{m_f}{v_a}\chi_P^f\, a^0 \bar{\psi}_f \gamma_5 \psi_f, \quad \text{with } \chi_P^f = \chi_A^f\,. \tag{93}$$

Phenomenologically, $\mathcal{L}^{\text{eff}}_R$ and $\mathcal{L}^{\text{eff}}_{NR}$ are indistinguishable once anomaly cancellation is enforced, even if $\mathcal{L}^{\text{eff}}_R$ has far fewer physical parameters. This can be understood as follows. Out of the initial ten parameters in $\mathcal{L}^{\text{eff}}_{NR}$ for massless neutrino ($\chi_V^\nu = -\chi_A^\nu$), four are eliminated by the anomaly cancellation constraints. Then, the naive Ward identities are valid again. The three $\chi_V^{u,d,e}$ parameters disappear thanks to $\partial_\mu V^\mu = 0$, and $\partial_\mu A^\mu = 2imP$ leaves only the three $\chi_P^f = \chi_A^f$ parameters as truly physical.

Now, starting from $\mathcal{L}^{\text{eff}}_R$, all the $a^0 \to V_1 V_2$ decay amplitudes are easily computed, and given by Eq. (43). In the $m_{u,d,e} \to \infty$ limit, keeping neutrino massless, these amplitudes match onto the effective couplings in Eq. (47), with

$$\mathcal{N}_C = \frac{1}{2}\left(\chi_A^u + \chi_A^d\right), \tag{94a}$$

$$\mathcal{N}_{em} = \frac{4}{9}N_C \chi_A^u + \frac{1}{9}N_C \chi_A^d + \chi_A^e, \tag{94b}$$

$$\mathcal{N}_0 = \frac{1}{4}\left(N_C\left(\frac{2}{3}\chi_A^u + \frac{1}{3}\chi_A^d\right) + \chi_A^e\right), \tag{94c}$$

$$\mathcal{N}_1 = \frac{1}{12}\left(N_C\left(\chi_A^u + \chi_A^d\right) + \chi_A^e\right), \tag{94d}$$

$$\mathcal{N}_2 = \frac{1}{12}\left(N_C\left(\chi_A^u + \chi_A^d\right) + \frac{3}{2}\chi_A^e\right). \tag{94e}$$

In this limit, we thus do not recover the manifestly $SU(2)_L \otimes U(1)_Y$ invariant pattern of anomalous interactions of $\mathcal{L}_{NR}^{\text{eff}}$. Finally, note that if one takes $\chi_A^u = \chi_P^u = x$, $\chi_A^d = \chi_P^d = \chi_A^e = \chi_P^e = 1/x$, the results in Eq. (48) are recovered. This shows that $\mathcal{L}_R^{\text{eff}}$ is the most generic implementation of the PQ (or DFSZ) axion model, where the $U(1)_{PQ}$ charges of the SM fermions can take any value.

5 Conclusion and perspective

In this paper, we derived the couplings of axions to gauge bosons. We started from a simplified toy model, and then extended the discussions to the KSVZ axion [13,14], the original PQ axion [1,2], and the DFSZ axion [26,27]. In each model, our strategy has been to match the axion decay modes computed using either a linear or a polar representation for the scalar field breaking the $U(1)_{PQ}$ symmetry. In this way, we were able to unravel the physics at play, and uncovered a number of interesting features:

- In the linear representation, the axion couplings to gauge bosons are not induced by the anomaly, but by non-anomalous pseudoscalar triangle graphs with fermions circulating in the loop. Though this has to our knowledge not been exploited before, these amplitudes can be identified with those of the pseudoscalar Higgs to gauge bosons calculated in the THDM, which have been known for a long time [21]. In the $m_{u,d,e} \to \infty$ limit, they match onto the effective interactions

$$\begin{aligned}\mathcal{L}^{\text{eff}} = \frac{a^0}{16\pi^2 v}\Big(& g_s^2 \mathcal{N}^{gg} G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} \\ &+ e^2\mathcal{N}^{\gamma\gamma}F_{\mu\nu}\tilde{F}^{\mu\nu} + \frac{2e^2}{c_W s_W}\left(\mathcal{N}_1^{\gamma Z} - s_W^2\mathcal{N}_2^{\gamma Z}\right)Z_{\mu\nu}\tilde{F}^{\mu\nu} \\ &+ \frac{e^2}{c_W^2 s_W^2}\left(\mathcal{N}_1^{ZZ} - 2s_W^2\mathcal{N}_2^{ZZ} + s_W^4\mathcal{N}_3^{ZZ}\right)Z_{\mu\nu}\tilde{Z}^{\mu\nu} \\ &+ 2\mathcal{N}^{WW} g^2 W^+_{\mu\nu}\tilde{W}^{-,\mu\nu}\Big),\end{aligned} \tag{95}$$

  with the coefficients shown in the first column of Table 1. In the opposite limit $m_{u,d,e} \to 0$, all these amplitudes vanish since the axion couplings to fermions are proportional to the fermion masses.
- In the context of axion models, it is customary to adopt a polar representation for the scalar fields. The pseudoscalar axion couplings to fermions are replaced by contact anomalous interactions to the gauge bosons and axion derivative interactions to the fermions. All these interactions are entirely fixed in terms of the assigned PQ charges of the SM fermions. As shown in Table 1, this parametrization is particularly convenient for $a^0 \to \gamma\gamma$ and $a^0 \to gg$ because the derivative interactions do not contribute in the $m_{u,d,e} \to \infty$ limit [10,11], and thus the strengths of the $a^0 \to \gamma\gamma$ and $a^0 \to gg$ processes can be read off the anomalous couplings $a^0 F_{\mu\nu}\tilde{F}^{\mu\nu}$ and $a^0 G^a_{\mu\nu}\tilde{G}^{a,\mu\nu}$. By contrast, and contrary to what is conjectured in Ref. [10], the central result of this paper is that for chiral gauge theories in which chiral fermions have non-trivial PQ charges, the strengths of the $a^0 \to \gamma Z$, ZZ, and W^+W^- processes do not match the anomalous couplings, even in the $m_{u,d,e} \to \infty$ limit, as evident comparing the first and second columns of Table 1. They thus cannot be encoded into the $SU(2)_L \otimes U(1)_Y$ invariant effective interactions $a^0 B_{\mu\nu}\tilde{B}^{\mu\nu}$ and $a^0 W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}$.
- The main reason for this mismatch is due to the presence of triangle graphs arising from the derivative interactions $\partial_\mu a^0 \bar{\psi}\gamma^\mu\gamma_5\psi$ and $\partial_\mu a^0 \bar{\psi}\gamma^\mu\psi$, which do not vanish in the $m_{u,d,e} \to \infty$ limit for chiral gauge theories. Importantly, even the vector couplings play a role since the anomalous breaking of the axionic vector current conservation enters through the VAV triangle graphs (see Figs. 4, 5). One important result of this paper is the proof that once all these triangle contributions (last three columns of Table 1) are summed with the local anomalous amplitudes (second column of Table 1), the THDM results are recovered. Without surprise, physical observables do not depend on the chosen parametrization, and this further confirms that the THDM calculation is correct.
- Our analysis also proves that the axion to gauge boson processes are not induced by the anomaly. This is obvious in the linear representation since the fermion loops driving these processes are not anomalous. In the polar representation, what happens is that the local anomalous interactions precisely cancel with the anomalies present in triangle graphs induced by the vector and axial derivative interactions, leaving the non-anomalous pseudoscalar triangles of the THDM as the only surviving mechanism driving all the axion to gauge bosons processes. Ultimately, this precise cancellation is ensured by the anomalous vector and axial Ward identities. It is required because all the amplitudes must vanish in the $m_{u,d,e,v} \to 0$ limit, but the anomalous contributions are independent of the fermion mass.[6]

[6] This interpretation is compatible with the matching observed in Table 1 in the $m_{u,d,e} \to \infty$ limit for $a^0 \to \gamma\gamma, gg$. Indeed, in the polar representation, these amplitudes receive two contributions, $\mathcal{M}_{ano}$ from the local anomalous term, and $\mathcal{M}_{der}$ from the triangle amplitudes built from $\partial_\mu a^0 \bar{\psi}\gamma^\mu\gamma_5\psi$. But in this case, the axial Ward identity translates as $\mathcal{M}_{der} = \mathcal{M}_{lin} - \mathcal{M}_{ano}$, with $\mathcal{M}_{lin}$ the THDM amplitude. So, even if parametrically, $\mathcal{M}_{lin} = \mathcal{M}_{ano}$ and $\mathcal{M}_{der} = 0$ when $m_{u,d,e} \to \infty$, $a^0 \to \gamma\gamma, gg$ are not induced by the anomaly since it cancels out in $\mathcal{M}_{ano} + \mathcal{M}_{der}$.

Table 1 Coefficients of the effective axion to gauge boson couplings of Eq. (95) in the $m_{u,d,e} \to \infty$ limit. For the linear representation, those are found directly from the THDM amplitudes of Ref. [21]. For the polar representation, the contributions of the local anomalous terms and that of the triangle amplitudes built on the axial (A) and vector (V) derivative interactions have to be added together. The fact that only the three independent coefficients $\mathcal{N}^{gg}, \mathcal{N}^{\gamma\gamma}$, and $\mathcal{N}_L$ occur for the local anomalous interactions comes from their $SU(2)_L \otimes U(1)_Y$ invariance. The coefficient $\mathcal{N}_L = -1/2(3\alpha + \beta)$, with α and β being the free parameters tuning the $U(1)_{\mathcal{B}}$ and $U(1)_{\mathcal{L}}$ components of $U(1)_{PQ}$. The explicit presence of β in the $a^0 \to ZZ, WW$ triangles is due to the peculiar nature of the neutrinos, which are kept massless. They do not contribute in the linear representation but have to explicitly appear in the anomalous interactions and derivative terms since those are $SU(2)_L \otimes U(1)_Y$ invariant

Linear $a^0\bar{\psi}\gamma_5\psi$	Anomalous interactions	Polar $\partial_\mu a^0\bar{\psi}\gamma^\mu\gamma_5\psi$		$\partial_\mu a^0\bar{\psi}\gamma^\mu\psi$
		AVV	AAA	VAV
$\mathcal{N}^{gg} = \frac{1}{2}\left(x + \frac{1}{x}\right)$	$\mathcal{N}^{gg}$	0	–	–
$\mathcal{N}^{\gamma\gamma} = \frac{4}{3}\left(x + \frac{1}{x}\right)$	$\mathcal{N}^{\gamma\gamma}$	0	–	–
$\mathcal{N}_1^{\gamma Z} = \frac{1}{2}\left(x + \frac{1}{x}\right)$	$\mathcal{N}_L$	0	–	$\mathcal{N}_1^{\gamma Z} - \mathcal{N}_L$
$\mathcal{N}_2^{\gamma Z} = \mathcal{N}^{\gamma\gamma}$	$\mathcal{N}^{\gamma\gamma}$	0	–	0
$\mathcal{N}_1^{ZZ} = \frac{1}{4}x + \frac{1}{3x}$	$\mathcal{N}_L$	$\frac{\beta}{16}$	$-\frac{1}{2}\mathcal{N}_1^{ZZ} + \frac{\beta}{16}$	$\frac{3}{2}\mathcal{N}_1^{ZZ} - \mathcal{N}_L - \frac{\beta}{8}$
$\mathcal{N}_2^{ZZ} = \mathcal{N}_1^{\gamma Z}$	$\mathcal{N}_L$	0	0	$\mathcal{N}_2^{ZZ} - \mathcal{N}_L$
$\mathcal{N}_3^{ZZ} = \mathcal{N}^{\gamma\gamma}$	$\mathcal{N}^{\gamma\gamma}$	0	0	0
$\mathcal{N}^{WW} = \frac{x}{4} + \frac{3}{8x}$	$\mathcal{N}_L$	$\frac{3}{2}\mathcal{N}^{WW} - \frac{3}{2}\mathcal{N}_1^{\gamma Z} + \frac{\beta}{16}$	$-\frac{1}{2}\mathcal{N}^{WW} + \frac{\beta}{16}$	$\frac{3}{2}\mathcal{N}_1^{\gamma Z} - \mathcal{N}_L - \frac{\beta}{8}$

- In all axion models where chiral fermions have PQ charges but conserve baryon and lepton numbers, both the vector couplings $\partial_\mu a^0\bar{\psi}\gamma^\mu\psi$ and the $a^0 W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}$, $a^0 B_{\mu\nu}\tilde{B}^{\mu\nu}$ anomalous contact interactions depend on some free parameters [10]. In Table 1, these free parameters enter into $\mathcal{N}_L$, and are related to the freedom to choose the PQ charge of left-handed fermions (there are also explicit occurrence of the free parameter β, but this is an artifact of keeping the neutrino massless). It is only once the failure of the naive vector Ward identity in the VAV triangle graphs is properly accounted for that these free parameters cancel out, as they should [24]. To understand how this cancellation is connected with the conservation of baryon and lepton numbers $\mathcal{B}$ and $\mathcal{L}$, notice first that the freedom to choose the PQ charges of the fermions is actually due to the invariance under $U(1)_{\mathcal{B}}$ and $U(1)_{\mathcal{L}}$. In general, the $U(1)_{PQ}$ symmetry has two arbitrary components aligned with $U(1)_{\mathcal{B}}$ and $U(1)_{\mathcal{L}}$. But then, these components translate into derivative couplings of the axion to the $\mathcal{B}$ and $\mathcal{L}$ fermionic vectorial currents together with a local $a^0 W^i_{\mu\nu}\tilde{W}^{i,\mu\nu}$ coupling from the $\mathcal{B}+\mathcal{L}$ anomaly. Clearly, all these couplings are spurious when the axion is a Goldstone boson living in a $\mathcal{B}$ and $\mathcal{L}$-invariant vacuum. Since $\mathcal{B}$ and $\mathcal{L}$ are not spontaneously broken, $\mathcal{N}_L$ must systematically drop out of physical observables, and Table 1 shows that this indeed occurs.
- In the linear representation, the calculation is done in the electroweak broken phase throughout. In the $m_{u,d,e} \to \infty$ limit, the amplitudes do not match onto $SU(2)_L \otimes U(1)_Y$ invariant operators, as shown in Table 1. Though this is somewhat expected, what is not obvious is the very peculiar way in which the $SU(2)_L \otimes U(1)_Y$ symmetry breaking seeps in. Indeed, in the non-linear representation, the anomalous interactions and the derivative interactions are manifestly $SU(2)_L \otimes U(1)_Y$ invariant once the SM fermions are assigned PQ charges in an $SU(2)_L \otimes U(1)_Y$ invariant way. Further, evidently, the anomalies present in the triangle graphs built on the derivative interactions do not break $SU(2)_L \otimes U(1)_Y$ since that would put the SM gauge invariance itself in jeopardy. So, in that representation, the fact that the chiral $SU(2)_L \otimes U(1)_Y$ symmetry is spontaneously broken enters only and entirely through the fermion masses. The triangle graph explicitly break the $SU(2)_L \otimes U(1)_Y$ symmetry because the $m_{u,d,e} \to \infty$ limit is incompatible with that symmetry. In the opposite limit $m_{u,d,e} \to 0$, the $SU(2)_L \otimes U(1)_Y$ symmetry is recovered but in a very trivial way: the triangle graphs precisely cancel the local anomalous term and all the amplitudes simply vanish.
- We derived generalized forms for all the triangle anomalies, including the terms proportional to the fermion masses, see Eqs. (30) and (31). This requires going beyond usual regularization procedures. To our knowledge, and except those based on dispersion relations [19], only that described by Weinberg in Ref. [17] offers sufficient freedom to choose which current in the AVV and AAA triangles is to carry the anomaly. This is crucial to deal with chiral gauge theories, in which some axial currents have to remain anomaly-free to preserve gauge invariance, as well as for treating the $\mathcal{B}$ and $\mathcal{L}$ vectorial currents [24].

In the present paper, we consider only true axion models, those designed to solve the strong CP puzzle and for which the axion ends up extremely light and extremely weakly coupled to SM fields. However, our results could have consequences for so-called Axion-Like Particle (ALP) searches (see e.g. Refs. [28–34]). Specifically, our analysis shows that an adequate effective parametrization is

$$\mathcal{L}_{ALP} = \frac{1}{2}(\partial_\mu a^0 \partial^\mu a^0 - m_a^2 a^0 a^0) + \mathcal{L}_{\text{DFSZ-like}} + \mathcal{L}_{\text{KSVZ-like}}, \tag{96}$$

with from Eq. (93),

$$\mathcal{L}_{\text{DFSZ-like}} = -i \sum_{f=u,d,e} \frac{m_f}{v_a} \chi_P^f \, a^0 \bar{\psi}_f \gamma_5 \psi_f, \tag{97}$$

for some free parameters χ_P^f, and from Eq. (25)

$$\mathcal{L}_{\text{KSVZ-like}} = \frac{a^0}{16\pi^2 v_a} \Big(g_s^2 \mathcal{N}_C^{eff} G^a_{\mu\nu} \tilde{G}^{a,\mu\nu} + g^2 \mathcal{N}_L^{eff} W^i_{\mu\nu} \tilde{W}^{i,\mu\nu} + g'^2 \mathcal{N}_Y^{eff} B_{\mu\nu} \tilde{B}^{\mu\nu} \Big), \tag{98}$$

for some free parameters $\mathcal{N}_{C,L,Y}^{eff}$. Typically, $\mathcal{L}_{\text{KSVZ-like}}$ encodes the effects of heavy vector-like fermions, while $\mathcal{L}_{\text{DFSZ-like}}$ encode the anomaly-free impact of charging the SM fermions under some global $U(1)_{PQ}$ symmetry whose current is coupled to the a^0 field. The impact of some new heavy chiral fermions (like a sequential fourth generation or a heavy neutrino) is not included, but could be by extending the sum in $\mathcal{L}_{\text{DFSZ-like}}$ to those states also. This parametrization of $\mathcal{L}_{\text{DFSZ-like}}$ ensures a^0 is truly axion-like: no $\mathcal{B}$ or $\mathcal{L}$ violating effects are implicit, only physical free parameters are introduced, and UV divergences should be under control. Still, it is important to stress that the $a^0 \to \gamma Z$, ZZ, and W^+W^- amplitudes induced by $\mathcal{L}_{\text{DFSZ-like}}$ are entirely tuned by triangle graphs, and in the $m_f \to \infty$ limit, they do not match the pattern of $\mathcal{L}_{\text{KSVZ-like}}$ but rather that in Eq. (47).

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: This manuscript has no associated data.]

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
Funded by SCOAP[3].

References

1. R.D. Peccei, H.R. Quinn, Phys. Rev. Lett. **38**, 1440 (1977)
2. R.D. Peccei, H.R. Quinn, Phys. Rev. D **16**, 1791 (1977)
3. S. Weinberg, Phys. Rev. Lett. **40**, 223 (1978)
4. F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978)
5. V. Shtabovenko, R. Mertig, F. Orellana, Comput. Phys. Commun. **207**, 432 (2016). arXiv:1601.01167 [hep-ph]
6. G. Grilli di Cortona, E. Hardy, J. Pardo Vega, G. Villadoro, JHEP **1601**, 034 (2016). arXiv:1511.02867 [hep-ph]
7. W.A. Bardeen, S.-H.H. Tye, Phys. Lett. B **74**, 229 (1978)
8. J.E. Kim, Phys. Rept. **150**, 1 (1987)
9. D.J.E. Marsh, Phys. Rept. **643**, 1 (2016). arXiv:1510.07633 [astro-ph.CO]
10. H. Georgi, D.B. Kaplan, L. Randall, Phys. Lett. B **169**, 73 (1986)
11. W.A. Bardeen, R.D. Peccei, T. Yanagida, Nucl. Phys. B **279**, 401 (1987)
12. G. Alonso-Álvarez, M.B. Gavela, P. Quilez, arXiv:1811.05466 [hep-ph]
13. J.E. Kim, Phys. Rev. Lett. **43**, 103 (1979)
14. M.A. Shifman, A.I. Vainshtein, V.I. Zakharov, Nucl. Phys. B **166**, 493 (1980)
15. S.L. Adler, W.A. Bardeen, Phys. Rev. **182**, 1517 (1969)
16. L. Ametller, L. Bergstrom, A. Bramon, E. Masso, Nucl. Phys. B **228**, 301 (1983)
17. S. Weinberg, in *The quantum theory of fields*, vol 2: Modern applications (Cambridge University Press, Cambridge, 1996)
18. A. Bilal, Lectures on Anomalies, arXiv:0802.0634 [hep-th]
19. Y. Frishman, A. Schwimmer, T. Banks, S. Yankielowicz, Nucl. Phys. B **177**, 157 (1981)
20. J.F. Gunion, H.E. Haber, G.L. Kane, S. Dawson, The Higgs Hunter's Guide. Front. Phys. **80**, 1 (2000)
21. J.F. Gunion, H.E. Haber, C. Kao, Phys. Rev. D **46**, 2907 (1992)
22. A. Djouadi, Phys. Rept. **457**, 1 (2008). arXiv:hep-ph/0503172
23. S. Kamefuchi, L. O'Raifeartaigh, A. Salam, Nucl. Phys. **28**, 529 (1961)
24. H.K. Dreiner, H.E. Haber, S.P. Martin, Phys. Rept. **494**, 1 (2010). arXiv:0812.1594 [hep-ph]
25. A. Latosinski, K.A. Meissner, H. Nicolai, Nucl. Phys. B **868**, 596 (2013). arXiv:1203.3886 [hep-ph]
26. M. Dine, W. Fischler, M. Srednicki, Phys. Lett. B **104**, 199 (1981)
27. A.R. Zhitnitsky, Sov. J. Nucl. Phys. **31**, 260 (1980) [Yad. Fiz. **31** (1980) 497]
28. A. Salvio, A. Strumia, W. Xue, JCAP **1401**, 011 (2014). arXiv:1310.6982 [hep-ph]
29. J. Jaeckel, M. Spannowsky, Phys. Lett. B **753**, 482 (2016). arXiv:1509.00476 [hep-ph]
30. M. Bauer, M. Neubert, A. Thamm, JHEP **1712**, 044 (2017). arXiv:1708.00443 [hep-ph]
31. J. Ebadi, S. Khatibi, M. Mohammadi Najafabadi, arXiv:1901.03061 [hep-ph]
32. I. Brivio, M.B. Gavela, L. Merlo, K. Mimasu, J.M. No, R. del Rey, V. Sanz, Eur. Phys. J. C **77**(8), 572 (2017). arXiv:1701.05379 [hep-ph]
33. M.B. Gavela, R. Houtz, P. Quilez, R. Del Rey, O. Sumensari, arXiv:1901.02031 [hep-ph]
34. M. Bauer, M. Heiles, M. Neubert, A. Thamm, Eur. Phys. J. C **79**(1), 74 (2019). arXiv:1808.10323 [hep-ph]