

Dirac spectrum in the chirally symmetric phase of a gauge theory. I

Matteo Giordano*

*Institute of Physics and Astronomy, ELTE Eötvös Loránd University,
Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*

(Dated: October 29, 2025)

I study the consequences of chiral symmetry restoration for the Dirac spectrum in finite-temperature gauge theories in the two-flavor chiral limit, using Ginsparg–Wilson fermions on the lattice. I prove that chiral symmetry is restored at the level of the susceptibilities of scalar and pseudoscalar bilinears if and only if all these susceptibilities do not diverge in the chiral limit $m \rightarrow 0$, with m the common mass of the light fermions. This implies in turn that they are infinitely differentiable functions of m^2 at $m = 0$, or m times such a function, depending on whether they contain an even or odd number of isosinglet bilinears. Expressing scalar and pseudoscalar susceptibilities in terms of the Dirac spectrum, I use their finiteness in the symmetric phase to derive constraints on the spectrum, and discuss the resulting implications for the fate of the anomalous $U(1)_A$ symmetry in the chiral limit. I also discuss the differentiability properties of spectral quantities with respect to m^2 , and show from first principles that the topological properties of the theory in the chiral limit are characterized by an instanton gas-like behavior if $U(1)_A$ remains effectively broken.

I. INTRODUCTION

Studies of the chiral limit of QCD, where N_f of the quark masses are taken to zero, have provided considerable insight into its real-world version describing the physics of strong interactions. In the chiral limit the theory has a classical $U(N_f)_L \times U(N_f)_R = SU(N_f)_L \times SU(N_f)_R \times U(1)_V \times U(1)_A$ continuous chiral symmetry, anomalously broken by quantum effects to $SU(N_f)_L \times SU(N_f)_R \times U(1)_V$ and, at low temperatures, spontaneously broken to its diagonal $SU(N_f)_V \times U(1)_V$ component. The chiral $SU(N_f)_L \times SU(N_f)_R$ symmetry, and the way it is realized, largely control the low-energy behavior of QCD at physical quark masses. Due to the lightness of the up and down quarks, the case $N_f = 2$ is particularly relevant from the phenomenological point of view. At zero temperature, the spontaneous breaking of $SU(2)_L \times SU(2)_R$ in the chiral limit explains the lightness of the pions and the absence of parity partners in the hadronic spectrum at the physical point, and its restoration at higher temperatures is behind the approximate chiral symmetry of QCD at temperatures above the crossover to the quark-gluon plasma [1–5].

Particularly important questions concerning the chiral limit are the nature of the $N_f = 2$ chiral transition, and the fate of the anomalous $U(1)_A$ symmetry in the chirally symmetric phase. These are not only of considerable theoretical interest, but can have phenomenological impact on important and diverse aspects of strongly interacting matter as described by QCD at the physical point, including heavy-ion collisions [6–13] and axion cosmology [14–19]. In spite of extensive theoretical and numerical studies, however, these questions remain currently open.

According to the effective-Lagrangian analysis of the seminal paper Ref. [6], for $N_f = 2$ the nature of the

transition and the fate of $U(1)_A$ are closely related, with the transition being second or first order depending on whether $U(1)_A$ remains effectively broken or is effectively restored at the transition. Here “effective breaking” and “effective restoration” mean respectively that even though $U(1)_A$ symmetry is, strictly speaking, always broken due to its anomalous nature, symmetry-breaking effects at a given nonzero temperature may be of comparable magnitude to those found at zero temperature, or strongly suppressed with respect to them. However, since Ref. [6], a number of additions to the original analysis [13, 20–23], and the use of alternative techniques such as chiral Lagrangians [24–27], the functional renormalization group [28, 29], and Schwinger–Dyson equations [30], have led to a larger set of possibilities, and so to a less tight (and less clear) relation between chiral symmetry restoration and the fate of $U(1)_A$.

Numerical calculations on the lattice, that could have in principle settled these issues, have resulted instead in contradictory claims. Using staggered fermions, the HotQCD collaboration concludes that $U(1)_A$ remains effectively broken at the transition, and that this is second order in the $O(4)$ class [31], in agreement with earlier studies [32]. This conclusion is further supported by dedicated studies of the effects of the anomaly [33–36], including with domain-wall and Möbius domain-wall fermions [4, 37–39]. On the other hand, using Möbius domain-wall fermions reweighted to overlap, the JLQCD collaboration concludes that $U(1)_A$ is restored in the symmetric phase [40], including close to the critical temperature [41], confirming earlier results using the same discretization [42, 43], as well as using improved Wilson fermions [44]. It is worth mentioning in passing that for $N_f = 3$ no sign of a first-order region has been found in the continuum limit, suggesting that the transition is second order [45, 46], a result in contrast with the analysis of Ref. [6], but compatible with other studies [28–30].

The approximate analytical methods discussed above are unfortunately affected by serious theoretical un-

* giordano@bodri.elte.hu

certainties. Studies based on the Ginzburg–Landau effective-Lagrangian approach typically ignore the role of gauge symmetries, which is still not fully understood (see Refs. [47, 48]). Both Ginzburg–Landau and chiral Lagrangians require the temperature dependence of the various coefficients in the Lagrangian as an input. Studies in the framework of the functional renormalization group or of Schwinger–Dyson equations suffer from the effects of hardly controllable approximations required to make the calculations tractable. Lattice studies have more solid foundations, as they are based on first principles and have fully controllable uncertainties, but are nonetheless affected by serious drawbacks. Dealing with chiral symmetry on the lattice is in fact notoriously difficult [49–52], and while fermion discretizations with good chiral properties (“Ginsparg–Wilson fermions”) [53–59] have been found [54, 55, 58, 60–73], they are very expensive from the numerical point of view, strongly limiting the lattice size that one can afford, and so how close one can reliably get to the thermodynamic and chiral limits.

A solution from first principles to the problem of the fate of $U(1)_A$ in the $N_f = 2$ chiral limit was proposed by Cohen in Ref. [74] (partly elaborating on the arguments of Ref. [9]). There he argued, using the formal continuum functional integral, that $U(1)_A$ must be effectively restored if $SU(2)_L \times SU(2)_R$ is. Here “effectively restored” takes a stronger meaning than in Ref. [6], and indicates that the effects of the anomaly become invisible in (at least the simplest) physical observables. This term will be understood in this sense in what follows. However, a loophole in the argument of Ref. [74] was pointed out in Refs. [75, 76], namely the incomplete treatment of the contributions of topologically nontrivial configurations to the path integral. While these configurations form a set of zero measure in the chiral limit, and were neglected in Ref. [74] for this reason, the contribution of the corresponding zero modes of the Dirac operator to the difference of correlators related by a $U(1)_A$ transformation can be non-vanishing in the chiral limit. The conclusion of Refs. [75, 76] was that $SU(2)_L \times SU(2)_R$ restoration does not necessarily imply $U(1)_A$ restoration, and that, in fact, $U(1)_A$ most likely remains effectively broken in the chiral limit also in the symmetric phase. A key assumption of Refs. [75, 76] is that in this phase the order in which one takes the thermodynamic and chiral limits should not matter, and so they can be interchanged.

A new strategy to study the relation between chiral symmetry restoration and the fate of $U(1)_A$ from first principles was proposed by Cohen in Ref. [77], and developed in full depth by Aoki, Fukaya, and Taniguchi in Ref. [78]. This strategy is to determine how chiral symmetry restoration constrains the behavior of the spectrum of the Dirac operator, and in turn what the resulting constraints imply for $U(1)_A$ in the symmetric phase. This strategy was also exploited in Refs. [79, 80], and has been recently revisited by myself in Refs. [81, 82] (see also Ref. [83]). Using Ginsparg–Wilson fermions on the lattice, as in Refs. [78, 80–82], this approach combines

analytic (and mathematically sound) methods with the first-principle and nonperturbative nature of the functional integral, to extract information from such a key object as the spectrum of the Dirac operator. In fact, the Dirac spectrum and the corresponding eigenvectors entirely encode the interactions of quarks with the gauge fields, and so in particular they should reflect the status of the various symmetries. (In this context, it is worth mentioning Ref. [84] about the relation between the spectrum and the scaling with mass and temperature of the cumulants of the chiral condensate.) While at zero and low temperatures the spontaneous breaking of chiral symmetry and the appearance of massless Goldstone bosons in the chiral limit allow one to exploit effective theories to gain considerable insight into the Dirac spectrum [85–92], such a luxury is generally not available in the symmetric phase, and one should content oneself with constraining the spectrum.

The central assumption of Ref. [78] is that in the symmetric phase the relevant observables (that can be expressed as expectation values of mass-independent functionals of gauge fields only) are analytic functions of m^2 , with m the common fermion mass, reflecting the analyticity and symmetry properties expected in the $N_f = 2$ chiral limit. Together with certain technical assumptions on the spectrum, this led the authors to conclude that in the symmetric phase the spectral density, $\rho(\lambda; m)$, as a function of the eigenvalue, λ , must vanish faster than λ^2 in the chiral limit, i.e., $\rho(\lambda; 0) = o(\lambda^2)$, and vanish at $\lambda = 0$ identically for small enough mass, i.e., $\rho(0; m) = 0$ for $|m| < m_0$; and that the topological susceptibility, χ_t , must vanish identically for small enough mass. This results in the conclusion that $U(1)_A$ must be effectively restored in the chiral limit in scalar and pseudoscalar correlation functions, partly supporting the original claim of Ref. [74]. Under assumptions similar to (but technically weaker than) those of Ref. [78], Ref. [79] proved with a simpler argument that $U(1)_A$ is effectively restored in the chiral limit in the symmetric phase at the level of the simplest order parameter, i.e., the difference $\chi_\pi - \chi_\delta$ of the usual pion and delta susceptibilities.

The predictions of Ref. [78] are supported by the numerical results of Refs. [40–44], but are in disagreement with the results of other studies [31–38]. While the chiral fermions used in Refs. [40–43] (but also in Ref. [38]) give one a better control of theoretical uncertainties, the staggered fermions used in Refs. [31–36] are computationally cheaper, and give one a better control of the finite-volume and finite-spacing systematics and of the statistical uncertainties of the numerical results. Stating that the issue of the fate of $U(1)_A$ in the symmetric phase remains unsettled seems a fair and balanced conclusion.

From the theoretical point of view, in order to avoid the conclusions of Refs. [78, 79] one needs to abandon at least one of their assumptions, the easiest choice being the technical assumptions on the spectral density. As a matter of fact, ensembles of sparse random matrices, of which the Dirac operator in the background of fluctuating gauge

fields constitutes an example, display a wide variety of properties concerning the dependence of the spectral density on λ . Technical assumptions on this dependence reflect more the experience with concrete models, mostly based on numerical simulations, than the results of rigorous theorems. This seems to leave some room for the possibility of effective $U(1)_A$ breaking in the symmetric phase. On the other hand, while the m^2 -analyticity assumption is certainly reasonable, it is arguably not more nor less reasonable *a priori* than commutativity of the thermodynamic and chiral limits, that leads to opposite conclusions concerning the fate of $U(1)_A$ [75, 76, 83], and it seems in fact quite reasonable to make both assumptions at once. Assuming m^2 -analyticity of the free energy density and commutativity of limits leads to severe restrictions on the functional form of the spectral density if $U(1)_A$ remains effectively broken [80]. These restrictions led Ref. [80] to conclude that effective $U(1)_A$ breaking in the chiral limit in the symmetric phase is possible only if the spectral density of non-zero modes develops in the thermodynamic limit a Dirac delta at $\lambda = 0$ for nonzero fermion mass, which is quite unlikely to happen on physical grounds. This seems to take away all the room left for effective $U(1)_A$ breaking, unless m^2 -analyticity or commutativity of limits fail and chiral symmetry is restored in some rather nontrivial way in the chiral limit, possibly not fully (see Refs. [80, 93–96] and Refs. [97–99] for alternative scenarios), or unless one can find a loophole in the analysis of Ref. [80].

The first question to address is then whether the m^2 -analyticity or commutativity assumptions are just reasonable assumptions, or necessary consequences of chiral symmetry restoration. A partial answer was provided in Ref. [81], where I proved that if chiral symmetry is restored in scalar and pseudoscalar susceptibilities, then these are C^∞ functions of m^2 at $m = 0$, i.e., functions of m^2 infinitely differentiable at zero (“ m^2 -differentiable”, for the sake of brevity), if they involve an even number of isosinglet scalar and pseudoscalar bilinears, and m times such a function if this number is odd. Moreover, if chiral symmetry is restored also in susceptibilities involving scalar and pseudoscalar bilinears and general (including nonlocal) functionals of the gauge fields only, m^2 -differentiability extends to the spectral density as well. This essentially implies the m^2 -analyticity assumptions of Refs. [78–80]. In fact, while different from analyticity, infinite differentiability actually suffices for (most of) their arguments, since they use only the existence of m^2 -derivatives at $m = 0$. On the other hand, although commutativity of the thermodynamic and chiral limits is supported by reasonable arguments (see footnote 5 in Ref. [75]), I am not aware of a proof that it necessarily follows from symmetry restoration, and one might have to abandon it.

The next question is whether the technical assumptions of Refs. [78, 79] on the spectral density are not too restrictive, excluding reasonable functional forms (different from a Dirac delta at the origin) that allow for effective

$U(1)_A$ breaking (which, if the conclusions of Ref. [80] are correct, would require abandoning the commutativity of limits). These assumptions are that $\rho(\lambda; m)$ is an analytic function of m^2 , and admits a power-law expansion in $|\lambda|$ sufficiently close to zero. In Ref. [78] a relaxed form of the second assumption was also considered, allowing the presence of a term $C|\lambda|^\alpha$ in the spectral density, with a mass-independent non-integer exponent $\alpha > 0$ and a mass-independent prefactor C .¹ Analyticity (or rather infinite differentiability) in m^2 is a consequence of symmetry restoration (in the extended sense), as discussed above, but the assumption of a regular behavior of the spectral density at the origin is called into question by recent (and less recent [100]) numerical results, indicating the presence of a possibly singular near-zero peak in the high-temperature phase [33–38, 40–43, 101–108].² It is possible that this peak goes away in the chiral limit without any visible effect and can be ignored. In fact, Refs. [40, 43] claim that it disappears entirely at a nonzero value of the quark mass, based on results obtained with chiral discretizations of the Dirac operator, and that its persistence observed in Refs. [33, 35] is an artefact due to the use of a mixed action, with overlap spectra computed on staggered backgrounds. On the other hand, Ref. [34] found persistent $U(1)_A$ -breaking effects in the chiral limit, originating in a near-zero peak, using exclusively staggered fermions. For physical values of the quark mass, Ref. [107] showed directly how a near-zero peak emerges in the staggered spectrum in the continuum limit, which supports the conclusion that it is a physical feature of the Dirac spectrum at nonzero quark mass. If a singular near-zero peak is indeed present, the questions are then how fast it has to disappear in the chiral limit in order to be compatible with chiral symmetry restoration, and if and how it can affect the fate of $U(1)_A$. Since a singular, mass-dependent peak was not explored in Refs. [78, 79], this behavior remained an interesting possibility to investigate.

In Ref. [81] I showed that a singular peak in the spectral density complying with chiral symmetry restoration and at the same time effectively breaking $U(1)_A$ is indeed technically possible. I also showed that in this case effective $U(1)_A$ breaking requires further peculiar features of the spectrum, including a close relation between the peak modes and topology, and the delocalization of these modes over the whole system; and that the first nontrivial cumulant of the topological charge is the same as in an ideal gas of instantons and anti-instantons, to leading order in m . This was shown to hold for all cumulants in Ref. [109] using an effective approach based on considerations of analyticity and symmetry. Of course, the

¹ In Ref. [78] these assumptions are actually made on the spectral density computed on individual gauge configurations of infinite size, assumed to be a well-defined ordinary function.

² In the chirally broken phase at low temperature, partially quenched chiral perturbation theory predicts a logarithmic divergence in ρ at $m \neq 0$ [88–90].

theoretical possibility of a singular peak with just the right features may simply be an unlikely edge case. It is then reassuring for its physical viability that an explicit mechanism leading to the right kind of behavior is provided by a very simple instanton-based random matrix model [110].

The purpose of this series of papers is to systematize the approach to the problem of chiral symmetry restoration and the fate of $U(1)_A$ based on the study of the Dirac spectrum [77–80], expanding the analysis of Ref. [81]. This requires first of all to work on the foundations of the approach, looking for a characterization of the chirally symmetric phase based on first principles rather than on plausible but unproven assumptions. This allows one to better disentangle conclusions that are fully justified by the nature of the symmetric phase from the consequences of more technical (and less controllable) assumptions. The conditions resulting from the request of chiral symmetry restoration are then translated into constraints on the Dirac spectrum, of very general nature. Characterizing the symmetric phase and deriving these constraints is the scope of the present paper. The main arguments, already outlined in Ref. [81], are discussed here in greater detail. The next step of the program is to work on the technical assumptions on the Dirac spectrum, extending and refining the analysis of Ref. [81], and generalizing some of its results. This includes carefully scrutinizing the conclusions of Ref. [80] concerning the consequences of commutativity of the thermodynamic and chiral limits. This will be discussed in a separate paper.

As in Ref. [81], I deal with the $N_f = 2$ chiral limit of general gauge theories on the lattice using Ginsparg–Wilson fermions. No particular restriction is made on the theory, besides assuming that it has a symmetric phase where $SU(2)_L \times SU(2)_R$ is fully realized. Results are obtained on the lattice, but there is no obstacle in extending them to the continuum, assuming that a continuum limit exists (which generally requires further restrictions on the theory). I work in the sector of the theory generated by scalar and pseudoscalar fermion bilinears, both flavor-singlet and flavor-triplet, referred to as the “scalar and pseudoscalar sector” for brevity, where susceptibilities can be expressed entirely in terms of Dirac eigenvalues only. I briefly summarize here the main results, of general nature, obtained in this paper.

Starting from the basic properties expected of the symmetric phase of a quantum field theory, I prove that chiral symmetry restoration at the level of scalar and pseudoscalar susceptibilities is equivalent to finiteness of these quantities in the chiral limit. In this context, by finite quantity I always mean a quantity that is not divergent, but possibly vanishing, in the chiral limit. This result implies in turn that in the symmetric phase the scalar and pseudoscalar susceptibilities must be m^2 -differentiable functions, or m times an m^2 -differentiable function. The discussion of the basic assumptions is more detailed than in Ref. [81], and the proof is simplified. As already pointed out above, this result essentially proves

that the analyticity assumptions of Refs. [77–80] (and of Ref. [109], see below) on susceptibilities and on the free energy density are necessary conditions for symmetry restoration, since the distinction between analytic and infinitely differentiable functions is of very limited practical relevance in this context.

These symmetry-restoration conditions are the starting point for the derivation of constraints on the Dirac spectrum, using an explicit expression for the generating function of scalar and pseudoscalar susceptibilities in terms of Dirac eigenvalues, reported in Ref. [81] and derived in detail here. This expression shows in particular how exact zero modes generally cannot be ignored even in the chiral limit, further supporting the criticism of this assumption of Ref. [74] made in Refs. [75, 76].

The approach of this paper is more general than, and subsumes that of Ref. [78], providing constraints not only on the spectral density and the topological susceptibility, but on eigenvalue correlations as well. The constraints derived here do not require any detailed assumption on the spectral density and other spectral quantities. At this stage, full restoration of $SU(2)_L \times SU(2)_R$ is compatible both with the effective breaking and with the effective restoration of $U(1)_A$. To obtain any insight into this issue, a more in-depth study is required, involving a detailed analysis of the properties of eigenvalue correlators, and requiring additional technical assumptions. This is discussed in the next paper of this series.

Under additional assumptions on how chiral symmetry restoration manifests in susceptibilities involving also nonlocal operators built entirely out of gauge fields, or in the presence of external fermion fields, I show that also the spectral density and other spectral quantities are m^2 -differentiable. This essentially justifies the m^2 -analyticity assumption on the spectral density made in Refs. [78, 79] starting from more fundamental assumptions on how chiral symmetry is realized, and provides further restrictions that can be exploited in constraining the behavior of the Dirac spectrum.

Finally, as anticipated in Ref. [81], using the formalism developed here I rederive the conclusion of Ref. [109] that in the chiral limit in the symmetric phase the cumulants of the topological charge are identical to those found in an ideal gas of (anti)instanton-like objects, to leading order in the fermion mass, if $U(1)_A$ is effectively broken. This result is put here on first-principles ground, justifying the assumptions of Ref. [109], and allowing for the study of corrections to the ideal-gas behavior.

The plan of this paper is the following. In Sec. II I summarize the relevant aspects of finite-temperature gauge theories on the lattice and of Ginsparg–Wilson fermions, including the spectral properties of the corresponding discretized Dirac operator. In Sec. III I discuss chiral symmetry restoration and its consequences, and derive a set of necessary and sufficient conditions for chiral symmetry restoration at the level of scalar and pseudoscalar susceptibilities, and their extension to susceptibilities involving also functionals of gauge fields. In Sec. IV I derive an

explicit expression for the generating function of scalar and pseudoscalar susceptibilities in terms of the eigenvalues of a Ginsparg–Wilson Dirac operator. In Sec. V I use the conditions of Sec. III and the results of Sec. IV to obtain constraints on the Dirac spectrum in the symmetric phase. In Sec. VI I show that in the chiral limit in the symmetric phase the topological charge behaves as in an ideal instanton gas if $U(1)_A$ remains effectively broken. In Sec. VII I draw my conclusions. Technical details are discussed in Appendices A, B, C, D, E, F, and G.

II. FINITE-TEMPERATURE LATTICE GAUGE THEORIES WITH GINSPARG–WILSON FERMIONS

In this series of papers I consider 3+1-dimensional finite-temperature lattice gauge theories based on some compact gauge group, with two flavors of dynamical light (eventually massless) fermions of mass m transforming in some irreducible N_c -dimensional representation of the gauge group, and any number of additional fermions, transforming in possibly different irreducible representations of the gauge group, that remain massive as $m \rightarrow 0$.

It is assumed that the gauge group, fermion content, and gauge group representations are such that a phase of the theory exists where in the limit $m \rightarrow 0$ the $SU(2)_L \times SU(2)_R$ chiral symmetry of the theory (see below Sec. IIA) is fully realized. Further restrictions are generally needed to ensure the existence of a continuum limit, but this does not affect the validity of the results as long as one works at finite lattice spacing.

The theory is discretized on a hypercubic lattice of linear spatial size L and temporal size $1/T$. Here and below I use lattice units. The spatial and spacetime lattice volumes are denoted by $V_3 = L^3$ and $V_4 = V_3/T$, respectively. The thermodynamic limit $V_4 \rightarrow \infty$ is taken by sending $V_3 \rightarrow \infty$ while keeping $1/T$ fixed.³ It is understood that when taking the chiral limit $m \rightarrow 0$, this is done after taking the thermodynamic limit, unless specified otherwise.

Gauge link variables taking values in the gauge group, denoted collectively by U , are associated with the lattice edges. Periodic boundary conditions both in time and in space are imposed on them by including additional edges that make the hypercubic lattice into a four-dimensional torus. The detailed form of the discretized gauge action does not play any role in this work; it is only assumed that it is invariant under the usual lattice spacetime symmetries (lattice translations, rotations, and reflections).

Two sets of Grassmann variables Ψ_{xacf} and $\bar{\Psi}_{xacf}$, representing the two light fermions, are associated with the lattice sites. These variables carry spacetime indices,

including the lattice coordinates x and a discrete Dirac index $\alpha = 1, \dots, 4$, a color (i.e., gauge-group representation) index, $c = 1, \dots, N_c$, and a flavor index, $f = 1, 2$, that will all be suppressed in the following. Boundary conditions periodic in space, and antiperiodic in time, are imposed on the fermionic variables.

Light fermions are coupled to the gauge links via a discretized massless Dirac operator D satisfying the Ginsparg–Wilson (GW) relation $\{D, \gamma_5\} = 2DR\gamma_5D$, where γ_5 is the usual Dirac matrix and R is a local operator that commutes with γ_5 [53–56, 58, 59].⁴ D carries only spacetime and color indices. It is again assumed that invariance under the usual lattice spacetime symmetries holds. Examples of operators obeying the GW relation are the domain wall operator [60–66], the overlap operator [58, 67–69], and the fixed-point action [54, 55, 70–73]. Massless Dirac operators obeying the GW relation possess an exact lattice chiral symmetry [54, 56–59, 69], that reduces to the usual chiral symmetry of continuum fermions in the continuum limit. The associated massive operator D_m is [111, 112]

$$D_m = D + m(\mathbf{1} - DR), \quad (1)$$

where $\mathbf{1}$ is the identity in spacetime (including Dirac) and color space. The mass is coupled here to the simplest proper order parameter for chiral symmetry breaking [55, 111, 112].

Massive fermions are similarly represented by pairs of sets of Grassmann variables, and are coupled to gauge fields via discretizations of the Dirac operator, possibly not of GW type. Again, their detailed form plays no role; invariance under the usual lattice spacetime symmetries is assumed, and the usual boundary conditions for fermions are imposed.

After integrating out the massive fermion fields, the partition function reads

$$\begin{aligned} Z &= \int DU \int D\Psi D\bar{\Psi} e^{-S_{\text{eff}}(U)} e^{-\bar{\Psi} D_m(U) \mathbf{1}_f \Psi} \\ &= \int DU e^{-S_{\text{eff}}(U)} [\det D_m(U)]^2, \end{aligned} \quad (2)$$

where S_{eff} includes the gauge action and the contribution of massive fermion fields, $\mathbf{1}_f$ is the identity in flavor space, DU denotes the product of the Haar measures associated with the link variables, and $D\Psi D\bar{\Psi}$ the product of the Berezin measures associated with the fermion fields. The measure is invariant under gauge transformations and lattice symmetry transformations. Expectation values of observables depending only on gauge and light-

³ The formalism remains unchanged at $T = 0$, the only difference being that the thermodynamic limit is taken by setting the temporal extension to L and sending $V_4 = L^4 \rightarrow \infty$.

⁴ The general form of the GW relation is $\{D, \gamma_5\} = D\{R', \gamma_5\}D$, with R' local [53, 59]. Without loss of generality one can replace $R' \rightarrow R = \frac{1}{2}(R' + \gamma_5 R' \gamma_5)$, with $[R, \gamma_5] = 0$ and $\{R', \gamma_5\} = 2R\gamma_5$.

fermion fields read

$$\langle \mathcal{O} \rangle = Z^{-1} \int DU \int D\Psi D\bar{\Psi} e^{-S_{\text{eff}}(U)} e^{-\bar{\Psi} D_m(U) \mathbf{1}_f \Psi} \times \mathcal{O}(\Psi, \bar{\Psi}, U). \quad (3)$$

These expectation values are understood to be evaluated in a finite spatial volume V_3 ; their dependence on T and V_3 is left implicit for notational simplicity.

Finally, I assume that the theory is invariant under a discrete “ CP transformation” of the form

$$U \rightarrow U_{CP}, \quad \Psi \rightarrow \mathcal{U}_{CP} \mathbf{1}_f \Psi, \quad \bar{\Psi} \rightarrow \bar{\Psi} \mathbf{1}_f \mathcal{U}_{CP}^\dagger, \quad (4)$$

where $\mathcal{U}_{CP}$ is a suitable unitary matrix (with spacetime and color indices only), and

$$S_{\text{eff}}(U_{CP}) = S_{\text{eff}}(U), \quad \mathcal{U}_{CP}^\dagger D(U_{CP}) \mathcal{U}_{CP} = D(U), \quad \mathcal{U}_{CP}^\dagger \gamma_5 \mathcal{U}_{CP} = -\gamma_5. \quad (5)$$

The integration measure is assumed to be invariant under the transformation Eq. (4). Calling this a CP transformation is a bit of a misnomer, although common in the literature: in fact, viable choices satisfying these requirements are the temporal reflection, reflections through a plane perpendicular to one of the spatial directions, or the spatial parity transformation. CP invariance is then guaranteed for all the most common discretizations of gauge and fermion actions. Notice, however, that the discussion in Secs. III and IV makes no use of CP invariance, and the results obtained there hold also if one includes a CP -violating topological term in the action (except of course for results where the use of CP is explicitly mentioned).

A. Chiral symmetry of the GW Dirac operator

Thanks to the chiral properties of GW Dirac operators, the system under consideration has at the classical level an exact $U(2)_L \times U(2)_R$ symmetry in the chiral limit [56, 57, 59]. Let $\vec{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ denote the usual Pauli matrices acting in flavor space, and let $\hat{\gamma}_5 \equiv (\mathbf{1} - 2DR)\gamma_5$. The GW relation can be recast as $D\gamma_5 + \hat{\gamma}_5 D = 0$, and implies $\hat{\gamma}_5^2 = \mathbf{1}$. Flavor nonsinglet, $SU(2)_L \times SU(2)_R$ chiral transformations, denoted by $\mathcal{U}(\vec{\alpha}_L, \vec{\alpha}_R)$, are defined by

$$\begin{aligned} \Psi &\rightarrow \Psi_{\mathcal{U}} = \left(U(\vec{\alpha}_L) \frac{1-\gamma_5}{2} + U(\vec{\alpha}_R) \frac{1+\gamma_5}{2} \right) \Psi, \\ \bar{\Psi} &\rightarrow \bar{\Psi}_{\mathcal{U}} = \bar{\Psi} \left(U(\vec{\alpha}_L)^\dagger \frac{1+\hat{\gamma}_5}{2} + U(\vec{\alpha}_R)^\dagger \frac{1-\hat{\gamma}_5}{2} \right), \end{aligned} \quad (6)$$

where $\vec{\alpha}_{L,R} \in \mathbb{R}^3$ and $U(\vec{\alpha}) \equiv e^{i\vec{\alpha} \cdot \frac{\vec{\sigma}}{2}} \in SU(2)$. Flavor-singlet, $U(1)_L \times U(1)_R$ chiral transformations, denoted by $\mathcal{U}^{(0)}(\alpha_L, \alpha_R)$, are similarly defined by

$$\begin{aligned} \Psi &\rightarrow \Psi_{\mathcal{U}^{(0)}} = \left(e^{i\alpha_L} \frac{1-\gamma_5}{2} + e^{i\alpha_R} \frac{1+\gamma_5}{2} \right) \Psi, \\ \bar{\Psi} &\rightarrow \bar{\Psi}_{\mathcal{U}^{(0)}} = \bar{\Psi} \left(e^{-i\alpha_L} \frac{1+\hat{\gamma}_5}{2} + e^{-i\alpha_R} \frac{1-\hat{\gamma}_5}{2} \right). \end{aligned} \quad (7)$$

Nonsinglet vector and axial transformations, $SU(2)_V$ and $SU(2)_A$, are defined respectively as $\mathcal{U}_V(\vec{\alpha}) \equiv \mathcal{U}(\vec{\alpha}, \vec{\alpha})$ and $\mathcal{U}_A(\vec{\alpha}) \equiv \mathcal{U}(-\vec{\alpha}, \vec{\alpha})$, and singlet vector and axial transformations, $U(1)_V$ and $U(1)_A$, are defined respectively as $\mathcal{U}_V^{(0)}(\alpha) \equiv \mathcal{U}^{(0)}(\alpha, \alpha)$ and $\mathcal{U}_A^{(0)}(\alpha) \equiv \mathcal{U}^{(0)}(-\alpha, \alpha)$, with $U(1)_L \times U(1)_R = U(1)_V \times U(1)_A$.

It is straightforward to show that $\bar{\Psi} D \mathbf{1}_f \Psi$ is invariant under the chiral $U(2)_L \times U(2)_R$ transformations Eqs. (6) and (7). The Berezin integration measure is invariant under $SU(2)_L \times SU(2)_R$ and $U(1)_V$ transformations, but not under $U(1)_A$ transformations, under which [56]

$$D\Psi D\bar{\Psi} \rightarrow D\Psi D\bar{\Psi} e^{-i4\alpha Q}, \quad (8)$$

where Q is the topological charge [see under Eq. (25)].⁵ This makes $U(1)_A$ an anomalous symmetry already on the lattice [53, 54, 56]. The full chiral symmetry of the classical lattice action for massless fermions is then broken by quantum effects to $SU(2)_L \times SU(2)_R \times U(1)_V$; a nonzero fermion mass m breaks this explicitly further, down to $SU(2)_V \times U(1)_V$. In the chiral limit $m \rightarrow 0$, $SU(2)_L \times SU(2)_R$ may not be recovered, but instead break down spontaneously to $SU(2)_V$: this is the case, e.g., for the two-flavor chiral limit of QCD and QCD-like theories based on the gauge group $SU(N_c)$ at zero and low temperatures. At higher temperatures, above a symmetry-restoring phase transition, chiral symmetry is instead fully realized (see, e.g., Refs. [31, 43]). The results of this work apply to gauge theories in such an $SU(2)_L \times SU(2)_R$ -symmetric phase.

Since the two-flavor chiral limit is taken along the line of mass-degenerate light fermions, $m_1 = m_2 = m$, $SU(2)_V$ symmetry is exactly realized for any m , although in principle the system may not be in a pure phase in the thermodynamic limit. However, this is guaranteed by the fact that $SU(2)_V$ cannot break down spontaneously if the integration measure is positive-definite: even if one breaks it explicitly by setting $m_1 \neq m_2$, one always recovers it when $m_{1,2} \rightarrow m$ (at least for nonzero m). The impossibility of spontaneous breaking of $SU(N_f)_V$ was shown in the continuum in Ref. [113], and on the lattice for various discretizations including GW fermions in Ref. [114] (see also Refs. [115, 116]).

B. Scalar and pseudoscalar bilinears

The quantities of interest in this work are the connected correlation functions of the following fermion bilinears,

$$\begin{aligned} S &\equiv \bar{\Psi}(\mathbf{1} - DR) \mathbf{1}_f \Psi, & P &\equiv \bar{\Psi}(\mathbf{1} - DR) \mathbf{1}_f \gamma_5 \Psi, \\ \vec{P} &\equiv \bar{\Psi}(\mathbf{1} - DR) \vec{\sigma} \gamma_5 \Psi, & \vec{S} &\equiv \bar{\Psi}(\mathbf{1} - DR) \vec{\sigma} \Psi, \end{aligned} \quad (9)$$

⁵ The proof of Eq. (8) in Ref. [56] is for the case $2R = 1$, but can be extended to general R without difficulty.

i.e., the spacetime integral of scalar and pseudoscalar isosinglet and isotriplet densities, including a suitable subtraction term [73, 111, 112]. Under infinitesimal chiral transformations of the fermion fields, the following four-component vectors of fermionic bilinears,

$$O_V \equiv \begin{pmatrix} S \\ i\vec{P} \end{pmatrix}, \quad O_W \equiv \begin{pmatrix} iP \\ -\vec{S} \end{pmatrix}, \quad (10)$$

transform by an infinitesimal $SO(4)$ rotation [73]. Using the exponential map, under a finite non-singlet chiral transformation $\mathcal{U}$, Eq. (6), one finds then

$$O_{V,W} \rightarrow (O_{V,W})_{\mathcal{U}} = \mathcal{R}(\mathcal{U}) O_{V,W}, \quad (11)$$

for some $\mathcal{R}(\mathcal{U}) \in SO(4)$, where $(O_{V,W})_{\mathcal{U}}$ denotes the chirally transformed scalar and pseudoscalar bilinears, i.e., the bilinears defined in Eq. (9) with Ψ and $\bar{\Psi}$ replaced by $\Psi_{\mathcal{U}}$ and $\bar{\Psi}_{\mathcal{U}}$. Since the exponential map is surjective for compact connected groups (see Ref. [117], ch. IV, theorem 2.2), the mapping $\mathcal{U} \rightarrow \mathcal{R}(\mathcal{U})$ provides a representation of $SU(2)_L \times SU(2)_R$ that is onto $SO(4)$.

Concerning flavor-singlet transformations, Eq. (7), the bilinears in Eq. (9) are manifestly invariant under vector transformations, while O_V and O_W get mixed under axial transformations: under the transformation $\mathcal{U}_A^{(0)}(\frac{\alpha}{2}) \in U(1)_A$ one finds

$$\begin{pmatrix} O_V \\ O_W \end{pmatrix} \rightarrow \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix} \begin{pmatrix} O_V \\ O_W \end{pmatrix}. \quad (12)$$

Finally, under the CP transformation, Eqs. (4) and (5), scalar and pseudoscalar bilinears transform as

$$S \rightarrow S, \quad \vec{P} \rightarrow -\vec{P}, \quad P \rightarrow -P, \quad \vec{S} \rightarrow \vec{S}. \quad (13)$$

C. Generating function

The correlation functions of scalar and pseudoscalar bilinears are conveniently handled through generating functions. Let

$$K(\Psi, \bar{\Psi}, U; V, W) \equiv j_S S + i\vec{j}_P \cdot \vec{P} + ij_P P - \vec{j}_S \cdot \vec{S} \quad (14) \\ = V \cdot O_V + W \cdot O_W,$$

with

$$V \equiv \begin{pmatrix} j_S \\ \vec{j}_P \end{pmatrix}, \quad W \equiv \begin{pmatrix} j_P \\ \vec{j}_S \end{pmatrix}, \quad (15)$$

where j_S and j_P are isosinglet scalar and pseudoscalar external sources, and similarly j_{Sa} and j_{Pa} , $a = 1, 2, 3$, are isotriplet scalar and pseudoscalar external sources, collected in the vectors $\vec{j}_S$ and $\vec{j}_P$. One defines the generating functions $\mathcal{Z}$ and $\mathcal{W}$ of full and connected correlation functions, respectively, as

$$\mathcal{Z}(V, W; m) \equiv \int DU \int D\Psi D\bar{\Psi} e^{-S_{\text{eff}}(U)} \\ \times e^{-\bar{\Psi} D_m(U) \mathbf{1}_f \Psi - K(\Psi, \bar{\Psi}, U; V, W)} \quad (16) \\ \equiv \exp \{V_4 \mathcal{W}(V, W; m)\}.$$

For notational simplicity, the dependence of $\mathcal{Z}$ and $\mathcal{W}$ on V_3 and T is omitted. Up to constant factors, the derivatives of $\mathcal{Z}$ with respect to the sources evaluated at vanishing sources are the full correlation functions of the scalar and pseudoscalar bilinears defined in Eq. (9) [see Eqs. (2) and (3)],

$$\left\langle S^{n_S} (iP)^{n_P} (i\vec{P})^{\vec{n}_P} \vec{S}^{\vec{n}_S} \right\rangle (-1)^{n_S + n_P + \sum_{a=1}^3 n_{Pa}} \\ = Z^{-1} (\partial_{j_S})^{n_S} (\partial_{j_P})^{n_P} \prod_{a=1}^3 [(\partial_{j_{Pa}})^{n_{Pa}} (\partial_{j_{Sa}})^{n_{Sa}}] \quad (17) \\ \times \mathcal{Z}(V, W; m)|_0,$$

where $\partial_x \equiv \partial/\partial x$, $|_0$ denotes setting $V = W = 0$, and I have used the shorthand $\vec{X}^{\vec{n}_X} \equiv \prod_{a=1}^3 X_a^{n_{Xa}}$. Similarly,

$$\frac{1}{V_4} \left\langle S^{n_S} (iP)^{n_P} (i\vec{P})^{\vec{n}_P} \vec{S}^{\vec{n}_S} \right\rangle_c (-1)^{n_S + n_P + \sum_{a=1}^3 n_{Pa}} \\ = (\partial_{j_S})^{n_S} (\partial_{j_P})^{n_P} \prod_{a=1}^3 [(\partial_{j_{Pa}})^{n_{Pa}} (\partial_{j_{Sa}})^{n_{Sa}}] \quad (18) \\ \times \mathcal{W}(V, W; m)|_0,$$

where $\langle \dots \rangle_c$ denotes connected correlation functions, defined recursively in the usual way (see Appendix A1). Up to the same constant factors as above, in the thermodynamic limit the derivatives of $\mathcal{W}$ at zero sources yield the scalar and pseudoscalar susceptibilities, $\chi \left(S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \right)$, where

$$\chi \left(\prod_i O_i \right) \equiv \lim_{V_4 \rightarrow \infty} \frac{1}{V_4} \langle \prod_i O_i \rangle_c. \quad (19)$$

For practical purposes, it is convenient to treat $\mathcal{W}$ as a formal power series in the sources, with the normalized connected correlators of scalar and pseudoscalar bilinears in Eq. (18) as coefficients, that can be truncated at any sufficiently high but finite order n if one is interested only in correlators involving no more than n bilinears. This guarantees the possibility to exchange taking derivatives with respect to the sources at zero sources with any other operation, in particular taking derivatives with respect to m , and taking the thermodynamic or chiral limit. Moreover, exchanging the thermodynamic limit with derivatives with respect to m is expected to be allowed at any nonzero m , where the finite correlation length of the system guarantees that the thermodynamic limit is uniform in m in any range of nonzero masses.⁶

⁶ Notice that in numerical simulations of lattice theories, the derivative of the thermodynamic limit of a quantity with respect to some parameter (such as the mass or the θ angle) is by practical necessity *defined* as the thermodynamic limit of the derivative. Lacking a rigorous proof of the properties of the relevant quantities in the thermodynamic limit, the other order of operations is currently unattainable, and one is forced to assume that derivatives and thermodynamic limit commute at $m \neq 0$ to make any progress.

With this in mind, it is convenient to denote with $\mathcal{W}_\infty = \lim_{V_4 \rightarrow \infty} \mathcal{W}$ the formal power series collecting the thermodynamic limit of the relevant correlators, i.e., the generating function of the susceptibilities, and write

$$\begin{aligned} & \chi \left(S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \right) (-1)^{n_S + n_P + \sum_{a=1}^3 n_{Pa}} \\ &= (\partial_{j_S})^{n_S} (\partial_{j_P})^{n_P} \prod_{a=1}^3 [(\partial_{j_{Pa}})^{n_{Pa}} (\partial_{j_{Sa}})^{n_{Sa}}] \\ & \times \mathcal{W}_\infty(V, W; m)|_0. \end{aligned} \quad (20)$$

It follows from the discussion in Sec. IIB that for a generic nonsinglet chiral transformation, Eq. (6), one has

$$K(\Psi_U, \bar{\Psi}_U, U; V, W) = K(\Psi, \bar{\Psi}, U; \mathcal{R}(\mathcal{U})^T V, \mathcal{R}(\mathcal{U})^T W), \quad (21)$$

with $\mathcal{R}(\mathcal{U}) \in \text{SO}(4)$ the rotation matrix associated with the chiral transformation, see Eq. (11). The generating function of the susceptibilities of the chirally transformed scalar and pseudoscalar bilinears, $(O_{V,W})_U$, is then simply $\mathcal{W}_\infty(\mathcal{R}(\mathcal{U})^T V, \mathcal{R}(\mathcal{U})^T W; m)$.

The transformation properties under CP , Eq. (13), imply also that

$$K(\mathcal{U}_{CP}\Psi, \bar{\Psi}\mathcal{U}_{CP}^\dagger, U_{CP}; V, W) = K(\Psi, \bar{\Psi}, U; CV, -CW), \quad (22)$$

where $\mathcal{C} \equiv \text{diag}(1, -1, -1, -1)$. Together with the invariance of the integration measure this implies that

$$\mathcal{Z}(CV, -CW; m) = \mathcal{Z}(V, W; m), \quad (23)$$

and similarly for $\mathcal{W}$ and $\mathcal{W}_\infty$. In the case of γ_5 -Hermitean GW operators with $2R = 1$ (see below Sec. IID), Eq. (23) implies that $\mathcal{Z}$ is real, see Appendix B, and therefore the derivatives of $\mathcal{W}_\infty$ at zero sources, i.e., the susceptibilities in Eq. (20), are real. (Since $\mathcal{Z}|_0 = \mathcal{Z}$ is positive at zero sources, the free energy density $-\mathcal{W}_\infty|_0$ is real as well.)

D. Spectrum of the GW Dirac operator

The main purpose of this work is to investigate the consequences of chiral symmetry restoration for the Dirac spectrum, more precisely for the eigenvalues of the massless GW operator,

$$D(U)\psi_n(U) = \mu_n(U)\psi_n(U). \quad (24)$$

The eigenvalues are generally complex, $\mu_n(U) \in \mathbb{C}$. The eigenvectors $\psi_n(U)$ carry spacetime and color indices, and the index n ranges over $N_{\text{tot}} = 4N_c V_4$ values. Both eigenvalues and eigenvectors depend on the gauge configuration; this dependence will be often omitted for notational simplicity.

Common realizations of GW fermions, such as domain wall [60–66] or overlap fermions [58, 67–69], have the additional properties that $2R = 1$ and that D is γ_5 -Hermitean, $\gamma_5 D \gamma_5 = D^\dagger$. In this case $1 - D$ is unitary, so

D and $D^\dagger$ have a common basis of orthonormal eigenvectors, with $D\psi_n = \mu_n\psi_n$, $\mu_n = 1 - e^{-i\varphi_n}$, $\varphi_n \in (-\pi, \pi]$, and $D^\dagger\psi_n = \mu_n^*\psi_n$, which implies that also $\gamma_5\psi_n$ is an eigenvector of D with eigenvalue μ_n^* . Complex modes, $\mu_n \neq \mu_n^*$, come then in conjugate pairs, with ψ_n and $\gamma_5\psi_n$ the corresponding orthogonal eigenvectors. If $\mu_n = \mu_n^*$ is real, it must be either $\mu_n = 0$ ($\varphi_n = 0$), or $\mu_n = 2$ ($\varphi_n = \pi$). These modes can be, and usually are, chosen to be chiral,

$$D\psi_n^{(r)} = r\psi_n^{(r)}, \quad \gamma_5\psi_n^{(r)} = \xi_n^{(r)}\psi_n^{(r)}, \quad (25)$$

where $r = 0, 2$ and $\xi_n^{(0,2)} = \pm 1$. I denote the number of zero modes with chirality ± 1 in configuration U by $N_\pm(U)$; the total number of zero modes by $N_0(U) \equiv N_+(U) + N_-(U)$; and the topological charge by $Q(U) \equiv N_+(U) - N_-(U)$. The identification of the topological charge with the index of D is justified by the lattice index theorem of Ref. [54]. One similarly has $N'_\pm(U)$ “doubler” modes with eigenvalue 2 and chirality ± 1 , and $N_2(U) \equiv N'_+(U) + N'_-(U)$ doubler modes in total, with $Q = N_+ - N_- = N'_+ - N'_-$ (since γ_5 is traceless).

The transformation property Eq. (5) implies that $\mathcal{U}_{CP}\psi_n(U)$ are eigenvectors of $D(U_{CP})$ with eigenvalue $\mu_n(U)$. The gauge action and the Dirac spectrum are then the same for the gauge configurations U and U_{CP} , that have therefore the same weight in the partition function. On the other hand, for the real chiral modes $\psi_n^{(r)}(U)$ one finds that $\mathcal{U}_{CP}\psi_n^{(r)}(U)$ has chirality opposite to that of $\psi_n^{(r)}(U)$, and so $Q(U_{CP}) = -Q(U)$, implying in particular that $\langle Q^{2k+1} \rangle = 0$ for any nonnegative integer k (this of course is not true anymore if a topological term is added to the action).

The density of complex modes and their correlation functions are conveniently expressed in terms of $\lambda_n \equiv 2 \sin \frac{\varphi_n}{2}$, with $\lambda_n^2 = |\mu_n|^2$. Denoting with $N(U) \equiv \frac{1}{2}[N_{\text{tot}} - N_0(U) - N_2(U)]$ the number of pairs of complex modes, it is convenient to label as μ_n , $n = 1, \dots, N$, the modes with $\text{Im} \mu_n > 0$ [$\varphi_n \in (0, \pi)$], and with $\mu_{-n} \equiv \mu_n^*$ their complex conjugates, so $\text{Im} \mu_{-n} < 0$, $\varphi_{-n} = -\varphi_n \in (-\pi, 0)$, and $\lambda_{-n} = -\lambda_n$. I define the spectral density in a given gauge configuration as

$$\rho_U(\lambda) \equiv \sum_n \delta(\lambda - \lambda_n(U)), \quad (26)$$

where the sum runs over $n = \pm 1, \dots, \pm N(U)$. This is a distribution supported in $(-2, 0) \cup (0, 2)$, symmetric about the origin, and normalized as

$$\int_{-2}^2 d\lambda \rho_U(\lambda) = 2 \int_0^2 d\lambda \rho_U(\lambda) = 2N(U). \quad (27)$$

(This differs by a factor V_4 from the normalization used in Ref. [81].) While ρ_U is a highly singular object, one expects that the (normalized) spectral density,

$$\rho_c^{(1)}(\lambda; m) \equiv \frac{1}{V_4} \langle \rho_U(\lambda) \rangle, \quad (28)$$

obtained after averaging over gauge configurations, is an ordinary function (although it may still develop distributional contributions in the thermodynamic limit).⁷ In analogy with ρ_U one defines also higher-point eigenvalue correlation functions in a fixed configuration, removing contact terms for coinciding arguments,

$$\begin{aligned} \rho_U^{(k)}(\lambda_1, \dots, \lambda_k) \\ \equiv \sum_{\substack{n_1, \dots, n_k \\ n_i \neq \pm n_j, \forall i \neq j}} \delta(\lambda_1 - \lambda_{n_1}(U)) \dots \delta(\lambda_k - \lambda_{n_k}(U)), \end{aligned} \quad (29)$$

where of course $\rho_U^{(1)} = \rho_U$. By construction $\rho_U^{(k)}$ is symmetric under permutations of its arguments; and is symmetric under reflection of any of its arguments, $\lambda_i \rightarrow -\lambda_i$, thanks to the symmetry of the spectrum. One then defines suitably normalized connected k -point eigenvalue correlators,

$$\rho_c^{(k)}(\lambda_1, \dots, \lambda_k; m) \equiv \frac{1}{V_4} \langle \rho_U^{(k)}(\lambda_1, \dots, \lambda_k) \rangle_c, \quad (30)$$

according to the usual recursive procedure, see Appendix A 1. The quantities $\rho_c^{(k)}$ are obviously symmetric under permutations and reflections of their arguments. One similarly defines normalized connected correlation functions involving the complex Dirac modes and other quantities O (see Appendix A 1),

$$\rho_{O_c}^{(k)}(\lambda_1, \dots, \lambda_k; m) \equiv \frac{1}{V_4} \langle O \rho_U^{(k)}(\lambda_1, \dots, \lambda_k) \rangle_c. \quad (31)$$

For $O = 1$ one gets back $\rho_{1_c}^{(k)} = \rho_c^{(k)}$; in this case the subscript O will be omitted. For $k = 0$ one gets the cumulants

$$b_O \equiv \rho_{O_c}^{(0)} = \frac{\langle O \rangle_c}{V_4}. \quad (32)$$

Examples relevant in the following are the correlation functions involving both zero and complex modes, i.e., $O = N_+^{k+} N_-^{k-}$, or $O = N_0^{k_0} Q^{k_1}$. Since N_+ , N_- and N_0, Q are linearly related, their connected correlation functions are related by the same linear transformation that relates their full correlation functions (see Appendix A 3). Relevant examples for $k = 1$ are

$$\begin{aligned} \rho_{N_0 c}^{(1)}(\lambda; m) &= \frac{\langle N_0 \rho_U(\lambda) \rangle_c}{V_4} = \frac{\langle N_0 \rho_U(\lambda) \rangle - \langle N_0 \rangle \langle \rho_U(\lambda) \rangle}{V_4}, \\ \rho_{Q^2 c}^{(1)}(\lambda; m) &= \frac{\langle Q^2 \rho_U(\lambda) \rangle_c}{V_4} = \frac{\langle Q^2 \rho_U(\lambda) \rangle - \langle Q^2 \rangle \langle \rho_U(\lambda) \rangle}{V_4}, \end{aligned} \quad (33)$$

⁷ The spectral density $\rho_c^{(1)}$ may still contain Dirac-delta terms if there are eigenvalues that appear repeatedly on a set of configurations of nonzero measure. In the system under consideration this should apply only to $\lambda_n = 0, 2$, that are excluded in Eq. (26).

where the last passage on the second line applies if CP invariance holds. The case $k = 0$ reduces to the cumulants of N_+ and N_- , or those of N_0 and Q ,

$$b_{N_0^{k_0} Q^{k_1}} = \rho_{N_0^{k_0} Q^{k_1} c}^{(0)} = \frac{\langle N_0^{k_0} Q^{k_1} \rangle_c}{V_4}. \quad (34)$$

For future utility, I define the following integrals associated with $\rho_{O_c}^{(k)}$,

$$I_O^{(k)}[g_1, \dots, g_k] \equiv \left[\prod_{i=1}^k \int_0^2 d\lambda_i g_i(\lambda_i) \right] \rho_{O_c}^{(k)}(\lambda_1, \dots, \lambda_k; m). \quad (35)$$

Their dependence on V_3 , T , and m is left implicit. The subscript O will again be omitted if $O = 1$. Of course $I_O^{(0)} = \rho_{O_c}^{(0)} = b_O$.

The spectral correlators $\rho_c^{(k)}$ measure the correlations between the number of modes in infinitesimal spectral intervals, and are expected to have a well-defined thermodynamic limit in the light of the typical behavior of random-matrix systems [118, 119],

$$\rho_{c\infty}^{(k)}(\lambda_1, \dots, \lambda_k; m) \equiv \lim_{V_4 \rightarrow \infty} \rho_c^{(k)}(\lambda_1, \dots, \lambda_k; m). \quad (36)$$

For $k = 1$ one obtains the spectral density in infinite volume (see Ref. [120]),

$$\rho(\lambda; m) \equiv \rho_{c\infty}^{(1)}(\lambda; m) = \lim_{V_4 \rightarrow \infty} \rho_c^{(1)}(\lambda; m), \quad (37)$$

and for $k = 2$ the connected two-point function,

$$\begin{aligned} \rho_{c\infty}^{(2)}(\lambda_1, \lambda_2; m) \\ = \left[\lim_{V_4 \rightarrow \infty} \frac{1}{V_4} (\langle \rho_U(\lambda_1) \rho_U(\lambda_2) \rangle - \langle \rho_U(\lambda_1) \rangle \langle \rho_U(\lambda_2) \rangle) \right] \\ - [\delta(\lambda_1 - \lambda_2) + \delta(\lambda_1 + \lambda_2)] \rho(\lambda_1; m). \end{aligned} \quad (38)$$

These quantities are expected to have at most integrable singularities.⁸ One similarly writes

$$\rho_{O_c\infty}^{(k)}(\lambda_1, \dots, \lambda_k; m) \equiv \lim_{V_4 \rightarrow \infty} \rho_{O_c}^{(k)}(\lambda_1, \dots, \lambda_k; m). \quad (39)$$

These quantities may generally have non-integrable singularities, and the thermodynamic limit of the integrals Eq. (35),

$$I_{O\infty}^{(k)}[g_1, \dots, g_k] \equiv \lim_{V_4 \rightarrow \infty} I_O^{(k)}[g_1, \dots, g_k], \quad (40)$$

⁸ The limit Eq. (36) should be understood in the distributional sense. For the spectral density, Eq. (37), one starts from the normalized mode number in a spectral interval $[\lambda_0, \lambda]$,

$$\mathcal{N}(\lambda_0, \lambda; m) \equiv \lim_{V_4 \rightarrow \infty} \frac{1}{V_4} \left\langle \int_{\lambda_0}^{\lambda} d\lambda' \rho_U(\lambda') \right\rangle,$$

and obtains the spectral density as the function (or more generally distribution) obeying $\rho(\lambda; m) = \partial_{\lambda} \mathcal{N}(\lambda_0, \lambda; m)$. By definition, integrating $\rho(\lambda; m)$ one gets back the mode density, so $\rho(\lambda; m)$ is integrable. Similar constructions are used to define precisely the quantities in Eq. (36).

may be divergent, depending on the choice of the functions g_i .⁹ Similarly, the thermodynamic limit of the cumulants b_O may be divergent. A finite thermodynamic limit is certainly expected for

$$\begin{aligned} n_0 &\equiv \lim_{V_4 \rightarrow \infty} b_{N_0} = \lim_{V_4 \rightarrow \infty} \frac{\langle N_0 \rangle}{V_4}, \\ \chi_t &\equiv \lim_{V_4 \rightarrow \infty} b_{Q^2} = \lim_{V_4 \rightarrow \infty} \frac{\langle Q^2 \rangle}{V_4}, \end{aligned} \quad (41)$$

which are the zero-mode density and the topological susceptibility, respectively (and I made again use of CP invariance in the last step). More generally, the existence of a thermodynamic limit is expected for the cumulants of the topological charge, b_{Q^k} , as Q is the sum of local quantities, $Q = \sum_x q(x)$ with $q(x) = -\text{tr} \{(DR)_{xx} \gamma_5\}$, with the trace running over Dirac and color indices only [54, 121, 122]; for more general $b_{N_0^{k_0} Q^{k_1}}$ this should be checked on a case-by-case basis.

It is argued, and well supported by numerical results, that $N_+(U)N_-(U) = 0$ almost everywhere (a.e.) in the space of gauge configurations. The argument is that a non-minimal realization of the index theorem, i.e., $Q(U) = N_+(U) - N_-(U)$ with $N_+(U)N_-(U) \neq 0$, requires that the gauge configuration be finely tuned to provide more zero modes than strictly required, and the set of such gauge configurations is of zero measure [123]. This implies in particular that $\langle N_0^2 \rangle = \langle Q^2 \rangle$, and so $n_0 = 0$ in the CP -invariant case (using $\langle N_0 \rangle^2 \leq \langle Q^2 \rangle$ and finiteness of χ_t). Moreover, since in this case $N_0 = |Q|$ a.e., one expects the cumulants $b_{N_0^k}$ to have a well-defined thermodynamic limit. By the same argument, one expects $N'_+(U)N'_-(U) = 0$ a.e., and so $N_2 = |Q| = N_0$ a.e., and the density of doubler modes to vanish, $n_2 \equiv \lim_{V_4 \rightarrow \infty} \frac{\langle N_2 \rangle}{V_4} = 0$, in the CP -invariant case.

III. CHIRAL SYMMETRY RESTORATION

The fundamental requirement for symmetry restoration in a local quantum field theory is that correlation functions of local operators that are related by a symmetry transformation become equal in the symmetric limit, i.e., when all the terms breaking the symmetry explicitly are removed from the theory. In the case at hand, this is the chiral limit in which the common mass m of the two flavors of light fermions is sent to zero, and chiral symmetry becomes exact at the classical level.

To obtain information on the Dirac spectrum, however, one would rather work with susceptibilities, specifically with the scalar and pseudoscalar susceptibilities in Eq. (20), that can be expressed solely in terms of the Dirac eigenvalues (see Sec. IV below). Restoration of chiral symmetry at the level of the local correlators implies restoration at the level of the susceptibilities (i.e., that susceptibilities related by a chiral transformation become equal as $m \rightarrow 0$), if the zero-momentum limit corresponding to integrating over the whole spacetime volume commutes with the chiral limit. This is the case if the correlation length of the system remains finite in the chiral limit, as one generally expects in the symmetric phase due to the expected absence of massless excitations, resulting in finite susceptibilities that are manifestly symmetric. A notable exception is the critical point of a continuous transition, where the correlation length diverges, and so do one or more of the susceptibilities. Symmetry at the level of the susceptibilities is not guaranteed in this case, although it may be possible in principle that the difference of symmetry-related susceptibilities still vanishes in the chiral limit, even if they separately diverge. Another notable exception are free continuum fermions at $T = 0$, for which chiral symmetry is restored in the chiral limit but (most of the) susceptibilities diverge due to the divergent correlation length, and the difference of symmetry-related susceptibilities does not always vanish in the chiral limit.¹⁰ Conversely, it may be mathematically possible that restoration of chiral symmetry at the level of susceptibilities does not reflect the actual restoration of the symmetry at the fundamental level of local correlators, and comes about only due to cancelations in the spacetime integrals of these correlators. This possibility, however, seems physically very unlikely, and will be ignored in the following.¹¹

In this work chiral symmetry restoration will be understood to mean restoration at the level of the susceptibilities, i.e., that for any chiral transformation $\mathcal{U}$ [see Eq. (19) and under Eqs. (11) and (17) for notation]

$$\begin{aligned} \lim_{m \rightarrow 0} \left[\chi \left(S^{n_S}(i\vec{P})^{\vec{n}_P}(iP)^{n_P} \vec{S}^{\vec{n}_S} \right) \right. \\ \left. - \chi \left(S_{\mathcal{U}}^{n_S}(i\vec{P}_{\mathcal{U}})^{\vec{n}_P}(iP_{\mathcal{U}})^{n_P} \vec{S}_{\mathcal{U}}^{\vec{n}_S} \right) \right] = 0. \end{aligned} \quad (42)$$

As formulated, the request of symmetry restoration, Eq. (42), allows in principle for divergent susceptibilities

⁹ More precisely, one obtains $\rho_{O_c \infty}^{(k)}$ from $I_{O_\infty}^{(k)}$ by taking for g_i the indicator functions of the spectral intervals $[\lambda_{0i}, \lambda_i]$, and then taking derivatives (in the distributional sense) with respect to λ_i (see footnote 8), as long as $I_{O_\infty}^{(k)}$ is finite for the given λ_{0i} and for λ_i in some small range. A divergent $I_{O_\infty}^{(k)}$ signals instead the presence of a non-integrable singularity of $\rho_{O_c \infty}^{(k)}$ in at least one of the spectral intervals.

¹⁰ The continuum massless fermion propagator decays only algebraically at large distances, $\phi^{-1} \propto \not{x}/|x|^4$. For susceptibilities involving b bilinears, integrating in the range $1/\Lambda \leq |x| \leq L$, this leads by power counting to a dependence Λ^{4-b} on the UV cutoff, and L^{b-4} on the IR cutoff, and so to divergent susceptibilities for $b \geq 4$ as $L \rightarrow \infty$ (at fixed Λ). In particular, $\chi(S^5)$ diverges as $m \rightarrow 0$ while $\chi((iP_a)^5) = 0$ identically, so that chiral symmetry is not realized at the level of susceptibilities.

¹¹ As will become clear below, in this case the consequences of the present analysis would still apply, even though chiral symmetry would not be restored in the sense of local quantum field theory.

in the chiral limit, as long as the divergences of susceptibilities related by a chiral transformation cancel out. On the other hand, while symmetric (and finite) local correlators and a finite correlation length in the chiral limit imply both finiteness and symmetry of the susceptibilities, finiteness of the susceptibilities in the chiral limit alone does not guarantee *a priori* that they will also be symmetric. (As already mentioned in Sec. I, in this context a finite quantity is a quantity that is non-divergent – including vanishing – in the chiral limit.) In principle, then, in the chiral limit susceptibilities could diverge yet be symmetric; or remain finite yet not be symmetric. I show in Sec. IIIB that in the scalar and pseudoscalar sector this is not possible: chiral symmetry is restored at the level of scalar and pseudoscalar susceptibilities if and only if these susceptibilities are finite. Finiteness of the susceptibilities then fully characterizes symmetry restoration in the scalar and pseudoscalar sector except when the correlation length diverges (and barring “accidental” restoration at the level of the susceptibilities but not at the level of local correlators). In particular, this characterization should apply within a finite-temperature symmetric phase.

As I will also show in Sec. IIIB, finiteness of the susceptibilities has the corollary that the “even” (respectively, “odd”) susceptibilities, i.e., those involving an even (respectively, odd) number of the isosinglet bilinears S and P , must be C^∞ (i.e., infinitely differentiable) functions of m^2 at $m = 0$ (respectively m times a C^∞ function), a property that I will refer to as “ m^2 -differentiability” for short. This is an equivalent characterization of symmetry restoration at the level of susceptibilities.

Thanks to the transformation properties of the relevant bilinears under chiral transformations, Eq. (11), the symmetry restoration condition Eq. (42) can be expressed in compact form in terms of the generating function $\mathcal{W}_\infty$ [see Eqs. (20) and (21) and the following discussion in Sec. IIC],

$$\lim_{m \rightarrow 0} [\mathcal{W}_\infty(V, W; m) - \mathcal{W}_\infty(RV, RW; m)] = 0, \quad \forall R \in \text{SO}(4). \quad (43)$$

This is the starting point of the analysis carried out in this work.

In a symmetric phase with a finite correlation length, the symmetry restoration condition Eq. (42) naturally extends to susceptibilities involving the spacetime integrals $G_i = \sum_x G_i(x)$ of local operators $G_i(x)$ built out entirely of gauge fields. Within the symmetric phase one has then

$$\lim_{m \rightarrow 0} \left[\chi \left(S^{n_S}(i\vec{P})^{\vec{n}_P}(iP)^{n_P} \tilde{S}^{\vec{n}_S} \prod_i G_i \right) - \chi \left(S_U^{n_S}(i\vec{P}_U)^{\vec{n}_P}(iP_U)^{n_P} \tilde{S}_U^{\vec{n}_S} \prod_i G_i \right) \right] = 0. \quad (44)$$

As shown in Sec. IIIC, also in this case Eq. (44) is satisfied if and only if all these susceptibilities are finite in the chiral limit, or equivalently if they are m^2 -differentiable

or m times an m^2 -differentiable function depending on whether they are even or odd. A relevant example are susceptibilities involving the topological charge, since Q admits a representation as the integral of a local density [see under Eq. (41)]. From this result follows in particular the m^2 -differentiability of its cumulants. Equation (44) can be reformulated in a manner similar to Eq. (43) by defining the augmented generating functions $\mathcal{Z}_G$, $\mathcal{W}_G$, and $\mathcal{W}_{G_\infty}$ [see Eq. (D5)], adding sources J_{G_i} for the gauge operators G_i and replacing $K \rightarrow K - \sum_i J_{G_i} G_i$ in the definition of $\mathcal{Z}$ and $\mathcal{W}$ [see Eqs. (14) and (16)]. Since G_i are unaffected by chiral transformations, the symmetry restoration condition reads

$$\lim_{m \rightarrow 0} [\mathcal{W}_{G_\infty}(V, W; J_G; m) - \mathcal{W}_{G_\infty}(RV, RW; J_G; m)] = 0, \quad \forall R \in \text{SO}(4), \quad (45)$$

where J_G denotes collectively the gauge-operator sources.

The fact that gauge fields are unaffected by chiral transformations suggests that one could reasonably expect that symmetry restoration is manifest also in susceptibilities involving scalar and pseudoscalar bilinears and general nonlocal operators built out of gauge fields only. In other words, Eq. (44) is expected to hold also if one includes (translation-invariant) nonlocal operators in the set $\{G_i\}$, leading to the same conclusions as above concerning finiteness and m^2 -differentiability of the susceptibilities. Of course, if one adopts a strictly local point of view, then whether or not symmetry is manifest in this kind of susceptibilities has no bearing on its being physically realized. On the other hand, nothing prevents one to use also these nonlocal functionals for a more detailed characterization of the phases of the theory. I will then treat this as an additional assumption, logically quite independent from the (essentially) local symmetry-restoration assumptions discussed above, and refer to it as “nonlocal restoration” when invoked.

For the purposes of this work, the main consequence of nonlocal restoration is the resulting m^2 -differentiability of spectral observables, such as the spectral density or the two-point function of Dirac eigenvalues, that are precisely of the relevant type – susceptibilities of operators that are (highly) nonlocal but involve only the gauge fields [see Eqs. (29) and (30)]. However, while sensible, the additional assumption of nonlocal restoration does not follow directly from the basic request of symmetry restoration for local correlators (and from the finiteness of the correlation length). For the spectral quantities of interest a perhaps more palatable argument can be obtained by making use of partially quenched theories. It is reasonable to expect that if one probes the system with external fields, coupled in such a way as not to break chiral symmetry explicitly, then in the symmetric phase chiral symmetry will remain manifest in local correlation functions involving both dynamical and external fields, as well as in the corresponding susceptibilities if the correlation length of the system is finite. In a partially quenched setup where both fermion and pseudofermion fields of the same mass M are added to the theory, canceling exactly

each other's contribution to the partition function, one expects then that scalar and pseudoscalar susceptibilities involving both bilinears built out of the dynamical fermion fields, Eq. (9), and their counterparts built out of the external fermion fields, will still display exact chiral symmetry in the chiral limit. This leads again to the same conclusions about m^2 -differentiability of (all) susceptibilities as in the original theory. As this should hold for arbitrary complex mass M , including when this approaches purely imaginary values where discontinuities appear; and since the spectral density, the two-point eigenvalue correlation function, and similar quantities can be obtained from these discontinuities (with no insertion of bilinears built with dynamical fields), it would then follow that they are m^2 -differentiable. While this approach still requires an additional assumption, it has the advantage of involving only susceptibilities of local operators in its formulation. This is discussed in Appendix C for γ_5 -Hermitean GW operators with $2R = 1$.

A. Functional form of the generating function

It is obvious from their definition, Eq. (16), that $\mathcal{Z}$ and $\mathcal{W}$ depend on the scalar isosinglet source j_S and on the fermion mass m only through the combination $j_S + m$,

$$\mathcal{W}(V, W; m) = \mathcal{W}(\tilde{V}(m), W; 0), \quad \tilde{V}(m) \equiv V + me_0, \quad (46)$$

where $e_0 \equiv (1, \vec{0})^T$, i.e., the generating functions at nonzero m equal those at $m = 0$ but with a shifted, mass-dependent source. A simple consequence of Eq. (46) is the relation

$$\partial_{j_S} \mathcal{W}(V, W; m) = \partial_m \mathcal{W}(V, W; m), \quad (47)$$

that implies the well-known fact that the mass derivatives of a susceptibility equal other susceptibilities involving additional scalar densities. Indeed, taking repeated derivatives at zero sources and taking the thermodynamic limit one finds

$$\begin{aligned} & \partial_m^k \chi \left(S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \right) \\ &= (-1)^k \chi \left(S^{n_S+k} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \right). \end{aligned} \quad (48)$$

Equation (46) combined with the full chiral symmetry of the massless theory strongly restricts the functional form of $\mathcal{W}$, and so of $\mathcal{W}_\infty$. In fact, since the integration measure in Eq. (16) is invariant under non-singlet chiral transformations, the exactly massless theory in a finite volume is invariant under $SU(2)_L \times SU(2)_R$ transformations. For arbitrary sources V and W one has then $\mathcal{W}(RV, RW; 0) = \mathcal{W}(V, W; 0)$, $\forall R \in SO(4)$, and so

$$\mathcal{W}(V, W; 0) \equiv \hat{\mathcal{W}}(V^2, W^2, 2V \cdot W), \quad (49)$$

i.e., it depends only on $SO(4)$ invariants. The factor of 2 in the third argument is purely conventional. By

Eq. (46), this implies

$$\begin{aligned} \mathcal{W}(V, W; m) &= \mathcal{W}(\tilde{V}(m), W; 0) \\ &= \hat{\mathcal{W}}(m^2 + u(V; m), w(W), \tilde{u}(V, W; m)), \end{aligned} \quad (50)$$

where

$$\begin{aligned} u(V; m) &\equiv \tilde{V}(m)^2 - m^2 = 2mj_S + V^2, \\ w(W) &\equiv W^2, \\ \tilde{u}(V, W; m) &\equiv 2\tilde{V}(m) \cdot W = 2(mj_P + V \cdot W). \end{aligned} \quad (51)$$

In the following, the source and mass dependences of these quantities will be mostly dropped for simplicity. Equations (46), (47), and (50) remain valid also in the thermodynamic limit, where they read

$$\mathcal{W}_\infty(V, W; m) = \mathcal{W}_\infty(\tilde{V}(m), W; 0), \quad (52)$$

$$\partial_{j_S} \mathcal{W}_\infty(V, W; m) = \partial_m \mathcal{W}_\infty(V, W; m), \quad (53)$$

and

$$\begin{aligned} & \mathcal{W}_\infty(V, W; m) \\ &= \hat{\mathcal{W}}_\infty(m^2 + u(V; m), w(W), \tilde{u}(V, W; m)) \\ &\equiv \lim_{V_4 \rightarrow \infty} \hat{\mathcal{W}}(m^2 + u(V; m), w(W), \tilde{u}(V, W; m)). \end{aligned} \quad (54)$$

These formal relations summarize exact relations between susceptibilities, such as Eq. (48), or the Ward-Takahashi identities (see below). The fact that the thermodynamic limit of the generating function at $m = 0$ appears on the right-hand side of Eq. (52) does not imply that the thermodynamic and chiral limits can be exchanged: since $\tilde{V}$ is m -dependent, the right-hand side still represents the same, massive theory as the left-hand side. Equation (54) then holds independently of the possibility of exchanging limits. Note also that while one can in practice treat $\mathcal{W}_\infty(V, W; m)$, and so $\hat{\mathcal{W}}_\infty(m^2 + u, w, \tilde{u}; m)$, as polynomials (of arbitrary order) in the sources, Eq. (52) does not imply that we can treat $\mathcal{W}_\infty(\tilde{V}, W; 0)$ as a polynomial in the shifted source $\tilde{V}$, and so $\hat{\mathcal{W}}_\infty(m^2 + u, w, \tilde{u}; m)$ as a polynomial of its arguments. Finally, notice that in the presence of CP symmetry, Eq. (23) implies that $\hat{\mathcal{W}}$ and so $\hat{\mathcal{W}}_\infty$ can depend only on $\tilde{u}^2$.

The relations Eqs. (50) and (54) are exact, and hold true independently of the fate of chiral symmetry in the chiral limit, as they follow only from the symmetries of the exactly massless theory in a finite volume. The functional forms of $\hat{\mathcal{W}}$ and $\hat{\mathcal{W}}_\infty$ imply that only powers of the three combinations u , w , and $\tilde{u}$ will appear in the expansion of $\mathcal{W}$ and $\mathcal{W}_\infty$ in the sources,¹² and so relations

¹² Since u and $\tilde{u}$ are not homogeneous in the sources, if one truncates the expansion of $\mathcal{W}$ or $\mathcal{W}_\infty$ to a fixed order in V and W the terms of highest order will violate this property. However, since one can take the order of the expansion to be arbitrarily high, this causes no problem in practice.

among susceptibilities will follow. Focussing on the thermodynamic limit and expanding $\hat{\mathcal{W}}_\infty$ in powers of $u, w, \tilde{u}$, one finds

$$\mathcal{W}_\infty(V, W; m) = \sum_{n_u, n_w, n_{\tilde{u}}=0}^{\infty} \frac{u^{n_u} w^{n_w} \tilde{u}^{n_{\tilde{u}}}}{n_u! n_w! n_{\tilde{u}}!} \mathcal{A}_{n_u n_w n_{\tilde{u}}}(m^2), \quad (55)$$

where

$$\mathcal{A}_{n_u n_w n_{\tilde{u}}}(m^2) \equiv \partial_u^{n_u} \partial_w^{n_w} \partial_{\tilde{u}}^{n_{\tilde{u}}} \hat{\mathcal{W}}_\infty(m^2 + u, w, \tilde{u})|_0; \quad (56)$$

for brevity I will often write $\mathcal{A}_n = \mathcal{A}_{n_u n_w n_{\tilde{u}}}$ and $\sum_{n=0}^{\infty} = \sum_{n_u, n_w, n_{\tilde{u}}=0}^{\infty}$. The manifest dependence on m^2 is only formal at this stage. Scalar and pseudoscalar susceptibilities are then finite linear combinations of the coefficients $\mathcal{A}_n$. It is clear from Eq. (55) that all the $\mathcal{A}_n$ can be obtained from the generating function at vanishing isosinglet sources, $j_S = j_P = 0$,

$$\begin{aligned} \mathcal{W}_\infty(V, W; m)|_{j_S, P=0} &= \hat{\mathcal{W}}_\infty(m^2 + \vec{j}_P^2, \vec{j}_S^2, 2\vec{j}_P \cdot \vec{j}_S) \\ &= \sum_{n=0}^{\infty} \frac{(\vec{j}_P^2)^{n_u} (\vec{j}_S^2)^{n_w} (2\vec{j}_P \cdot \vec{j}_S)^{n_{\tilde{u}}}}{n_u! n_w! n_{\tilde{u}}!} \mathcal{A}_n(m^2), \end{aligned} \quad (57)$$

since $\vec{j}_P^2$, $\vec{j}_S^2$, and $\vec{j}_P \cdot \vec{j}_S$ are independent variables. Using this, one can show that the $\mathcal{A}_n$ are equivalent to a subset of susceptibilities, involving only the bilinears $\vec{P}$ and $\vec{S}$, from which they are obtained as finite linear combinations with m -independent coefficients (see Appendix D). The properties of the coefficients $\mathcal{A}_n$ are then easily translated to those of the scalar and pseudoscalar susceptibilities and vice versa. Moreover, since $\hat{\mathcal{W}}_\infty$ at $j_{S,P} = 0$ depends only on $m^2 + \vec{j}_P^2$, from Eq. (57) one proves that

$$\begin{aligned} \partial_{m^2} \hat{\mathcal{W}}_\infty(m^2 + \vec{j}_P^2, \vec{j}_S^2, 2\vec{j}_P \cdot \vec{j}_S) \\ = \partial_{\vec{j}_P^2} \hat{\mathcal{W}}_\infty(m^2 + \vec{j}_P^2, \vec{j}_S^2, 2\vec{j}_P \cdot \vec{j}_S), \end{aligned} \quad (58)$$

which implies

$$\partial_{m^2} \mathcal{A}_{n_u n_w n_{\tilde{u}}}(m^2) = \mathcal{A}_{n_u+1 n_w n_{\tilde{u}}}(m^2). \quad (59)$$

All these results hold also for the finite-volume generating function, $\mathcal{W}$, and the coefficients of its expansion in powers of $u, w, \tilde{u}$, of which $\mathcal{A}_n$ represent the thermodynamic limit. (More precisely, the properties above are first shown to hold in a finite volume simply as a consequence of the functional form of $\mathcal{W}$, and hold for $\mathcal{W}_\infty$ since one can commute the thermodynamic limit with source and mass derivatives.)

The construction leading to Eqs. (50)–(54) admits a geometric interpretation. The resulting functional forms reflect the invariance of $\mathcal{Z}$, $\mathcal{W}$, and $\mathcal{W}_\infty$ under the affine transformation

$$V \rightarrow RV + m(R - \mathbf{1}_4)e_0, \quad W \rightarrow RW, \quad (60)$$

where $R \in \text{SO}(4)$ and $\mathbf{1}_4$ is the four-dimensional identity matrix. This is the leftover at nonzero m of the chiral symmetry of the exactly massless theory. For infinitesimal chiral transformations, this invariance implies the well-known integrated Ward-Takahashi identities. This is shown in Appendix E.

B. Necessary and sufficient conditions for chiral symmetry restoration

The behavior in the chiral limit of the coefficients $\mathcal{A}_n$, Eq. (56), allows one to fully characterize the restoration of chiral symmetry in the scalar and pseudoscalar sector [in the sense of Eq. (42), or equivalently Eq. (43)]. In fact, a necessary and sufficient condition for chiral symmetry restoration in this sector is that all $\mathcal{A}_n(m^2)$, including the free energy density $-\mathcal{A}_{000}(m^2) = -\mathcal{W}_\infty(0, 0; m)$, have a finite chiral limit [81], i.e., chiral symmetry is restored if and only if

$$\mathcal{A}_{n*} \equiv \lim_{m \rightarrow 0} \mathcal{A}_n(m^2) \quad (61)$$

exist and are finite, $\forall n$. Here I provide a simpler and shorter proof than that of Ref. [81], also filling in some details omitted there (see also Ref. [82]).

Sufficiency is an obvious consequence of the functional form Eq. (55): if $\mathcal{A}_{n*}$ exist and are finite, then

$$\begin{aligned} \lim_{m \rightarrow 0} \mathcal{W}_\infty(V, W; m) \\ = \sum_{n=0}^{\infty} \frac{(V^2)^{n_u} (W^2)^{n_w} (2V \cdot W)^{n_{\tilde{u}}}}{n_u! n_w! n_{\tilde{u}}!} \mathcal{A}_{n*}, \end{aligned} \quad (62)$$

which is manifestly symmetric. Since the available $\text{SO}(4)$ invariants, i.e., V^2 , W^2 , and $2V \cdot W$, are also $\text{O}(4)$ invariants, $\mathcal{W}_\infty$ is actually $\text{O}(4)$ -symmetric in the symmetric phase.

To prove necessity, the first step is to notice that the symmetry restoration condition Eq. (43) implies that in the chiral limit the generating function $\mathcal{W}_\infty$ depends only on $\text{SO}(4)$ -invariant combinations of the sources. This would be trivial to show if one assumed the existence and finiteness of the chiral limit of $\mathcal{W}_\infty$, but this is precisely what one wants to prove here. The proof of the statement above (without assuming existence and finiteness of the chiral limit) is given in Appendix F. In the chiral limit $\mathcal{W}_\infty$, or equivalently $\hat{\mathcal{W}}_\infty$, can then depend on the sources only through V^2 , W^2 , and $2V \cdot W$, although at this stage it is not guaranteed that the coefficients of an expansion in powers of these invariants have a finite limit as $m \rightarrow 0$. Nonetheless, one finds that

$$\begin{aligned} \lim_{m \rightarrow 0} [\partial_{j_S} - 2(j_S \partial_{V^2} + j_P \partial_{2V \cdot W})] \\ \times \hat{\mathcal{W}}_\infty(m^2 + 2mj_S + V^2, W^2, 2mj_P + V \cdot W) = 0, \end{aligned} \quad (63)$$

and so

$$\lim_{m \rightarrow 0} m \partial_u \hat{\mathcal{W}}_\infty(m^2 + u, w, \tilde{u}) = 0. \quad (64)$$

The second step is to use this result to characterize the behavior of the mass derivative of $\hat{\mathcal{W}}_\infty$ in the chiral limit. Using Eqs. (47) and (64) one finds

$$\begin{aligned} \lim_{m \rightarrow 0} \partial_m \mathcal{W}_\infty(V, W; m) &= \lim_{m \rightarrow 0} \partial_{j_S} \mathcal{W}_\infty(V, W; m) \\ &= 2 \lim_{m \rightarrow 0} (j_S \partial_u + j_P \partial_{\tilde{u}}) \hat{\mathcal{W}}_\infty(m^2 + u, w, \tilde{u}), \end{aligned} \quad (65)$$

and so for $j_{S,P} = 0$

$$\lim_{m \rightarrow 0} \partial_m \mathcal{W}_\infty(V, W; m)|_{j_{S,P}=0} = 0, \quad (66)$$

that readily implies¹³

$$\lim_{m \rightarrow 0} \partial_m \mathcal{A}_n(m^2) = 0, \quad (67)$$

for every $n = (n_u, n_w, n_{\bar{u}})$. Then $\mathcal{A}_n$ in the chiral limit,

$$\mathcal{A}_{n*} = \mathcal{A}_n(m_0^2) + \lim_{m \rightarrow 0} \int_{m_0}^m d\bar{m} \partial_{\bar{m}} \mathcal{A}_n(\bar{m}^2), \quad (68)$$

exists finite $\forall n$, for an arbitrary choice of the integration limit m_0 , since the integrand is regular as $\bar{m} \rightarrow 0$ (indeed, it vanishes there) and so the integral in Eq. (68) has a finite limit as $m \rightarrow 0$.¹⁴ This completes the proof.

The equivalence of $\mathcal{A}_n$ and scalar and pseudoscalar susceptibilities (see Appendix D) allows one to characterize chiral symmetry restoration directly in terms of the latter: chiral symmetry is restored if and only if all scalar and pseudoscalar susceptibilities are finite in the chiral limit.¹⁵ Barring the unlikely case of symmetric susceptibilities in the absence of chiral symmetry restoration, spontaneous chiral symmetry breaking requires then one or more susceptibilities to diverge in the chiral limit, in agreement with the appearance of massless excitations and of a divergent correlation length implied by Goldstone's theorem both at zero [124] and nonzero temperature [125–129] (see also Ref. [99] for the specific case of chiral symmetry).

It follows from the finiteness of all the $\mathcal{A}_{n*}$ and from Eq. (59) that

$$\lim_{m \rightarrow 0} \partial_{m^2} \mathcal{A}_{n_u n_w n_{\bar{u}}}(m^2) = \mathcal{A}_{n_u+1 n_w n_{\bar{u}}} * \quad (69)$$

is finite, i.e., $\mathcal{A}_n$ are m^2 -differentiable. This implies that all the odd mass-derivatives of $\mathcal{A}_n$ vanish in the chiral limit. Since m^2 -differentiability implies finiteness, an equivalent characterization of the symmetric phase is then the following: chiral symmetry is restored [in the sense of Eq. (42)] if and only if $\mathcal{A}_n$ are m^2 -differentiable for all n . As a consequence of the functional form Eq. (55) of the generating function, even (respectively, odd) susceptibilities are linear combinations of the $\mathcal{A}_n(m^2)$ with

coefficients that are even (respectively, odd) polynomials in m . It follows that chiral symmetry is restored in the chiral limit if and only if even susceptibilities are m^2 -differentiable, and odd susceptibilities are m times an m^2 -differentiable function, which is another characterization of the symmetric phase for what concerns the scalar and pseudoscalar sector.

C. Remarks

The extension of the proof in Sec. IIIB to susceptibilities involving translation-invariant operators G_i built out of gauge fields is straightforward. Since G_i are unaffected by chiral transformations, the corresponding sources J_{G_i} in the augmented generating function, $\mathcal{W}_{G_\infty}$ [see Eq. (D5)], act merely as spectators in the symmetry-restoration condition Eq. (45), and all the results derived above still hold. These include the functional form Eq. (54) for $\mathcal{W}_{G_\infty}$, that can be expanded as in Eq. (55), with coefficients $\mathcal{A}_n(m^2; J_G)$ that are now generating functions themselves (that can be re-expressed in terms of a subset of generating functions involving $\{G_i\}$, $\vec{P}$, and $\vec{S}$, see Appendix D) that obey Eq. (59). Expanding $\mathcal{A}_n(m^2; J_G)$ in powers of J_{G_i} , the resulting susceptibilities must all be finite in the chiral limit, by the same argument as in Sec. IIIB, and therefore m^2 -differentiable or m times an m^2 -differentiable function, depending on whether they are even or odd in the number of isosinglet bilinears. Of course, finiteness is also a sufficient condition for symmetry restoration at the level of susceptibilities, involving now gauge operators as well.

All the above applies also to nonlocal gauge operators if one assumes nonlocal restoration, which in particular implies the m^2 -differentiability of the spectral density ρ , Eq. (37), in the symmetric phase. Indeed, $\rho = \chi(\rho_U)$ in the notation of Eq. (19), with ρ_U , Eq. (26), a translation-invariant nonlocal functional of the gauge fields.¹⁶ As already mentioned in the introductory remarks of this section, an alternative proof of m^2 -differentiability of spectral quantities under the assumption that chiral symmetry remains manifest when probing the system with external fields is discussed in Appendix C.

Clearly, m^2 -differentiability is a weaker property than the m^2 -analyticity assumed in Refs. [77–80], as it only implies that the quantities of interest can be written as $\mathcal{O} = z(m^2) + \sum_{k=0}^{\infty} a_k m^{2k}$, with $|a_k| < \infty$ and z vanishing with all its derivatives at $m = 0$, and with no guarantee

¹³ More directly, Eq. (64) implies $\lim_{m \rightarrow 0} 2m \mathcal{A}_{n_u+1, n_w, n_{\bar{u}}}(m^2) = 0$, as one sees by computing the derivative, setting $j_{S,P} = 0$, and using the independence of $\vec{j}_P^2$, $\vec{j}_S^2$, and $\vec{j}_P \cdot \vec{j}_S$. By Eq. (59) this implies $\lim_{m \rightarrow 0} 2m \partial_{m^2} \mathcal{A}_n(m^2) = \lim_{m \rightarrow 0} \partial_m \mathcal{A}_n(m^2) = 0$.

¹⁴ The integrand in Eq. (68) is a regular function of m for $m \neq 0$, since $\mathcal{A}_n(m^2)$ are linear combinations of physical susceptibilities (see Appendix D), and therefore so are $\partial_m \mathcal{A}_n(m^2) = 2m \mathcal{A}_{n_u+1 n_w n_{\bar{u}}}(m^2)$ [see Eq. (59)].

¹⁵ More precisely, since the $\mathcal{A}_n$ are equivalent to a subset of susceptibilities (see Appendix D), chiral symmetry is restored if and only if these are finite. As all the other susceptibilities are obtained from $\mathcal{A}_n$, finiteness of all susceptibilities follows from that of those in the subset.

¹⁶ To proceed rigorously, one starts from the normalized mode number $\mathcal{N}(\lambda_0, \lambda; m)$ (see footnote 8), which is the expectation value of a real-valued translation-invariant nonlocal functional of the gauge fields, and so is m^2 -differentiable under the assumption of nonlocal restoration. The spectral density is then obtained by differentiating $\mathcal{N}$ with respect to λ , and similarly its m^2 -derivatives are obtained, by definition, by differentiating the m^2 -derivatives of $\mathcal{N}$ with respect to λ .

that the sum has a finite radius of convergence. Nonetheless, m^2 -differentiability suffices for almost all the arguments of Refs. [77–80], and so for most practical purposes their m^2 -analyticity assumptions are essentially justified as a necessary consequence of symmetry restoration.

In the discussion above the lattice regularization plays no specific role, other than putting quantum field theories on sound mathematical footing. If a continuum, infinite-volume generating function $\mathcal{W}_{\text{cont}}$ can be defined at nonzero light-fermion mass after suitable renormalization, one would still require Eq. (43) to hold for chiral symmetry to be restored in the chiral limit. The chiral symmetry of the massless theory is exact at any nonzero lattice spacing and so it can be preserved under renormalization (see Refs. [55, 130]), implying in turn that the resulting functional form Eq. (54) for $\mathcal{W}_{\text{cont}}$ is also preserved. All the consequences derived above will then still hold in the continuum, including the result that chiral symmetry is restored in the scalar and pseudoscalar sector if and only if all the (renormalized) susceptibilities remain finite in the chiral limit, implying in turn the appropriate m^2 -differentiability property for even and odd susceptibilities.

IV. GENERATING FUNCTION FROM THE DIRAC SPECTRUM

In this section I compute explicitly the fermionic determinant in the presence of scalar and pseudoscalar sources for $N_f = 2$ degenerate flavors in terms of Dirac eigenvalues, and obtain an exact expression for the generating function in terms of eigenvalue correlation functions, in the case of γ_5 -Hermitean GW Dirac operators D with $2R = 1$. In the rest of this paper I restrict to this case, whose special properties have been discussed in Sec. IID. The result confirms the general analysis of the functional form of the partition function discussed in Sec. III A, and is the starting point for deriving constraints on the Dirac spectrum in the chirally symmetric phase.

A. Fermionic determinant in the presence of sources

After integrating out the light fermion fields in Eq. (16), the generating function of full correlators reads

$$\mathcal{Z}(V, W; m) = \int DU e^{-S_{\text{eff}}(U)} \det \mathcal{M}(U; V, W; m), \quad (70)$$

where

$$\begin{aligned} \mathcal{M}(U; V, W; m) \\ \equiv D(U) \mathbf{1}_f + \left(1 - \frac{1}{2} D(U)\right) (A(V, W) + iB(V, W) \gamma_5), \end{aligned} \quad (71)$$

with

$$\begin{aligned} A(V, W) &\equiv (j_S + m) \mathbf{1}_f - \vec{j}_S \cdot \vec{\sigma}, \\ B(V, W) &\equiv j_P \mathbf{1}_f + \vec{j}_P \cdot \vec{\sigma}, \end{aligned} \quad (72)$$

is a matrix carrying spacetime (including Dirac), color, and flavor indices. Its determinant is most easily obtained in the orthonormal basis $\{\psi_n \phi_f\}$, where ψ_n are the orthonormal eigenvectors of D , and ϕ_f , $f = 1, 2$, is the canonical basis of flavor space, $(\phi_f)_i = \delta_{fi}$. Zero modes and doubler modes are chosen with definite chirality. For pairs of conjugate complex modes, μ_n and μ_n^* , the corresponding eigenvectors are chosen to be ψ_n and $\gamma_5 \psi_n$, respectively. In this basis the 2×2 blocks $(\mathcal{M}_{n'n})_{ff} = (\psi_{n'} \phi_f, \mathcal{M} \psi_n \phi_f)$ of $\mathcal{M}$ read

$$\mathcal{M}_{n'n} = \mu_n \mathbf{1}_f \delta_{n'n} + \left(1 - \frac{1}{2} \mu_{n'}\right) [A \delta_{n'n} + iB(\gamma_5)_{n'n}], \quad (73)$$

where $(\gamma_5)_{n'n} \equiv (\psi_{n'}, \gamma_5 \psi_n)$, and so $\mathcal{M}$ has a simple block structure in the Dirac eigenmode indices n, n' , with nonzero matrix elements only on the diagonal for chiral zero modes and doubler modes, and off-diagonal for n, n' a pair of conjugate complex modes. It follows then

$$\begin{aligned} \det \mathcal{M} &= 2^{2N_2} [\det(A + iB)]^{N_+} [\det(A - iB)]^{N_-} \\ &\times \prod_{n, \text{Im } \mu_n > 0} \det M(\mu_n), \end{aligned} \quad (74)$$

where for $z \in \mathbb{C}$

$$M(z) \equiv \begin{pmatrix} z \mathbf{1}_f + \left(1 - \frac{z}{2}\right) A & i \left(1 - \frac{z}{2}\right) B \\ i \left(1 - \frac{z^*}{2}\right) B & z^* \mathbf{1}_f + \left(1 - \frac{z^*}{2}\right) A \end{pmatrix}. \quad (75)$$

A zero mode of chirality ξ contributes a factor

$$\begin{aligned} \det(A + i\xi B) &= \tilde{V}^2 - W^2 + 2i\xi \tilde{V} \cdot W \\ &= m^2 + u - w + i\xi \tilde{u}, \end{aligned} \quad (76)$$

with $\tilde{V}$ defined in Eq. (46), and $u, w, \tilde{u}$ defined in Eq. (51). For the contribution of a pair of complex modes μ_n, μ_n^* , one finds (see Appendix G for details)

$$\begin{aligned} \det M(\mu_n) &= \lambda_n^4 + 2\lambda_n^2 h(\lambda_n) (\tilde{V}^2 + W^2) \\ &\quad + h(\lambda_n)^2 \left((\tilde{V}^2 - W^2)^2 + (2\tilde{V} \cdot W)^2 \right) \\ &= [\lambda_n^2 + m^2 h(\lambda_n)]^2 \\ &\quad + 2h(\lambda_n) [(\lambda_n^2 + m^2 h(\lambda_n))u + (\lambda_n^2 - m^2 h(\lambda_n))w] \\ &\quad + h(\lambda_n)^2 [(u - w)^2 + \tilde{u}^2], \end{aligned} \quad (77)$$

where λ_n is defined above Eq. (26), with $\lambda_n^2 = |\mu_n|^2$, and

$$h(\lambda) \equiv 1 - \frac{\lambda^2}{4}. \quad (78)$$

The first expression in Eq. (77) shows that $\det M(\mu_n) > 0$. Equations (76) and (77) show explicitly that the partition function $\mathcal{Z}$ depends only on SO(4)-invariant combinations of $\tilde{V}$ and W , as anticipated [see Eq. (50)]. In conclusion,

$$\begin{aligned} \det \mathcal{M} &= (\det D_m)^2 (1 + X_0)^{N_+} (1 + X_0^*)^{N_-} \\ &\times \prod_{n, \lambda_n > 0} [1 + X(\lambda_n)], \end{aligned} \quad (79)$$

where $(\det D_m)^2$ is the fermionic determinant at zero sources,

$$\det D_m = m^{N_0} 2^{N_2} \prod_{n, \lambda_n > 0} [\lambda_n^2 + m^2 h(\lambda_n)], \quad (80)$$

and

$$\begin{aligned} X_0 &\equiv \frac{u - w + i\tilde{u}}{m^2}, \\ X(\lambda) &\equiv 2 \left(f(\lambda; m)u + \tilde{f}(\lambda; m)w \right) \\ &\quad + f(\lambda; m)^2 ((u - w)^2 + \tilde{u}^2) \\ &= 2\hat{f}(\lambda; m)(u + w) \\ &\quad + m^4 f(\lambda; m)^2 (2\text{Re } X_0 + |X_0|^2), \end{aligned} \quad (81)$$

where

$$\begin{aligned} f(\lambda; m) &\equiv \frac{h(\lambda)}{\lambda^2 + m^2 h(\lambda)}, \\ \tilde{f}(\lambda; m) &\equiv f(\lambda; m) - 2m^2 f(\lambda; m)^2, \\ \hat{f}(\lambda; m) &\equiv f(\lambda; m) - m^2 f(\lambda; m)^2. \end{aligned} \quad (82)$$

The dependence of X_0 and $X(\lambda)$ on the sources and on m is left implicit for notational simplicity.

For certain choices of gauge group and gauge-group representation [e.g., $\text{SU}(N_c)$ and adjoint representation], the Dirac spectrum has Kramers degeneracy due to the existence of an antiunitary operator $\mathcal{T}$ that obeys $[\mathcal{T}, D] = 0$ and $\mathcal{T}^2 = -1$ (see Ref. [91]). This implies that complex modes are doubly degenerate, and $N_{\pm}$ are even. One can then replace $1 + X(\lambda_n) \rightarrow [1 + X(\lambda_n)]^2$ in Eq. (79), and $\lambda_n^2 + m^2 h(\lambda_n) \rightarrow [\lambda_n^2 + m^2 h(\lambda_n)]^2$ in Eq. (80), while limiting the product to the reduced spectrum. In the rest of this paper, when explicit expressions in terms of the Dirac spectrum are provided for the relevant quantities, it is assumed that the Dirac spectrum is not degenerate (see also footnote 18).

B. Cumulant expansion of the partition function

Using the results above, labeling the eigenvalues for a given gauge configuration U so that the $N = (N_{\text{tot}} - N_0 - N_2)/2$ complex eigenvalues with positive imaginary part correspond to $n = 1, \dots, N$, one finds

$$\frac{\mathcal{Z}(V, W; m)}{\mathcal{Z}(0, 0; m)} = \left\langle e^{N_+ S(X_0)} e^{N_- S(X_0^*)} \prod_{n=1}^N [1 + X(\lambda_n)] \right\rangle, \quad (83)$$

where $S(t) \equiv \ln(1 + t)$.¹⁷ Writing

$$\prod_{n=1}^N [1 + X(\lambda_n)] = \sum_{k=0}^{\infty} \frac{1}{k!} Y_k, \quad (84)$$

where $Y_0 \equiv 1$, $Y_k \equiv 0$ for $k > N$, and

$$Y_k \equiv \sum_{\substack{n_1, \dots, n_k=1 \\ n_i \neq n_j, \forall i, j}}^N X(\lambda_{n_1}) \dots X(\lambda_{n_k}), \quad 1 \leq k \leq N, \quad (85)$$

expanding the exponentials in Eq. (83) in power series, and setting $A_{\vec{k}} \equiv N_+^{k_1} N_-^{k_2} Y_{k_3}$, where $\vec{k} = (k_1, k_2, k_3)$, one finds by standard combinatorics (see Appendix A 2)

$$\frac{\mathcal{Z}(V, W; m)}{\mathcal{Z}(0, 0; m)} = \exp \left\{ \sum_{\vec{k} \neq \vec{0}} \left(\prod_{j=1}^3 \frac{t_j^{k_j}}{k_j!} \right) \langle A_{\vec{k}} \rangle_c \right\}, \quad (86)$$

where $t_1 = S(X_0)$, $t_2 = S(X_0^*)$, $t_3 = 1$. The connected correlation functions, $\langle A_{\vec{k}} \rangle_c$, are defined in the usual way (see Appendices A 1 and A 4), and in terms of spectral quantities they read

$$\langle A_{\vec{k}} \rangle_c = I_{N_+^{k_1} N_-^{k_2}}^{(k_3)} [X, \dots, X], \quad (87)$$

see Eq. (35). Moreover (see Ref. [131], §24.1.3)

$$\frac{S(x)^k}{k!} = \sum_{n=k}^{\infty} s(n, k) \frac{x^n}{n!}, \quad (88)$$

where $s(n, k)$ are the Stirling numbers of the first kind, and since $s(n, 0) = s(0, n) = \delta_{n0}$, one can finally write [see Eqs. (30), (31), and (35)]

$$\begin{aligned} \mathcal{W}(V, W; m) - \mathcal{W}(0, 0; m) &= \frac{1}{V_4} \ln \frac{\mathcal{Z}(V, W; m)}{\mathcal{Z}(0, 0; m)} \\ &= \sum_{\vec{n} \neq \vec{0}} \frac{X_0^{n_1} X_0^{*n_2}}{n_1! n_2! n_3!} \sum_{k_1=0}^{n_1} \sum_{k_2=0}^{n_2} s(n_1, k_1) s(n_2, k_2) \\ &\quad \times I_{N_+^{k_1} N_-^{k_2}}^{(n_3)} [X, \dots, X]. \end{aligned} \quad (89)$$

This result provides an explicit representation of the generating function in terms of spectral correlators of the GW Dirac operator, involving both complex and zero modes; and therefore a spectral representation of all scalar and pseudoscalar susceptibilities once the thermodynamic limit is taken. While individual terms may not have a well-defined thermodynamic limit (see end of Sec. II D), the combinations corresponding to the various susceptibilities certainly do at $m \neq 0$ (and in the symmetric phase also as $m \rightarrow 0$, see Sec. III B). If desired, one can straightforwardly re-express $I_{N_+^{k_1} N_-^{k_2}}^{(n_3)}$ in terms of $I_{N_0^{k_0} Q^{k_1}}^{(n_3)}$ [see Eq. (A21)]. Notice that $I_{N_+^{k_1} N_-^{k_2}}^{(n_3)}$ generally do not vanish in the thermodynamic limit, even under the additional assumption $N_+ N_- = 0$ a.e.; and that there is *a priori* no reason for their contribution to $\mathcal{W}_{\infty}$ to be negligible in the chiral limit, even though zero modes are suppressed, as they are multiplied by factors $X_0^{n_1} X_0^{*n_2} \propto m^{-2(n_1+n_2)}$ (with $n_i \geq k_i$).

¹⁷ For complex arguments $\ln$ denotes the principal value of the complex logarithm, $\ln z = \ln |z| + i \arg z$, with $\arg z \in (-\pi, \pi]$.

An alternative route to a spectral representation of $\mathcal{W}$ is to directly expand $(1 + X_0)^{N_+}$ and $(1 + X_0^*)^{N_-}$ in Eq. (79) in powers of X_0 and X_0^* . Setting

$$s_k(t) \equiv t(t-1)\dots(t-k+1) = \sum_{l=1}^k s(k, l) t^l, \quad (90)$$

for $k \geq 1$, and $s_0(t) \equiv 1$, one has

$$(1+x)^n = \sum_{k=0}^n \frac{x^k}{k!} s_k(n) = \sum_{k=0}^{\infty} \frac{x^k}{k!} s_k(n), \quad (91)$$

since $s_k(n) = 0$ for $n \in \mathbb{N}_0$ if $k > n$. One still finds Eq. (86) but this time with $A_{\vec{k}} = s_{k_1}(N_+)s_{k_2}(N_-)Y_{k_3}$, and $t_1 = X_0$, $t_2 = X_0^*$, $t_3 = 1$, and one obtains a more compact expression for the generating function,

$$\begin{aligned} \mathcal{W}(V, W; m) - \mathcal{W}(0, 0; m) \\ = \sum_{\vec{n} \neq 0} \frac{X_0^{n_1} X_0^{*n_2}}{n_1! n_2! n_3!} I_{s_{n_1}(N_+)s_{n_2}(N_-)}^{(n_3)}[X, \dots, X], \end{aligned} \quad (92)$$

where $I_{s_{n_1}(N_+)s_{n_2}(N_-)}^{(k)}$ is defined according to Eq. (35) in terms of the spectral correlator $\rho_{s_{n_1}(N_+)s_{n_2}(N_-)}^{(k)} = V_4^{-1} \langle s_{s_1}(N_+)s_{s_2}(N_-)\rho_U^{(k)} \rangle_c$, obtained recursively by applying Eq. (A1) to $A_s^{(k)} = s_{s_1}(N_+)s_{s_2}(N_-)\rho_U^{(k)}$. Equations (92) and (89) are equivalent by virtue of the following combinatorial result,

$$\begin{aligned} \sum_{\sigma_1=0}^{n_1} \sum_{\sigma_2=0}^{n_2} s(n_1, \sigma_1)s(n_2, \sigma_2)\rho_{N_+^{\sigma_1}N_-^{\sigma_2}}^{(n_3)}(\lambda_1, \dots, \lambda_{n_3}; m) \\ = \rho_{s_{n_1}(N_+)s_{n_2}(N_-)}^{(n_3)}(\lambda_1, \dots, \lambda_{n_3}; m), \end{aligned} \quad (93)$$

which is a consequence of Eq. (88) (see Appendix A 3).¹⁸

Since X_0 is linear and $X(\lambda)$ is quadratic in $u, w, \tilde{u}$, the order of the polynomial in these variables that multiplies the correlation function $\rho_{s_{n_1}(N_+)s_{n_2}(N_-)}^{(n_3)}$ is $n_1 + n_2 + 2n_3$, and so this quantity appears in coefficients $\mathcal{A}_n$ of order at most $n_u + n_w + n_{\tilde{u}} = n_1 + n_2 + 2n_3$. In particular, $\rho_c^{(n_3)}$ appears at most at order $2n_3$. If CP symmetry holds, a more natural expansion is in powers of u, w , and $\tilde{u}^2$. In such an expansion, the spectral density $\rho = \rho_c^{(1)}$ appears in first-order coefficients, and in coefficients that are second order in u and w , so one cannot get any constraint on the spectral density from coefficients of order higher than 2. One can in fact show that there are no new constraints even from the second-order coefficients (see end of Sec. V A).

CP invariance was not used to obtain Eqs. (89) and (92), which remain unchanged also in the presence of CP -breaking terms, such as a θ -term, in the action S_{eff} . In a CP -symmetric theory the roles of N_+ and N_- can be interchanged, and so in Eqs. (89) and (92) one can replace

$$\begin{aligned} X_0^{n_1} X_0^{*n_2} &\rightarrow \text{Re}(X_0^{n_1} X_0^{*n_2}) \\ &= |X_0|^{n_1+n_2} T_{|n_1-n_2|}(\text{Re } X_0/|X_0|), \end{aligned} \quad (94)$$

with $T_n(x)$ the Chebyshev polynomial of order n . Since $T_n(-x) = (-1)^n T_n(x)$, and $n_1 + n_2$ and $|n_1 - n_2|$ have the same parity, one sees that $\text{Re}(X_0^{n_1} X_0^{*n_2})$ depends only on $\text{Re } X_0$ and $|X_0|^2$, and so that $\mathcal{W}$ involves only powers of $u + w$, $|X_0|^2 = (u - w)^2 + \tilde{u}^2$ (so only $\tilde{u}^2$ appears), and $\text{Re } X_0 = u - w$. From Eq. (12) follows that the generating function of the $U(1)_A$ -rotated scalar and pseudoscalar susceptibilities is obtained by replacing $(V, W) \rightarrow (V \cos \alpha - W \sin \alpha, V \sin \alpha + W \cos \alpha)$. In the limit $m \rightarrow 0$, both $u + w \rightarrow V^2 + W^2$ and $|X_0|^2 \rightarrow (V^2 + W^2)^2 - 4(V^2 W^2 - (V \cdot W)^2)$ become $U(1)_A$ invariants, so $U(1)_A$ -breaking effects originate in a nontrivial dependence on $\text{Re } X_0 = u - w \rightarrow V^2 - W^2$ surviving the chiral limit, since this quantity is not $U(1)_A$ -invariant.

C. Lowest orders

I now obtain explicitly the lowest-order terms of $\mathcal{W}$ in the CP -symmetric case. As already pointed out, CP invariance implies that $\mathcal{W}$ contains only even powers of $\tilde{u}$, and so it is natural to treat $\tilde{u}^2$ on the same level as u and w by setting

$$\begin{aligned} \mathcal{C}(u, w, \tilde{u}^2; m) &\equiv \mathcal{W}(V, W; m) - \mathcal{W}(0, 0; m) \\ &= \hat{\mathcal{W}}(m^2 + u, w, \tilde{u}) - \hat{\mathcal{W}}(m^2, 0, 0), \end{aligned} \quad (95)$$

and expanding in powers of u, w , and $\tilde{u}^2$,

$$\begin{aligned} \mathcal{C}(u, w, \tilde{u}^2; m) \\ = u\mathcal{C}_u(m^2) + w\mathcal{C}_w(m^2) + \tilde{u}^2\mathcal{C}_{\tilde{u}^2}(m^2) \\ + \frac{1}{2} [u^2\mathcal{C}_{uu}(m^2) + 2uw\mathcal{C}_{uw}(m^2) + w^2\mathcal{C}_{ww}(m^2) \\ + 2u\tilde{u}^2\mathcal{C}_{u\tilde{u}^2}(m^2) + 2w\tilde{u}^2\mathcal{C}_{w\tilde{u}^2}(m^2) \\ + \tilde{u}^4\mathcal{C}_{\tilde{u}^2\tilde{u}^2}(m^2)] + \dots, \end{aligned} \quad (96)$$

with omitted terms of order three or higher. The thermodynamic limit of these quantities are denoted by $\mathcal{C}_X^\infty \equiv \lim_{V_4 \rightarrow \infty} \mathcal{C}_X$, and $\mathcal{C}^\infty$ denotes the corresponding generating function.

The coefficients $\mathcal{C}_X^\infty$ can be expressed as linear combinations (with m -independent coefficients) of a restricted set of susceptibilities (see Appendix D), but they have simpler expressions in terms of the full set, Eq. (20), that can be obtained using the functional form of $\mathcal{W}_\infty$, Eq. (54), and the consequences of the anomalous $U(1)_A$ symmetry

¹⁸ For a spectrum with Kramers degeneracy, the final results Eqs. (89) and (92) should be modified by constructing the spectral quantities entering $I_O^{(k)}$ using ρ_U defined in terms of the reduced spectrum, and replacing $X(\lambda) \rightarrow 2X(\lambda) + X(\lambda)^2$.

(see Sec. VI). For the first-order coefficients one has

$$\begin{aligned}\chi_\pi &\equiv \chi((iP_a)^2) = 2\mathcal{C}_u^\infty, \\ \chi_\delta &\equiv \chi((S_a)^2) = 2\mathcal{C}_w^\infty,\end{aligned}\quad (97)$$

where χ_π and χ_δ are the usual pion and delta susceptibilities, and

$$\begin{aligned}\chi_{\pi\delta} &\equiv \chi((iP_a)S_a(iP_b)S_b)|_{a \neq b} = 8\mathcal{C}_{\tilde{u}^2}^\infty, \\ \chi((iP_a)(iP)S_a) &= -8m\mathcal{C}_{\tilde{u}^2}^\infty.\end{aligned}\quad (98)$$

Notice that for the chiral condensate one has

$$\Sigma \equiv -\chi(S) = 2m\mathcal{C}_u^\infty = m\chi_\pi. \quad (99)$$

For the η susceptibility one has

$$\chi_\eta \equiv \chi((iP)^2) = 2\mathcal{C}_w^\infty + 8m^2\mathcal{C}_{\tilde{u}^2}^\infty = \chi_\delta + m^2\chi_{\pi\delta}. \quad (100)$$

Using the anomalous Ward-Takahashi identity $\frac{\chi_\pi - \chi_\eta}{4} = \frac{\chi_t}{m^2}$ [see, e.g., Ref. [132], and Eq. (152) below], the first equation in Eq. (98) becomes

$$8\mathcal{C}_{\tilde{u}^2}^\infty = \chi_{\pi\delta} = \frac{\chi_\pi - \chi_\delta}{m^2} - \frac{4\chi_t}{m^4} = \frac{4}{m^2} \left(\frac{\mathcal{C}_u^\infty - \mathcal{C}_w^\infty}{2} - \frac{\chi_t}{m^2} \right). \quad (101)$$

For the σ susceptibility one finds

$$\chi_\sigma \equiv \chi(S^2) = 2\mathcal{C}_u^\infty + 4m^2\mathcal{C}_{uu}^\infty, \quad (102)$$

involving a second-order coefficient. For the second-order coefficients one has

$$\begin{aligned}\chi((iP_a)^2(iP_b)^2)|_{a \neq b} &= 4\mathcal{C}_{uu}^\infty, \\ \chi(S_a^2 S_b^2)|_{a \neq b} &= 4\mathcal{C}_{ww}^\infty, \\ \chi((iP_a)^2 S_b^2)|_{a \neq b} &= 4\mathcal{C}_{uw}^\infty,\end{aligned}\quad (103)$$

and moreover

$$\begin{aligned}\chi((iP_a)S_a(iP_b)S_b(iP_c)^2)|_{a \neq b \neq c \neq a} &= 16\mathcal{C}_{u\tilde{u}^2}^\infty, \\ \chi((iP_a)S_a(iP_b)S_b S_c^2)|_{a \neq b \neq c \neq a} &= 16\mathcal{C}_{w\tilde{u}^2}^\infty.\end{aligned}\quad (104)$$

For $\mathcal{C}_{\tilde{u}^2 \tilde{u}^2}$ the simplest relation is

$$\chi\left(\left(\prod_{a=1}^3 (iP_a)S_a\right)(iP)\right) = 192m\mathcal{C}_{\tilde{u}^2 \tilde{u}^2}^\infty, \quad (105)$$

where also m appears (an m -independent expression for $\mathcal{C}_{\tilde{u}^2 \tilde{u}^2}$ can be obtained in terms of susceptibilities involving eight fermion bilinears, using the results in Appendix D).

Specializing now to a non-degenerate Dirac spectrum in the CP -symmetric case, it is straightforward to compute the coefficients in Eq. (96). Recalling Eqs. (34) and (35), one has for the first-order coefficients

$$\begin{aligned}\mathcal{C}_u &= \frac{b_{N_0}}{m^2} + 2I^{(1)}[f], \\ \mathcal{C}_w &= -\frac{b_{N_0}}{m^2} + 2I^{(1)}[\tilde{f}], \\ \mathcal{C}_{\tilde{u}^2} &= \frac{1}{m^2} \left(\frac{b_{N_0} - b_{Q^2}}{2m^2} + m^2 I^{(1)}[f^2] \right) \\ &= \frac{1}{2m^2} \left(\frac{\mathcal{C}_u - \mathcal{C}_w}{2} - \frac{b_{Q^2}}{m^2} \right),\end{aligned}\quad (106)$$

while the second-order coefficients are

$$\begin{aligned}\mathcal{C}_{uu} &= \mathcal{T}_1 + 4I^{(2)}[f, f] + \frac{4}{m^2} I_{N_0}^{(1)}[f], \\ \mathcal{C}_{uw} &= -\mathcal{T}_1 + 4I^{(2)}[f, f] - 8m^2 I^{(2)}[f, f^2] - 4I_{N_0}^{(1)}[f^2], \\ \mathcal{C}_{ww} &= \mathcal{T}_1 + 4I^{(2)}[\tilde{f}, \tilde{f}] - \frac{4}{m^2} I_{N_0}^{(1)}[\tilde{f}], \\ \mathcal{C}_{u\tilde{u}^2} &= \mathcal{T}_2 + 2I^{(2)}[f, f^2] + \frac{1}{m^2} I_{N_0}^{(1)}[f^2] \\ &\quad + \frac{1}{m^4} I_{N_0}^{(1)}[f] - \frac{1}{m^4} I_{Q^2}^{(1)}[f], \\ \mathcal{C}_{w\tilde{u}^2} &= -\mathcal{T}_2 + 2I^{(2)}[\tilde{f}, f^2] - \frac{1}{m^2} I_{N_0}^{(1)}[f^2] \\ &\quad + \frac{1}{m^4} I_{N_0}^{(1)}[\tilde{f}] - \frac{1}{m^4} I_{Q^2}^{(1)}[\tilde{f}], \\ \mathcal{C}_{\tilde{u}^2 \tilde{u}^2} &= \mathcal{T}_3 + I^{(2)}[f^2, f^2] + \frac{1}{m^4} I_{N_0}^{(1)}[f^2] - \frac{1}{m^4} I_{Q^2}^{(1)}[f^2],\end{aligned}\quad (107)$$

where

$$\begin{aligned}\mathcal{T}_1 &\equiv \frac{1}{m^4} (b_{N_0^2} - b_{N_0} + 2m^4 I^{(1)}[f^2]) \\ &= 2\mathcal{C}_{\tilde{u}^2} + \frac{1}{m^4} (b_{N_0^2} + b_{Q^2} - 2b_{N_0}), \\ \mathcal{T}_2 &\equiv \frac{1}{m^6} \left(-\frac{1}{2} b_{N_0 Q^2} + b_{Q^2} + \frac{1}{2} b_{N_0^2} - b_{N_0} \right), \\ \mathcal{T}_3 &\equiv \frac{1}{m^8} \left(\frac{b_{Q^4}}{12} - \frac{b_{N_0 Q^2}}{2} + \frac{2b_{Q^2}}{3} + \frac{b_{N_0^2}}{4} - \frac{b_{N_0}}{2} \right).\end{aligned}\quad (108)$$

The second-order coefficients $\mathcal{C}_{uu}$, $\mathcal{C}_{uw}$, and $\mathcal{C}_{u\tilde{u}^2}$ can be written in a more compact form by recognizing the presence of m^2 -derivatives in their expressions. This also allows one to show explicitly that they are the m^2 -derivative of $\mathcal{C}_u$, $\mathcal{C}_w$, and $\mathcal{C}_{\tilde{u}^2}$ [see Eq. (59)]. The mass derivative of the expectation value of any mass-independent observable $\mathcal{O}$ reads

$$\partial_m \langle \mathcal{O} \rangle = -[\langle \mathcal{O} S \rangle - \langle \mathcal{O} \rangle \langle S \rangle]. \quad (109)$$

For $\mathcal{O} = \mathcal{O}(U)$ depending only on the gauge fields, after integrating fermions out, and exploiting the properties of the spectrum, one finds after a short calculation

$$\begin{aligned}\partial_m \langle \mathcal{O} \rangle &= \frac{2}{m} (\langle \mathcal{O} N_0 \rangle - \langle \mathcal{O} \rangle \langle N_0 \rangle) \\ &\quad + 4m \int_0^2 d\lambda f(\lambda; m) [\langle \mathcal{O} \rho_U(\lambda) \rangle - \langle \mathcal{O} \rangle \langle \rho_U(\lambda) \rangle].\end{aligned}\quad (110)$$

For $\mathcal{O} = N_0/V_4$, $\mathcal{O} = Q^2/V_4$, and $\mathcal{O} = \rho_U/V_4$, using $2m\partial_m = \partial_m$, one finds [see Eqs. (33)–(35)]

$$\begin{aligned}\partial_{m^2} b_{N_0} &= \frac{b_{N_0^2}}{m^2} + 2I_{N_0}^{(1)}[f], \\ \partial_{m^2} b_{Q^2} &= \frac{b_{N_0 Q^2}}{m^2} + 2I_{Q^2}^{(1)}[f],\end{aligned}\quad (111)$$

and

$$\begin{aligned} \partial_{m^2} \rho_c^{(1)}(\lambda; m) &= \frac{1}{m^2} \rho_{N_0 c}^{(1)}(\lambda; m) + 2f(\lambda; m) \rho_c^{(1)}(\lambda; m) \\ &+ 2 \int_0^2 d\lambda' f(\lambda'; m) \rho_c^{(2)}(\lambda, \lambda'; m). \end{aligned} \quad (112)$$

Moreover, since $\partial_{m^2} f = -f^2$, one has

$$\partial_{m^2} I^{(1)}[f^2] = \frac{1}{m^2} I_{N_0}^{(1)}[f^2] + 2I^{(2)}[f^2, f]. \quad (113)$$

Notice that under the assumption $N_+ N_- = 0$ a.e., in the thermodynamic limit one finds $n_0 = 0$ and so $\partial_m n_0 = 0$, and the first equation in Eq. (111) results in

$$I_{N_0 \infty}^{(1)}[f] = -\frac{b_{N_0^2 \infty}}{2m^2} \leq 0, \quad b_{N_0^2 \infty} \equiv \lim_{V_4 \rightarrow \infty} b_{N_0^2}. \quad (114)$$

This shows that exact zero modes and complex modes generally repel each other, independently of the status of chiral symmetry. Note that in this case $b_{N_0^2 \infty}$ is surely finite, as $\langle N_0^2 \rangle$ and $\langle N_0 \rangle^2$ are both (no more than) $O(V_4)$, and $\rho_{c N_0 \infty}^{(1)}(\lambda)$ is integrable,

$$\begin{aligned} \int_0^2 d\lambda \rho_{c N_0 \infty}^{(1)}(\lambda) &= \lim_{V_4 \rightarrow \infty} \int_0^2 d\lambda \frac{\langle N_0 \rho_U(\lambda) \rangle_c}{V_4} \\ &= -\lim_{V_4 \rightarrow \infty} \frac{\langle N_0^2 \rangle_c}{V_4} = -b_{N_0^2 \infty}. \end{aligned} \quad (115)$$

V. CONSTRAINTS ON THE DIRAC SPECTRUM

As shown in Sec. IIIB, chiral symmetry restoration requires the finiteness in the chiral limit of the coefficients $\mathcal{A}_n$ in Eq. (56), or equivalently of the coefficients $\mathcal{C}_X^\infty$, see Eq. (96) (and of the free energy density $\mathcal{A}_{000} = -\mathcal{W}_\infty|_0$). Using Eq. (89) or Eq. (92), this requirement translates into constraints on the Dirac spectrum. In this section I discuss the constraints obtained imposing finiteness in the chiral limit of the lowest-order coefficients $\mathcal{C}_X^\infty$, i.e., the thermodynamic limit of Eqs. (106) and (107). This is done from a general point of view, without making additional technical assumptions on the spectrum.

A. Constraints from first-order coefficients

The constraints from the first-order coefficients of the generating function, Eq. (106), amount to imposing finiteness in the chiral limit of χ_π , χ_δ , and $\chi_{\pi\delta}$ [see Eqs. (97), (98), and (101)],

$$\begin{aligned} \frac{\chi_\pi}{2} &= \mathcal{C}_u^\infty = \frac{n_0}{m^2} + 2I_\infty^{(1)}[f], \\ \frac{\chi_\delta}{2} &= \mathcal{C}_w^\infty = -\frac{n_0}{m^2} + 2I_\infty^{(1)}[\tilde{f}], \\ \frac{\chi_{\pi\delta}}{8} &= \mathcal{C}_{\tilde{u}^2}^\infty = \frac{1}{2m^2} \left(\frac{\chi_\pi - \chi_\delta}{4} - \frac{\chi_t}{m^2} \right). \end{aligned} \quad (116)$$

Independently of the request of finiteness, since $n_0 \geq 0$ and $I_\infty^{(1)}[f^k] \geq 0$ one has $\chi_\pi \geq 0$. Moreover, since $m^2 f^2 \leq f$, one has $\chi_\pi \pm \chi_\delta \geq 0$, implying $|\chi_\delta| \leq \chi_\pi$. Finiteness of χ_δ then follows after imposing finiteness of χ_π , and since both terms in χ_π are positive this requires that they be separately finite. Requiring finiteness of χ_π in the chiral limit is then equivalent to requiring that

$$\lim_{m \rightarrow 0} \frac{n_0}{m^2} < \infty, \quad \lim_{m \rightarrow 0} I_\infty^{(1)}[f] < \infty. \quad (117)$$

Finiteness of $\chi_{\pi\delta}$ in the chiral limit requires that

$$\frac{\chi_\pi - \chi_\delta}{4} = \frac{\chi_t}{m^2} + O(m^2), \quad (118)$$

which since the left-hand side must be finite requires in turn the finiteness of $\frac{\chi_t}{m^2}$ in the chiral limit. Using the explicit expression for χ_π and χ_δ in terms of the spectrum, Eq. (118) reads

$$\frac{n_0}{m^2} + 2m^2 I_\infty^{(1)}[f^2] = \frac{\chi_t}{m^2} + O(m^2). \quad (119)$$

For the usual $U(1)_A$ order parameter Δ ,

$$\Delta \equiv \lim_{m \rightarrow 0} \frac{\chi_\pi - \chi_\delta}{4}, \quad (120)$$

Eq. (119) implies that

$$\Delta = \lim_{m \rightarrow 0} \left(\frac{n_0}{m^2} + 2m^2 I_\infty^{(1)}[f^2] \right) = \lim_{m \rightarrow 0} \frac{\chi_t}{m^2}. \quad (121)$$

Equations (117) and (119) fully summarize all the constraints from the first-order coefficients. Under the additional assumption $N_+ N_- = 0$ a.e., one has $n_0 = 0$, so the first-order constraints boil down to requiring finiteness of $I_\infty^{(1)}[f]$ and $\frac{\chi_t}{m^2}$ in the chiral limit, and that $2m^2 I_\infty^{(1)}[f^2] - \frac{\chi_t}{m^2} = O(m^2)$, with Eq. (121) simplifying to

$$\Delta = \lim_{m \rightarrow 0} 2m^2 I_\infty^{(1)}[f^2] = \lim_{m \rightarrow 0} \frac{\chi_t}{m^2}. \quad (122)$$

As shown in Section IIIB, the coefficients of the generating function in the thermodynamic limit, and so the even susceptibilities, must be not only finite but also m^2 -differentiable in the chiral limit. From the last equation in Eq. (116) recast as $\frac{\chi_t}{m^2} = \frac{1}{4} (\chi_\pi - \chi_\delta - m^2 \chi_{\pi\delta})$, or directly from the anomalous Ward-Takahashi identity $\frac{\chi_t}{m^2} = \frac{\chi_\pi - \chi_\delta}{4}$ [see Eq. (152) below], follows then that $\frac{\chi_t}{m^2}$ must be m^2 -differentiable, a stronger result than m^2 -differentiability of χ_t [see discussion after Eq. (44) in Sec. III], and obtained using only symmetry restoration in the scalar and pseudoscalar sector.

Using the representation of susceptibilities in terms of the Dirac spectrum one can obtain a lower bound on $\chi_\pi - \chi_\delta$, that implies that effective restoration of $U(1)_A$ symmetry in the scalar and pseudoscalar sector at a nonzero value of the quark mass is impossible on the lattice. Making use of the assumption $N_+ N_- = 0$ a.e., so

that $n_0 = 0$ (as well as $n_2 = 0$), one finds for the chiral condensate Σ , Eq. (99),

$$\Sigma = 4mI_\infty^{(1)}[f], \quad (123)$$

and using the Cauchy-Schwarz inequality one shows that

$$2I_\infty^{(1)}[f]^2 \leq \nu I_\infty^{(1)}[f^2], \quad (124)$$

where $\nu \equiv 2 \int_0^2 d\lambda \rho(\lambda; m) = 4N_c$ is the average number of complex modes per unit four-volume [see Eq. (27)]. One has then

$$\chi_\pi - \chi_\delta \geq \frac{\Sigma^2}{\nu}. \quad (125)$$

The quantity on the right-hand side is strictly positive for any nonzero mass, so $\chi_\pi - \chi_\delta \neq 0$ at $m \neq 0$, and $U(1)_A$ cannot be effectively restored at nonzero quark mass on the lattice. Concerning the chiral limit, in the broken phase $\Sigma(0^\pm) = \pm\Sigma(0^+)$ with $\Sigma(0^+) > 0$, χ_π diverges, and $\Delta \neq 0$ (in fact one expects it to diverge, since χ_δ is expected to remain finite), so $U(1)_A$ is effectively broken; in the symmetric phase $\Sigma(0^\pm) = 0$, Eq. (125) reduces to $\Delta \geq 0$, and so both effective breaking (by $\Delta > 0$) and effective restoration (that requires $\Delta = 0$) are allowed. Extending Eq. (125) to the continuum limit is hampered by the additive and multiplicative UV divergences affecting both of its sides. Note, however, that Eq. (124) still holds if the integrals defining $I_\infty^{(1)}$ and ν are cut off at the

same point. If one deals with the additive divergences of χ_π , χ_δ , and Σ by cutting off the corresponding integrals over λ , then Eq. (125) still holds provided all quantities (including ν) are suitably redefined. Since the remaining multiplicative renormalization affects both sides in the same way, one can then extend the modified inequality to the continuum limit, showing that $U(1)_A$ cannot be effectively restored at nonzero m .

As already pointed out [see comments after Eq. (93) in Sec. IV B], since coefficients of order higher than two involve only eigenvalue correlation functions of order at least two, there are no further coefficients involving the spectral density other than those in Eq. (106) and (107). Moreover, in the second-order coefficients $\rho_c^{(1)}$ appears only in $\mathcal{T}_1$, Eq. (108), where it enters through the first-order coefficient $\mathcal{C}_{\tilde{u}^2}$. In the thermodynamic limit, ρ appears in the resulting coefficients only through $\chi_{\pi\delta}$, which is finite if the constraints discussed in this subsection are satisfied. It follows that no new direct constraint on the spectral density can be obtained, other than those coming from the first-order coefficients, Eqs. (117) and (119).

B. Constraints from second-order coefficients

Instead of using directly the second-order coefficients in Eq. (107), to obtain constraints on the spectrum it is more convenient to work with an equivalent set of quantities obtained by an invertible linear transformation, namely (in the thermodynamic limit)

$$\mathcal{C}_{uu}^\infty = \frac{1}{m^2} \left(4m^2 I^{(2)}[f, f] + \frac{\chi_t - b_{N_0^\infty}}{m^2} + 2m^2 \partial_{m^2} \frac{n_0}{m^2} + 2m^2 \mathcal{C}_{\tilde{u}^2}^\infty \right) = \partial_{m^2} \mathcal{C}_u^\infty \quad (126)$$

$$\frac{1}{2} (\mathcal{C}_{uu}^\infty + 2\mathcal{C}_{uw}^\infty + \mathcal{C}_{ww}^\infty) = 8I_\infty^{(2)}[\hat{f}, \hat{f}], \quad (127)$$

$$\frac{1}{2} (\mathcal{C}_{uu}^\infty - \mathcal{C}_{uw}^\infty) = \partial_{m^2} \left(2m^2 \mathcal{C}_{\tilde{u}^2}^\infty + \frac{\chi_t}{m^2} \right) = \frac{1}{2} \partial_{m^2} (\mathcal{C}_u^\infty - \mathcal{C}_w^\infty), \quad (128)$$

$$\begin{aligned} \mathcal{C}_{u\tilde{u}^2}^\infty &= \frac{1}{2m^2} \left[-\partial_{m^2} \frac{\chi_t}{m^2} + \frac{1}{2} (\mathcal{C}_{uu}^\infty - \mathcal{C}_{uw}^\infty) - 2\mathcal{C}_{\tilde{u}^2}^\infty \right] = \partial_{m^2} \mathcal{C}_{\tilde{u}^2}^\infty \\ &= \frac{1}{2m^4} \left[(1 - m^2 \partial_{m^2}) \frac{\chi_t - n_0}{m^2} + 2m^2 I_{N_0^\infty}^{(1)}[f^2] + 4m^4 I_\infty^{(2)}[f^2, f] \right], \end{aligned} \quad (129)$$

$$\frac{1}{4} (\mathcal{C}_{uu}^\infty - \mathcal{C}_{ww}^\infty) - m^2 (\mathcal{C}_{u\tilde{u}^2}^\infty + \mathcal{C}_{w\tilde{u}^2}^\infty) = \frac{2}{m^2} I_{Q^2}^{(1)}[\hat{f}], \quad (130)$$

$$\begin{aligned} \mathcal{C}_{\tilde{u}^2\tilde{u}^2}^\infty &= \frac{1}{m^4} \left[\frac{\kappa_{4t} - \chi_t}{12m^4} - \frac{1}{16} (\mathcal{C}_{uu}^\infty - 2\mathcal{C}_{uw}^\infty + \mathcal{C}_{ww}^\infty - 8\mathcal{C}_{\tilde{u}^2}^\infty) + \frac{m^2}{2} (\mathcal{C}_{u\tilde{u}^2}^\infty - \mathcal{C}_{w\tilde{u}^2}^\infty) \right] \\ &= \frac{1}{12m^4} \left[\frac{\kappa_{4t} - \chi_t}{m^4} - 3 \left(\partial_{m^2} \frac{\chi_t}{m^2} - \frac{2}{m^2} I_{Q^2}^{(1)}[\hat{f}] \right) + 3m^2 (\mathcal{C}_{u\tilde{u}^2}^\infty - \mathcal{C}_{w\tilde{u}^2}^\infty) \right], \end{aligned} \quad (131)$$

where $b_{N_0^\infty}$ is defined in Eq. (114),

$$\kappa_{4t} \equiv \lim_{V_4 \rightarrow \infty} b_{Q^4}, \quad (132)$$

and I have made use of Eq. (106) and of the deriva-

tive formulas in Eqs. (111)–(113). Imposing finiteness of the quantities in Eqs. (126)–(131) in the chiral limit is of course equivalent to imposing finiteness of those in Eq. (107) in the thermodynamic limit followed by the chi-

ral limit. From a general point of view, finiteness of the left-hand side of Eqs. (126), (128), and (129) corresponds to the existence of the first m^2 -derivative of $\mathcal{C}_{u,w,\tilde{u}^2}^\infty$ at $m = 0$, which is expected (for all m^2 -derivatives) in the symmetric phase (see Sec. III B). Notice that Eqs. (128) and (129) provide the same expression for $\partial_{m^2} \frac{\chi_t}{m^2}$ in terms of first- and second-order coefficients.

Finiteness in the chiral limit of the quantities in Eqs. (126)–(131) implies a number of constraints on the Dirac spectrum. The first two are

$$4m^2 I_\infty^{(2)}[f, f] = \frac{b_{N_0^\infty} - \chi_t}{m^2} - 2m^2 \partial_{m^2} \frac{n_0}{m^2} + O(m^2),$$

$$I_\infty^{(2)}[\hat{f}, \hat{f}] = O(1). \quad (133)$$

These follow from Eqs. (126) and (127), and provide constraints on the two-point eigenvalue correlator, $\rho_{c\infty}^{(2)}$, Eq. (38), with the first one relating it to topological properties of the theory. Making use of the assumption $N_+ N_- = 0$ a.e., from this constraint one finds in particular

$$-\lim_{m \rightarrow 0} 4m^2 I_\infty^{(2)}[f, f] = \lim_{m \rightarrow 0} \lim_{V_4 \rightarrow \infty} \frac{\langle N_0 \rangle^2}{m^2 V_4} \equiv \Delta'. \quad (134)$$

It is straightforward to show that $0 \leq \Delta' \leq \Delta$, so Δ' must be finite in the symmetric phase.¹⁹ Since the left-hand side should be dominated by the near-zero modes, one expects that these predominantly repel each other if $\Delta' \neq 0$. Deeper insight will be obtained in the second paper of this series under further technical assumptions on $\rho_{c\infty}^{(2)}$.

Two more constraints are

$$(1 - m^2 \partial_{m^2}) \frac{n_0 - \chi_t}{m^2} = 2m^2 I_{N_0^\infty}^{(1)}[f^2]$$

$$+ 4m^4 I_\infty^{(2)}[f^2, f] + O(m^4),$$

$$\partial_{m^2} \frac{\chi_t}{m^2} = O(1). \quad (135)$$

The first one follows from the request of finiteness of $\mathcal{C}_{u\tilde{u}^2}^\infty = \partial_{m^2} \mathcal{C}_{\tilde{u}^2}^\infty$, Eq. (129), in the chiral limit. The second one follows from finiteness of the quantity appearing in Eq. (128), using the finiteness of $\mathcal{C}_{\tilde{u}^2}^\infty$, and the required finiteness of $\partial_{m^2} \mathcal{C}_{\tilde{u}^2}^\infty$. This constraint is just a special case of the all-order result for $\frac{\chi_t}{m^2}$ already discussed above [see under Eq. (122)]. Using this constraint and the assumption $N_+ N_- = 0$ a.e., one finds $(1 - m^2 \partial_{m^2}) \frac{\chi_t}{m^2} = \Delta + O(m^4)$, which leads to

$$-\Delta = \lim_{m \rightarrow 0} \left(2m^2 I_{N_0^\infty}^{(1)}[f^2] + 4m^4 I_\infty^{(2)}[f^2, f] \right), \quad (136)$$

with the quantity in brackets deviating from $-\Delta$ only at order $O(m^4)$. This constraint involves both $\rho_{N_0 c \infty}^{(1)}$

and $\rho_{c\infty}^{(2)}$, Eqs. (33) and (38), or equivalently ρ and $\partial_{m^2} \rho$ [see Eq. (112)], again put into a relation with topological properties of the theory.

Finiteness of the quantity appearing in Eq. (130) requires

$$\frac{1}{m^2} I_{Q^2 \infty}^{(1)}[\hat{f}] = O(1). \quad (137)$$

This request is made more precise by using Eq. (103) to express the left-hand side of Eq. (130) in terms of susceptibilities. One finds

$$\lim_{m \rightarrow 0} \frac{2}{m^2} I_{Q^2 \infty}^{(1)}[\hat{f}] = \frac{1}{4} \lim_{m \rightarrow 0} (\mathcal{C}_{uu}^\infty - \mathcal{C}_{ww}^\infty)$$

$$= \frac{1}{16} \lim_{m \rightarrow 0} [\chi((iP_a)^2 (iP_b)^2) - \chi(S_a^2 S_b^2)]|_{a \neq b} \equiv \Delta_2, \quad (138)$$

and Eq. (137) requires $|\Delta_2| < \infty$. This quantity is an order parameter for $U(1)_A$ that depends on the correlation between zero and complex modes, as encoded in $\rho_{Q^2 c}^{(1)}$, Eq. (33). This can be written as

$$\rho_{Q^2 c}^{(1)}(\lambda; m) = \frac{\langle Q^2 \rho_U(\lambda) \rangle_c}{V_4} = - \partial_\theta^2 \frac{\langle \rho_U(\lambda) \rangle_\theta}{V_4} \Big|_{\theta=0} \quad (139)$$

$$= - \partial_\theta^2 \rho_c^{(1)}(\lambda; m; \theta) \Big|_{\theta=0},$$

where $\langle \dots \rangle_\theta$ denotes the expectation value in the presence of a θ term, defined by replacing $-S_{\text{eff}} \rightarrow -S_{\text{eff}} + i\theta Q$ in Eqs. (2) and (3) [see Eq. (147) below], and $\rho_c^{(1)}(\lambda; m; \theta) \equiv \frac{1}{V_4} \langle \rho_U(\lambda) \rangle_\theta$ is the normalized spectral density (in a finite volume) of the GW Dirac operator in this case. Then $I_{Q^2}^{(1)}[\hat{f}] = -\partial_\theta^2 I^{(1)}[\hat{f}; \theta]|_{\theta=0}$, where

$$I^{(1)}[g; \theta] \equiv \int_0^2 d\lambda g(\lambda) \rho_c^{(1)}(\lambda; m; \theta), \quad (140)$$

and so in the thermodynamic limit [that can be exchanged with the derivatives with respect to θ at $m \neq 0$, see discussion after Eq. (19)]

$$I_{Q^2 \infty}^{(1)}[\hat{f}] = -\frac{1}{2} \partial_\theta^2 \left(I_\infty^{(1)}[f; \theta] + I_\infty^{(1)}[\tilde{f}; \theta] \right) \Big|_{\theta=0}$$

$$= -\frac{1}{4} \partial_\theta^2 [\mathcal{C}_u^\infty(\theta) + \mathcal{C}_w^\infty(\theta)] \Big|_{\theta=0} \quad (141)$$

$$= -\frac{1}{8} \partial_\theta^2 [\chi_\pi(\theta) + \chi_\delta(\theta)] \Big|_{\theta=0},$$

where $\mathcal{C}_{u,w}^\infty(\theta)$ denote the thermodynamic limit of the expansion coefficients $\mathcal{C}_{u,w}(\theta)$ of $\mathcal{C}(\theta)$, i.e., $\mathcal{C}$ of Eq. (95) evaluated in the presence of a θ term; $\chi_\pi(\theta)$ and $\chi_\delta(\theta)$ are the pion and delta susceptibilities in this case; and I have used $2\hat{f} = f + \tilde{f}$ and Eq. (106), that holds unchanged provided all quantities are evaluated at nonzero θ . Notice that although in this case the density of zero modes, b_{N_0} , generally does not vanish in the thermodynamic limit, it still exactly cancels out in $\mathcal{C}_u(\theta) + \mathcal{C}_w(\theta)$.

¹⁹ In the large-volume limit the distribution of Q is expected to be Gaussian, and from $N_0 = |Q|$ follows that $\Delta' = \frac{2}{\pi} \Delta$.

From Eq. (138) follows then that in the symmetric phase the relation

$$\Delta_2 = - \lim_{m \rightarrow 0} \frac{1}{4m^2} \partial_\theta^2 (\chi_\pi(\theta) + \chi_\delta(\theta))|_{\theta=0} \quad (142)$$

holds between Δ_2 and the pion and delta susceptibilities.

Finally, finiteness of $\mathcal{C}_{\tilde{u}^2\tilde{u}^2}$, Eq. (131), requires

$$\frac{\kappa_{4t} - \chi_t}{m^4} = 3 \left(\partial_{m^2} \frac{\chi_t}{m^2} \Big|_{m=0} - \Delta_2 \right) + O(m^2), \quad (143)$$

where I used Eq. (138). If $\chi_t \propto m^2$, and so $U(1)_A$ remains effectively broken, this requires that κ_{4t} and χ_t be equal to leading order in m , and so that the distribution of the topological charge be indistinguishable from that of an ideal gas of instantons and anti-instantons, with identical and vanishingly small density $\chi_t/2$, to lowest order in the fermion mass and at the level of the first non-trivial cumulant. [If χ_t vanishes faster than m^2 , and so is at least $O(m^4)$, this is not necessarily the case, as Eq. (143) would generally imply that κ_{4t} and χ_t differ at leading order.] Leading corrections to the ideal behavior are $O(m^4)$, and encoded in the first term on the right-hand side of Eq. (143). A more general result was obtained in Ref. [109], namely that an ideal instanton gas behavior holds to lowest order in the fermion mass for all cumulants, under the assumption that in the symmetric phase the free energy density at finite θ angle is analytic in m^2 and $U(1)_A$ remains effectively broken by $\Delta \neq 0$. This conclusion can be obtained straightforwardly using the formalism of the present paper, as I show below in Sec. VI. Moreover, the m^2 -differentiability of susceptibilities required in the symmetric phase (see Sec. IIIB) justifies the expansion in powers of m^2 used in Ref. [109].

C. Remarks

The constraints Eqs. (117) and (119) [and Eq. (121)] are not new, and have appeared in various forms (at least implicitly) in the literature [77–80, 83]. Here, however, they have been fully justified from the theoretical point of view. Moreover, the present approach shows that these constraints, derived from the first-order coefficients, are the only constraints that involve the spectral density directly [see comments at the end of Sec. VA].

The constraints from the second-order coefficients, Eqs. (133), (135), (137), and (143), are instead new,²⁰ and involve the thermodynamic limit of the two-point correlation function of complex modes, $\rho_c^{(2)}$, and of the correlation functions $\rho_{N_0 c}^{(1)}$ and $\rho_{Q^2 c}^{(1)}$ involving zero and complex modes. These are only indirectly related to the

spectral density, through derivatives with respect to the mass [see Eq. (112)] or the θ angle [see Eq. (139)].

Without further assumptions, at this stage full restoration of $SU(2)_L \times SU(2)_R$ symmetry is compatible both with effective breaking and with effective restoration of $U(1)_A$, as the order parameters Δ , Eq. (120), and Δ_2 , Eq. (138), can be nonzero without contradicting the other requirements. In order to achieve effective breaking with $\Delta \neq 0$, the topological susceptibility must be proportional to m^2 (with a nonzero proportionality constant) in the chiral limit [see Eq. (121)], in which case one finds an ideal gas-like behavior for the topological charge [see Eq. (143) and Sec. VI], and sufficiently strong repulsion between near-zero modes to satisfy Eq. (134). A nonzero Δ_2 requires instead that correlations between zero and complex modes do not vanish too fast in the chiral limit [see Eq. (138)]. On the other hand, restoration of $U(1)_A$ requires $\Delta = \Delta_2 = 0$, and therefore not only that χ_π and χ_δ become equal at $\theta = 0$ in the chiral limit, but also that their sum is independent of θ up to $O(\theta^2)$ and $O(m^2)$ [see Eq. (142)].

D. Comparison with Ref. [78]

In Ref. [78] Aoki, Fukaya, and Taniguchi provided a detailed discussion of spectral constraints resulting from chiral symmetry restoration in the scalar and pseudoscalar sector. Their requirements for symmetry restoration, however, differ from the one used here, i.e., the finiteness of scalar and pseudoscalar susceptibilities. Since the latter is a necessary and sufficient condition for chiral symmetry restoration at the level of susceptibilities (see Sec. III), it should automatically imply that the symmetry-restoration conditions of Ref. [78] are satisfied, lest these are more restrictive than necessary. If so, the approach of Ref. [78] should not be able to provide more constraints on the spectrum than the ones obtained here. I now show that this is indeed the case.

The requirements of Ref. [78] for symmetry restoration are the following. (i.) For operators $\mathcal{O} = \prod_{i \in I_{\mathcal{O}}} O_i$, with $I_{\mathcal{O}} = \{1, \dots, N_{\mathcal{O}}\}$ and O_i chosen from the set $\{S, i\vec{P}, i\vec{P}, \vec{S}\}$, the (suitably normalized) expectation values $V_4^{-n_{\delta_{Aa}} \mathcal{O}} \langle \delta_{Aa} \mathcal{O} \rangle$ of their infinitesimal axial transformations $\delta_{Aa} \mathcal{O}$ [see Eq. (E9)] vanish in the chiral limit (taken after the thermodynamic limit). (ii.) Expectation values of m -independent observables that depend only on gauge fields are analytic functions of m^2 . As I show below, condition (i.) follows from finiteness of $\mathcal{C}_{u,w,\tilde{u}^2}^\infty$ in the chiral limit. Condition (ii.), relaxed to the relevant quantities being C^∞ functions of m^2 without major practical effects on the approach of Ref. [78], follows from the m^2 -differentiability properties proved in Secs. IIIB and IIIC, if also nonlocal restoration or restoration in external fields is required.

In condition (i.), $n_{\delta_{Aa}} \mathcal{O}$ are appropriate powers of the volume matching the leading volume dependence of

²⁰ An incomplete form of the first constraint in Eq. (133) appeared in Eq. (4.28) of Ref. [79]. The constraint Eq. (143) follows from the results of Ref. [109]; here it is obtained from first principles.

$\langle \delta_{Aa} \mathcal{O} \rangle$ as $V_4 \rightarrow \infty$. These powers are determined using the cluster property of correlation functions, encoded in the finiteness of $\frac{1}{V_4} \ln \mathcal{Z} = \mathcal{W}$ in the thermodynamic limit, which allows one to write for a generic observable $\mathcal{O}$ of the type above

$$\langle \mathcal{O} \rangle = V_4^{n_{\mathcal{O}}} \sum_{\pi \in \Pi(I_{\mathcal{O}})} \prod_{p \in \pi} \chi(\mathcal{O}(p)) + o(V_4^{n_{\mathcal{O}}}), \quad (144)$$

where the sum is over the partitions of $I_{\mathcal{O}}$ with the maximal amount $n_{\mathcal{O}}$ of parts p such that $\chi(\mathcal{O}(p)) \neq 0$, and $\mathcal{O}(p) = \prod_{i \in p} O_i$. These partitions involve only the “irreducible” correlation functions coinciding with their connected part, i.e., those for which $\langle \mathcal{O} \rangle = \langle \mathcal{O} \rangle_c$. By inspection of $\mathcal{W}$ as constrained by the symmetries of the theory, Eq. (50), and taking into account CP invariance, one sees that correlation functions not identically vanishing must be even under $P, \vec{P} \rightarrow -P, -\vec{P}$, as well as under $P_a, S_a \rightarrow -P_a, -S_a$ for each $a = 1, 2, 3$ separately. This allows one to obtain the following exhaustive list of irreducible correlation functions,

$$\begin{aligned} \langle S \rangle &= -2m\mathcal{C}_u^{\infty} V_4 + o(V_4), \\ \langle (iP_a)^2 \rangle &= 2\mathcal{C}_u^{\infty} V_4 + o(V_4), \\ \langle (iP)^2 \rangle &= (2\mathcal{C}_w^{\infty} + 8m^2\mathcal{C}_{\tilde{u}^2}^{\infty}) V_4 + o(V_4), \\ \langle (S_a)^2 \rangle &= 2\mathcal{C}_w^{\infty} V_4 + o(V_4), \\ \langle (iP_a)(iP)S_a \rangle &= -8m\mathcal{C}_{\tilde{u}^2}^{\infty} V_4 + o(V_4), \\ \langle (iP_a)(iP_b)S_aS_b \rangle &= 8\mathcal{C}_{\tilde{u}^2}^{\infty} V_4 + o(V_4) \quad (a \neq b), \end{aligned} \quad (145)$$

see Eqs. (97)–(100).²¹

Using the Ward-Takahashi identities Eq. (E13), condition (i.) is equivalent to requiring

$$\lim_{m \rightarrow 0} m \lim_{V_4 \rightarrow \infty} V_4^{-n_{P_a} \mathcal{O}} \langle P_a \mathcal{O} \rangle = 0, \quad (146)$$

with $n_{P_a} \mathcal{O} = n_{\delta_{Aa}} \mathcal{O}$. The cluster property allows one to write $V_4^{-n_{P_a} \mathcal{O}} \langle P_a \mathcal{O} \rangle$ as a product of scalar and pseudoscalar susceptibilities, that are necessarily finite in the symmetric phase, as shown above in Sec. IIIB, and so

²¹ Invariance of $\langle S^{n_S} (i\vec{P})^{\tilde{n}_P} (iP)^{n_P} \vec{S}^{\tilde{n}_S} \rangle$ under $P, \vec{P} \rightarrow -P, -\vec{P}$ and $P_a, S_a \rightarrow -P_a, -S_a$ requires

$$(-1)^{n_P + n_{S_a}} = 1, \quad \forall a, \quad (-1)^{n_P + \sum_{a=1}^3 n_{P_a}} = 1.$$

These conditions are solved by

$$\begin{aligned} n_P &= 2N_P + \tilde{n}_P, & n_{P_a} &= 2N_{P_a} + \tilde{o}_a, \\ n_{S_a} &= 2N_{S_a} + \tilde{o}_a, \end{aligned}$$

with N_P, N_{P_a} , and N_{S_a} non-negative integers, $\tilde{n}_P = 0, 1$, $\tilde{o}_a = 0, 1$, $a = 1, 2, 3$, and with $\tilde{n}_P + \sum_{a=1}^3 \tilde{o}_a = 0, 2, 4$. The last condition has the solutions $\tilde{n}_P = 0, \tilde{o}_a = 0 \forall a$, leading to the first four irreducible correlation functions in Eq. (145); $\tilde{n}_P = 1, \sum_{a=1}^3 \tilde{o}_a = 1$, leading to the fifth one; $\tilde{n}_P = 0, \sum_{a=1}^3 \tilde{o}_a = 2$, leading to the sixth one; and $\tilde{n}_P = 1, \sum_{a=1}^3 \tilde{o}_a = 3$, which does not lead to new irreducible correlation functions.

also this condition is automatically satisfied. Again, only the irreducible correlation functions of Eq. (145) appear in this quantity, to leading order in the volume. The approach of Ref. [78] cannot then yield more constraints on the Dirac spectrum than those following from the finiteness and m^2 -differentiability of $\mathcal{C}_{u,w,\tilde{u}^2}^{\infty}$ (i.e., finiteness of $\mathcal{A}_{n_u n_w n_{\tilde{u}}}$ with arbitrary $n_u, n_w = 0, 1$, and $n_{\tilde{u}} = 0, 2$), which is only a subset of the constraints following from finiteness of all the susceptibilities, and is therefore a less general approach than the one used here. In particular, it can lead only to the same direct constraints Eqs. (117) and (119) on the spectral density obtained in this paper, that follow from finiteness of $\mathcal{C}_{u,w,\tilde{u}^2}^{\infty}$ in the chiral limit.

VI. IDEAL INSTANTON GAS-LIKE BEHAVIOR

Under the assumption that the free energy density at finite θ angle is analytic in m^2 in the symmetric phase, and that $U(1)_A$ remains effectively broken in the chiral limit (which throughout this section will be synonymous with $\Delta \neq 0$), Ref. [109] showed that to lowest order in the fermion mass, the cumulants of the topological charge are the same found for an ideal gas of instantons and anti-instantons, with identical and vanishingly small densities $\chi_t/2 \propto m^2$. Here I rederive this conclusion directly from the assumption of chiral symmetry restoration at the level of susceptibilities. As shown in Sec. IIIB, this necessarily leads to m^2 -differentiability (rather than analyticity) of the free energy density and of scalar and pseudoscalar susceptibilities at $\theta = 0$; I show below that this also implies the m^2 -differentiability of the derivatives of the free energy density with respect to θ at $\theta = 0$, i.e., of the cumulants of Q , as already argued [see under Eq. (44)].

Consider the usual partition function in the presence of a topological term,

$$Z(\theta; m) \equiv \int DU e^{-S_{\text{eff}}(U) + i\theta Q(U)} \int D\Psi D\bar{\Psi} e^{-\bar{\Psi} D_m(U) \Psi}. \quad (147)$$

Using the transformation properties of the action and of the measure under the $U(1)_A$, flavor-singlet axial transformation $\mathcal{U}_A^{(0)}(\frac{\theta}{4}) = \mathcal{U}^{(0)}(-\frac{\theta}{4}, \frac{\theta}{4})$ [see Eqs. (7), (8), and (12)], one readily finds the identity

$$Z(\theta; m) = \tilde{\mathcal{Z}}(j_S(\theta; m), j_P(\theta; m); m), \quad (148)$$

with $\tilde{\mathcal{Z}}(j_S, j_P; m) \equiv \mathcal{Z}(V, W; m)|_{\vec{j}_P, S=0}$ the generating function Eq. (16) for vanishing isotriplet sources, and

$$j_S(\theta; m) \equiv m(\cos \frac{\theta}{2} - 1), \quad j_P(\theta; m) \equiv m \sin \frac{\theta}{2}, \quad (149)$$

are m - and θ -dependent isoscalar sources.²² The free energy density in the presence of a topological term is

²² Setting $\theta = 2\pi$ in Eqs. (147) and (148) one finds the identity $Z(0; m) = Z(2\pi; m) = Z(0; -m)$, that implies an additional non-anomalous discrete $\mathbb{Z}_{2A}$ symmetry in the chiral limit.

then related to $\tilde{\mathcal{W}}_\infty \equiv \mathcal{W}_\infty|_{j_P, s=\vec{0}}$ as

$$\begin{aligned} F(\theta; m) &\equiv - \lim_{V_4 \rightarrow \infty} \frac{1}{V_4} \ln Z(\theta; m) \\ &= -\tilde{\mathcal{W}}_\infty(j_S(\theta; m), j_P(\theta; m); m). \end{aligned} \quad (150)$$

Since $j_S(\theta; m) = O(\theta^2)$ and $j_P(\theta; m) = O(\theta)$ for small θ , the expansion of the right-hand side of Eq. (150) in powers of the sources corresponds to an expansion for small θ . More precisely, θ derivatives of F at $\theta = 0$ equal finite linear combinations of derivatives of $\tilde{\mathcal{W}}_\infty$ with respect to scalar and pseudoscalar isosinglet sources at zero sources (i.e., of isosinglet susceptibilities), with coefficients the θ -derivatives of powers of $j_{S,P}(\theta; m)$ at $\theta = 0$.²³ Since terms odd in j_P vanish thanks to CP invariance of the theory at $\theta = 0$ and $m \neq 0$, and since $F(0; m) = -\tilde{\mathcal{W}}_\infty(0, 0; m) = -\mathcal{W}_\infty(0, 0, m)$, one finds

$$\begin{aligned} F(\theta; m) &= F(0; m) - j_S(\theta; m) \partial_{j_S} \mathcal{W}_\infty(m)|_0 \\ &\quad - \frac{1}{2} j_P(\theta; m)^2 \partial_{j_P}^2 \mathcal{W}_\infty(m)|_0 \\ &\quad - \frac{1}{2} j_S(\theta; m)^2 \partial_{j_S}^2 \mathcal{W}_\infty(m)|_0 \\ &\quad - \frac{1}{2} j_S(\theta; m) j_P(\theta; m)^2 \partial_{j_S} \partial_{j_P}^2 \mathcal{W}_\infty(m)|_0 \\ &\quad - \frac{1}{24} j_P(\theta; m)^4 \partial_{j_P}^4 \mathcal{W}_\infty(m)|_0 + O(\theta^6), \end{aligned} \quad (151)$$

where the mass dependence of the derivatives of $\mathcal{W}_\infty$ at zero sources is shown explicitly. This expression holds independently of the fate of chiral symmetry in the chiral limit, and provides the correct mass dependence of the θ -derivatives of $F(\theta; m)$ and of its mass-derivatives evaluated at $\theta = 0$, including in the chiral limit (i.e., with the chiral limit taken after setting $\theta = 0$). In particular, the topological susceptibility $\chi_t = \partial_\theta^2 F(\theta; m)|_{\theta=0}$ reads

$$\chi_t = \frac{1}{4} (m\Sigma - m^2\chi_\eta) = \frac{m^2}{4} (\chi_\pi - \chi_\eta), \quad (152)$$

which is a well-known integrated Ward-Takahashi identity for the anomalous $U(1)_A$ symmetry [132]. Together with Eq. (100), this relation implies Eq. (101).

Using instead the functional form Eq. (54) for the generating function $\mathcal{W}_\infty$ one finds

$$F(\theta; m) = -\hat{\mathcal{W}}_\infty(m^2 + u(\theta; m), w(\theta; m), \tilde{u}(\theta; m)), \quad (153)$$

where

$$\begin{aligned} u(\theta; m) &= -w(\theta; m) \equiv -m^2 \left(\sin \frac{\theta}{2}\right)^2, \\ \tilde{u}(\theta; m) &\equiv m^2 \sin \theta. \end{aligned} \quad (154)$$

Independently of the fate of chiral symmetry as $m \rightarrow 0$, one can expand $\tilde{\mathcal{W}}_\infty$ around zero sources in powers of $u(\theta; m)$, $w(\theta; m)$, and $\tilde{u}(\theta; m)$, or rather $\tilde{u}(\theta; m)^2 = 4m^4 \left(\sin \frac{\theta}{2}\right)^2 [1 - \left(\sin \frac{\theta}{2}\right)^2]$ thanks to CP invariance, finding [see Eqs. (95) and (96)]

$$\begin{aligned} F(\theta; m) - F(0; m) &= m^2 \left(\sin \frac{\theta}{2}\right)^2 (\mathcal{C}_u^\infty(m^2) - \mathcal{C}_w^\infty(m^2) - 4m^2 \mathcal{C}_{\tilde{u}^2}^\infty(m^2)) \\ &\quad - \frac{m^4}{2} \left(\sin \frac{\theta}{2}\right)^4 [\mathcal{C}_{uu}^\infty(m^2) - 2\mathcal{C}_{uw}^\infty(m^2) + \mathcal{C}_{ww}^\infty(m^2) \\ &\quad - 8\mathcal{C}_{\tilde{u}^2}^\infty(m^2) - 8m^2 (\mathcal{C}_{u\tilde{u}^2}^\infty(m^2) - \mathcal{C}_{w\tilde{u}^2}^\infty(m^2)) \\ &\quad + 16m^4 \mathcal{C}_{\tilde{u}^2\tilde{u}^2}^\infty(m^2)] + O\left(\left(\sin \frac{\theta}{2}\right)^6\right). \end{aligned} \quad (155)$$

This is generally not a systematic expansion in powers of $m^{2n} \left(\sin \frac{\theta}{2}\right)^2$, with $n \geq 1$: in the broken phase $\mathcal{C}_u^\infty = \frac{\chi_\pi}{2m} = \frac{\Sigma}{2m}$ diverges like $1/m$ in the chiral limit, implying also that $(m^2)^n \partial_u^n \mathcal{C}_u^\infty|_0 \sim m^{-1}$ for all $n \geq 1$. Identifying the coefficient of a given power of m requires then resumming infinitely many terms.

In the symmetric phase, on the other hand, the coefficients of the expansion Eq. (155) are finite, and actually m^2 -differentiable in the chiral limit, so F can be truly expanded in powers of $m^2 \left(\sin \frac{\theta}{2}\right)^2$ [or equivalently in powers of $m^2 \cos \theta$, or in a Fourier series in $m^{2n} \cos(n\theta)$], with $O(1)$ coefficients that can be further expanded in powers of m^2 around zero (possibly with zero radius of convergence, and up to a function vanishing at $m = 0$ with all its derivatives). This shows that the θ -derivatives of the free energy are m^2 -differentiable. In particular, the omitted terms in Eq. (155) are then both $O(\theta^6)$ and $O(m^6)$. Turning the argument around, one has that $F(\theta; m)$ can be expanded in powers of m^2 , with coefficients that are finite polynomials in $\cos \theta$, and with the expansion being valid for arbitrary θ .

To lowest order in m^2 , one finds in the symmetric phase

$$\begin{aligned} F(\theta; m) - F(0; m) &= (1 - \cos \theta) \frac{m^2}{2} (\mathcal{C}_u^\infty(0) - \mathcal{C}_w^\infty(0)) + O(m^4) \\ &= (1 - \cos \theta) m^2 \Delta + O(m^4) \\ &= (1 - \cos \theta) \chi_t + O(m^4), \end{aligned} \quad (156)$$

to all orders in θ , and having used the constraint Eq. (121) in the last passage. If $\Delta \neq 0$, this is the free energy density of an ideal gas of instantons and anti-instantons of equal densities $\chi_t/2 = m^2 \Delta/2 + O(m^4)$. One concludes that an ideal instanton gas-like behavior of the topological charge distribution in the chiral limit is a necessary condition for chiral symmetry restoration if $U(1)_A$ remains effectively broken.

To find the corrections to the ideal gas behavior, one expands Eq. (155) up to order $O(\theta^4)$, obtaining

$$F(\theta; m) - F(0; m) = \frac{\theta^2}{2} \chi_t(m^2) - \frac{\theta^4}{24} \kappa_{4t}(m^2) + O(m^6 \theta^6), \quad (157)$$

²³ Since the expansion in powers of the sources is formal, so is the expansion in θ , that may have zero radius of convergence, and may miss terms vanishing at zero with all their derivatives. This does not affect the correctness of the result for the θ derivatives of F at $\theta = 0$ [as long as they commute with the thermodynamic limit – see discussion after Eq. (19) and footnote 6].

where to all orders in m^2 [see also Eqs. (101) and (131)]

$$\begin{aligned} \frac{\chi_t(m^2)}{m^2} &= \frac{1}{2} (\mathcal{C}_u^\infty(m^2) - \mathcal{C}_w^\infty(m^2)) - 2m^2 \mathcal{C}_{\bar{u}^2}^\infty(m^2), \\ \frac{\kappa_{4t}(m^2)}{m^4} &= \frac{\chi_t(m^2)}{m^4} + \frac{3}{4} [\mathcal{C}_{uu}^\infty(m^2) - 2\mathcal{C}_{uw}^\infty(m^2) \\ &\quad + \mathcal{C}_{ww}^\infty(m^2) - 8\mathcal{C}_{\bar{u}^2}^\infty(m^2) \\ &\quad - 8m^2 (\mathcal{C}_{u\bar{u}^2}^\infty(m^2) - \mathcal{C}_{w\bar{u}^2}^\infty(m^2)) \\ &\quad + 16m^4 \mathcal{C}_{\bar{u}^2\bar{u}^2}^\infty(m^2)] . \end{aligned} \quad (158)$$

One finds for the coefficient b_2 [133] [see Eq. (143)]

$$\begin{aligned} b_2 &\equiv -\frac{1}{12} \frac{\kappa_{4t}(m^2)}{\chi_t(m^2)} \\ &= -\frac{1}{12} \left[1 + \frac{3m^2}{\Delta + O(m^2)} \left(\partial_{m^2} \frac{\chi_t}{m^2} \Big|_{m=0} - \Delta_2 \right) \right. \\ &\quad \left. + O\left(\frac{m^4}{\Delta + O(m^2)} \right) \right] . \end{aligned} \quad (159)$$

If $\Delta \neq 0$, to leading order one finds the ideal-gas result $b_2 = -\frac{1}{12}$. If $\Delta = 0$ the second term in square brackets is generally $O(1)$ and one does not find an ideal-gas behavior.

Since QCD at the physical point is rather close to the chiral limit, the results above lead one to expect the onset of an instanton gas-like behavior for the topological charge not far above the chiral crossover temperature, T_c , similarly to what has been observed in the pure gauge case [134], if $U(1)_A$ remains effectively broken. The results of Ref. [14], however, indicate the persistence of large deviations of b_2 from $-1/12$ up to $T \sim 2T_c$ on fine lattices. It is possible, of course, that the large deviation is actually physical, which would indicate that either $U(1)_A$ is effectively restored in the chiral limit, or that, although $U(1)_A$ is effectively broken, the coefficient of the $O(m^2)$ correction is large. On the other hand, it is known that taking the continuum limit of topology-related observables is difficult, and algorithmic improvements have led to significant revisions of some of the results of Ref. [14] (see Refs. [19, 135, 136]), so it seems safe to say that the situation is not yet settled.

To avoid misunderstandings, and as already pointed out in Ref. [109], it is worth stressing that the analysis above by no means imply that if $U(1)_A$ is effectively broken in the symmetric phase, then the relevant topological degrees of freedom in the chiral limit are the usual instantons and anti-instantons, or more precisely their finite-temperature analogs, i.e., calorons and anti-calorons [137–149]. This also means that the required instanton gas-like behavior need not be the same behavior found in the usual semiclassical dilute instanton gas [150, 151]. What is required is that the topological properties of the system can be described (at least formally) in terms of effective degrees of freedom corresponding to objects carrying unit topological charge, of vanishingly small density, and fluctuating independently of each other.

If localized objects of integer charge (of some sort) were indeed the relevant topological degrees of freedom, thanks to the index theorem they would support exact chiral zero modes if isolated from each other. In the random matrix model for the low-lying Dirac spectrum of Ref. [110], based on a dilute instanton gas interacting via the fermionic determinant, the mixing of these modes leads to a singular peak in the spectral density. This, in turn, affects the instanton density through the fermionic determinant leading to $\chi_t \propto m^2$. Both these effects lead to effective $U(1)_A$ breaking in the chiral limit, fulfilling the constraint Eq. (122) in a nontrivial way. The presence of a dilute gas of instanton-like objects in typical gauge configurations is then likely to be a sufficient condition for effective $U(1)_A$ breaking in the symmetric phase.

For completeness, I conclude this section discussing briefly the spontaneously broken phase. In this case it is convenient to stick to the expansion of $F(\theta; m)$ in powers of $j_{S,P}(\theta; m)$, Eq. (151). This expression contains only susceptibilities involving the scalar and pseudoscalar singlet operators, that are expected to remain finite in the chiral limit since the corresponding particles (i.e., σ and η) should remain massive. To leading order in m one finds then for small θ

$$\begin{aligned} F(\theta; m) - F(0; m) &= -j_S(\theta; m) \partial_{j_S} \mathcal{W}_\infty(m)|_0 + o(m) \\ &= |m| \left(1 - \cos \frac{\theta}{2} \right) \Sigma(0^+) + o(m) , \end{aligned} \quad (160)$$

that provides the correct expression for the leading behavior in the chiral limit of all the θ derivatives of $F(\theta; m)$ at $\theta = 0$. One has then $\chi_t = \frac{1}{4}|m|\Sigma(0^+) + o(m)$ and $b_2 = -1/48 + o(m)$, up to corrections that vanish in the chiral limit, matching the expectations obtained using chiral Lagrangians [85, 152]. Amusingly, Eq. (160) is equal to the free energy density of an ideal gas of topological objects of charge $\pm \frac{1}{2}$, of equal densities $\frac{1}{2}|m|\Sigma(0^+)$. This effective description, however, holds only in the vicinity of $\theta = 0$, as Eq. (160) has no reason to be valid for larger θ , and even lacks the required 2π -periodicity. Enforcing this property one finds $F(\theta; m) - F(0; m) = |m| \left(1 - |\cos \frac{\theta}{2}| \right) \Sigma(0^+)$ to leading order in m , in agreement with the analysis of Refs. [85, 152], although the present derivation guarantees the correctness of this expression only in the vicinity of $\theta = 0 \pmod{2\pi}$.

VII. CONCLUSIONS

The nature of the chiral transition in the $N_f = 2$ chiral limit of a gauge theory and the fate of the anomalous $U(1)_A$ symmetry in the symmetric phase are still open problems in spite of extensive investigations, both analytical [6–9, 13, 20–30, 74–83, 93–99, 109] and numerical [4, 31–44, 100–108, 110]. In this paper I have revisited the first-principles approach to this problem based on the study of the Dirac spectrum [77–82] using lattice gauge theory and chiral (Ginsparg–Wilson) fermions [53–73, 111, 112, 114, 116, 121, 130, 132], putting its founda-

tions on solid ground and developing it in full generality. The main results are the following.

(1.) I have proved that chiral symmetry is restored at the level of scalar and pseudoscalar susceptibilities if and only if these are finite (i.e., non-divergent) in the chiral limit; or equivalently, if and only if susceptibilities involving an even number of isosinglet scalar and pseudoscalar bilinears are m^2 -differentiable (i.e., functions of m^2 infinitely differentiable at $m = 0$), and m times an m^2 -differentiable function if this number is odd (Sec. III B). A symmetry being manifest at the level of susceptibilities is a general property of a quantum field theory within the symmetric phase, where the correlation length is finite, and so the finiteness of scalar and pseudoscalar susceptibilities in the chiral limit is a general property of the chirally symmetric phase of a gauge theory.

(2.) Under the extended assumption that chiral symmetry is restored in susceptibilities involving scalar and pseudoscalar bilinears and general, possibly nonlocal operators containing only gauge fields (nonlocal restoration), I have proved that also the spectral density of the Dirac operator and similar spectral quantities are m^2 -differentiable in the symmetric phase (Sec. III C). This follows also if chiral symmetry is restored in scalar and pseudoscalar susceptibilities involving additional bilinears of external fermion fields (Appendix C). These extended assumptions are still based on the essential features of symmetry restoration in a quantum field theory. Together with those in (1.), these results essentially turn the analyticity assumptions of Refs. [77–80, 109] into a necessary consequence of symmetry restoration.

(3.) After making simplifying assumptions on the Ginsparg–Wilson–Dirac operator that are met by its most common realizations, I have obtained an explicit expression for the generating function of scalar and pseudoscalar susceptibilities in terms of the Dirac eigenvalues (Sec. IV). Imposing finiteness of these susceptibilities I have then obtained a set of constraints on the Dirac spectrum (Sec. V). In particular, I have shown that the only constraints involving the spectral density directly are those coming from finiteness in the chiral limit of the pion susceptibility, χ_π , that implies also finiteness of the delta susceptibility, χ_δ ; and from $\frac{\chi_\pi - \chi_\delta}{4} - \frac{\chi_t}{m^2} = O(m^2)$, with χ_t the topological susceptibility, implying $\Delta = \lim_{m \rightarrow 0} \frac{\chi_\pi - \chi_\delta}{4} = \lim_{m \rightarrow 0} \frac{\chi_t}{m^2}$ (Sec. V A). These constraints are generally compatible with both effective breaking and effective restoration of $U(1)_A$. Moreover, I have proved that $\frac{\chi_t}{m^2}$ is m^2 -differentiable in the symmetric phase, and obtained a lower bound on $\chi_\pi - \chi_\delta$ that shows the impossibility of effective $U(1)_A$ restoration at nonzero m (Sec. V A).

(4.) I have also obtained further constraints involving two-point eigenvalue correlation functions, showing that the correlations among near-zero complex modes, and between zero and complex modes, are closely connected with topology and the fate of $U(1)_A$. In particular, an order parameter for $U(1)_A$ is related to the second derivative of $\chi_\pi + \chi_\delta$ with respect to the “vacuum angle” θ at

$\theta = 0$, which in turn is determined by the correlation between zero and near-zero modes (Sec. V B).

(5.) I have shown that in the chiral limit the cumulants of the topological charge must be identical to those found in an ideal gas of instantons and anti-instantons of vanishingly small total density $\chi_t \propto m^2$, to leading order in m , if $U(1)_A$ remains effectively broken by $\Delta \neq 0$ (Sec. VI).

The results in (1.) and (2.) are of very general nature, based only on the properties expected of the symmetric phase of a quantum field theory, and on the symmetry properties of a gauge theory with two degenerate fermions. These results allow one to use the Dirac spectrum to study chiral symmetry restoration and the fate of $U(1)_A$ in a systematic and truly first-principles way. They also subsume and extend the approach of Ref. [78], allowing one to obtain information not only on the spectral density, but also on the correlations among eigenvalues (Sec. V D).

The results discussed in (3.) and (4.) are also very general. In particular, the representation of the generating function in terms of spectral quantities applies independently of the status of chiral symmetry. When translating the finiteness of susceptibilities in the symmetric phase into constraints on the Dirac spectrum, no assumption is made on how the various spectral quantities depend on the position in the spectrum or on m ; in particular, the assumption of nonlocal restoration is not used.

Finally, (5.) is completely general, requiring only the finiteness condition (1.) on susceptibilities and the known (anomalous) symmetry properties of gauge theories under $U(1)_A$ transformations, besides the effective breaking of $U(1)_A$ by $\Delta \neq 0$. This result was already proved in Ref. [109] using an effective-theory approach under the assumption of m^2 -analyticity of the free energy density in the presence of a θ term. Here the assumptions of Ref. [109] are put on firmer ground, and the emergence of an ideal instanton-gas behavior in the chiral limit if $U(1)_A$ remains effectively broken is shown rigorously, allowing also the study of corrections to the ideal gas behavior. Also in this case the assumption of nonlocal restoration is not needed.

The results of this paper set the stage for a detailed study of the Dirac spectrum in the symmetric phase of a gauge theory, once that more detailed properties of the spectral density and eigenvalue correlators are taken into account. This will be the purpose of the second paper of this series.

ACKNOWLEDGMENTS

I thank V. Azcoiti, C. Bonanno, G. Endrődi, I. Horváth, S. D. Katz, D. Nógrádi, A. Patella, A. Pásztor, Zs. Szép, and especially T. G. Kovács for discussions. This work was partially supported by the NKFIH grants K-147396, NKKP Excellence 151482, and TKP2021-NKTA-64.

Appendix A: Connected correlation functions

1. Recursive formula

Consider families of observables of the general form $A_{\vec{s}}^{(k)}(\Lambda_k)$, labeled by discrete indices $\vec{s} = (s_1, \dots, s_d) \in \mathbb{N}_0^d$ and $k \in \mathbb{N}_0$, and by continuous variables $\Lambda_k = \{\lambda_1, \dots, \lambda_k\}$, $\lambda_j \in I \subseteq \mathbb{R}$. Denote $\vec{k} \equiv (\vec{s}, k)$. Connected correlation functions for these observables are defined recursively from the correlation functions $\langle A_{\vec{s}}^{(k)}(\Lambda_k) \rangle$ via

$$\begin{aligned} \langle A_{\vec{s}}^{(k)}(\Lambda_k) \rangle_c &\equiv \langle A_{\vec{s}}^{(k)}(\Lambda_k) \rangle \\ &- \sum_{\substack{\pi \in \Pi(S_{\vec{k}}) \\ |\pi| > 1}} \prod_{p \in \pi} \langle A_{\vec{s}(p)}^{(k(p))}(\Lambda_{k(p)}(p)) \rangle_c, \end{aligned} \quad (\text{A1})$$

where the sum runs over the partitions $\Pi(S_{\vec{k}})$ of a set $S_{\vec{k}}$ of $s_1, s_2, \dots, s_d, k$ elements of type $1, 2, \dots, d, d+1$ into parts, p (i.e., disjoint subsets whose union is the whole set), containing $s(p)_j$ elements of type $j = 1, \dots, d$, and the $k(p)$ elements of type $d+1$ in the subset $\Lambda_{k(p)}(p) \subseteq \Lambda_k$. Only nontrivial partitions are included, i.e., the number of parts, $|\pi|$, obeys $|\pi| > 1$. Clearly $\sum_{p \in \pi} \vec{s}(p) = \vec{s}$, $\sum_{p \in \pi} k(p) = k$, and $\cup_{p \in \pi} \Lambda_{k(p)}(p) = \Lambda_k$. Integrating over Λ_k ,

$$A_{\vec{k}}[g] \equiv \int_I d\lambda_1 g(\lambda_1) \dots \int_I d\lambda_k g(\lambda_k) A_{\vec{s}}^{(k)}(\Lambda_k), \quad (\text{A2})$$

with g continuous (and integrable over I), one obtains quantities of the same general form, i.e., $A_{\vec{k}}[g] = A_{\vec{s}'}^{(k')}$ with $\vec{s}' = \vec{k}$ and $k' = 0$ (and with trivial dependence on Λ_k). The corresponding connected correlation functions

$$\langle A_{\vec{k}}[g] \rangle_c \equiv \langle A_{\vec{k}}[g] \rangle - \sum_{\substack{\pi \in \Pi(S_{\vec{k}}) \\ |\pi| > 1}} \prod_{p \in \pi} \langle A_{\vec{k}(p)}[g] \rangle_c, \quad (\text{A3})$$

defined according to the general rule, Eq. (A1), equal for $k \geq 1$

$$\langle A_{\vec{k}}[g] \rangle_c = \int_I d\lambda_1 g(\lambda_1) \dots \int_I d\lambda_k g(\lambda_k) \langle A_{\vec{s}}^{(k)}(\Lambda_k) \rangle_c. \quad (\text{A4})$$

The proof by induction is straightforward.

2. Partition function

The partition function associated with the correlation functions $\langle A_{\vec{k}}[g] \rangle$ is

$$Z(t) = 1 + \sum_{\vec{k} \neq \vec{0}} \left(\prod_{j=1}^{d+1} \frac{t_j^{k_j}}{k_j!} \right) \langle A_{\vec{k}}[g] \rangle, \quad (\text{A5})$$

where t denotes collectively the complex variables $t_1, \dots, t_{d+1}$. Using the decomposition in connected components,

$$Z(t) = 1 + \sum_{\vec{k} \neq \vec{0}} \left(\prod_{j=1}^{d+1} \frac{t_j^{k_j}}{k_j!} \right) \sum_{\pi \in \Pi(S_{\vec{k}})} \prod_{p \in \pi} \langle A_{\vec{k}(p)}[g] \rangle_c; \quad (\text{A6})$$

taking into account that contributions to Z are entirely characterized by an integer-valued, non-identically vanishing function $m(\vec{\sigma}) \in \mathbb{N}_0$, $m(\vec{\sigma}) \neq 0$, counting the number of parts labeled by $\vec{\sigma} \neq \vec{0}$; and taking into account that there are $\prod_{j=1}^{d+1} k_j! / \left[\prod_{\vec{\sigma} \neq \vec{0}} m(\vec{\sigma})! \left(\prod_{j=1}^{d+1} \sigma_j! \right)^{m(\vec{\sigma})} \right]$ partitions giving the same contribution, one finds

$$Z(t) = \exp \left\{ \sum_{\vec{\sigma} \neq \vec{0}} \left(\prod_{j=1}^{d+1} \frac{t_j^{\sigma_j}}{\sigma_j!} \right) \langle A_{\vec{\sigma}}[g] \rangle_c \right\}. \quad (\text{A7})$$

The connected correlation functions $\langle A_{\vec{k}}[g] \rangle_c$ are then obtained by differentiation with respect to the t_j at $t_1 = \dots = t_{d+1} = 0$, and $\langle A_{\vec{s}}^{(k)}(\Lambda_k) \rangle_c$ are further obtained by functional differentiation of these quantities with respect to $g(\lambda_1), \dots, g(\lambda_k)$.

3. Relations between cumulants

For the observables

$$\begin{aligned} A_{\vec{s}}^{(k)}(\Lambda_k) &= \left[\prod_{i=1}^d O_i^{s_i} \right] \gamma^{(k)}(\Lambda_k), \\ A_{\vec{s}}'^{(k)}(\Lambda_k) &= \left[\prod_{i=1}^d O_i'^{s_i} \right] \gamma^{(k)}(\Lambda_k), \end{aligned} \quad (\text{A8})$$

with $O_i = \sum_{j=1}^d \mathcal{C}_{ij} O_j'$, one has $A_{\vec{s}}^{(k)} = \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} A_{\vec{s}'}'^{(k)}$ for suitable $C_{\vec{s}\vec{s}'}$. One has for the corresponding partition functions

$$\begin{aligned} Z[t; g] &= \sum_{K, k=0}^{\infty} \frac{1}{K!k!} \left\langle \left(\sum_{i=1}^d t_i O_i \right)^K \Gamma^{(k)}[g] \right\rangle \\ &= \sum_{K, k=0}^{\infty} \frac{1}{K!k!} \left\langle \left(\sum_{i=1}^d t_i' O_i' \right)^K \Gamma^{(k)}[g] \right\rangle \\ &= Z'[t'; g], \end{aligned} \quad (\text{A9})$$

with $t_i' = \sum_{j=1}^d \mathcal{C}_{ji} t_j$, $1 \leq i \leq d$, and

$$\Gamma^{(k)}[g] \equiv \int_I d\lambda_1 g(\lambda_1) \dots \int_I d\lambda_k g(\lambda_k) \gamma^{(k)}(\lambda_1, \dots, \lambda_k). \quad (\text{A10})$$

(Here I set $t_{d+1} = t_{d+1}' = 1$ without loss of generality.) Full correlation functions for the two sets of quantities are obtained by applying the differential operator $D_{\vec{s}} \equiv \prod_{i=1}^d \partial_{t_i}^{s_i}$ or $D_{\vec{s}}' \equiv \prod_{i=1}^d \partial_{t_i'}^{s_i}$ to Z or Z' and

then setting $t_i = 0$ or $t'_i = 0$, $\forall i$, respectively. Connected correlation functions are similarly obtained by applying the same operators to $\ln Z$ or $\ln Z'$. Since $D_{\vec{s}} = \prod_{i=1}^d \left(\sum_{j=1}^d C_{ji} \partial_{t'_j} \right)^{s_i} = \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} D'_{\vec{s}'}$, it follows that the same linear relation holds between the connected correlation functions of $A_{(\vec{s},k)}[g] = \left[\prod_{i=1}^d O_i^{s_i} \right] \Gamma^{(k)}[g]$ and $A'_{(\vec{s},k)}[g] = \left[\prod_{i=1}^d O_i^{s_i} \right] \Gamma^{(k)}[g]$ as between their full correlation functions and between the observables themselves,

$$\begin{aligned} \langle A_{(\vec{s},k)}[g] \rangle &= \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} \langle A'_{(\vec{s}',k)}[g] \rangle, \\ \langle A_{(\vec{s},k)}[g] \rangle_c &= \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} \langle A'_{(\vec{s}',k)}[g] \rangle_c. \end{aligned} \quad (\text{A11})$$

By functional differentiation one sees that the same holds for the correlation functions of $A_{\vec{s}}^{(k)}(\Lambda_k)$ and $A'_{\vec{s}}^{(k)}(\Lambda_k)$.

If instead the following relation holds between the integrated quantities $A_{(\vec{s},k)}[g]$ and $A'_{(\vec{s},k)}[g]$,

$$\begin{aligned} A_{(\vec{s},k)}[g] &= \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} A'_{(\vec{s}',k)}[g], \\ C_{\vec{s}\vec{s}'} &= \prod_{i=1}^d c_{ss'}^{(i)}, \quad c_{ss'}^{(i)} = \left(\frac{d}{dx} \right)^s \frac{\omega_i(x)^{s'}}{s'!} \Big|_{x=0}, \end{aligned} \quad (\text{A12})$$

for some functions $\omega_i(x)$ with $\omega_i(0) = 0$, $i = 1, \dots, d$, then

$$\begin{aligned} Z[t; g] &= \sum_{k=0}^{\infty} \sum_{\vec{s}} \left(\prod_{j=1}^d \frac{t_j^{s_j}}{s_j!} \right) \langle A_{(\vec{s},k)}[g] \rangle \\ &= \sum_{k=0}^{\infty} \sum_{\vec{s}'} \left(\prod_{j=1}^d \frac{\omega_j(t_j)^{s'_j}}{s'_j!} \right) \langle A'_{(\vec{s}',k)}[g] \rangle \\ &= Z'[\omega(t); g], \end{aligned} \quad (\text{A13})$$

and so one finds for any function $\Omega(Z)$

$$\begin{aligned} D_{\vec{s}} \Omega(Z[t; g])|_{t=0} &= D_{\vec{s}} \Omega(Z'[\omega(t); g])|_{t=0} \\ &= \sum_{\vec{s}'} D_{\vec{s}} \left(\prod_{j=1}^d \frac{\omega_j(t_j)^{s'_j}}{s'_j!} \right) \Big|_{t=0} D'_{\vec{s}'} \Omega(Z'[t'; g])|_{t'=0} \\ &= \sum_{\vec{s}'} C_{\vec{s}\vec{s}'} D'_{\vec{s}'} \Omega(Z'[t'; g])|_{t'=0}, \end{aligned} \quad (\text{A14})$$

with $D_{\vec{s}}$ and $D'_{\vec{s}}$ defined as above, so that both full and connected correlation functions of $A_{(\vec{s},k)}$ and $A'_{(\vec{s}',k)}$ are linearly related by the same $C_{\vec{s}\vec{s}'}$, Eq. (A12). By functional differentiation, one shows that the same holds for $A_{\vec{s}}^{(k)}(\Lambda_k)$ and $A'_{\vec{s}}^{(k)}(\Lambda_k)$ and their full and connected correlation functions.

4. Relevant observables

The correlation functions considered in this paper involve observables of the form [see Eqs. (9) and (29), and under Eq. (17) for notation]

$$A_{\vec{s}}^{(0)} = S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S}, \quad (\text{A15})$$

with $\vec{s} = (n_S, \vec{n}_P, n_P, \vec{n}_S)$, where additional functionals G_i of gauge fields only may be included; and observables of the form

$$A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) = N_+^{n_+} N_-^{n_-} \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \quad (\text{A16})$$

with $\vec{s} = (n_+, n_-)$, including the case $k = 0$ where $\rho_U^{(0)} \equiv 1$, or their linear combinations

$$A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) = N_0^{n_0} Q^{n_1} \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \quad (\text{A17})$$

with $\vec{s} = (n_0, n_1)$, or the nonlinear combinations

$$A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) = s_{n_1}(N_+) s_{n_2}(N_-) \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \quad (\text{A18})$$

with $\vec{s} = (n_1, n_2)$ and $s_n(t)$ defined in Eq. (90), and the integrals of the quantities in Eqs. (A16)–(A18) over $\lambda_1, \dots, \lambda_k$. Finally, in the discussion in Appendix C, observables of the following more general form are involved, namely

$$\begin{aligned} A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) &= S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \\ &\quad \times N_+^{n_+} N_-^{n_-} \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \end{aligned} \quad (\text{A19})$$

with $\vec{s} = (n_S, \vec{n}_P, n_P, \vec{n}_S, n_+, n_-)$, or their linear combinations

$$\begin{aligned} A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) &= S^{n_S} (i\vec{P})^{\vec{n}_P} (iP)^{n_P} \vec{S}^{\vec{n}_S} \\ &\quad \times N_0^{n_0} Q^{n_1} \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \end{aligned} \quad (\text{A20})$$

with $\vec{s} = (n_S, \vec{n}_P, n_P, \vec{n}_S, n_0, n_1)$. These observables are precisely of the form discussed above in Appendix A 1, and the scalar and pseudoscalar connected correlation functions, Eq. (18), and the spectral correlators in Eqs. (30) and (31) for $O = N_+^{n_+} N_-^{n_-}$, $O = N_0^{n_0} Q^{n_1}$, or $O = s_{n_1}(N_+) s_{n_2}(N_-)$, are then defined by Eq. (A1). Moreover, full and connected correlation functions of the quantities in Eqs. (A16) and (A17) [and those of the quantities in Eqs. (A19) and (A20)] are related by the same linear transformation. The same applies to the correlation functions of their integrals over $\lambda_1, \dots, \lambda_k$, $I_{N_+^{k_1} N_-^{k_2}}^{(n_3)}$ and $I_{N_0^{k_1} Q^{k_2}}^{(n_3)}$ [see Eq. (35)]. In particular,

$$\begin{aligned} I_{N_+^{k_1} N_-^{k_2}}^{(n_3)} &= \frac{1}{2^{k_1+k_2}} \sum_{j_1=0}^{k_1} \sum_{j_2=0}^{k_2} \binom{k_1}{j_1} \binom{k_2}{j_2} (-1)^{j_2} \\ &\quad \times I_{N_0^{k_1+k_2-j_1-j_2} Q^{j_1+j_2}}^{(n_3)}. \end{aligned} \quad (\text{A21})$$

Finally, $A_{\vec{s}}^{(k)} = s_{n_1}(N_+)s_{n_2}(N_-)\rho_U^{(k)}$, $\vec{s} = (n_1, n_2)$, and $A_{\vec{s}'}^{(k)} = N_+^{n_+}N_-^{n_-}\rho_U^{(k)}$, $\vec{s}' = (n_+, n_-)$, are related by [recall $s(n, k) = 0$ if $k > n$]

$$A_{\vec{s}}^{(k)}(\lambda_1, \dots, \lambda_k) = \sum_{n_+=0}^{\infty} \sum_{n_-=0}^{\infty} s(n_+, n_+)s(n_2, n_-) \times A_{\vec{s}'}^{(k)}(\lambda_1, \dots, \lambda_k). \quad (\text{A22})$$

Since [Eq. (88)]

$$\sum_{n=0}^{\infty} s(n, n') \frac{x^n}{n!} = \frac{[\ln(1+x)]^{n'}}{n'!} = \frac{S(x)^{n'}}{n'!}, \quad (\text{A23})$$

the relation Eq. (A22) is of the form Eq. (A12) with $\omega_1(x) = \omega_2(x) = S(x)$ [and of course with $A_{(\vec{s}, k)}$ and $A'_{(\vec{s}', k)}$ replaced by $A_{\vec{s}}^{(k)}$ and $A'_{\vec{s}'}^{(k)}$], and Eq. (93) follows.

The generating function $\mathcal{Z}/\mathcal{Z}_0$ of full scalar and pseudoscalar correlation functions, normalized by the partition function and expressed in terms of Dirac eigenvalues, Eq. (83), is of the form Eq. (A5) with $A_{\vec{k}}[g] = N_+^{n_+}N_-^{n_-}Y_{\vec{k}}$, $\vec{k} = (n_+, n_-, k)$, where

$$Y_{\vec{k}} = \int_0^2 d\lambda_1 X(\lambda_1) \dots \int_0^2 d\lambda_k X(\lambda_k) \rho_U^{(k)}(\lambda_1, \dots, \lambda_k), \quad (\text{A24})$$

so $g = X$ and $I = [0, 2]$, and $t_1 = S(X_0)$, $t_2 = S(X_0^*)$, $t_3 = 1$. From Eq. (A7) follows then Eq. (86); taking the logarithm to obtain the generating function $\mathcal{W} - \mathcal{W}_0$, Eq. (89) follows. Alternatively, $\mathcal{Z}/\mathcal{Z}_0$ can be put in the form Eq. (A5) with $A_{\vec{k}}[g] = s_{n_1}(N_+)s_{n_2}(N_-)Y_{\vec{k}}$, $\vec{k} = (n_1, n_2, k)$, $t_1 = X_0$, $t_2 = X_0^*$, $t_3 = 1$, from which Eq. (92) follows.

Appendix B: Reality of the partition function

For γ_5 -Hermitean GW operators with $2R = \mathbf{1}$, the transformation properties Eq. (22) imply the reality of $\mathcal{Z}$, and so of the derivatives of $\mathcal{W}$ at zero sources. Setting $\mathcal{K}(V, W) \equiv j_S \mathbf{1}_f + i\vec{j}_P \cdot \vec{\sigma} \gamma_5 + i j_P \mathbf{1}_f \gamma_5 - \vec{j}_S \cdot \vec{\sigma}$, since $\mathcal{K}(\mathcal{C}V, -\mathcal{C}W) = \mathcal{K}(V, W)^\dagger$ and $[\gamma_5, \mathcal{K}(V, W)] = 0$, from Eq. (23) and $\det \gamma_5 = 1$ one readily finds

$$\mathcal{Z}(V, W; m)^* = \int DU e^{-S_{\text{eff}}(U)} \det [D_m(U) \mathbf{1}_f + \mathcal{K}(V, W) (\mathbf{1} - \frac{1}{2} D(U))] . \quad (\text{B1})$$

Since D_m is invertible ($\det D_m > 0$), one can exchange the order of factors in the second term in square brackets by virtue of Sylvester's theorem, $\det[\mathbf{1} + XY] = \det[\mathbf{1} + YX]$, and one concludes that

$$\mathcal{Z}(V, W; m)^* = \mathcal{Z}(V, W; m). \quad (\text{B2})$$

Since $\mathcal{Z}(0, 0; m) = \int DU e^{-S_{\text{eff}}(U)} [\det D_m(U)]^2 > 0$, it also follows that $\mathcal{W}(0, 0; m) \in \mathbb{R}$.

Appendix C: Proof of m^2 -differentiability of spectral quantities using partially quenched theories

In this Appendix I argue that spectral quantities such as the spectral density $\rho(\lambda; m)$ are m^2 -differentiable, i.e., finite with finite m^2 -derivatives in the chiral limit, under the assumption that in this limit chiral symmetry becomes manifest in scalar and pseudoscalar susceptibilities involving not only bilinears built with the physical fermionic fields, but also their analogs built with external fermionic fields. This is done in a partially quenched (PQ) setup, including both fermion and suitable scalar fields in the partition function in order for the corresponding determinants to cancel out exactly.

1. Spectral density

Define the following quantities,

$$\Upsilon^{(k)}(z^2; V, W; m) \equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \left\langle (iP_a^{\text{PQ}})^2 \right\rangle_{k; V, W}, \quad (\text{C1})$$

with $k = 0, \pm 1$, where for observables independent of the light-fermion fields

$$\begin{aligned} \langle \mathcal{O} \rangle_{k; V, W} &\equiv \frac{1}{Z^{(k)}} \int dU e^{-S_{\text{eff}}(U)} \det \mathcal{M}(U; V, W; m) \\ &\times \int d\omega d\bar{\omega} \int d\varphi d\varphi^* e^{-S^{(k)}(\omega, \bar{\omega}, \varphi, \varphi^*, U; z)} \\ &\times \mathcal{O}(\omega, \bar{\omega}, \varphi, \varphi^*, U), \\ Z^{(k)} &\equiv \int dU e^{-S_{\text{eff}}(U)} \det \mathcal{M}(U; V, W; m) \\ &\times \int d\omega d\bar{\omega} \int d\varphi d\varphi^* e^{-S^{(k)}(\omega, \bar{\omega}, \varphi, \varphi^*, U; z)}, \end{aligned} \quad (\text{C2})$$

with $\mathcal{M}$ the fermionic matrix defined in Eq. (71), and

$$iP_a^{\text{PQ}}(\omega, \bar{\omega}, U) \equiv i\bar{\omega} \left(\mathbf{1} - \frac{D(U)}{2} \right) \gamma_5 \sigma_a \omega. \quad (\text{C3})$$

Here $\omega_{\alpha cf}$ and $\bar{\omega}_{\alpha cf}$, $f = 1, 2$, are a “flavor” doublet of Grassmann field variables, and $\varphi_{\alpha c}$ are c -number complex field variables, carrying spacetime coordinate x , Dirac index $\alpha = 1, \dots, 4$, and color index $c = 1, \dots, N_c$. Both ω and ϕ transform in the same representation of the gauge group as the physical light-fermion fields, and $\bar{\omega}$ (and ϕ^*) in its complex conjugate. All indices are suppressed in Eqs. (C2) and (C3) and in the following. Moreover, $S^{(k)}$, $k = 0, \pm 1$, are the partially quenched actions

$$\begin{aligned} S^{(k)}(\omega, \bar{\omega}, \varphi, \varphi^*, U; z) \\ \equiv \bar{\omega} \left[D(U) + z \left(\mathbf{1} - \frac{D(U)}{2} \right) \right] \mathbf{1}_f \omega \\ + (-i)^k \varphi^\dagger [D(U) D(U)^\dagger + z^2 H(U)] \varphi, \end{aligned} \quad (\text{C4})$$

with $D(U)$ a γ_5 -Hermitean GW Dirac operator with $2R = \mathbf{1}$, and $H(U) \equiv \mathbf{1} - \frac{D(U)D(U)^\dagger}{4}$. The independence

of $\Upsilon^{(k)}$ of the index $a = 1, 2, 3$ follows from the vector flavor symmetry of the partially quenched actions. The quantities $\Upsilon^{(k)}$ provide three representations of the following resolvent,

$$G(z^2; V, W; m) \equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle \mathcal{B}(U; z^2) \rangle_{V, W}, \quad (C5)$$

$$\mathcal{B}(U; z^2) \equiv \text{tr} \frac{H(U)}{D(U)D(U)^\dagger + z^2 H(U)},$$

where the trace runs over coordinate, Dirac, and color indices, in terms of expectation values in a partially quenched theory, that are valid in three different domains of the complex mass z where the corresponding path integrals are convergent. In Eq. (C5) $\langle \dots \rangle_{V, W}$ denotes the expectation value in the presence of source terms,

$$\langle \mathcal{O}(U) \rangle_{V, W} \equiv \frac{\left\langle \mathcal{O}(U) \frac{\det \mathcal{M}(U; V, W; m)}{[\det D_m(U)]^2} \right\rangle}{\left\langle \frac{\det \mathcal{M}(U; V, W; m)}{[\det D_m(U)]^2} \right\rangle}, \quad (C6)$$

with $\langle \dots \rangle$ defined in Eq. (3). The relation between G and $\Upsilon^{(k)}$ reads

$$2G(z^2; V, W; m) = \begin{cases} \Upsilon^{(+1)}(z^2; V, W; m), & \text{Im } z^2 > 0, \\ \Upsilon^{(-1)}(z^2; V, W; m), & \text{Im } z^2 < 0, \\ \Upsilon^{(0)}(z^2; V, W; m), & \text{Re } z^2 > 0. \end{cases} \quad (C7)$$

The proof of this statement is straightforward, and uses only the properties of D , and the fact that $\det [(-i)^k \mathbf{1}] = (-i)^{4N_c V_4 k} = 1$, for any k . Since G is analytic in the cut complex plane, $\mathbb{C} \setminus \{\text{Re } z^2 \leq 0, \text{Im } z^2 = 0\}$, this shows that each $\Upsilon^{(k)}$ is analytic in its domain of definition. The domains of definition (and analyticity) of $\Upsilon^{(\pm 1)}$ overlap with that of $\Upsilon^{(0)}$, and the union of the three domains covers the whole cut complex plane. Writing the trace explicitly in terms of the eigenvalues of D , one finds

$$G(z^2; V, W; m) = \frac{n_0(V, W; m)}{z^2} + 2 \int_0^2 d\lambda \rho(\lambda; V, W; m) f(\lambda; z), \quad (C8)$$

where f is defined in Eq. (82), and²⁴

$$n_0(V, W; m) \equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle N_0 \rangle_{V, W}, \quad (C9)$$

$$\rho(\lambda; V, W; m) \equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle \rho_U(\lambda) \rangle_{V, W},$$

see Eqs. (37) and (41). Instead of ρ it is convenient to use the spectral density r associated with $x_n \equiv \lambda_n / \sqrt{1 - \lambda_n^2/4}$ [see under Eq. (26) for notation],

$$r(x; V, W; m) \equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle r_U(x) \rangle_{V, W}, \quad (C10)$$

$$r_U(x) \equiv \sum_n \delta(x - x_n),$$

related to ρ as $\rho = \frac{dx}{d\lambda} r$. One has

$$G(z^2; V, W; m) = \frac{n_0(V, W; m)}{z^2} + 2 \int_0^\infty dx \frac{r(x; V, W; m)}{x^2 + z^2}. \quad (C11)$$

For $z^2 = w + i\epsilon$ with $\epsilon > 0$ and $w \in \mathbb{R}$ one finds

$$\lim_{\epsilon \rightarrow 0} [G(w + i\epsilon; V, W; m) - G(w - i\epsilon; V, W; m)] = -2\pi i \theta(-w) \frac{r(\sqrt{-w}; V, W; m)}{\sqrt{-w}}. \quad (C12)$$

The discontinuity of G along the cut yields then the spectral density.

One makes now the symmetry-restoration assumption that the susceptibilities involving $i\bar{P}^{PQ}$ as well as physical scalar and pseudoscalar bilinears are symmetric in the chiral limit, for arbitrary complex mass z in the domains of definition of the three partially quenched theories defined by $S^{(k)}$, as well as on the boundaries of these domains. In the various domains of z^2 these susceptibilities are obtained from $\Upsilon^{(k)}(z^2; V, W; m)$ by taking derivatives with respect to the sources at zero sources, and so the symmetry requirement reads

$$\lim_{m \rightarrow 0} [\Upsilon^{(k)}(z^2; V, W; m) - \Upsilon^{(k)}(z^2; RV, RW; m)] = 0, \quad (C13)$$

for $k = 0, \pm 1$ and $\forall R \in \text{SO}(4)$. This extends by analytic continuation, patching the three $\Upsilon^{(k)}$ together, to

$$\lim_{m \rightarrow 0} [G(z^2; V, W; m) - G(z^2; RV, RW; m)] = 0, \quad (C14)$$

$\forall R \in \text{SO}(4)$, with the susceptibilities obtained from $G(z^2; V, W; m)$ defined in the whole cut complex z^2 -plane. By the same argument as in Appendix IIIB, Eq. (C13), or equivalently Eq. (C14), imply that these susceptibilities are finite in the chiral limit, with finite m^2 -derivatives of arbitrary order if they contain an even number of isoscalar bilinears; the same applies to the susceptibilities divided by m if this number is odd. From Eq. (C14) one obtains a similar relation for the discontinuity of G , and so for the spectral density one has

$$\lim_{m \rightarrow 0} [r(x; V, W; m) - r(x; RV, RW; m)] = 0, \quad (C15)$$

$\forall R \in \text{SO}(4)$. By the same argument as above one has that the derivatives of r with respect to the sources

²⁴ At $W = 0$, the symmetry-breaking term $mS + V \cdot O_V = (m + j_S)S + i\vec{j}_P \cdot \vec{P}$ can be rotated by a suitable chiral transformation to $m_V S$, with $m_V^2 \equiv (m + j_S)^2 + \vec{j}_P^2 = m^2 + u$. In this case $\det \mathcal{M}$ is then positive-definite, as can also be seen directly from its explicit expression, Eqs. (79)–(82). For $j_S = 0$ and $\vec{j}_P = m_t \hat{\alpha}$, this shows the equivalence of a theory including also a twisted-mass term, $im_t \hat{\alpha} \cdot \vec{P}$, besides the usual term mS , with a theory including only the usual mass term but with mass $(m^2 + m_t^2)^{1/2}$. More generally, since the Dirac spectrum is unaffected by the transformation as it depends only on U , under the usual assumption $N_+ N_- = 0$ a.e. one finds $\lim_{V_4 \rightarrow \infty} V_4^{-1} \langle N_0 \rangle_{V, 0} = 0$ also at $V \neq 0$.

at zero sources are finite quantities (possibly distributions) in the chiral limit, m^2 -differentiable if they contain an even number of isoscalar bilinear and m times an m^2 -differentiable quantity if this number is odd.²⁵ This applies in particular to $r(x; m) \equiv r(x; V, W; m)|_0$, which is just the spectral density for the variable x , from which one recovers the usual spectral density $\rho(\lambda; m) = \rho(\lambda; V, W; m)|_0$ as

$$\rho(\lambda; m) = \left(1 - \frac{\lambda^2}{4}\right)^{-\frac{3}{2}} r\left(\lambda \left(1 - \frac{\lambda^2}{4}\right)^{-\frac{1}{2}}; m\right). \quad (\text{C16})$$

The validity of Eq. (C15) relies upon the assumption that symmetry of the susceptibilities in the chiral limit holds all the way to the boundary of the domain of definition of the partially quenched theories defined by the actions $S^{(\pm 1)}$. One could relax this assumption to that of symmetry restoration only in the interior of these analyticity domains, provided one also assumes that the chiral limit $m \rightarrow 0$ and the limit $\epsilon \rightarrow 0$ defining the discontinuity can be exchanged. Alternatively, one can assume that G can be further analytically extended beyond the cut onto some suitable Riemann surface, and that the symmetry restoration condition holds there as well.

2. Higher-order correlation functions

One can prove the m^2 -differentiability of higher-order eigenvalue correlation functions using a similar construc-

tion, adding to the theory several partially quenched flavor doublets of fermion fields $\omega = (\omega_1, \dots, \omega_{N_{\text{PQ}}})$, $\bar{\omega} = (\bar{\omega}_1, \dots, \bar{\omega}_{N_{\text{PQ}}})$, and complex scalar fields $\varphi = (\varphi_1, \dots, \varphi_{N_{\text{PQ}}})$, of masses $z = (z_1, \dots, z_{N_{\text{PQ}}})$, and defining the partially quenched actions

$$S^{(\vec{k})}(\omega, \bar{\omega}, \varphi, \varphi^*, U; z) \equiv \sum_{j=1}^{N_{\text{PQ}}} S^{(k_j)}(\omega_j, \bar{\omega}_j, \varphi_j, \varphi_j^*, U; z_j), \quad (\text{C17})$$

where $\vec{k} = (k_1, \dots, k_{N_{\text{PQ}}})$, $k_j = 0, \pm 1$. One then makes the symmetry-restoration assumption on the correlators

$$\Upsilon^{(\vec{k})}(z^2; V, W; m) = \left\langle \prod_{j=1}^{N_{\text{PQ}}} (iP_{j a_j})^2 \right\rangle_{\vec{k}; V, W}, \quad (\text{C18})$$

where $z^2 = (z_1^2, \dots, z_{N_{\text{PQ}}}^2)$,

$$iP_{j a} = i\bar{\omega}_j \left(\mathbf{1} - \frac{D(U)}{2} \right) \gamma_5 \sigma_a \omega_j, \quad (\text{C19})$$

and the expectation values $\langle \dots \rangle_{\vec{k}; V, W}$ are defined as in Eq. (C2), replacing $S^{(k)}$ with $S^{(\vec{k})}$. These quantities are again independent of the indices $a_j = 1, 2, 3$ thanks to the vector flavor symmetry of the partially quenched fermion doublets. For two-point spectral correlation functions, one sets $N_{\text{PQ}} = 2$ and defines the following resolvent,

$$\begin{aligned} G(z_1^2, z_2^2; V, W; m) &\equiv \lim_{V_4 \rightarrow \infty} V_4^{-1} \left[\langle \mathcal{B}(U; z_1^2) \mathcal{B}(U; z_2^2) \rangle_{V, W} - \langle \mathcal{B}(U; z_1^2) \rangle_{V, W} \langle \mathcal{B}(U; z_2^2) \rangle_{V, W} \right], \\ 4G(z_1^2, z_2^2; V, W; m) &= \lim_{V_4 \rightarrow \infty} V_4^{-1} \left\langle (iP_{1a}^{\text{PQ}})^2 (iP_{2b}^{\text{PQ}})^2 \right\rangle_{\vec{k}; V, W; c}, \end{aligned} \quad (\text{C20})$$

where c denotes the connected part defined in the usual way (see Appendix A), and with the representation in terms of the partially quenched theory labeled by $\vec{k} = (k_1, k_2)$ holding within the convergence domains dis-

cussed above in Appendix C 1 ($k_i = \pm 1$ if $\text{Im } z_i^2 \geq 0$, $k_i = 0$ if $\text{Re } z_i^2 > 0$). Writing the traces in terms of the eigenvalues of D one finds

$$\begin{aligned} G(z_1^2, z_2^2; V, W; m) &= \frac{b_{N_0^2 \infty}(V, W; m)}{z_1^2 z_2^2} + 2 \int_0^\infty dx \left[\frac{1}{z_1^2 (x^2 + z_2^2)} + \frac{1}{(x^2 + z_1^2) z_2^2} \right] r_{N_0 c \infty}(x; V, W; m) \\ &\quad + 4 \int_0^\infty dx \int_0^\infty dx' \frac{r_c^{(2)}(x, x'; V; m) + (\delta(x - x') + \delta(x + x')) r(x; V, W; m)}{(x^2 + z_1^2)(x'^2 + z_2^2)}, \end{aligned} \quad (\text{C21})$$

²⁵ To proceed rigorously, one should consider the discontinuity of G integrated over an interval $[x_0, x]$ of nonzero length, that yields

the following generating function (see footnote 8)

$$\mathcal{N}^{(r)}(x_0, x; V, W; m) = \lim_{V_4 \rightarrow \infty} V_4^{-1} \left\langle \int_{x_0}^x dx' r_U(x') \right\rangle_{V, W}.$$

For any interval this has the same symmetry property as in Eq. (C15), and so it is an ordinary generating function with finite m^2 -derivatives at $m = 0$. The statement for $r(V, W; m)$ follows by taking the derivative of $\mathcal{N}^{(r)}$ with respect to x .

where [see Eqs. (34) and (114), (33) and (39), and (38)]

$$\begin{aligned} b_{N_0^\infty}(V, W; m) &= \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle N_0^2 \rangle_{V, W, c}, \\ r_{N_0 c \infty}(x; V, W; m) &= \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle N_0 r_U(x) \rangle_{V, W, c}, \\ r_{c \infty}^{(2)}(x, x'; V, W; m) &= \lim_{V_4 \rightarrow \infty} V_4^{-1} \langle r_U(x) r_U(x') \rangle_{V, W, c} \\ &\quad - (\delta(x - x') + \delta(x + x')) \\ &\quad \times r(x; V, W; m), \end{aligned} \quad (\text{C22})$$

understood as generating functions. Approaching $z_{1,2}^2 = 0$ from positive real values, $z_{1,2}^2 = w_{1,2}^2 \in \mathbb{R}^+$, one finds

$$\lim_{w_{1,2} \rightarrow 0} w_1^2 w_2^2 G(w_1^2, w_2^2; V, W; m) = b_{N_0^\infty}(V, W; m), \quad (\text{C23})$$

while approaching $z_2^2 = 0$ from positive real values, $z_2^2 = w_2^2 \in \mathbb{R}^+$, and computing the discontinuity on the negative real axis of z_1^2 , one gets

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \lim_{w_2 \rightarrow 0} w_2^2 [G(-w_1^2 + i\epsilon, w_2^2; V, W; m) \\ - G(-w_1^2 - i\epsilon, w_2^2; V, W; m)] \\ = -2\pi i \frac{r_{N_0 c \infty}(|w_1|; V, W; m)}{|w_1|}. \end{aligned} \quad (\text{C24})$$

Finally, from the double discontinuity on the negative real axes of $z_{1,2}^2$ one gets

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \lim_{\epsilon' \rightarrow 0} [G(-w_1^2 + i\epsilon, -w_2^2 + i\epsilon'; V, W; m) \\ - G(-w_1^2 - i\epsilon, -w_2^2 + i\epsilon'; V, W; m) \\ - G(-w_1^2 + i\epsilon, -w_2^2 - i\epsilon'; V, W; m) \\ + G(-w_1^2 - i\epsilon, -w_2^2 - i\epsilon'; V, W; m)] \\ = \frac{4\pi^2}{|w_1||w_2|} \left[r_c^{(2)}(|w_1|, |w_2|; V, W; m) \right. \\ \left. + \delta(|w_1| - |w_2|) r(|w_1|; V, W; m) \right]. \end{aligned} \quad (\text{C25})$$

Assuming as above in Appendix C 1 that chiral symmetry is realized in the chiral limit in the partially quenched theories labeled by $\vec{k}$, one proves that G has the same m^2 -differentiability properties at $m = 0$ as its counterpart in the previous subsection, in the whole domain of analyticity and at its boundary, and this property is inherited by $b_{N_0^\infty}(0, 0; m)$, $r_{N_0 c \infty}(x; 0, 0; m)$ and $r_{c \infty}^{(2)}(x, x'; 0, 0; m)$, and so by $\rho_{N_0 c \infty}(\lambda; m)$ and $\rho_{c \infty}^{(2)}(\lambda, \lambda'; m)$, that are obtained by a mass-independent change of variables.

Appendix D: $\mathcal{A}_n$ and susceptibilities

Consider the generating function for vanishing isosinglet sources, $j_{S,P} = 0$, Eq. (57). I consider for notational simplicity the CP -invariant case, in which $\mathcal{A}_{n_u n_w n_{\bar{u}}}(m^2) \neq 0$ only for $n_{\bar{u}} = 2n_u$. For $(n_u, n_w, n_{\bar{u}}) \neq$

$(0, 0, 0)$, setting $B_{n_u n_w n_{\bar{u}}}(m^2) \equiv 4^{n_{\bar{u}}} \mathcal{A}_{n_u n_w 2n_{\bar{u}}}(m^2)$, and further choosing $j_{P3} = j_{S2} = j_{S3} = 0$, one finds after simple algebra

$$\begin{aligned} \mathcal{W}_\infty(V, W; m)|_{j_S=j_P=j_{P3}=j_{S2}=j_{S3}=0} \\ = \sum_{a,b,c=0}^{\infty} \frac{(j_{P1}^2)^a (j_{P2}^2)^b (j_{S1}^2)^c}{(2a)!(2b)!(2c)!} \chi_{abc}(m^2), \end{aligned} \quad (\text{D1})$$

where for $(a, b, c) \neq (0, 0, 0)$

$$\begin{aligned} \chi_{abc}(m) &\equiv \sum_{d=0}^{\min(a,c)} C_{abcd} B_{a-d+b \ c-d \ d}(m^2), \\ C_{abcd} &\equiv \frac{(2a)!(2b)!(2c)!}{(a-d)!b!(c-d)!(2d)!}, \end{aligned} \quad (\text{D2})$$

are just the usual susceptibilities in the restricted subset

$$\chi_{abc}(m) = \chi((iP_1)^{2a} (iP_2)^{2b} S_1^{2c}), \quad (\text{D3})$$

while $\chi_{000} = \mathcal{A}_{000} = \mathcal{W}_\infty|_0$ is minus the free energy density. For $a = z$, $b = x$, $c = y + z$ one has from Eq. (D2)

$$\begin{aligned} B_{xyz} &= \frac{x!y!\chi_{z \ x \ y+z}}{(2x)![2(y+z)]!} \\ &\quad - \sum_{k=1}^z \frac{y!(2z)!B_{x+k \ y+k \ z-k}}{k!(y+k)![2(z-k)]!}, \quad z \geq 1, \\ B_{xy0} &= \frac{x!y!\chi_{0xy}}{(2x)!(2y)!}, \quad x + y > 0. \end{aligned} \quad (\text{D4})$$

By recursion on z , one shows that B_{xyz} , $x, y, z \geq 0$, $x + y + z > 0$, are finite linear combinations of χ_{abc} with $c = y + z$ and $a + b = x + z$, with mass-independent coefficients. To see this, notice that for the term $B_{x'y'z'} = B_{x+k \ y+k \ z-k}$ inside the summation in Eq. (D4) one has $0 \leq z' < z$, i.e., the maximal z' decreases at every step of the recursion, while $y' + z' = y + z$ and $x' + z' = x + z$ remain constant, and $x' + y' + z' > x + y + z > 0$. The extension to the non- CP -invariant case presents no difficulty.

The same result applies to the coefficients of the generating function of susceptibilities involving also functionals of gauge fields only, $G_i(U)$, obtained from

$$\begin{aligned} \mathcal{Z}_G(V, W; J_G; m) &\equiv \int DU \int D\Psi D\bar{\Psi} e^{-S_{\text{eff}}(U)} \\ &\quad \times e^{-\bar{\Psi} D_m(U) \mathbf{1}_f \Psi - K(\Psi, \bar{\Psi}, U; V, W)} \\ &\quad \times e^{\sum_{i=1}^N J_{G_i} G_i(U)}, \\ \mathcal{W}_G(V, W; J_G; m) &\equiv \frac{1}{V_4} \ln \mathcal{Z}_G(V, W; J_G; m), \\ \mathcal{W}_{G^\infty}(V, W; J_G; m) &\equiv \lim_{V_4 \rightarrow \infty} \mathcal{W}_G(V, W; J_G; m). \end{aligned} \quad (\text{D5})$$

Since gauge fields are unaffected by chiral transforma-

tions one has again

$$\begin{aligned}\mathcal{W}_{G\infty}(V, W; J_G; m) &= \hat{\mathcal{W}}_{G\infty}(m^2 + u, w, \tilde{u}; J_G) \\ &= \sum_{n_u, n_w, n_{\tilde{u}}=0}^{\infty} \frac{u^{n_u} w^{n_w} \tilde{u}^{n_{\tilde{u}}}}{n_u! n_w! n_{\tilde{u}}!} \mathcal{A}_{n_u n_w n_{\tilde{u}}}(m^2; J_G) \\ &= \sum_{n_u, n_w, n_{\tilde{u}}=0}^{\infty} \frac{u^{n_u} w^{n_w} \left(\frac{\tilde{u}}{2}\right)^{2n_{\tilde{u}}}}{n_u! n_w! (2n_{\tilde{u}})!} B_{n_u n_w n_{\tilde{u}}}(m^2; J_G),\end{aligned}\quad (\text{D6})$$

having used CP invariance in the last passage, and having set $B_{n_u n_w n_{\tilde{u}}}(m^2; J_G) = 4^{n_{\tilde{u}}} \mathcal{A}_{n_u n_w 2n_{\tilde{u}}}(m^2; J_G)$. One finds again

$$\begin{aligned}\mathcal{W}_{G\infty}(V, W; J_G; m)|_{j_S=j_P=j_{P3}=j_{S2}=j_{S3}=0} \\ = \sum_{a,b,c=0}^{\infty} \frac{(j_{P1}^2)^a (j_{P2}^2)^b (j_{S1}^2)^c}{(2a)!(2b)!(2c)!} \chi_{abc}(m^2; J_G),\end{aligned}\quad (\text{D7})$$

where now

$$\chi_{abc}(m; J_G) \equiv \sum_{d=0}^{\min(a,c)} C_{abcd} B_{a-d+b} c-d(m^2; J_G) \quad (\text{D8})$$

are the generating functions of susceptibilities involving scalar and pseudoscalar bilinears and gauge operators,

$$\prod_{i=1}^N \partial_{J_{Gi}}^{n_i} \chi_{abc}(m; J_G)|_0 = \chi \left((iP_1)^{2a} (iP_2)^{2b} S_1^{2c} \prod_{i=1}^N G_i^{n_i} \right). \quad (\text{D9})$$

Proceeding as above, one shows recursively that $B_{xyz}(m; J_G)$ can be reconstructed from $\chi_{abc}(m; J_G)$, so they contain equivalent information.

Appendix E: Ward-Takahashi identities

Using Eq. (46) and the exact chiral invariance of the massless theory, one finds after simple algebra

$$\begin{aligned}\mathcal{Z}(R^T V, R^T W; m) &= \mathcal{Z}(R^T V + m e_0, R^T W; 0) \\ &= \mathcal{Z}(V + m R e_0, W; 0) = \mathcal{Z}(V + m(R - \mathbf{1}_4) e_0, W; m),\end{aligned}\quad (\text{E1})$$

where $e_0 = (1, \vec{0})^T$ and $\mathbf{1}_4$ is the 4×4 identity matrix, or equivalently

$$\mathcal{Z}(V, W; m) = \mathcal{Z}(RV + m(R - \mathbf{1}_4) e_0, RW; m). \quad (\text{E2})$$

The same relation holds replacing $\mathcal{Z}$ with $\mathcal{W}$. The generating functions are therefore invariant under the mass-dependent affine transformation Eq. (60),

$$V \rightarrow RV + m(R - \mathbf{1}_4) e_0, \quad W \rightarrow RW. \quad (\text{E3})$$

Expanding Eq. (E1) for small $R - \mathbf{1}_4$, one finds to leading order

$$\begin{aligned}\mathcal{Z}(R^T V, R^T W; m) - \mathcal{Z}(V, W; m) \\ = m[(R - \mathbf{1}_4) e_0] \cdot \partial_V \mathcal{Z}(V, W; m) + \dots\end{aligned}\quad (\text{E4})$$

The quantity $\mathcal{Z}(R^T V, R^T W; m)$ is the generating function of the correlators of the chirally transformed bilinears $RO_{V,W}$ [see Eqs. (10) and (11)]. Denoting with O_i the components of the vector

$$O \equiv (O_V, O_W)^T = \left(S, i\vec{P}, iP, -\vec{S} \right)^T, \quad (\text{E5})$$

after taking derivatives with respect to the sources, denoted collectively as $J \equiv (V, W)$, and setting these to zero, one finds for the left-hand side of Eq. (E4)

$$\begin{aligned}\left(\prod_{s=1}^k (-\partial_{J_{is}}) \right) [\mathcal{Z}(R^T V, R^T W; m) - \mathcal{Z}(V, W; m)]|_0 \\ = \left\langle \left(\prod_{s=1}^k (RO)_{i_s} \right) - \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle,\end{aligned}\quad (\text{E6})$$

which is the expectation value of the variation of $\prod_{s=1}^k O_{i_s}$ under a chiral transformation. Specializing now to infinitesimal vector and axial nonsinglet transformations [see under Eq. (7)] this leads to the well-known integrated Ward-Takahashi identities. For general vector and axial transformations $\mathcal{U}_{V,A}(\vec{\alpha})$, one correspondingly finds the $SO(4)$ matrices $R_{V,A}(\vec{\alpha}) \equiv \mathcal{R}(\mathcal{U}_{V,A}(\vec{\alpha}))$, with

$$R_V(\vec{\alpha}) = \begin{pmatrix} 1 & \vec{0}^T \\ \vec{0} & \tilde{R}(\vec{\alpha}) \end{pmatrix}, \quad (\text{E7})$$

$$\tilde{R}(\vec{\alpha}) \vec{v} \equiv \Pi_{\hat{\alpha}} \vec{v} + \cos|\vec{\alpha}| \Pi_{\hat{\alpha}^\perp} \vec{v} + \sin|\vec{\alpha}| \hat{\alpha} \wedge \vec{v},$$

and

$$R_A(\vec{\alpha}) = \begin{pmatrix} \cos(|\vec{\alpha}|) & \sin(|\vec{\alpha}|) \hat{\alpha}^T \\ -\sin(|\vec{\alpha}|) \hat{\alpha} & \Pi_{\hat{\alpha}^\perp} + \cos(|\vec{\alpha}|) \Pi_{\hat{\alpha}} \end{pmatrix}. \quad (\text{E8})$$

Here $\tilde{R} \in SO(3)$, $\Pi_{\hat{\alpha}} \equiv \hat{\alpha} \hat{\alpha}^T$ and $\Pi_{\hat{\alpha}^\perp} \equiv \mathbf{1}_3 - \Pi_{\hat{\alpha}}$, with $\hat{\alpha} = \vec{\alpha}/|\vec{\alpha}|$ and $\mathbf{1}_3$ the 3×3 identity matrix. Denoting for a generic function $\mathcal{F}(O_V, O_W)$

$$\begin{aligned}\mathcal{F}(R_{V,A} O_V, R_{V,A} O_W) - \mathcal{F}(O_V, O_W) \\ \equiv i\vec{\alpha} \cdot \vec{\delta}_{V,A} \mathcal{F}(O_V, O_W) + O(\vec{\alpha}^2),\end{aligned}\quad (\text{E9})$$

Eq. (E4) reads for $R = R_{V,A}$

$$\begin{aligned}\left(\prod_{s=1}^k (-\partial_{J_{is}}) \right) [\mathcal{Z}(R_{V,A}^T V, R_{V,A}^T W; m) - \mathcal{Z}(V, W; m)]|_0 \\ = i\vec{\alpha} \cdot \left\langle \vec{\delta}_{V,A} \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle + O(\vec{\alpha}^2) \\ = m[(R_{V,A} - \mathbf{1}_4) e_0] \cdot \partial_V \left(\prod_{s=1}^k (-\partial_{J_{is}}) \right) \mathcal{Z}(V, W; m)|_0 \\ + O(\vec{\alpha}^2).\end{aligned}\quad (\text{E10})$$

Since $(R_V - \mathbf{1}_4) e_0 = 0$ and $(R_A - \mathbf{1}_4) e_0 = -(0, \vec{\alpha})^T$, one finds for the right-hand side to order $|\vec{\alpha}|$

$$m[(R_V - \mathbf{1}_4) e_0] \cdot \partial_V \left(\prod_{s=1}^k (-\partial_{J_{is}}) \right) \mathcal{Z}(V, W; m)|_0 = 0, \quad (\text{E11})$$

for a vector transformation, and

$$\begin{aligned}m[(R_A - \mathbf{1}_4) e_0] \cdot \partial_V \left(\prod_{s=1}^k (-\partial_{J_{is}}) \right) \mathcal{Z}(V, W; m)|_0 \\ = m\vec{\alpha} \cdot \left\langle (i\vec{P}) \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle,\end{aligned}\quad (\text{E12})$$

for an axial transformation, and one concludes

$$\begin{aligned} \left\langle \delta_{Va} \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle &= 0, \\ \left\langle \delta_{Aa} \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle &= m \left\langle P_a \left(\prod_{s=1}^k O_{i_s} \right) \right\rangle. \end{aligned} \quad (\text{E13})$$

Replacing $\mathcal{Z}$ with $\mathcal{W}$ in Eq. (E4) one shows that the same identities hold for the connected correlation functions, i.e., for $\langle \dots \rangle$ replaced with $\langle \dots \rangle_c$ in Eq. (E13).

Appendix F: Chiral limit of the generating function in the symmetric phase

The symmetry-restoration condition Eq. (43) formally collects the symmetry-restoration conditions Eq. (42) for the whole set of susceptibilities. Denoting the susceptibilities with a fixed number n_V of O_V -type bilinears and n_W of O_W -bilinears compactly as $\chi_{\vec{n}} \equiv \chi((\otimes_{n_V} O_V) \otimes (\otimes_{n_W} O_W))$, $\vec{n} = (n_V, n_W)$, one has

$$\mathcal{W}_{\infty}(V, W; m) = \sum_{\vec{n}} \frac{J_{\vec{n}} \cdot \chi_{\vec{n}}(m)}{n_V! n_W!}, \quad (\text{F1})$$

where $J_{\vec{n}} \equiv (\otimes_{n_V} V) \otimes (\otimes_{n_W} W)$. Symmetry restoration requires that for each $\vec{n}$ separately

$$\lim_{m \rightarrow 0} [\mathbf{I} - \mathbf{R}(R)] \chi_{\vec{n}}(m) = 0, \quad \forall R \in \text{SO}(4), \quad (\text{F2})$$

where $\mathbf{I} \equiv \otimes_n \mathbf{1}_4$ and $\mathbf{R}(R) \equiv \otimes_n R$, with $n = n_V + n_W$. The representation space $\otimes_n \mathbb{R}^4$ can be decomposed into the invariant subspace $\mathbb{I}$ formed by vectors invariant under any transformation, $\mathbb{I} \equiv \{x \in \otimes_n \mathbb{R}^4 \mid \mathbf{R}(R)x = x \ \forall R \in \text{SO}(4)\}$, and its orthogonal complement, $\mathbb{I}_{\perp}$, which is also an invariant subspace. One has then $\otimes_n \mathbb{R}^4 = \mathbb{I} \oplus \mathbb{I}_{\perp}$, and one can write $\chi_{\vec{n}}(m) = x_{\vec{n}}(m) + v_{\vec{n}}(m)$ with $x_{\vec{n}}(m) \in \mathbb{I}$ and $v_{\vec{n}}(m) \in \mathbb{I}_{\perp}$. It is easy to show that $x \in \mathbb{I}$ if and only if $\mathbf{t}_a x = 0$ for $a = 1, \dots, d = \dim \text{SO}(4)$, where the Hermitean matrices $\mathbf{t}_a$ are the representatives of the group generators t_a in the n -fold tensor-product representation.²⁶ Since this is reducible for $n > 1$, $\mathbb{I}$ generally does not have to be trivial.

Consider now $\mathbf{R}_a = \mathbf{R}(e^{i\epsilon_a t_a}) = e^{i\epsilon_a \mathbf{t}_a}$ (no summation over a), $a = 1, \dots, d$. Choosing ϵ_a so that $\epsilon_a \|\mathbf{t}_a\| < 2\pi$, the matrices $\mathbf{M}_a = \mathbf{I} - \mathbf{R}_a$ are positive semidefinite,

and $\mathbf{M}_a w = 0$ if and only if $\mathbf{t}_a w = 0$. Then for $\mathbf{M} \equiv \sum_{a=1}^d \mathbf{M}_a$ one has $\mathbf{M}w = 0$ if and only if $\mathbf{t}_a w = 0$ for $a = 1, \dots, d$, so if and only if $w \in \mathbb{I}$, and so the restriction of $\mathbf{M}$ to $\mathbb{I}_{\perp}$ is invertible. Defining $\tilde{\mathbf{M}}w$ for $w \in \otimes_n \mathbb{R}^4$ by decomposing (uniquely) $w = x + v$ with $x \in \mathbb{I}$ and $v \in \mathbb{I}_{\perp}$, and setting $\tilde{\mathbf{M}}w \equiv x + (\mathbf{M}|_{\mathbb{I}_{\perp}})^{-1} v$, one has then

$$\begin{aligned} 0 &= \tilde{\mathbf{M}} \sum_{a=1}^d \left(\lim_{m \rightarrow 0} \mathbf{M}_a \chi_{\vec{n}}(m) \right) = \lim_{m \rightarrow 0} \tilde{\mathbf{M}} \left(\sum_{a=1}^d \mathbf{M}_a \right) v_{\vec{n}}(m) \\ &= \lim_{m \rightarrow 0} \tilde{\mathbf{M}} \mathbf{M} v_{\vec{n}}(m) = \lim_{m \rightarrow 0} v_{\vec{n}}(m), \end{aligned} \quad (\text{F3})$$

as a consequence of Eq. (F2). Setting

$$\mathcal{W}_{\infty}^{\text{inv}}(V, W; m) \equiv \sum_{\vec{n}} \frac{J_{\vec{n}} \cdot x_{\vec{n}}(m)}{n_V! n_W!}, \quad (\text{F4})$$

with $\mathcal{W}_{\infty}^{\text{inv}}(RV, RW; m) = \mathcal{W}_{\infty}^{\text{inv}}(V, W; m)$, one has then

$$\lim_{m \rightarrow 0} [\mathcal{W}_{\infty}(V, W; m) - \mathcal{W}_{\infty}^{\text{inv}}(V, W; m)] = 0, \quad (\text{F5})$$

and in particular

$$\begin{aligned} &\lim_{m \rightarrow 0} [\partial_{j_S} - 2(j_S \partial_{V^2} + j_P \partial_{2V \cdot W})] \mathcal{W}_{\infty}(V, W; m) \\ &= \lim_{m \rightarrow 0} [\partial_{j_S} - 2(j_S \partial_{V^2} + j_P \partial_{2V \cdot W})] \mathcal{W}_{\infty}^{\text{inv}}(V, W; m) = 0. \end{aligned} \quad (\text{F6})$$

In Refs. [81, 82] it is implicitly made use of the relation $\lim_{m \rightarrow 0} (\mathbf{I} - \mathbf{R}) \chi_{\vec{n}}(m) = (\mathbf{I} - \mathbf{R}) \lim_{m \rightarrow 0} \chi_{\vec{n}}(m)$ [see Eqs. (S7) and (S8) of the Supplemental Material of Ref. [81], and Eq. (9) of Ref. [82]], that is true if one assumes the existence of $\lim_{m \rightarrow 0} \chi_{\vec{n}}(m)$, but is not warranted otherwise. The argument above shows that this assumption is not necessary to prove Eq. (F5), upon which the arguments of Refs. [81, 82] rely.

Appendix G: Contribution of complex modes to the fermionic determinant

For the contribution $\det M(\mu)$ to the fermionic determinant in the presence of sources of a pair of complex modes μ, μ^* , with $|\mu|^2 = 2\text{Re} \mu$ and $0 < |\mu|^2 < 4$, see Eqs. (74) and (75), one finds

²⁶ If $\mathbf{R}w = w \ \forall R \in \text{SO}(4)$ then in particular $e^{i\epsilon \mathbf{t}_a} w = w \ \forall \epsilon$, and so $0 = -i\partial_{\epsilon} e^{i\epsilon \mathbf{t}_a} w|_{\epsilon=0} = \mathbf{t}_a w$, for $a = 1, \dots, d$. Conversely, since $\text{SO}(4)$ is a connected Lie group one can write any

$R \in \text{SO}(4)$ as $R = \prod_j R_j$ with $R_j = e^{i \sum_{a=1}^d \alpha_a^{(j)} t_a}$ for suitable $\alpha_a^{(j)}$. If $\mathbf{t}_a w = 0$ for $a = 1, \dots, d$ then $\mathbf{R}w = \left(\prod_j \mathbf{R}(R_j) \right) w = \left(\prod_j e^{i \sum_{a=1}^d \alpha_a^{(j)} \mathbf{t}_a} \right) w = w$.

$$\begin{aligned}
\det M(\mu) &= \det \begin{pmatrix} (1 - \frac{\mu}{2}) \mathbf{1}_f & 0 \\ 0 & (1 - \frac{\mu^*}{2}) \mathbf{1}_f \end{pmatrix} \det \begin{pmatrix} \frac{\mu}{1 - \frac{\mu}{2}} \mathbf{1}_f + A & iB \\ iB & \frac{\mu^*}{1 - \frac{\mu^*}{2}} \mathbf{1}_f + A \end{pmatrix} = |1 - \frac{\mu}{2}|^4 \det \begin{pmatrix} A + iB & i \frac{\text{Im} \mu}{1 - \frac{|\mu|^2}{4}} \mathbf{1}_f \\ i \frac{\text{Im} \mu}{1 - \frac{|\mu|^2}{4}} \mathbf{1}_f & A - iB \end{pmatrix} \\
&= \left(1 - \frac{|\mu|^2}{4}\right)^2 \det \left((A + iB)(A - iB) + \frac{|\mu|^2}{1 - \frac{|\mu|^2}{4}} \right) = \det \left(|\mu|^2 + \left(1 - \frac{|\mu|^2}{4}\right) (A + iB)(A - iB) \right),
\end{aligned} \tag{G1}$$

having used the following identity for determinants of block square matrices with square blocks of equal size,

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} \frac{a+b+c+d}{2} & \frac{a-b+c-d}{2} \\ \frac{a+b-c-d}{2} & \frac{a-b+c+d}{2} \end{pmatrix}, \tag{G2}$$

the following identities satisfied by the complex eigenvalues,

$$\begin{aligned}
|1 - \frac{\mu}{2}|^2 &= 1 - \frac{|\mu|^2}{4}, & \text{Re} \frac{\mu}{1 - \frac{\mu}{2}} &= 0, \\
\text{Im} \frac{\mu}{1 - \frac{\mu}{2}} &= \frac{\text{Im} \mu}{1 - \frac{|\mu|^2}{4}}, & (\text{Im} \mu)^2 &= |\mu|^2 \left(1 - \frac{|\mu|^2}{4}\right),
\end{aligned} \tag{G3}$$

and the properties of determinants of block matrices with commuting blocks [153]. Setting now $|\mu|^2 = \lambda^2$ [see above

Eq. (26)] and $h(\lambda) = 1 - \frac{\lambda^2}{4}$ [see Eq. (78)], since

$$\begin{aligned}
(A + iB)(A - iB) &= (\tilde{V}^2 + W^2) \mathbf{1}_f + 2\vec{L} \cdot \vec{\sigma} \\
&= (m^2 + u + w) \mathbf{1}_f + 2\vec{L} \cdot \vec{\sigma},
\end{aligned} \tag{G4}$$

with $\tilde{V}$ defined in Eq. (46), and

$$\begin{aligned}
\vec{L} &\equiv j_P \vec{J}_P - (m + j_S) \vec{J}_S + \vec{J}_P \wedge \vec{J}_S, \\
\vec{L}^2 &= \tilde{V}^2 W^2 - (\tilde{V} \cdot W)^2 = (m^2 + u)w - \frac{1}{4}\tilde{u}^2,
\end{aligned} \tag{G5}$$

one has

$$\begin{aligned}
\det M(\mu) &= \det \left((\lambda^2 + h(\lambda)(m^2 + u + w)) \mathbf{1}_f + 2h(\lambda) \vec{L} \cdot \vec{\sigma} \right) = (\lambda^2 + h(\lambda)(m^2 + u + w))^2 - 4h(\lambda)^2 \vec{L}^2 \\
&= (\lambda^2 + h(\lambda)(m^2 + u + w))^2 - 4h(\lambda)^2 ((m^2 + u)w - \frac{1}{4}\tilde{u}^2) \\
&= \lambda^4 + 2\lambda^2 h(\lambda)(m^2 + u + w) + h(\lambda)^2 ((m^2 + u - w)^2 + \tilde{u}^2) \\
&= [\lambda^2 + m^2 h(\lambda)]^2 + 2h(\lambda)[(\lambda^2 + m^2 h(\lambda))u + (\lambda^2 - m^2 h(\lambda))w] + h(\lambda)^2 [(u - w)^2 + \tilde{u}^2].
\end{aligned} \tag{G6}$$

-
- [1] Y. Aoki, G. Endrődi, Z. Fodor, S. D. Katz, and K. K. Szabó, *Nature* **443**, 675 (2006), arXiv:hep-lat/0611014.
 - [2] S. Borsányi, Z. Fodor, C. Hoelbling, S. D. Katz, S. Krieg, C. Ratti, and K. K. Szabó (Wuppertal-Budapest collaboration), *J. High Energy Phys.* **09**, 073 (2010), arXiv:1005.3508 [hep-lat].
 - [3] A. Bazavov, T. Bhattacharya, M. Cheng, C. DeTar, H.-T. Ding, S. Gottlieb, R. Gupta, P. Hegde, U. M. Heller, F. Karsch, E. Laermann, L. Levkova, S. Mukherjee, P. Petreczky, C. Schmidt, R. A. Soltz, W. Soeldner, R. Sugar, D. Toussaint, W. Unger, and P. Vranas, *Phys. Rev. D* **85**, 054503 (2012), arXiv:1111.1710 [hep-lat].
 - [4] T. Bhattacharya *et al.*, *Phys. Rev. Lett.* **113**, 082001 (2014), arXiv:1402.5175 [hep-lat].
 - [5] A. Bazavov, N. Brambilla, H. T. Ding, P. Petreczky, H. P. Schadler, A. Vairo, and J. H. Weber, *Phys. Rev. D* **93**, 114502 (2016), arXiv:1603.06637 [hep-lat].
 - [6] R. D. Pisarski and F. Wilczek, *Phys. Rev. D* **29**, 338 (1984).
 - [7] K. Rajagopal and F. Wilczek, *Nucl. Phys. B* **399**, 395 (1993), arXiv:hep-ph/9210253.
 - [8] K. Rajagopal and F. Wilczek, *Nucl. Phys. B* **404**, 577 (1993), arXiv:hep-ph/9303281.
 - [9] E. V. Shuryak, *Comments Nucl. Part. Phys.* **21**, 235 (1994), arXiv:hep-ph/9310253.
 - [10] J. I. Kapusta, D. Kharzeev, and L. D. McLerran, *Phys. Rev. D* **53**, 5028 (1996), arXiv:hep-ph/9507343.
 - [11] Z. Huang and X.-N. Wang, *Phys. Rev. D* **53**, 5034 (1996), arXiv:hep-ph/9507395.
 - [12] S. D. Bass and P. Moskal, *Rev. Mod. Phys.* **91**, 015003 (2019), arXiv:1810.12290 [hep-ph].
 - [13] R. D. Pisarski and F. Rennecke, *Phys. Rev. Lett.* **132**, 251903 (2024), arXiv:2401.06130 [hep-ph].
 - [14] C. Bonati, M. D'Elia, M. Mariti, G. Martinelli, M. Mesiti, F. Negro, F. Sanfilippo, and G. Villadoro, *J. High Energy Phys.* **03** (2016), 155, arXiv:1512.06746 [hep-lat].
 - [15] P. Petreczky, H.-P. Schadler, and S. Sharma, *Phys. Lett. B* **762**, 498 (2016), arXiv:1606.03145 [hep-lat].
 - [16] J. Frison, R. Kitano, H. Matsufuru, S. Mori, and N. Yamada, *J. High Energy Phys.* **09** (2016), 021, arXiv:1606.07175 [hep-lat].

- [17] S. Borsányi *et al.*, Nature **539**, 69 (2016), arXiv:1606.07494 [hep-lat].
- [18] M. P. Lombardo and A. Trunin, Int. J. Mod. Phys. A **35**, 2030010 (2020), arXiv:2005.06547 [hep-lat].
- [19] A. Athenodorou, C. Bonanno, C. Bonati, G. Clemente, F. D'Angelo, M. D'Elia, L. Maio, G. Martinelli, F. Sanfilippo, and A. Todaro, J. High Energy Phys. **10** (2022), 197, arXiv:2208.08921 [hep-lat].
- [20] A. Butti, A. Pelissetto, and E. Vicari, J. High Energy Phys. **08** (2003), 029, arXiv:hep-ph/0307036.
- [21] F. Basile, A. Pelissetto, and E. Vicari, J. High Energy Phys. **02** (2005), 044, arXiv:hep-th/0412026.
- [22] A. Pelissetto and E. Vicari, Phys. Rev. D **88**, 105018 (2013), arXiv:1309.5446 [hep-lat].
- [23] F. Giacosa, G. Kovács, P. Kovács, R. D. Pisarski, and F. Rennecke, Phys. Rev. D **111**, 016014 (2025), arXiv:2410.08185 [hep-ph].
- [24] E. Meggiolaro, Z. Phys. C **62**, 669 (1994).
- [25] M. Marchi and E. Meggiolaro, Nucl. Phys. B **665**, 425 (2003), arXiv:hep-ph/0301247.
- [26] E. Meggiolaro and A. Mordà, Phys. Rev. D **88**, 096010 (2013), arXiv:1309.4598 [hep-ph].
- [27] E. Meggiolaro, Phys. Rev. D **99**, 114009 (2019), arXiv:1903.10307 [hep-ph].
- [28] G. Fejős, Phys. Rev. D **105**, L071506 (2022), arXiv:2201.07909 [hep-ph].
- [29] G. Fejős and T. Hatsuda, Phys. Rev. D **110**, 016021 (2024), arXiv:2404.00554 [hep-ph].
- [30] J. Bernhardt and C. S. Fischer, Phys. Rev. D **108**, 114018 (2023), arXiv:2309.06737 [hep-ph].
- [31] H.-T. Ding *et al.* (HotQCD collaboration), Phys. Rev. Lett. **123**, 062002 (2019), arXiv:1903.04801 [hep-lat].
- [32] S. Chandrasekharan, D. Chen, N. H. Christ, W.-J. Lee, R. Mawhinney, and P. M. Vranas, Phys. Rev. Lett. **82**, 2463 (1999), arXiv:hep-lat/9807018.
- [33] V. Dick, F. Karsch, E. Laermann, S. Mukherjee, and S. Sharma, Phys. Rev. D **91**, 094504 (2015), arXiv:1502.06190 [hep-lat].
- [34] H. T. Ding, S. T. Li, S. Mukherjee, A. Tomiya, X. D. Wang, and Y. Zhang, Phys. Rev. Lett. **126**, 082001 (2021), arXiv:2010.14836 [hep-lat].
- [35] O. Kaczmarek, L. Mazur, and S. Sharma, Phys. Rev. D **104**, 094518 (2021), arXiv:2102.06136 [hep-lat].
- [36] O. Kaczmarek, R. Shanker, and S. Sharma, Phys. Rev. D **108**, 094501 (2023), arXiv:2301.11610 [hep-lat].
- [37] A. Bazavov *et al.* (HotQCD collaboration), Phys. Rev. D **86**, 094503 (2012), arXiv:1205.3535 [hep-lat].
- [38] M. I. Buchoff *et al.*, Phys. Rev. D **89**, 054514 (2014), arXiv:1309.4149 [hep-lat].
- [39] R. V. Gavai, M. E. Jaensch, O. Kaczmarek, F. Karsch, M. Sarkar, R. Shanker, S. Sharma, S. Sharma, and T. Ueding, Phys. Rev. D **111**, 034507 (2025), arXiv:2411.10217 [hep-lat].
- [40] S. Aoki, Y. Aoki, G. Cossu, H. Fukaya, S. Hashimoto, T. Kaneko, C. Rohrhofer, and K. Suzuki (JLQCD collaboration), Phys. Rev. D **103**, 074506 (2021), arXiv:2011.01499 [hep-lat].
- [41] S. Aoki, Y. Aoki, H. Fukaya, S. Hashimoto, I. Kanamori, T. Kaneko, Y. Nakamura, C. Rohrhofer, K. Suzuki, and D. Ward (JLQCD collaboration), PoS LATTICE2023, 185 (2024), arXiv:2401.14022 [hep-lat].
- [42] G. Cossu, S. Aoki, H. Fukaya, S. Hashimoto, T. Kaneko, H. Matsufuru, and J.-I. Noaki, Phys. Rev. D **87**, 114514 (2013), [Erratum: Phys. Rev. D **88**, 019901 (2013)], arXiv:1304.6145 [hep-lat].
- [43] A. Tomiya, G. Cossu, S. Aoki, H. Fukaya, S. Hashimoto, T. Kaneko, and J. Noaki, Phys. Rev. D **96**, 034509 (2017), [Addendum: Phys. Rev. D **96**, 079902 (2017)], arXiv:1612.01908 [hep-lat].
- [44] B. B. Brandt, A. Francis, H. B. Meyer, O. Philipsen, D. Robaina, and H. Wittig, J. High Energy Phys. **12** (2016), 158, arXiv:1608.06882 [hep-lat].
- [45] F. Cuteri, O. Philipsen, and A. Sciarra, J. High Energy Phys. **11** (2021), 141, arXiv:2107.12739 [hep-lat].
- [46] L. Dini, P. Hegde, F. Karsch, A. Lahiri, C. Schmidt, and S. Sharma, Phys. Rev. D **105**, 034510 (2022), arXiv:2111.12599 [hep-lat].
- [47] A. Pelissetto, A. Tripodo, and E. Vicari, Phys. Rev. D **96**, 034505 (2017), arXiv:1706.04365 [hep-lat].
- [48] C. Bonati, A. Pelissetto, and E. Vicari, Phys. Rev. D **101**, 034505 (2020), arXiv:2001.01132 [cond-mat.stat-mech].
- [49] H. B. Nielsen and M. Ninomiya, Phys. Lett. B **105**, 219 (1981).
- [50] H. B. Nielsen and M. Ninomiya, Nucl. Phys. B **185**, 20 (1981), [Erratum: Nucl. Phys. B **195**, 541 (1982)].
- [51] H. B. Nielsen and M. Ninomiya, Nucl. Phys. B **193**, 173 (1981).
- [52] D. Friedan, Commun. Math. Phys. **85**, 481 (1982).
- [53] P. H. Ginsparg and K. G. Wilson, Phys. Rev. D **25**, 2649 (1982).
- [54] P. Hasenfratz, V. Laliena, and F. Niedermayer, Phys. Lett. B **427**, 125 (1998), arXiv:hep-lat/9801021.
- [55] P. Hasenfratz, Nucl. Phys. B **525**, 401 (1998), arXiv:hep-lat/9802007.
- [56] M. Lüscher, Phys. Lett. B **428**, 342 (1998), arXiv:hep-lat/9802011.
- [57] Y. Kikukawa and A. Yamada, Nucl. Phys. B **547**, 413 (1999), arXiv:hep-lat/9808026.
- [58] H. Neuberger, Phys. Lett. B **427**, 353 (1998), arXiv:hep-lat/9801031.
- [59] I. Horváth, Phys. Rev. D **60**, 034510 (1999), arXiv:hep-lat/9901014.
- [60] D. B. Kaplan, Phys. Lett. B **288**, 342 (1992), arXiv:hep-lat/9206013.
- [61] Y. Shamir, Nucl. Phys. B **406**, 90 (1993), arXiv:hep-lat/9303005.
- [62] V. Furman and Y. Shamir, Nucl. Phys. B **439**, 54 (1995), arXiv:hep-lat/9405004.
- [63] A. Boriçi, Nucl. Phys. B Proc. Suppl. **83**, 771 (2000), arXiv:hep-lat/9909057.
- [64] T.-W. Chiu, Phys. Rev. Lett. **90**, 071601 (2003), arXiv:hep-lat/0209153.
- [65] R. C. Brower, H. Neff, and K. Orginos, Nucl. Phys. B Proc. Suppl. **153**, 191 (2006), arXiv:hep-lat/0511031.
- [66] R. C. Brower, H. Neff, and K. Orginos, Comput. Phys. Commun. **220**, 1 (2017), arXiv:1206.5214 [hep-lat].
- [67] R. Narayanan and H. Neuberger, Nucl. Phys. B **412**, 574 (1994), arXiv:hep-lat/9307006.
- [68] R. Narayanan and H. Neuberger, Phys. Rev. Lett. **71**, 3251 (1993), arXiv:hep-lat/9308011.
- [69] H. Neuberger, Phys. Lett. B **417**, 141 (1998), arXiv:hep-lat/9707022.
- [70] P. Hasenfratz and F. Niedermayer, Nucl. Phys. B **414**, 785 (1994), arXiv:hep-lat/9308004.
- [71] W. Bietenholz and U. J. Wiese, Nucl. Phys. B **464**, 319 (1996), arXiv:hep-lat/9510026.
- [72] T. A. DeGrand, A. Hasenfratz, P. Hasenfratz, and

- F. Niedermayer, Nucl. Phys. B **454**, 587 (1995), arXiv:hep-lat/9506030.
- [73] P. Hasenfratz, S. Hauswirth, T. Jörg, F. Niedermayer, and K. Holland, Nucl. Phys. B **643**, 280 (2002), arXiv:hep-lat/0205010.
- [74] T. D. Cohen, Phys. Rev. D **54**, R1867 (1996), arXiv:hep-ph/9601216.
- [75] N. J. Evans, S. D. H. Hsu, and M. Schwetz, Phys. Lett. B **375**, 262 (1996), arXiv:hep-ph/9601361.
- [76] S. H. Lee and T. Hatsuda, Phys. Rev. D **54**, R1871 (1996), arXiv:hep-ph/9601373.
- [77] T. D. Cohen, in *APCTP Workshop on Astro-Hadron Physics: Properties of Hadrons in Matter* (1997) pp. 100–114, arXiv:nucl-th/9801061.
- [78] S. Aoki, H. Fukaya, and Y. Taniguchi, Phys. Rev. D **86**, 114512 (2012), arXiv:1209.2061 [hep-lat].
- [79] T. Kanazawa and N. Yamamoto, J. High Energy Phys. **01** (2016), 141, arXiv:1508.02416 [hep-th].
- [80] V. Azcoiti, Phys. Rev. D **107**, 114516 (2023), arXiv:2304.14725 [hep-lat].
- [81] M. Giordano, Phys. Rev. D **110**, L091504 (2024), arXiv:2404.03546 [hep-lat].
- [82] M. Giordano, PoS **LATTICE2024**, 188 (2025), arXiv:2412.02517 [hep-lat].
- [83] N. Carabba and E. Meggiolaro, Phys. Rev. D **105**, 054034 (2022), arXiv:2106.10074 [hep-ph].
- [84] H.-T. Ding, W.-P. Huang, S. Mukherjee, and P. Petreczky, Phys. Rev. Lett. **131**, 161903 (2023), arXiv:2305.10916 [hep-lat].
- [85] H. Leutwyler and A. V. Smilga, Phys. Rev. D **46**, 5607 (1992).
- [86] A. V. Smilga and J. Stern, Phys. Lett. B **318**, 531 (1993).
- [87] J. J. M. Verbaarschot, Phys. Rev. Lett. **72**, 2531 (1994), arXiv:hep-th/9401059.
- [88] J. C. Osborn, D. Toublan, and J. J. M. Verbaarschot, Nucl. Phys. B **540**, 317 (1999), arXiv:hep-th/9806110.
- [89] P. H. Damgaard and H. Fukaya, J. High Energy Phys. **01** (2009), 052, arXiv:0812.2797 [hep-lat].
- [90] L. Giusti and M. Lüscher, J. High Energy Phys. **03**, 013 (2009), arXiv:0812.3638 [hep-lat].
- [91] J. J. M. Verbaarschot and T. Wettig, Ann. Rev. Nucl. Part. Sci. **50**, 343 (2000), arXiv:hep-ph/0003017.
- [92] D. Toublan and J. J. M. Verbaarschot, Nucl. Phys. B **603**, 343 (2001), arXiv:hep-th/0012144.
- [93] V. Azcoiti, Phys. Rev. D **94**, 094505 (2016), arXiv:1609.01230 [hep-lat].
- [94] V. Azcoiti, Phys. Rev. D **96**, 014505 (2017), arXiv:1704.04906 [hep-lat].
- [95] V. Azcoiti, Phys. Rev. D **100**, 074511 (2019), arXiv:1907.01872 [hep-lat].
- [96] V. Azcoiti, Symmetry **13**, 209 (2021), arXiv:2101.06439 [hep-lat].
- [97] M. Giordano, J. Phys. A **54**, 37LT01 (2021), arXiv:2009.00486 [hep-th].
- [98] M. Giordano, PoS **LATTICE2021**, 401 (2022), arXiv:2110.12250 [hep-lat].
- [99] M. Giordano, J. High Energy Phys. **12** (2022), 103, arXiv:2206.11109 [hep-th].
- [100] R. G. Edwards, U. M. Heller, J. E. Kiskis, and R. Narayanan, Phys. Rev. D **61**, 074504 (2000), arXiv:hep-lat/9910041.
- [101] A. Alexandru and I. Horváth, Phys. Rev. D **92**, 045038 (2015), arXiv:1502.07732 [hep-lat].
- [102] T. G. Kovács and R. Á. Vig, Phys. Rev. D **97**, 014502 (2018), arXiv:1706.03562 [hep-lat].
- [103] A. Alexandru and I. Horváth, Phys. Rev. D **100**, 094507 (2019), arXiv:1906.08047 [hep-lat].
- [104] R. Á. Vig and T. G. Kovács, Phys. Rev. D **103**, 114510 (2021), arXiv:2101.01498 [hep-lat].
- [105] T. G. Kovács, PoS **LATTICE2021**, 238 (2022), arXiv:2112.05454 [hep-lat].
- [106] X.-L. Meng, P. Sun, A. Alexandru, I. Horváth, K.-F. Liu, G. Wang, and Y.-B. Yang (χ QCD, CLQCD), J. High Energy Phys. **12** (2024), 101, arXiv:2305.09459 [hep-lat].
- [107] A. Alexandru, C. Bonanno, M. D’Elia, and I. Horváth, Phys. Rev. D **110**, 074515 (2024), arXiv:2404.12298 [hep-lat].
- [108] Z. Fodor, A. Y. Kotov, T. G. Kovács, and K. K. Szabó, in *16th Conference on Quark Confinement and the Hadron Spectrum* (2025) arXiv:2503.22243 [hep-lat].
- [109] T. Kanazawa and N. Yamamoto, Phys. Rev. D **91**, 105015 (2015), arXiv:1410.3614 [hep-th].
- [110] T. G. Kovács, Phys. Rev. Lett. **132**, 131902 (2024), arXiv:2311.04208 [hep-lat].
- [111] S. Chandrasekharan, Phys. Rev. D **60**, 074503 (1999), arXiv:hep-lat/9805015.
- [112] F. Niedermayer, Nucl. Phys. B Proc. Suppl. **73**, 105 (1999), arXiv:hep-lat/9810026.
- [113] C. Vafa and E. Witten, Nucl. Phys. B **234**, 173 (1984).
- [114] M. Giordano, Phys. Rev. D **107**, 114509 (2023), arXiv:2303.03109 [hep-lat].
- [115] R. Aloisio, V. Azcoiti, G. Di Carlo, A. Galante, and A. F. Grillo, Nucl. Phys. B **606**, 322 (2001), arXiv:hep-lat/0011079.
- [116] V. Azcoiti, G. Di Carlo, E. Follana, and A. Vaquero, J. High Energy Phys. **07** (2010), 047, arXiv:1004.3463 [hep-lat].
- [117] T. Bröcker and T. tom Dieck, *Representations of Compact Lie Groups*, 1st ed., Graduate Texts in Mathematics, Vol. 98 (Springer, Berlin, Heidelberg, 1985).
- [118] M. L. Mehta, *Random matrices*, 3rd ed., Pure and Applied Mathematics, Vol. 142 (Academic Press, New York, 2004).
- [119] L. A. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, Vol. 171 (American Mathematical Society, Providence, 2011).
- [120] J. L. van Hemmen, J. Phys. A: Math. Gen. **15**, 3891 (1982).
- [121] M. Lüscher, Phys. Lett. B **593**, 296 (2004), arXiv:hep-th/0404034.
- [122] E. Vicari and H. Panagopoulos, Phys. Rept. **470**, 93 (2009), arXiv:0803.1593 [hep-th].
- [123] T. Blum, N. H. Christ, C. Cristian, C. Dawson, X. Liao, G. Liu, R. Mawhinney, L. Wu, and Y. Zhestkov, Phys. Rev. D **65**, 014504 (2002), arXiv:hep-lat/0105006.
- [124] J. Goldstone, A. Salam, and S. Weinberg, Phys. Rev. **127**, 965 (1962).
- [125] R. Lange, Phys. Rev. Lett. **14**, 3 (1965).
- [126] D. Kastler, D. W. Robinson, and A. Swieca, Commun. Math. Phys. **2**, 108 (1966).
- [127] J. A. Swieca, Commun. Math. Phys. **4**, 1 (1967).
- [128] G. Morchio and F. Strocchi, Commun. Math. Phys. **99**, 153 (1985).
- [129] G. Morchio and F. Strocchi, J. Math. Phys. **28**, 622

- (1987).
- [130] C. Alexandrou, E. Follana, H. Panagopoulos, and E. Vicari, Nucl. Phys. B **580**, 394 (2000), arXiv:hep-lat/0002010.
 - [131] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed., Applied Mathematics Series, Vol. 55 (United States Department of Commerce, National Bureau of Standards, United States, 1972).
 - [132] L. Giusti, G. C. Rossi, and M. Testa, Phys. Lett. B **587**, 157 (2004), arXiv:hep-lat/0402027.
 - [133] L. Del Debbio, H. Panagopoulos, and E. Vicari, J. High Energy Phys. **09** (2004), 028, arXiv:hep-th/0407068.
 - [134] C. Bonati, M. D’Elia, H. Panagopoulos, and E. Vicari, Phys. Rev. Lett. **110**, 252003 (2013), arXiv:1301.7640 [hep-lat].
 - [135] C. Bonati, M. D’Elia, G. Martinelli, F. Negro, F. Sanfilippo, and A. Todaro, J. High Energy Phys. **11** (2018), 170, arXiv:1807.07954 [hep-lat].
 - [136] C. Bonanno, G. Clemente, M. D’Elia, L. Maio, and L. Parente, J. High Energy Phys. **08** (2024), 236, arXiv:2404.14151 [hep-lat].
 - [137] B. J. Harrington and H. K. Shepard, Phys. Rev. D **17**, 2122 (1978).
 - [138] B. J. Harrington and H. K. Shepard, Phys. Rev. D **18**, 2990 (1978).
 - [139] T. C. Kraan and P. van Baal, Phys. Lett. B **428**, 268 (1998), arXiv:hep-th/9802049.
 - [140] T. C. Kraan and P. van Baal, Nucl. Phys. B **533**, 627 (1998), arXiv:hep-th/9805168.
 - [141] T. C. Kraan and P. van Baal, Phys. Lett. B **435**, 389 (1998), arXiv:hep-th/9806034.
 - [142] K.-M. Lee and P. Yi, Phys. Rev. D **56**, 3711 (1997), arXiv:hep-th/9702107.
 - [143] K.-M. Lee, Phys. Lett. B **426**, 323 (1998), arXiv:hep-th/9802012.
 - [144] K.-M. Lee and C.-h. Lu, Phys. Rev. D **58**, 025011 (1998), arXiv:hep-th/9802108.
 - [145] M. García Pérez, A. González-Arroyo, C. Pena, and P. van Baal, Phys. Rev. D **60**, 031901 (1999), arXiv:hep-th/9905016.
 - [146] M. N. Chernodub, T. C. Kraan, and P. van Baal, Nucl. Phys. B Proc. Suppl. **83**, 556 (2000), arXiv:hep-lat/9907001.
 - [147] D. Diakonov, Proc. Int. Sch. Phys. Fermi **130**, 397 (1996), arXiv:hep-ph/9602375.
 - [148] T. Schäfer and E. V. Shuryak, Rev. Mod. Phys. **70**, 323 (1998), arXiv:hep-ph/9610451.
 - [149] D. Diakonov, Nucl. Phys. B Proc. Suppl. **195**, 5 (2009), arXiv:0906.2456.
 - [150] D. J. Gross, R. D. Pisarski, and L. G. Yaffe, Rev. Mod. Phys. **53**, 43 (1981).
 - [151] A. Boccaletti and D. Nógrádi, J. High Energy Phys. **03** (2020), 045, arXiv:2001.03383 [hep-ph].
 - [152] P. Di Vecchia and G. Veneziano, Nucl. Phys. B **171**, 253 (1980).
 - [153] J. R. Silvester, Math. Gaz. **84**, 460 (2000).