Canonical quantization of nonlinear sigma models with a theta term and applications to symmetry-protected topological phases

Matthew F. Lapa[*] and Taylor L. Hughes

Department of Physics and Institute for Condensed Matter Theory,
University of Illinois at Urbana-Champaign, Champaign, Illinois 61801-3080, USA

We canonically quantize $O(D+2)$ nonlinear sigma models (NLSMs) with a theta term on arbitrary smooth, closed, connected, oriented D-dimensional spatial manifolds $\mathcal{M}$, with the goal of proving the suitability of these models for describing symmetry-protected topological (SPT) phases of bosons in D spatial dimensions. We show that in the disordered phase of the NLSM, and when the coefficient θ of the theta term is an integer multiple of 2π, the theory on $\mathcal{M}$ has a unique ground state and a finite energy gap to all excitations. We also construct the ground state wave functional of the NLSM in this parameter regime, and we show that it is independent of the metric on $\mathcal{M}$ and given by the exponential of a *Wess-Zumino* term for the NLSM field, in agreement with previous results on flat space. Our results show that the NLSM in the disordered phase and at $\theta = 2\pi k, k \in \mathbb{Z}$, has a symmetry-preserving ground state but no topological order (i.e., no topology-dependent ground state degeneracy), making it an ideal model for describing SPT phases of bosons. Thus, our work places previous results on SPT phases derived using NLSMs on solid theoretical ground. To canonically quantize the NLSM on $\mathcal{M}$, we use Dirac's method for the quantization of systems with second class constraints, suitably modified to account for the curvature of space. In a series of four Appendixes, we provide the technical background needed to follow the discussion in the main sections of the paper.

I. INTRODUCTION

Nonlinear sigma models (NLSMs), quantum field theories in which the field is a map from spacetime to a *target manifold* $\mathcal{T}$, have a long history of study in both high-energy and condensed matter physics [1–13]. The earliest example of such a model was introduced in Ref. [1] by Gell-Mann and Lévy and applied to the theory of β-decay. Some time later, these models were brought to the attention of condensed matter physicists when, for example, Haldane showed that the $O(3)$ NLSM with a *theta term* and coefficient ("theta angle") $\theta = 2\pi S$ described the continuum limit of a spin-S Heisenberg chain in one spatial dimension [8]. The theta term is a particular topological term that can be added to the NLSM action when the dimension of the target manifold $\mathcal{T}$ is the same as the dimension of spacetime. Very recently, interest in NLSMs with a theta term has experienced a revival due to the proposal, formalized in Ref. [14], that the disordered phase of an $O(D+2)$ NLSM with a theta term and $\theta = 2\pi k, k \in \mathbb{Z}$, can describe (a subset of) symmetry-protected topological (SPT) phases of bosons in D spatial dimensions (for general references on SPT phases, we refer the reader to Refs. [15–20]). The work of many authors has provided a trove of evidence supporting this proposal [21–31]. However, despite the many successes of the NLSM description, several outstanding issues still require clarification. In particular, the ground state of the $O(D+2)$ NLSM, in the parameter regime relevant for the description of SPT

phases, has only been studied on flat space [26]. A thorough study of the ground state (or states) of this theory, as well as the energy gap to the first excited states, should be carried out on *arbitrary* curved spatial manifolds (with various topologies) in order to establish the suitability of the $O(D+2)$ NLSM with a theta term (in the disordered phase and at $\theta = 2\pi k, k \in \mathbb{Z}$) as a model of bosonic SPT phases. For example, it is important to check that the model has a unique ground state no matter the underlying spatial manifold, as befits a system without topological order. It is the purpose of this paper to provide such a study.

Let us begin by providing the setup for the description of bosonic SPT phases in D spatial dimensions using $O(D+2)$ NLSMs [14]. The target manifold of the $O(D+2)$ NLSM is $\mathcal{T} = S^{D+1}$, the unit $(D+1)$-sphere, and the NLSM field is a $(D+2)$-component unit vector field n with components n^a, $a = 1, \ldots, D+2$. If the SPT phase is protected by a symmetry group G, then the symmetry transformation information is naturally encoded in the NLSM description of that phase via a homomorphism $\sigma: G \to SO(D+2)$, which assigns to each group element $g \in G$ an $SO(D+2)$ matrix $\sigma(g)$ which rotates the NLSM field n.[1] The reason for mapping G into $SO(D+2)$

---

[1]For symmetries which also have a nontrivial action on spacetime, for example time-reversal symmetry, the homomorphism instead takes the form $\sigma: G \to O(D+2)$. In other words, for symmetries with a nontrivial action on spacetime, one must allow for the possibility of orientation-reversing transformations on the target space of the NLSM.

*lapa2@illinois.edu

is that the $O(D+2)$ NLSM *with* a theta term has an $SO(D+2)$ global symmetry, and so embedding G inside $SO(D+2)$ immediately guarantees the invariance of the NLSM description of the SPT phase under the action of G, at least at the classical level. At the quantum level, the symmetry is expected to be unbroken only in the disordered phase of the NLSM, which is the phase of interest for the description of SPT phases. For a given group G, many distinct homomorphisms σ are possible, and the different possibilities (modulo a notion of "independent NLSMs" explained in Ref. [14]) correspond to different SPT phases with the symmetry of the group G.

The NLSM description has been shown to capture many of the physical properties of bosonic SPT phases. For example, it can predict the structure of the ground state wave functional [26,32] and the braiding statistics of point and looplike excitations in gauged SPT phases [22]. In addition, in Ref. [27], an explicit connection was made between the NLSM classification of SPT phases of Refs. [14,21] and the group cohomology classification of bosonic SPT phases of Ref. [19]. Very recently, the present authors demonstrated how the NLSM description of bosonic SPT phases can be combined with the theory of gauged Wess-Zumino actions to compute the topological electromagnetic response of some bosonic SPT phases in all dimensions [31].

All of these works strongly support the idea that the $O(D+2)$ NLSM in its disordered phase, with a theta term and $\theta = 2\pi k$, $k \in \mathbb{Z}$, and a suitable symmetry assignment $\sigma : G \to SO(D+2)$, can describe nontrivial SPT phases with symmetry group G. However, several of the defining properties of an SPT phase are based on the behavior of the SPT phase when it is placed on (closed) spatial manifolds $\mathcal{M}$ of arbitrary topology, and the NLSM description of SPT phases has not yet been tested in this setting. To be precise, let $\mathcal{M}$ be an arbitrary smooth, closed,[2] connected, oriented, D-dimensional manifold. We also equip $\mathcal{M}$ with a Riemannian metric. In this paper, we prove the following three properties of the $O(D+2)$ NLSM when formulated on spatial manifolds $\mathcal{M}$ of this kind:

(1) The ground state of the $O(D+2)$ NLSM in the disordered phase and at $\theta = 2\pi k$, $k \in \mathbb{Z}$, on $\mathcal{M}$ is *unique*.

(2) The ground state wave functional of the NLSM on $\mathcal{M}$ is *independent* of the metric on $\mathcal{M}$ and is proportional to the exponential of a suitably defined *Wess-Zumino* term[3] for the NLSM field n, just as in the case on flat space [26].

(3) There is a finite energy gap between the ground state and the first excited state of the $O(D+2)$ NLSM (in the disordered phase and at $\theta = 2\pi k$, $k \in \mathbb{Z}$) on $\mathcal{M}$.

---

[2] A *closed* manifold is a compact manifold without boundary.
[3] For the case of the $O(3)$ NLSM on flat $(1+1)$-dimensional spacetime, the WZ form of the ground state wave functional was originally derived in Ref. [33].

These three properties together imply that the $O(D+2)$ NLSM in the disordered phase and at $\theta = 2\pi k$, $k \in \mathbb{Z}$, represents a system with SPT order but *not* topological order (no topology-dependent ground state degeneracy). In particular, the fact that the ground state wave functional involves a Wess-Zumino (WZ) term for **n** implies that the ground state is invariant under the action of the group G which protects the SPT phase. This is equivalent to the statement that the ground state does not spontaneously break the symmetry of the group G, which is a crucial property of an SPT phase (see, for example, the discussion in the introduction of Ref. [18]).

In order to prove these statements, we canonically quantize the $O(D+2)$ NLSM with a theta term on $(D+1)$-dimensional spacetimes of the form $\mathcal{M} \times \mathbb{R}$, where $\mathcal{M}$ is a D-dimensional spatial manifold and $\mathbb{R}$ represents time. As stated above, we assume that the spatial manifold $\mathcal{M}$ is smooth, closed, connected, and oriented, and we equip $\mathcal{M}$ with a Riemannian metric. The canonical quantization of the $O(N)$ NLSM on flat space and for various N, in various dimensions and with various topological terms has been considered previously in Refs. [7,34–38]. In particular, Ref. [37] considered the $O(3)$ NLSM with a theta term in one spatial dimension, and Refs. [36,38] considered the $O(3)$ NLSM with *Hopf term* in two spatial dimensions. To carry out the quantization, these references used Dirac's method [39,40] for the quantization of systems with second class constraints, and we follow the same route in this paper (with suitable modifications to account for the curvature of the space $\mathcal{M}$). This formalism is necessary to handle the constraint that the NLSM field **n** be a unit vector field, which is equivalent to the statement that the target space $\mathcal{T}$ of the $O(D+2)$ NLSM is the unit sphere S^{D+1}. After the quantization, we study this theory in its disordered limit, and in that limit, we prove the three properties of this model which are stated above.

This paper is organized as follows. In Sec. II, we introduce the NLSM with a theta term and discuss the canonical quantization of this model on flat space. In Sec. III, we compute the ground state wave functional and the energy gap of the NLSM in the disordered phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$ on flat space. In Sec. IV, we quantize the NLSM on curved spaces $\mathcal{M}$ and then compute the ground state wave functional and the energy gap as well as prove the uniqueness of the ground state of the theory on $\mathcal{M}$. Section V presents our conclusions. The paper also includes several Appendixes containing additional background material necessary for the discussion in the main sections of the paper. In Appendix A, we review the solution of the quantum mechanics problem of a particle constrained to the surface of a sphere S^{N-1}, equivalent to the $O(\mathbf{n})$ NLSM in one spacetime dimension. In Appendix B, we explain the need for regularization of the NLSM Hamiltonian, and we also discuss an alternative

regularization scheme from the one used in Sec. III. In Appendix C, we review the symplectic geometry approach to the Hamiltonian mechanics of a continuum field theory. Finally, in Appendix D, we explain a simple example of an intrinsic construction of a Wess-Zumino term for the NLSM field on $\mathcal{M}$ which does not require the use of a higher-dimensional manifold $\mathcal{B}$ which has $\mathcal{M}$ as its boundary.

II. NONLINEAR SIGMA MODELS, HAMILTONIAN FORMALISM, AND CANONICAL QUANTIZATION

In this section, we introduce the NLSM with a theta term and discuss its canonical quantization on flat space $\mathbb{R}^D$ using Dirac's method for quantization in the presence of constraints. We show in some detail that this system possesses only two second class constraints, regardless of the value of the coefficient θ of the theta term. We then compute the classical *Dirac brackets* for the NLSM using the two second class constraints. Finally, following Dirac's prescription, we obtain the commutation relations for the quantum theory from the Dirac brackets in the same way in which one obtains the commutation relations from the Poisson brackets of an unconstrained classical theory. We also discuss a functional Schrödinger representation of these commutation relations, previously used in a field theory context in Ref. [38], which we use throughout the paper for concrete calculations.

A. NLSM and the theta term

The $O(\mathbf{n})$ NLSM in $D + 1$ spacetime dimensions is a theory of an N-component vector field n with components n^a, $a = 1, …, N$, subject to the constraint $\mathbf{n} \cdot \mathbf{n} = n_a n^a = 1$ (so $\mathbf{n}$ is a *unit* vector field). The action for the NLSM takes the form

$$S[\mathbf{n}] = \int d^{D+1}x \frac{1}{2f} (\partial^\mu n_a)(\partial_\mu n^a), \qquad (2.1)$$

where x^μ, $\mu = 0, …, D$, $(x^0 = t)$ are the spacetime coordinates, $d^{D+1}x = dx^0 \cdots dx^D$, and we sum over all indices (latin or greek) which appear once as a superscript and once as a subscript. Also, we use the notation $\partial_\mu \equiv \frac{\partial}{\partial x^\mu}$. Latin indices are raised and lowered using the metric δ_{ab} and its inverse δ^{ab}, while greek indices are raised and lowered using the "mostly minus" Minkowski metric $\eta_{\mu\nu}$ [i.e., as a matrix $\eta = \text{diag}(1, -1, …, -1)$] and its inverse $\eta^{\mu\nu}$. We use units in which $c = \hbar = 1$. In this section of the paper, we work on flat, $(D + 1)$-dimensional Minkowksi spacetime, denoted by $\mathbb{R}^{D,1}$. Finally, the quantity f is a positive coupling constant with units of $(\text{length})^{D-1}$ (so that the bare NLSM field is dimensionless). For $D > 1$, the model is in an ordered phase for small f and a disordered phase

for large f.[4] For $D = 1$, there is no ordered phase, and the renormalization group flow at any scale is always toward the disordered phase [6].

The target manifold $\mathcal{T}$ of the $O(\mathbf{n})$ NLSM is the unit sphere S^{N-1}, and in the special case that $N = D + 2$, the dimension of the target manifold is the same as the dimension of spacetime. For this particular value of N, there is an interesting topological term, called the theta term, which can be added to the action for the NLSM. This topological term is simply the pullback to spacetime of the volume form on S^{D+1} via the map $n : \mathbb{R}^{D,1} \to S^{D+1}$. We now explain this in more detail.

Let ω_{D+1} be the volume form on S^{D+1} (with the radius of the sphere set to 1). If the sphere is parametrized by the coordinates n^a, $a = 1, …, D + 2$, subject to the constraint $n^a n_a = 1$, then the volume form in these coordinates is

$$\omega_{D+1} = \sum_{a=1}^{D+2} (-1)^{a-1} n^a dn^1 \wedge \cdots \wedge \overline{dn^a} \wedge \cdots \wedge dn^{D+2}, \qquad (2.2)$$

where the overline means to omit that term from the wedge product. In what follows, we also use the notation $\mathcal{A}_{D+1} \equiv \text{Area}[S^{D+1}] = \frac{2\pi^{\frac{D+2}{2}}}{\Gamma(\frac{D+2}{2})}$ for the area of S^{D+1} (so $\mathcal{A}_1 = 2\pi$, $\mathcal{A}_2 = 4\pi$, etc.). With this notation, the theta term can be written compactly as

$$S_\theta[\mathbf{n}] = \frac{1}{\mathcal{A}_{D+1}} \int_{\mathbb{R}^{D,1}} \mathbf{n}^* \omega_{D+1}, \qquad (2.3)$$

where $\mathbf{n}^* \omega_{D+1}$ denotes the pullback to spacetime of the form ω_{D+1} via the map $\mathbf{n} : \mathbb{R}^{D,1} \to S^{D+1}$. In coordinates, the theta term can be written as

$$S_\theta[\mathbf{n}] = \frac{1}{\mathcal{A}_{D+1}} \int d^{D+1}x \epsilon_{a_1 \cdots a_{D+2}} n^{a_1} \partial_t n^{a_2} \partial_1 n^{a_3} \cdots \partial_D n^{a_{D+2}}. \qquad (2.4)$$

The full action for the NLSM with a theta term takes the form

$$S[\mathbf{n}] = \int d^{D+1}x \frac{1}{2f} (\partial^\mu n_a)(\partial_\mu n^a) - \theta S_\theta[\mathbf{n}], \qquad (2.5)$$

---

[4]For $D > 1$, the coupling constant f has units, so the magnitude of the coupling constant can only be established with respect to a reference scale f^*. Such an f^* is naturally provided by the location of the zero of the beta function $\beta(f)$ of the coupling constant f for $D > 1$ (Ref. [6] showed that there is no zero in $D = 1$). The value $f = f^*$ is the point at which the renormalization group flow of f crosses over from a flow toward the ordered phase to a flow toward the disordered phase at low energies.

where the dimensionless parameter θ is the theta angle discussed in Sec. I. The Lagrangian which follows from this action is

$$\mathcal{L} = \frac{1}{2f}(\partial^\mu n_a)(\partial_\mu n^a) + \frac{\theta}{\mathcal{A}_{D+1}}B_a(\partial_t n^a), \quad (2.6)$$

where we introduced

$$B_a = \epsilon_{aa_1\cdots a_{D+1}}n^{a_1}\partial_1 n^{a_2}\cdots\partial_D n^{a_{D+1}}. \quad (2.7)$$

At this point, we are now in a position to proceed with the canonical quantization of this system.

B. Quantization of constrained systems

Due to the constraint $n^a n_a = 1$ on the NLSM field, the canonical quantization of the NLSM requires Dirac's theory of constrained Hamiltonian systems and the use of Dirac brackets in particular [39]. Let us briefly sketch, following Ref. [40], the steps involved in the quantization of a constrained system. We first recall some basic definitions. A constraint ϕ is a function on phase space which is to be set equal to zero. Two functions f and g on phase space are *strongly equivalent* if they are equal throughout phase space. This is just the ordinary equality of functions, $f = g$. Two functions f and g are called *weakly equivalent* if they become equal when all constraints ϕ are set to zero. Weak equivalence of two functions f and g is denoted by $f \approx g$.[5]

The first step in the quantization of a constrained system is to find all of the constraints. This step involves the construction of a modified Hamiltonian $\tilde{H}$ such that the time derivative of any constraint ϕ, as given by the modified Hamiltonian, weakly vanishes. In other words, we have $\frac{d}{dt}\phi = \{\phi, \tilde{H}\} \approx 0$, where $\{\cdot, \cdot\}$ denotes the ordinary Poisson bracket. This is a consistency condition on the time evolution of the dynamical system, as constraints should not change with time. In general, the modified Hamiltonian $\tilde{H}$ is distinct from the original Hamiltonian H obtained from the Lagrangian via a Legendre transformation.

The next step in the quantization is to isolate the *second class constraints* and then use these constraints to construct the Dirac bracket. Recall that the second class constraints are those constraints which have nonvanishing Poisson brackets with each other. Let $\psi_I(\mathbf{x})$, $I = 1, \ldots, N_c$, denote the second class constraints in our system,[6] and define the functions $M_{IJ}(\mathbf{x}, \mathbf{y})$ by

$$M_{IJ}(\mathbf{x}, \mathbf{y}) = \{\psi_I(\mathbf{x}), \psi_J(\mathbf{y})\}. \quad (2.8)$$

The functions $M_{IJ}(\mathbf{x}, \mathbf{y})$ should be viewed as the components of a matrix $\mathbf{M}$ with discrete indices I, J and continuous spatial indices $(\mathbf{x}, \mathbf{y})$. In terms of this matrix, the Dirac bracket for two functions $f(\mathbf{x})$ and $g(\mathbf{y})$ on phase space is given by

$$\{f(\mathbf{x}), g(\mathbf{y})\}_D = \{f(\mathbf{x}), g(\mathbf{y})\}$$
$$- \sum_{I,J=1}^{N_c}\int d^D\mathbf{z}d^D\mathbf{z}'\{f(\mathbf{x}), \psi_I(\mathbf{z})\}$$
$$\times M_{IJ}^{-1}(\mathbf{z}, \mathbf{z}')\{\psi_J(\mathbf{z}'), g(\mathbf{y})\}, \quad (2.9)$$

where the functions $M_{IJ}^{-1}(\mathbf{x}, \mathbf{y})$ are the components of the inverse matrix to $\mathbf{M}$ in the sense that

$$\sum_{J=1}^{N_c}\int d^D\mathbf{y}M_{IJ}(\mathbf{x}, \mathbf{y})M_{JK}^{-1}(\mathbf{y}, \mathbf{z}) = \delta_{IK}\delta^{(D)}(\mathbf{x} - \mathbf{z}). \quad (2.10)$$

Finally, the equal-time commutators in the quantized theory are obtained from the classical Dirac brackets according to the rule

$$\{f(\mathbf{x}), g(\mathbf{y})\}_D \to -i[\hat{f}(\mathbf{x}), \hat{g}(\mathbf{y})], \quad (2.11)$$

where $\hat{f}(\mathbf{x})$ is the operator in the quantum theory corresponding to the classical function $f(\mathbf{x})$. The quantum theory is then defined by the operator $\hat{H}$ corresponding to the *original* Hamiltonian H obtained from the Legendre transformation of the Lagrangian (not the modified Hamiltonian $\tilde{H}$ discussed above), combined with the equal-time commutators obtained from the classical Dirac brackets.[7]

C. Canonical quantization of the $O(D+2)$ NLSM with a theta term

We now carry out the program outlined in the last section for the NLSM with a theta term. To start, the momentum conjugate to n^a is

$$\pi_a \equiv \frac{\partial\mathcal{L}}{\partial(\partial_t n^a)} = \frac{1}{f}(\partial_t n_a) + \frac{\theta}{\mathcal{A}_{D+1}}B_a, \quad (2.12)$$

and the Hamiltonian obtained from the Lagrangian via the Legendre transformation $H = \int d^D\mathbf{x}(\pi_a\partial_t n^a - \mathcal{L})$ is

$$H = \int d^D\mathbf{x}\left\{\frac{f}{2}\left(\pi_a - \frac{\theta}{\mathcal{A}_{D+1}}B_a\right)^2 + \frac{1}{2f}(\nabla n^a)^2\right\}, \quad (2.13)$$

---

[5]In this paper, the symbol "$\approx$" is *only* used to denote weak equivalence.

[6]In a field theory, the constraints are usually functions of the position $\mathbf{x}$ in space.

[7]By construction, in the quantum theory defined in this way, one has $[\hat{\tilde{H}}, \hat{\mathcal{O}}] \approx [\hat{H}, \hat{\mathcal{O}}]$ for any operator $\hat{\mathcal{O}}$. Thus, the operator $\hat{H}$ corresponding to the original Hamiltonian may be used in the quantum theory instead of the operator $\hat{\tilde{H}}$ corresponding to the modified Hamiltonian.

where $x = (x^1, \ldots, x^D)$ is the vector of spatial coordinates, $d^D\mathbf{x} = \prod_{j=1}^D dx^j$, ∇ is the spatial gradient, $(\nabla n^a)^2 \equiv (\nabla n^a) \cdot (\nabla n_a)$, etc. For this system, the Poisson bracket of any two functionals F_1 and F_2 of the fields $n^a(\mathbf{x})$ and their conjugate momenta $\pi_a(\mathbf{x})$ is given by

$$\{F_1, F_2\} = \int d^D\mathbf{x}\left(\frac{\delta F_1}{\delta n^a(\mathbf{x})}\frac{\delta F_2}{\delta \pi_a(\mathbf{x})} - \frac{\delta F_1}{\delta \pi_a(\mathbf{x})}\frac{\delta F_2}{\delta n^a(\mathbf{x})}\right), \tag{2.14}$$

where $\frac{\delta}{\delta n^a(\mathbf{x})}$ is a functional derivative.

We now move on to the problem of finding all of the constraints for this system. To begin with, we have only the single constraint

$$\psi_1(\mathbf{x}) = n_a(\mathbf{x})n^a(\mathbf{x}) - 1. \tag{2.15}$$

Setting $\psi_1(\mathbf{x}) = 0$ enforces the condition that $\mathbf{n}$ is a unit vector field. Following Dirac's procedure, we now use this constraint to construct a modified Hamiltonian $\tilde{H}$ such that $\{\psi_1(\mathbf{x}), \tilde{H}\} \approx 0$. As a first attempt toward the construction of $\tilde{H}$, we define the modified Hamiltonian H' by

$$H' = H + \int d^D y u_1(\mathbf{y})\psi_1(\mathbf{y}), \tag{2.16}$$

where $u_1(\mathbf{x})$ is an as yet undetermined function. Note that $H' \approx H$ since the constraint weakly vanishes, $\psi_1(\mathbf{x}) \approx 0$. Using the product rule for the Poisson bracket, we find that

$$\{\psi_1(\mathbf{x}), H'\} = \{\psi_1(\mathbf{x}), H\} + \int d^D\mathbf{y}(\{\psi_1(\mathbf{x}), u_1(\mathbf{y})\}\psi_1(\mathbf{y}) + u_1(\mathbf{y})\{\psi_1(\mathbf{x}), \psi_1(\mathbf{y})\}). \tag{2.17}$$

A short computation shows that $\{\psi_1(\mathbf{x}), \psi_1(\mathbf{y})\} = 0$. Then, since the constraint $\psi_1(\mathbf{y})$ is weakly equivalent to zero, we find that

$$\{\psi_1(\mathbf{x}), H'\} \approx \{\psi_1(\mathbf{x}), H\}. \tag{2.18}$$

Finally, due to the identity $n^a(\mathbf{x})B_a(\mathbf{x}) = 0$, we have $\{\psi_1(\mathbf{x}), H\} = 2fn^a(\mathbf{x})\pi_a(\mathbf{x})$ for any value of θ. This means that

$$\{\psi_1(\mathbf{x}), H'\} \approx 2fn^a(\mathbf{x})\pi_a(\mathbf{x}), \tag{2.19}$$

and so we find a second constraint,

$$\psi_2(\mathbf{x}) = n^a(\mathbf{x})\pi_a(\mathbf{x}), \tag{2.20}$$

which must also be set to zero for consistent time evolution of this system.

We now make a further modification to the Hamiltonian and define

$$H'' = H + \sum_{I=1}^2 \int d^D y u_I(\mathbf{y})\psi_I(\mathbf{y}), \tag{2.21}$$

where we introduced a second undetermined function $u_2(\mathbf{y})$ and investigate the conditions under which $\{\psi_2(\mathbf{x}), H''\} \approx 0$. After some algebra, we find that

$$\{\psi_2(\mathbf{x}), H''\} \approx \{\psi_2(\mathbf{x}), H\} + \int d^D y u_1(\mathbf{y})\{\psi_2(\mathbf{x}), \psi_1(\mathbf{y})\}. \tag{2.22}$$

At this point, it is possible that, depending on the form of $\{\psi_2(\mathbf{x}), \psi_1(\mathbf{y})\}$, the equation

$$\{\psi_2(\mathbf{x}), H\} + \int d^D y u_1(\mathbf{y})\{\psi_2(\mathbf{x}), \psi_1(\mathbf{y})\} = 0 \tag{2.23}$$

can be solved to yield a function $u_1(\mathbf{x})$ such that $\{\psi_2(\mathbf{x}), H''\} \approx 0$. Below, we show that this is indeed the case. This means that the constraints ψ_1 and ψ_2 account for all of the constraints in this problem, and it also means that the additional function $u_2(\mathbf{x})$ is not needed for the construction of the modified Hamiltonian. Therefore, we can set $u_2(\mathbf{x}) = 0$ at this point. However, we note here that the fact that $u_2(\mathbf{x})$ can be set to zero is specific to this particular problem. It is easy to imagine a scenario in which the function $u_2(\mathbf{x})$ would not be zero, for example, if the Poisson bracket of ψ_1 and ψ_2 were to vanish; then, requiring $\{\psi_2(\mathbf{x}), H''\} \approx 0$ would yield a new constraint $\psi_3(\mathbf{x}) = \{\psi_2(\mathbf{x}), H\}$. If that were the case, then it is very likely that we would need a nonzero $u_2(\mathbf{x})$ to construct a final modified Hamiltonian $\tilde{H}$ such that $\{\psi_3(\mathbf{x}), \tilde{H}\} \approx 0$. However, for the problem considered here, we can safely set $u_2(\mathbf{x}) = 0$, and so we find that the final modified Hamiltonian is given by

$$\tilde{H} = H + \int d^D y u_1(\mathbf{y})\psi_1(\mathbf{y}), \tag{2.24}$$

where $u_1(\mathbf{x})$ solves Eq. (2.23). Finally, we note that for the purposes of the quantization we do not need to know the exact form of $\tilde{H}$. This is fortunate because the Poisson bracket $\{\psi_2(\mathbf{x}), H\}$ is fairly complicated in the case that the parameter θ is nonzero.

Now that we know all of the constraints in the problem, we can look at their Poisson brackets with each other. For this system, we find that the function $M_{IJ}(\mathbf{x}, \mathbf{y})$ defined in Eq. (2.8) has the explicit form

$$M_{IJ}(\mathbf{x}, \mathbf{y}) = 2i(\sigma^y)_{IJ}r^2(\mathbf{x})\delta^{(D)}(\mathbf{x} - \mathbf{y}), \tag{2.25}$$

where $(\sigma^y)_{IJ}$ is the (I, J) element of the second Pauli matrix σ^y and where we defined the radial coordinate

$$r^2(\mathbf{x}) = n_a(\mathbf{x})n^a(\mathbf{x}). \tag{2.26}$$

In terms of $r^2(\mathbf{x})$, the first second class constraint in the NLSM problem reads as $\psi_1(\mathbf{x}) = r^2(\mathbf{x}) - 1$. The inverse of $M_{IJ}(\mathbf{x}, \mathbf{y})$ is

$$M_{IJ}^{-1}(\mathbf{x}, \mathbf{y}) = -\frac{i}{2}(\sigma^y)_{IJ}\frac{1}{r^2(\mathbf{x})}\delta^{(D)}(\mathbf{x} - \mathbf{y}). \tag{2.27}$$

The fact that the inverse exists means that Eq. (2.23) can indeed be solved for the function $u_1(\mathbf{x})$, as we claimed above.

We can now use the components $M_{IJ}^{-1}(\mathbf{x}, \mathbf{y})$ from Eq. (2.27) to construct the classical Dirac brackets for this theory [see Eq. (2.9) for the definition of the Dirac bracket]. To quantize the theory, we then replace all functions $f(\mathbf{x})$ with operators $\hat{f}(\mathbf{x})$ and replace the Dirac brackets with commutators as shown in Eq. (2.11). After following these steps, we find that the equal-time commutation relations for the $O(D + 2)$ NLSM with a theta term, *for any value* of θ, are given by

$$[\hat{n}^a(\mathbf{x}), \hat{n}^b(\mathbf{y})] = 0 \tag{2.28a}$$

$$[\hat{n}^a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = i\left(\delta^a{}_b - \frac{\hat{n}^a(\mathbf{x})\hat{n}_b(\mathbf{y})}{\hat{r}^2(\mathbf{x})}\right)\delta^{(D)}(\mathbf{x} - \mathbf{y}) \tag{2.28b}$$

$$[\hat{\pi}_a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = \frac{i}{\hat{r}^2(\mathbf{x})}(\hat{\pi}_a(\mathbf{x})\hat{n}_b(\mathbf{y}) - \hat{\pi}_b(\mathbf{y})\hat{n}_a(\mathbf{x})) \\ \times \delta^{(D)}(\mathbf{x} - \mathbf{y}). \tag{2.28c}$$

At this point, we note that by construction $\hat{r}^2(\mathbf{x})$ commutes with $\hat{\pi}_a(\mathbf{y})$ (in fact, it commutes with any operator on the Hilbert space). Therefore, $\hat{r}^2(\mathbf{x})$ is in the center of the algebra defined by the commutation relations in Eqs. (2.28), and it is consistent to plug in the constraint $\hat{r}^2(\mathbf{x}) = 1$. An explicit discussion about this point can be found in Ref. [41] in the context of the quantum mechanical problem of a free particle constrained to move on the surface of a sphere. The fact that $\hat{r}^2(\mathbf{x})$ commutes with any operator on the Hilbert space simply follows from the fact that the Dirac bracket of any functional F on phase space with a constraint $\psi_I(\mathbf{x})$ is *strongly equal* to zero, $\{F, \psi_I(\mathbf{x})\}_D = 0$. From now on, we work with the commutation relations obtained after this substitution. These have the form

$$[\hat{n}^a(\mathbf{x}), \hat{n}^b(\mathbf{y})] = 0 \tag{2.29a}$$

$$[\hat{n}^a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = i(\delta^a{}_b - \hat{n}^a(\mathbf{x})\hat{n}_b(\mathbf{y}))\delta^{(D)}(\mathbf{x} - \mathbf{y}) \tag{2.29b}$$

$$[\hat{\pi}_a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = i(\hat{\pi}_a(\mathbf{x})\hat{n}_b(\mathbf{y}) - \hat{\pi}_b(\mathbf{y})\hat{n}_a(\mathbf{x}))\delta^{(D)}(\mathbf{x} - \mathbf{y}), \tag{2.29c}$$

and they have appeared in several papers on the canonical quantization of the $O(N)$ NLSM [34–38]. However, we emphasize that we have explicitly shown here that these commutators are valid for the NLSM with a theta term for any value of the parameter θ. We also note that there is an operator ordering ambiguity in the commutation relation for two momenta; however, we can say that the two terms on the right-hand side of Eq. (2.29c) should have the *same* ordering, so that the commutator has the important property that $[\pi_a(\mathbf{x}), \pi_b(\mathbf{y})] = -[\pi_b(\mathbf{y}), \pi_a(\mathbf{x})]$.

To make progress in analyzing the NLSM, we employ a functional Schrödinger representation of the commutation relations of Eqs. (2.29) in which $\hat{n}^a(\mathbf{x})$ acts as multiplication by the function $n^a(\mathbf{x})$ and $\hat{\pi}_a(\mathbf{x})$ is given in terms of a functional derivative with respect to $n^a(\mathbf{x})$ as

$$\hat{\pi}_a(\mathbf{x}) = -i(\delta_a{}^b - n_a(\mathbf{x})n^b(\mathbf{x}))\frac{\delta}{\delta n^b(\mathbf{x})}. \tag{2.30}$$

This choice reproduces *all* of the commutators shown in Eqs. (2.29) (with the operator ordering indicated there). This Schrödinger representation was used previously in Ref. [38] to construct soliton operators in the $O(3)$ NLSM with Hopf term in three spacetime dimensions. It has also been used in the study of the $O(N)$ NLSM in one spacetime dimension [41–49], which is equivalent to the quantum mechanics problem of a free particle in $\mathbb{R}^N$ confined to the surface of the sphere S^{N-1}. In Appendix A, we review the solution of this quantum mechanical model using this Schrödinger representation. We use the results of Appendix A in Secs. III and IV to study the energy gap in the $O(D + 2)$ NLSM in the limit of infinitely large coupling f on flat and curved space, respectively.

III. GROUND STATE WAVE FUNCTIONAL AND THE ENERGY GAP ON FLAT SPACE

In this section, we study the $O(D + 2)$ NLSM with a theta term in the disordered ($f \to \infty$) phase with $\theta = 2\pi k$, $k \in \mathbb{Z}$, on *flat* space $\mathbb{R}^D$. We give an alternative derivation, within the canonical formalism, of the result of Ref. [26] for the ground state wave functional of the NLSM at these parameter values. Finally, we use a lattice regularization of the NLSM to prove the uniqueness of the ground state and the existence of an energy gap in the disordered phase of the model at $\theta = 2\pi k$. This section should be viewed as a warm-up for Sec. IV in which we discuss the ground state wave functional and the energy gap of the NLSM on an arbitrary spatial manifold $\mathcal{M}$.

A. Ground state wave functional at large f and $\theta = 2\pi k$, $k \in \mathbb{Z}$

We first discuss the construction of the ground state wave functional. As discussed above, we consider the disordered

phase of the model in which $f \to \infty$. In this limit, the Hamiltonian operator is approximately given by

$$\hat{H} = \frac{f}{2} \int d^D \mathbf{x} \left(\hat{\pi}_a - \frac{\theta}{\mathcal{A}_{D+1}} \hat{B}_a \right)^2, \qquad (3.1)$$

where we ignore terms proportional to $\frac{1}{f}$. Since the Hamiltonian in this limit is expressed as an integral over space of the square of a local operator, the lowest possible energy of any eigenstate is zero. This means that the ground state wave functional of the NLSM in this limit is determined only by the property that it is annihilated by the operators

$$\hat{\mathcal{D}}_a^{(\theta)} = \hat{\pi}_a - \frac{\theta}{\mathcal{A}_{D+1}} \hat{B}_a, \qquad a = 1, ..., D+2. \qquad (3.2)$$

On the other hand, because of the specific form of the operator $\hat{\pi}^a(\mathbf{x})$ in the Schrödinger representation Eq. (2.30) used in this paper, the Hamiltonian needs to be regularized in some way before any excited states can be constructed. We discuss the need for regularization of the Hamiltonian in more detail in Appendix B. For now, however, we are only interested in the construction of the ground state wave functional, and so we can delay the issue of regularization of the Hamiltonian until the next subsection.

To start, consider the case where $\theta = 0$. Since $\hat{\pi}_a$ is proportional to a functional derivative with respect to n^a, and since the energy is bounded below by zero, we can see that the ground state wave functional is just a constant, $\Psi_{\theta=0}[\mathbf{n}] = 1$. A general state $|\Psi\rangle$ in the Hilbert space of the NLSM can be expanded in the "position basis" $\{|\mathbf{n}\rangle\}$, which contains a state $|\mathbf{n}\rangle$ for every possible configuration of the NLSM field on the space $\mathbb{R}^D$. The field operator $\hat{n}^a(\mathbf{x})$ is diagonal in this basis, $\hat{n}^a(\mathbf{x})|\mathbf{n}\rangle = n^a(\mathbf{x})|\mathbf{n}\rangle$, where $n^a(\mathbf{x})$ is the *function* corresponding to the particular state $|\mathbf{n}\rangle$. For a more precise formulation, we should restrict the set $\{|\mathbf{n}\rangle\}$ to include only those field configurations on $\mathbb{R}^D$ with finite (classical) potential energy. This restriction implies a choice of boundary condition on the field configurations at spatial infinity; for example, we could choose $\mathbf{n}(\mathbf{x}) \to \mathbf{n}_0$, a particular constant field configuration, as $|\mathbf{x}| \to \infty$.

A general state in this basis takes the form

$$|\Psi\rangle \propto \int [D\mathbf{n}] \Psi[\mathbf{n}] |\mathbf{n}\rangle, \qquad (3.3)$$

where $\Psi[\mathbf{n}]$ is the wave functional (i.e., the amplitude of the basis state $|\mathbf{n}\rangle$ in the full state $|\Psi\rangle$) and the integration is over all possible configurations of the field $\mathbf{n}$ at every point in space (possibly subject to a boundary condition at spatial infinity ensuring finite energy). We define the measure $[D\mathbf{n}]$ to be the product over all points $\mathbf{x}$ in space of the volume form ω_{D+1} on the sphere S^{D+1}. Since the ground

state wave functional of the NLSM at $\theta = 0$ is just $\Psi_{\theta=0}[\mathbf{n}] = 1$, it follows that the state vector for the ground state is just an equal weight superposition of all basis states,

$$|\Psi_{\theta=0}\rangle \propto \int [D\mathbf{n}] |\mathbf{n}\rangle. \qquad (3.4)$$

This state can be thought of as a continuum analog of a trivial paramagnetic state.

Next, we look at the ground state for nonzero θ in the particular case that θ is an integer multiple of 2π. In this case, it is possible to remove the term $\frac{\theta}{\mathcal{A}_{D+1}} \hat{B}_a$ from the operator $\hat{\mathcal{D}}_a^{(\theta)}$ via a well-defined unitary transformation, which means that the ground state in this case can be obtained by multiplying the ground state at $\theta = 0$ by a unitary operator. As we discussed in Sec. I, the case $\theta = 2\pi k$, $k \in \mathbb{Z}$, is also interesting from a physical point of view because for these values of θ the $O(D+2)$ NLSM has been shown to capture many of the physical properties of SPT phases of bosons in D spatial dimensions [14,21–31].

The ground state wave functional at $\theta = 2\pi k$ can be constructed using a WZ term for the NLSM field $\mathbf{n}$. Recall now that we are working in $(D+1)$-dimensional Minkowski spacetime $\mathbb{R}^{D,1}$, so that the physical space is just $\mathbb{R}^D$. The WZ term is written as an integral over the extended space $\mathcal{B} = [0, 1] \times \mathbb{R}^D$, where $\mathbb{R}^D$ represents the original D-dimensional space, and $[0, 1]$ is an auxiliary direction of space used in the construction of the WZ term. We use the notation $\tilde{n}^a(\mathbf{x}, s)$ to denote the extension of the NLSM field $n^a(\mathbf{x})$ into the extra direction, where $\mathbf{x} \in \mathbb{R}^D$ and $s \in [0, 1]$. Typically, one chooses boundary conditions in the auxiliary direction so that $\tilde{n}^a(\mathbf{x}, 0) = \delta^a_{D+2}$ (i.e., a trivial configuration) and $\tilde{n}^a(\mathbf{x}, 1) = n^a(\mathbf{x})$, so that the physical space sits at $s = 1$.

We now show that for $\theta = 2\pi k$, $k \in \mathbb{Z}$, the ground state wave functional is

$$\Psi_{\theta=2\pi k}[\mathbf{n}] = e^{-ikS_{\text{WZ}}[\mathbf{n}]}, \qquad (3.5)$$

where $S_{\text{WZ}}[\mathbf{n}]$ is the WZ term,

$$\begin{aligned} S_{\text{WZ}}[\mathbf{n}] &= \frac{2\pi}{\mathcal{A}_{D+1}} \int_{\mathcal{B}} \tilde{\mathbf{n}}^* \omega_{D+1} \\ &= \frac{2\pi}{\mathcal{A}_{D+1}} \int_0^1 ds \\ &\quad \times \int d^D \mathbf{x} \epsilon_{a_1 \cdots a_{D+2}} \tilde{n}^{a_1} \partial_s \tilde{n}^{a_2} \partial_1 \tilde{n}^{a_3} \cdots \partial_D \tilde{n}^{a_{D+2}}. \end{aligned}$$
$$(3.6)$$

The WZ term involves the pullback of the volume form ω_{D+1} on S^{D+1} to the extended space $\mathcal{B}$ via the map $\tilde{\mathbf{n}} : \mathcal{B} \to S^{D+1}$. To prove Eq. (3.5), we first recall the formula for the variation of the WZ term,

$$\delta S_{\mathrm{WZ}}[\mathbf{n}] = -\frac{2\pi}{\mathcal{A}_{D+1}} \int d^D\mathbf{x} \epsilon_{a_1 \cdots a_{D+2}} \delta n^{a_1} n^{a_2} \partial_1 n^{a_3} \cdots \partial_D n^{a_{D+2}},$$

(3.7)

which is an integral only over the physical space $\mathbb{R}^D$ (in the case that we can neglect terms coming from the boundary of physical space). Then, we have

$$\frac{\delta}{\delta n^a(\mathbf{x})} \Psi_{\theta=2\pi k}[\mathbf{n}]$$

$$= i\frac{2\pi k}{\mathcal{A}_{D+1}} (\epsilon_{aa_2 \cdots a_{D+2}} n^{a_2} \partial_1 n^{a_3} \cdots \partial_D n^{a_{D+2}}) \Psi_{\theta=2\pi k}[\mathbf{n}]$$

$$= i\frac{\theta}{\mathcal{A}_{D+1}} B_a(\mathbf{x}) \Psi_{\theta=2\pi k}[\mathbf{n}].$$

(3.8)

Then, using $\hat{\pi}_a(\mathbf{x}) = -i(\delta_a{}^b - n_a(\mathbf{x})n^b(\mathbf{x}))\frac{\delta}{\delta n^b(\mathbf{x})}$ and the fact that $n^b B_b = 0$, we find that

$$\hat{\mathcal{D}}_a^{(\theta=2\pi k)} \Psi_{\theta=2\pi k}[\mathbf{n}] = 0,$$

(3.9)

which completes the proof. The state vector for the ground state at $\theta = 2\pi k$ then takes the form

$$|\Psi_{\theta=2\pi k}\rangle \propto \int [D\mathbf{n}] e^{-ikS_{\mathrm{WZ}}[\mathbf{n}]} |\mathbf{n}\rangle.$$

(3.10)

Thus, we have succeeding in rederiving the result of Ref. [26] for the ground state wave functional of this system within the canonical formalism.

The relationship between the ground state wave functionals at $\theta = 0$ and $\theta = 2\pi k$ can be understood in terms of a unitary transformation of the Hamiltonian by the *operator*

$$\hat{\mathcal{U}}^{(k)} = e^{-ikS_{\mathrm{WZ}}[\hat{\mathbf{n}}]}.$$

(3.11)

In the Schrödinger representation, and using a suitable test functional, one can show that

$$\hat{\mathcal{U}}^{(k),\dagger} \hat{\mathcal{D}}_a^{(\theta=2\pi k)} \hat{\mathcal{U}}^{(k)} = \hat{\mathcal{D}}_a^{(\theta=0)} = \hat{\pi}_a,$$

(3.12)

which means that

$$\hat{\mathcal{U}}^{(k),\dagger} \hat{H}_{(\theta=2\pi k)} \hat{\mathcal{U}}^{(k)} = \hat{H}_{(\theta=0)}$$

(3.13)

and that

$$|\Psi_{\theta=2\pi k}\rangle = \hat{\mathcal{U}}^{(k)} |\Psi_{\theta=0}\rangle.$$

(3.14)

Note also that since $\hat{\mathcal{U}}^{(k)}$ commutes with the potential energy term $\frac{1}{2f}(\nabla n^a)^2$ Eq. (3.13) holds for the full Hamiltonian of Eq. (2.13) (i.e., not just in the large f limit). In fact, for any values of f and θ, the full Hamiltonian obeys the relation

$$\hat{\mathcal{U}}^{(1),\dagger} \hat{H}_{(\theta)} \hat{\mathcal{U}}^{(1)} = \hat{H}_{(\theta-2\pi)},$$

(3.15)

which shows that the spectrum of the $O(D+2)$ NLSM with a theta term is 2π periodic in the value of the parameter θ. This is a crucial result since it will let us simultaneously study the energy spectra for any values of θ related by a 2π shift.

We see that the theta angle of the NLSM enters into the Hamiltonian of Eq. (2.13) as something like a gauge field, and the derivative operator $\hat{\mathcal{D}}_a^{(\theta)}$ looks like a covariant derivative. In the case that $\theta = 2\pi k$, $k \in \mathbb{Z}$, we can interpret the phase of the ground state wave functional as being obtained from a gauge transformation which removes the "gauge field" term $\frac{\theta}{\mathcal{A}_{D+1}} \hat{B}_a$ from $\hat{\mathcal{D}}_a^{(\theta)}$ at the expense of an additional phase in the wave functional. This gauge transformation, however, can only be performed when θ is an integer multiple of 2π. This is because the exponential $e^{-ikS_{\mathrm{WZ}}[\mathbf{n}]}$ of the WZ term, which involves an extension of the field n^a into an auxiliary direction, is only well-defined (i.e., independent of the extension) when k is an integer [50]. To be precise, we note here that to apply the argument of Ref. [50] on the quantization of k we must replace flat space $\mathbb{R}^D$ with a D-dimensional sphere so that space is a compact manifold (the radius of the sphere can be taken to be very large so that the curvature is nearly zero). The original infinite space $\mathbb{R}^D$ is then obtained in the limit that the radius of the D-sphere goes to infinity. We now move on to a discussion of the uniqueness of the ground state and the calculation of the energy gap in the NLSM at $\theta = 2\pi k$ and $f \to \infty$.

B. Uniqueness of the ground state and the energy gap at large f and $\theta = 2\pi k$, $k \in \mathbb{Z}$

In the previous subsection, we showed that the NLSM Hamiltonians at $\theta = 2\pi k$ and $\theta = 0$ are related by a unitary transformation, which means that the energy spectrum in this model at $\theta = 2\pi k$ is identical to the spectrum at $\theta = 0$. In the context of applications to SPT phases, one of the most important properties of the NLSM at large f that we would like to verify is the uniqueness of the ground state and the existence of an energy gap between the ground state and all of the excited states. In this subsection, we use a lattice regularization of the NLSM at large f to prove the uniqueness of the ground state and the existence of an energy gap at $\theta = 0$. Since the NLSM Hamiltonian at $\theta = 2\pi k$ is related to the Hamiltonian at $\theta = 0$ by a unitary transformation, the uniqueness of the ground state and the existence of an energy gap at $\theta = 2\pi k$ follow immediately from this result at $\theta = 0$. In Appendix B, we also present an alternative regularization procedure for the NLSM Hamiltonian in the disordered limit, and we show that this alternative procedure gives a result for the energy gap which is consistent with the result derived in this section using a lattice regularization. Therefore, we expect that our

result for the energy gap of the $O(D+2)$ NLSM in its disordered limit is independent of the specific details of the regularization scheme used in the calculation.

To start, we consider a hypercubic lattice with spacing a and coordinates which are vectors with integer entries and denoted by boldface latin letters $\mathbf{j}, \mathbf{k}$, etc. The continuum coordinate $\mathbf{x}$ is given in terms of the lattice coordinate $\mathbf{j}$ by $\mathbf{x} = a\mathbf{j}$. In the lattice regularization, the Dirac delta function is represented by $\delta^{(D)}(\mathbf{x} - \mathbf{y}) \rightarrow \frac{1}{a^D} \delta_{\mathbf{jk}}$ if $\mathbf{x} = a\mathbf{j}$ and $\mathbf{y} = a\mathbf{k}$. If we define lattice operators $\hat{n}_{\mathbf{j}}^a$ and $\hat{\pi}_{a,\mathbf{j}}$ by

$$\hat{n}^a(a\mathbf{j}) = \hat{n}_{\mathbf{j}}^a \qquad (3.16a)$$

$$\hat{\pi}_a(a\mathbf{j}) = \frac{1}{a^D} \hat{\pi}_{a,\mathbf{j}}, \qquad (3.16b)$$

where $\hat{n}^a(a\mathbf{j})$ and $\hat{\pi}_a(a\mathbf{j})$ are the continuum field operators at $\mathbf{x} = a\mathbf{j}$, then the NLSM commutation relations of Eqs. (2.29) become

$$[\hat{n}_{\mathbf{j}}^a, \hat{n}_{\mathbf{k}}^b] = 0 \qquad (3.17a)$$

$$[\hat{n}_{\mathbf{j}}^a, \hat{\pi}_{b,\mathbf{k}}] = i(\delta^a{}_b - \hat{n}_{\mathbf{j}}^a \hat{n}_{b,\mathbf{k}})\delta_{\mathbf{jk}} \qquad (3.17b)$$

$$[\hat{\pi}_{a,\mathbf{j}}, \hat{\pi}_{b,\mathbf{k}}] = i(\hat{\pi}_{a,\mathbf{j}} \hat{n}_{b,\mathbf{k}} - \hat{\pi}_{b,\mathbf{k}} \hat{n}_{a,\mathbf{j}})\delta_{\mathbf{jk}}. \qquad (3.17c)$$

The integration over space becomes $\int d^D\mathbf{x} \rightarrow a^D \sum_{\mathbf{j}}$, and so the regularized Hamiltonian at large f and $\theta = 0$ takes the form

$$\hat{H}(a) = \frac{f}{2a^D} \sum_{\mathbf{j}} \hat{\pi}_{a,\mathbf{j}} \hat{\pi}^a{}_{\mathbf{j}}. \qquad (3.18)$$

Here, we have written $\hat{H}(a)$ to indicate the explicit dependence of the Hamiltonian on the cutoff a.

The regularized Hamiltonian Eq. (3.18) is a sum of many identical Hamiltonians for an $O(N)$ NLSM in one spacetime dimension, with $N = D + 2$. In Appendix A, we review the solution of this quantum mechanics problem using Dirac's formalism for quantizing constrained systems. Using the results from Appendix A, we can rewrite the Hamiltonian as

$$\hat{H}(a) = \frac{f}{2a^D} \sum_{\mathbf{j}} \hat{\mathcal{C}}_{\mathbf{j}}, \qquad (3.19)$$

where $\hat{\mathcal{C}}_{\mathbf{j}}$ is the quadratic Casimir of $so(D+2)$ formed from the conserved charge operators in the $O(D+2)$ NLSM on site $\mathbf{j}$. We immediately deduce that the unique ground state of this system is the state with

$$\hat{\mathcal{C}}_{\mathbf{j}} = 0, \quad \forall \; \mathbf{j}, \qquad (3.20)$$

i.e., the state which is the tensor product of the trivial representation of $SO(D+2)$ on all sites $\mathbf{j}$. The energy gap, which is equal to the energy of the first excited state, is ("m" stands for mass)

$$m(a) = \frac{f}{2a^D}(D+1). \qquad (3.21)$$

This energy corresponds to the case that one site in the lattice is excited to a state in the fundamental representation of $SO(D+2)$. In the theory at large f, the first excited state is highly degenerate, but this degeneracy will be broken by the inclusion of a small kinetic energy term (with coefficient $\frac{1}{f}$), which will cause the energies of all degenerate states in the first excited state manifold to disperse.

We see that for a fixed bare coupling constant f the energy gap $m(a)$ goes to infinity as we take the continuum limit $a \rightarrow 0$. On the other hand, it is more physical to make the coupling constant cutoff dependent, $f \rightarrow f(a)$, and demand that $f(a)$ depend on the cutoff a in such a way as to make the mass gap $m(a)$ independent of the cutoff a used to define the theory. Following the procedure of Ref. [6], we demand that $\frac{dm(a)}{da} = 0$, which yields the renormalization group equation for $f(a)$ in the regime of large f,

$$a\frac{df(a)}{da} = Df(a). \qquad (3.22)$$

We find that $f(a) \rightarrow \infty$ in the infrared (i.e., low-energy) limit $a \rightarrow \infty$, which confirms the validity of our expansion of the Hamiltonian in powers of $\frac{1}{f}$. Integrating Eq. (3.22) from some reference scale a_0 in the ultraviolet, at which $f = f_0$, to the scale a, we find that $f(a)$ is given in terms of f_0 as $f(a) = (\frac{a}{a_0})^D f_0$, so that the mass gap m (which is now independent of a) is given in terms of f_0 and the reference scale a_0 by

$$m = \frac{f_0}{2a_0^D}(D+1). \qquad (3.23)$$

IV. QUANTIZATION, GROUND STATE WAVE FUNCTIONAL, AND ENERGY GAP ON CURVED SPACE

In this section, we repeat the analysis of Secs. II and III in the case that the spacetime takes the form $\mathcal{M} \times \mathbb{R}$, where $\mathbb{R}$ represents the time direction and $\mathcal{M}$ is a curved, D-dimensional manifold representing space (the precise assumptions on the properties of $\mathcal{M}$ were stated in Sec. I and are repeated below). In particular, we will accomplish the goal of the paper, which is to prove the three properties of the NLSM on curved space which are stated in Sec. I. That is, we prove the uniqueness of the ground state and the existence of an energy gap in the $O(D+2)$ NLSM in the disordered ($f \rightarrow \infty$) phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$, on arbitrary

spatial manifolds $\mathcal{M}$, and we also explicitly construct the ground state wave functional on $\mathcal{M}$. We find that the wave functional takes the form of an exponential of a WZ term for n^a, just as in the case on flat space [26]. To prove the uniqueness of the ground state and the existence of an energy gap in the NLSM on $\mathcal{M}$, we use a *triangulation* of the manifold to set up a latticelike regularization of the NLSM Hamiltonian at large f. Within this regularization scheme, the demonstration of the uniqueness of the ground state and the computation of the energy gap can be done in a way which is very similar to the calculation on flat space from Sec. III.

The results of this section prove that the $O(D+2)$ NLSM, in the parameter regime studied in this paper, possesses SPT order, but not topological order, and is therefore a suitable model for SPT phases. One interesting aspect of the theory on a curved space $\mathcal{M}$ is that for certain choices of manifold $\mathcal{M}$ the standard construction of the WZ term fails, and so alternative constructions are needed. We discuss the standard construction of the WZ term and one type of alternative construction in some detail in this section. Then, in Appendix D, we give an explicit example of a third construction which can be used when the other two constructions fail. Before we discuss these details, however, we need to first explain the modifications to the canonical quantization procedure of Sec. II which are needed to study the NLSM in the canonical formalism on the curved space $\mathcal{M}$.

A. Canonical quantization of the NLSM on a curved space

In this subsection, we discuss the canonical quantization of the $O(D+2)$ NLSM with a theta term on a spacetime of the form $\mathcal{M} \times \mathbb{R}$, where $\mathbb{R}$ represents the time direction and $\mathcal{M}$ is a smooth, closed, connected, oriented, D-dimensional manifold. We take the metric on spacetime to have the form (we use a mostly minus signature for the metric)

$$g = dt \otimes dt - G_{ij}(\mathbf{x})dx^i \otimes dx^j, \qquad (4.1)$$

where $i, j = 1, \dots, D$ (and a sum over repeated indices is implied). On flat Minkowski space, we have $G_{ij}(\mathbf{x}) = \delta_{ij}$, but in the general case, $G_{ij}(\mathbf{x})$ are the components of a Riemannian metric on $\mathcal{M}$. In addition, we have

$$\det[g] = (-1)^D \det[G]. \qquad (4.2)$$

By a common abuse of notation, we will also use the letters g and G to denote $\det[g]$ and $\det[G]$, respectively, for the remainder of the paper.

To start, we use the formalism of Appendix C to understand how to quantize a free scalar field on a curved space. The key piece of information we need is the appropriate form of the Poisson bracket for a free scalar field on a curved space. With this information in hand, we

can then use Dirac's procedure to quantize the NLSM on a curved space, since the $O(D+2)$ NLSM consists of $D+2$ scalar fields, but subject to the additional constraint $n^a(\mathbf{x})n_a(\mathbf{x}) = 1$. At this point, we suggest that the reader skim through Appendix C to understand our notation for the symplectic geometry approach to studying field theories in the Hamiltonian formalism.

First, we outline our general strategy for determining the correct symplectic form Ω to use to describe a field theory on a curved space. Suppose that the system we would like to study on a curved space has a definition in terms of an action

$$S = \int d^{D+1}x\sqrt{(-1)^D g}\mathcal{L} = \int d^{D+1}x\sqrt{G}\mathcal{L}, \qquad (4.3)$$

where $\mathcal{L}$ is the Lagrangian and we assumed a metric on spacetime of the form of Eq. (4.1). In this case, our strategy for determining the appropriate symplectic form is to choose Ω such that the Hamilton equations of motion obtained from Ω via Eq. (C10) coincide with the Euler-Lagrange equations of motion obtained from the action for our system on a curved space. Once we know the correct Ω, we can use it to find the correct Poisson brackets from Eq. (C9). These Poisson brackets will then give us the information we need to find the commutation relations for the fields in the quantum field theory on the curved space $\mathcal{M}$.

Let us see how this all plays out in the case of a free scalar field ϕ. In this case, the Lagrangian is

$$\mathcal{L} = \frac{1}{2}(\partial^\mu\phi)(\partial_\mu\phi) = \frac{1}{2}[(\partial_t\phi)^2 - G^{ij}\partial_i\phi\partial_j\phi], \qquad (4.4)$$

where in the first line $\mu = 0, 1, \dots, D$ (and $x^0 = t$). In the second line, we specialized to the case of curved space only [i.e., a metric of the form shown in Eq. (4.1)], and we used the tensor G^{ij} which satisfies the relation $G^{ij}G_{jk} = \delta^i{}_k$. The momentum conjugate to ϕ is $\pi = \frac{\partial\mathcal{L}}{\partial(\partial_t\phi)} = \partial_t\phi$, and the Hamiltonian is

$$H = \int d^D\mathbf{x}\sqrt{G}(\pi\partial_t\phi - \mathcal{L})$$
$$= \frac{1}{2}\int d^D\mathbf{x}\sqrt{G}(\pi^2 + G^{ij}\partial_i\phi\partial_j\phi). \qquad (4.5)$$

Starting from the action $S = \int d^{D+1}\sqrt{G}\mathcal{L}$, we can derive the Euler-Lagrange equation of motion for ϕ,

$$\partial_t^2\phi - \frac{1}{\sqrt{G}}\partial_i(\sqrt{G}G^{ij}\partial_j\phi) = 0. \qquad (4.6)$$

Now that we know the Euler-Lagrange equation of motion for ϕ, we can look for a choice of Ω so that the Hamilton equations obtained from it are equivalent to this

Euler-Lagrange equation. We find that the choice of Ω which yields the correct equations of motion is

$$\Omega = \int d^D\mathbf{x}\sqrt{G(\mathbf{x})}\delta\pi(\mathbf{x}) \wedge \delta\phi(\mathbf{x}). \qquad (4.7)$$

Indeed, using this form of Ω with Eq. (C10), we find that

$$\partial_t\phi = \pi \qquad (4.8a)$$

$$\partial_t\pi = \frac{1}{\sqrt{G}}\partial_i(\sqrt{G}G^{ij}\partial_j\phi), \qquad (4.8b)$$

which is clearly equivalent to the equation of motion Eq. (4.6) derived from the action. Using the correct form of Ω we can now derive the form of the Poisson bracket for ϕ and π on curved space. First, using Eq. (C8), we find that the vector fields on the phase space corresponding to the functionals $\phi(\mathbf{x})$ and $\pi(\mathbf{x})$ are

$$\underline{V}_{\phi(\mathbf{x})} = -\frac{1}{\sqrt{G(\mathbf{x})}}\frac{\delta}{\delta\pi(\mathbf{x})} \qquad (4.9a)$$

$$\underline{V}_{\pi(\mathbf{x})} = \frac{1}{\sqrt{G(\mathbf{x})}}\frac{\delta}{\delta\phi(\mathbf{x})}. \qquad (4.9b)$$

From these, we find that

$$\{\phi(\mathbf{x}), \pi(\mathbf{y})\} = i_{\underline{V}_{\phi(\mathbf{x})}} i_{\underline{V}_{\pi(\mathbf{y})}}\Omega$$

$$= \frac{1}{\sqrt{G(\mathbf{x})}}\delta^{(D)}(\mathbf{x} - \mathbf{y}). \qquad (4.10)$$

This then tells us that the correct commutation relation for the operators $\hat{\phi}(\mathbf{x})$ and $\hat{\pi}(\mathbf{y})$ in the quantized theory on curved space is

$$[\hat{\phi}(\mathbf{x}), \hat{\pi}(\mathbf{y})] = \frac{i}{\sqrt{G(\mathbf{x})}}\delta^{(D)}(\mathbf{x} - \mathbf{y}). \qquad (4.11)$$

Given this form of Ω, we can also work out a general formula for the Poisson bracket of any two functionals F_1 and F_2 of the phase space variables. To do this, we need to first solve Eq. (C8) for the vector field $\underline{V}_F$ corresponding to a given functional F. If we write the vector field $\underline{V}_F$ as

$$\underline{V}_F = \int d^D\mathbf{x}\left(V_F^\phi\frac{\delta}{\delta\phi(\mathbf{x})} + V_F^\pi\frac{\delta}{\delta\pi(\mathbf{x})}\right), \qquad (4.12)$$

then the solution of Eq. (C8) for the components of $\underline{V}_F$ is

$$V_F^\phi = \frac{1}{\sqrt{G(\mathbf{x})}}\frac{\delta F}{\delta\pi(\mathbf{x})} \qquad (4.13a)$$

$$V_F^\pi = -\frac{1}{\sqrt{G(\mathbf{x})}}\frac{\delta F}{\delta\phi(\mathbf{x})}. \qquad (4.13b)$$

Plugging into Eq. (C9), we find that the Poisson bracket of any two functionals F_1 and F_2 in the theory of a free scalar field on a curved space is given by

$$\{F_1, F_2\} = \int d^D\mathbf{x}\frac{1}{\sqrt{G(\mathbf{x})}}\left(\frac{\delta F_1}{\delta\phi(\mathbf{x})}\frac{\delta F_2}{\delta\pi(\mathbf{x})} - \frac{\delta F_1}{\delta\pi(\mathbf{x})}\frac{\delta F_2}{\delta\phi(\mathbf{x})}\right). \qquad (4.14)$$

The only modification from the usual Poisson bracket on flat space is the extra factor of $\frac{1}{\sqrt{G(\mathbf{x})}}$.

Now, we combine this information with Dirac's procedure for dealing with constraints in the Hamiltonian formalism to derive the commutation relations for the NLSM with a theta term on the curved space $\mathcal{M}$. The action for the $O(D + 2)$ NLSM with a theta term on curved space is $S = \int d^{D+1}x\sqrt{G}\mathcal{L}$ with

$$\mathcal{L} = \frac{1}{2f}(\partial^\mu n_a)(\partial_\mu n^a) + \frac{1}{\sqrt{G}}\frac{\theta}{\mathcal{A}_{D+1}}B_a(\partial_t n^a), \qquad (4.15)$$

where the contraction of greek (spacetime) indices is now done with the metric $g_{\mu\nu}$ from Eq. (4.1) and B_a was defined in Eq. (2.7). The momentum conjugate to n^a is now

$$\pi_a = \frac{\partial\mathcal{L}}{\partial(\partial_t n^a)} = \frac{1}{f}(\partial_t n_a) + \frac{1}{\sqrt{G}}\frac{\theta}{\mathcal{A}_{D+1}}B_a, \qquad (4.16)$$

and the Hamiltonian on curved space takes the form

$$H = \int d^D\mathbf{x}\sqrt{G}\left\{\frac{f}{2}\left(\pi_a - \frac{1}{\sqrt{G}}\frac{\theta}{\mathcal{A}_{D+1}}B_a\right)^2 + \frac{1}{2f}G^{ij}\partial_i n^a\partial_j n_a\right\}. \qquad (4.17)$$

Finally, from our discussion above on the canonical formalism for a single scalar field on curved space, we know that the correct Poisson bracket for two functionals F_1 and F_2 in the NLSM on curved space is

$$\{F_1, F_2\} = \int d^D\mathbf{x}\frac{1}{\sqrt{G(\mathbf{x})}}\left(\frac{\delta F_1}{\delta n^a(\mathbf{x})}\frac{\delta F_2}{\delta\pi_a(\mathbf{x})} - \frac{\delta F_1}{\delta\pi_a(\mathbf{x})}\frac{\delta F_2}{\delta n^a(\mathbf{x})}\right). \qquad (4.18)$$

Using this Poisson bracket, we may now proceed as in Sec. II and use Dirac's procedure for handling constraints to quantize the NLSM on curved space. We skip the details as they are very similar to those in Sec. II and just present the results. The NLSM with a theta term on curved space is again characterized by two second class constraints,

$$\psi_1(\mathbf{x}) = n^a(\mathbf{x})n_a(\mathbf{x}) - 1 \qquad (4.19a)$$

$$\psi_2(\mathbf{x}) = n^a(\mathbf{x})\pi_a(\mathbf{x}). \qquad (4.19b)$$

The Poisson bracket of these constraints, computed using the Poisson bracket of Eq. (4.18) for the NLSM on curved space, is $\{\psi_I(\mathbf{x}), \psi_J(\mathbf{y})\} = M_{IJ}(\mathbf{x}, \mathbf{y})$ with

$$M_{IJ}(\mathbf{x}, \mathbf{y}) = \frac{2i}{\sqrt{G(\mathbf{x})}}(\sigma^y)_{IJ} r^2(\mathbf{x})\delta^{(D)}(\mathbf{x} - \mathbf{y}), \qquad (4.20)$$

where $r^2(\mathbf{x}) = n_a(\mathbf{x})n^a(\mathbf{x})$. Its inverse, which is needed to compute the Dirac brackets for the NLSM on curved space, is

$$M_{IJ}^{-1}(\mathbf{x}, \mathbf{y}) = -\frac{i}{2}\sqrt{G(\mathbf{x})}(\sigma^y)_{IJ}\frac{1}{r^2(\mathbf{x})}\delta^{(D)}(\mathbf{x} - \mathbf{y}). \qquad (4.21)$$

The components $M_{IJ}^{-1}(\mathbf{x}, \mathbf{y})$ can now be used to construct the classical Dirac brackets for the NLSM on curved space. Then, to quantize the NLSM on curved space, we replace all functions $f(\mathbf{x})$ with operators $\hat{f}(\mathbf{x})$ on the Hilbert space, and we obtain the quantum commutation relations for the NLSM on curved space by replacing the Dirac brackets with commutators as in Sec. II for the NLSM on flat space. In addition, as in Sec. II, we set the operator $\hat{r}^2(\mathbf{x}) = 1$, which is consistent since this operator commutes with all other operators in the Hilbert space. Therefore, we find that the commutation relations for the NLSM with a theta term on curved space are

$$[\hat{n}^a(\mathbf{x}), \hat{n}^b(\mathbf{y})] = 0 \qquad (4.22a)$$

$$[\hat{n}^a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = \frac{i}{\sqrt{G(\mathbf{x})}}(\delta^a{}_b - \hat{n}^a(\mathbf{x})\hat{n}_b(\mathbf{y}))\delta^{(D)}(\mathbf{x} - \mathbf{y}) \qquad (4.22b)$$

$$[\hat{\pi}_a(\mathbf{x}), \hat{\pi}_b(\mathbf{y})] = \frac{i}{\sqrt{G(\mathbf{x})}}(\hat{\pi}_a(\mathbf{x})\hat{n}_b(\mathbf{y}) - \hat{\pi}_b(\mathbf{y})\hat{n}_a(\mathbf{x}))$$
$$\times \delta^{(D)}(\mathbf{x} - \mathbf{y}). \qquad (4.22c)$$

Again, the only modification from the case of flat space is the extra factor of $\frac{1}{\sqrt{G(\mathbf{x})}}$. As in the case on flat space, these commutation relations also admit a functional Schrödinger representation in which $\hat{n}^a(\mathbf{x})$ acts as multiplication by $n^a(\mathbf{x})$ and now $\hat{\pi}_a(\mathbf{x})$ acts as the functional derivative operator

$$\hat{\pi}_a(\mathbf{x}) = -\frac{i}{\sqrt{G(\mathbf{x})}}(\delta_a{}^b - n_a(\mathbf{x})n^b(\mathbf{x}))\frac{\delta}{\delta n^b(\mathbf{x})}. \qquad (4.23)$$

In the next subsection, we use this Schrödinger representation to solve for the ground state wave functional of the

$O(D+2)$ NLSM in the disordered ($f \to \infty$) phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$.

B. Ground state wave functional at large f and $\theta = 2\pi k$, $k \in \mathbb{Z}$

In the large f limit, the Hamiltonian operator for the $O(D+2)$ NLSM with a theta term on the curved space $\mathcal{M}$ takes the form

$$\hat{H} = \int d^D\mathbf{x}\sqrt{G}\frac{f}{2}\left(\hat{\pi}_a - \frac{1}{\sqrt{G}}\frac{\theta}{\mathcal{A}_{D+1}}\hat{B}_a\right)^2, \qquad (4.24)$$

where we again dropped the potential energy term with coefficient proportional to $\frac{1}{f}$. We now investigate the form of the ground state wave functional of this theory in the case where $\theta = 2\pi k$, $k \in \mathbb{Z}$, which is the case where the NLSM is expected to describe an SPT phase on the curved space $\mathcal{M}$. As in the case on flat space, the ground state wave functional is determined by the condition that it be annihilated by the operators

$$\hat{\mathcal{D}}_a^{(\theta)} = \hat{\pi}_a - \frac{1}{\sqrt{G}}\frac{\theta}{\mathcal{A}_{D+1}}\hat{B}_a, \qquad a = 1, \ldots, D+2. \qquad (4.25)$$

In the functional Schrödinger representation used in this paper, this operator takes the form

$$\hat{\mathcal{D}}_a^{(\theta)}(\mathbf{x}) = \frac{1}{\sqrt{G(\mathbf{x})}}\left\{-i(\delta_a{}^b - n_a(\mathbf{x})n^b(\mathbf{x}))\frac{\delta}{\delta n^b(\mathbf{x})}\right.$$
$$\left. - \frac{\theta}{\mathcal{A}_{D+1}}B_a(\mathbf{x})\right\}. \qquad (4.26)$$

We see that the dependence of this operator on the metric of space is only through the overall factor of $\frac{1}{\sqrt{G(\mathbf{x})}}$. This a consequence of the fact that the Dirac brackets for the NLSM on curved space have an explicit dependence on $\sqrt{G(\mathbf{x})}$, while the theta term in the NLSM action is independent of the metric. This property of $\hat{\mathcal{D}}_a^{(\theta)}(\mathbf{x})$ is very important. It implies that the ground state wave functional at large f and $\theta = 2\pi k$ on the curved space $\mathcal{M}$ is independent of the metric on $\mathcal{M}$ and can be constructed in the exact same way as on flat space; i.e., the ground state wave functional is the exponential of a WZ term for the NLSM field, $\Psi_{\theta=2\pi k}[\mathbf{n}] = e^{-ikS_{WZ}[\mathbf{n}]}$. As we mentioned at the beginning of this section, for certain choices of $\mathcal{M}$, the standard construction of the WZ term fails, and so alternative constructions are needed. We now turn to a discussion of this issue.

The crucial property of the WZ term $S_{WZ}[\mathbf{n}]$, which allows for the construction of a functional annihilated by $\hat{\mathcal{D}}_a^{(\theta=2\pi k)}$, is the formula Eq. (3.7) for its variation with respect to $n^a(\mathbf{x})$. We now review two different methods for constructing an action of which the variation is given by

Eq. (3.7), and then we discuss specific examples of manifolds $\mathcal{M}$ where both constructions fail. In these cases, a third construction of the WZ term is available using the methods outlined in Ref. [51]. In Appendix D, we give an explicit example of the construction of the WZ term using the methods of Ref. [51] in the simple case where the dimension of the space $\mathcal{M}$ is $D = 1$.

The first construction of the WZ term that we discuss is the standard construction that appears in the literature [50]. This construction uses a higher-dimensional manifold $\mathcal{B}$ which has $\mathcal{M}$ as its boundary. In Sec. III, we discussed this construction on flat space $\mathcal{M} = \mathbb{R}^D$, and we now discuss how it works for a general curved spatial manifold $\mathcal{M}$. For the standard construction of the WZ term for the NLSM field, we first look for a $(D + 1)$-dimensional manifold $\mathcal{B}$ which has $\mathcal{M}$ as its boundary, $\partial\mathcal{B} = \mathcal{M}$. Then, for a given NLSM field configuration n^a on $\mathcal{M}$, we construct an extension $\tilde{n}^a$ of the NLSM field configuration into the bulk of the manifold $\mathcal{B}$ such that $\tilde{n}^a|_{\partial\mathcal{B}} = n^a$. Finally, using the extended manifold $\mathcal{B}$ and the extension $\tilde{n}^a$ of the NLSM field, the standard construction of the WZ term for $\mathbf{n}$ is given by

$$S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi}{\mathcal{A}_{D+1}} \int_{\mathcal{B}} \tilde{\mathbf{n}}^* \omega_{D+1}, \qquad (4.27)$$

where $\tilde{\mathbf{n}}^* \omega_{D+1}$ denotes the pullback of the volume form ω_{D+1} of S^{D+1} to the extended space $\mathcal{B}$ via the map $\tilde{\mathbf{n}} : \mathcal{B} \to S^{D+1}$. Since in this construction the WZ term depends on the choice of the manifold $\mathcal{B}$ and the choice of extension of the NLSM field $\tilde{\mathbf{n}}$, we need to check that the exponential $e^{-ikS_{\text{WZ}}[\mathbf{n}]}$ is independent of these choices in order for the wave functional to be well defined.

The exponential of the WZ term constructed in this way will be well defined if it is independent of the specific choices of extended manifold $\mathcal{B}$ and field extension $\tilde{n}^a$. To check this, suppose we have two different choices of extended manifold $\mathcal{B}$ and $\mathcal{B}'$, with $\partial\mathcal{B} = \partial\mathcal{B}' = \mathcal{M}$, and two different field extensions $\tilde{\mathbf{n}}$ and $\tilde{\mathbf{n}}'$, with $\tilde{\mathbf{n}}|_{\partial\mathcal{B}} = \tilde{\mathbf{n}}'|_{\partial\mathcal{B}'} = \mathbf{n}$. Let $S_{\text{WZ}}[\mathbf{n}]$ and $S'_{\text{WZ}}[\mathbf{n}]$ be the WZ terms defined using $(\mathcal{B}, \tilde{\mathbf{n}})$ and $(\mathcal{B}', \tilde{\mathbf{n}}')$, respectively. Then, we can write

$$e^{-ikS_{\text{WZ}}[\mathbf{n}]} = e^{-ik(S_{\text{WZ}}[\mathbf{n}] - S'_{\text{WZ}}[\mathbf{n}])} e^{-ikS'_{\text{WZ}}[\mathbf{n}]}. \qquad (4.28)$$

If follows from this expression that the exponential of the WZ term will be well defined if the difference $S_{\text{WZ}}[\mathbf{n}] - S'_{\text{WZ}}[\mathbf{n}]$ of the two WZ terms is an integer multiple of 2π (we assume $k \in \mathbb{Z}$), since in that case we have $e^{-ikS_{\text{WZ}}[\mathbf{n}]} = e^{-ikS'_{\text{WZ}}[\mathbf{n}]}$. The difference of WZ terms is in turn equivalent to a single integral

$$I[\tilde{\mathbf{n}}''] = \frac{2\pi}{\mathcal{A}_{D+1}} \int_{\mathcal{X}} \tilde{\mathbf{n}}''^* \omega_{D+1}, \qquad (4.29)$$

where $\mathcal{X}$ is a closed $(D + 1)$-dimensional manifold formed by gluing $\mathcal{B}$ to $\mathcal{B}'$ along their common boundary $\mathcal{M}$ and where $\tilde{\mathbf{n}}''$ is the NLSM field configuration on the entire $(D + 1)$-dimensional manifold $\mathcal{X}$ formed in this way ($\tilde{\mathbf{n}}$ and $\tilde{\mathbf{n}}'$ agree at the boundary where the gluing takes place, and on the rest of $\mathcal{X}$, they define the configuration $\tilde{\mathbf{n}}''$). Since we are dealing with orientable manifolds, we must specify the orientation of the boundaries of $\mathcal{B}$ and $\mathcal{B}'$ when we glue them together to construct $\mathcal{X}$. In the construction of $\mathcal{X}$ discussed here, the manifolds $\mathcal{B}$ and $\mathcal{B}'$ are glued together in a such a way that the orientation of $\partial\mathcal{B}'$ is opposite to the orientation of $\partial\mathcal{B}$. This choice of orientations is forced on us because we are considering the *difference* of WZ terms.

We see that, in order to determine whether the exponential of the WZ term is well defined, it suffices to check that the integral in Eq. (4.29) is an integer multiple of 2π for *any* closed $(D + 1)$-dimensional manifold $\mathcal{X}$ and any NLSM field configuration $\tilde{\mathbf{n}}''$ on $\mathcal{X}$. To see that this is indeed the case, we note that

$$\int_{\mathcal{X}} \tilde{\mathbf{n}}''^* \omega_{D+1} = \deg[\tilde{\mathbf{n}}''] \int_{S^{D+1}} \omega_{D+1}, \qquad (4.30)$$

where $\deg[\tilde{\mathbf{n}}''] \in \mathbb{Z}$ is the *degree* of the map $\tilde{\mathbf{n}}'' : \mathcal{X} \to S^{D+1}$. It is an integer which counts how many times the space $\mathcal{X}$ is "wrapped" around the sphere S^{D+1} by the map $\tilde{\mathbf{n}}''$ (see, for example, Chap. VI of Ref. [52]). Combining this with the fact that $\int_{S^{D+1}} \omega_{D+1} = \mathcal{A}_{D+1}$, we find that

$$I[\tilde{\mathbf{n}}''] = 2\pi \deg[\tilde{\mathbf{n}}''], \qquad (4.31)$$

which proves that the exponential $e^{-ikS_{\text{WZ}}[\mathbf{n}]}$ of the WZ term is well defined for integer k.

Besides the standard construction of the WZ term using the higher-dimensional manifold $\mathcal{B}$, it is also possible to define $S_{\text{WZ}}[\mathbf{n}]$ as a functional integral in a theory of fermions defined on the manifold $\mathcal{M}$. This construction relies on a result of Abanov and Wiegmann, who constructed theories of fermions coupled to an NLSM field $\mathbf{n}$ which produce a WZ term for $\mathbf{n}$ after integrating out the fermions [11]. The coupling of the fermions to the NLSM field involves a mass parameter M, and the partition function $Z[\mathbf{n}]$ for the theory of fermions coupled to $\mathbf{n}$ can be computed in a gradient expansion in powers of $\frac{1}{M}$. From this partition function, one can define an effective action for the NLSM field via $S_{\text{eff}}[\mathbf{n}] = -\ln(Z[\mathbf{n}])$. In Ref. [11], the authors calculated the variation of $S_{\text{eff}}[\mathbf{n}]$ with respect to $n^a(\mathbf{x})$, and they showed that the imaginary part of this variation has exactly the form of Eq. (3.7). Therefore, the results of Ref. [11] imply that one can *define* the WZ term using the partition function $Z[\mathbf{n}]$ as

$$S_{\text{WZ}}[\mathbf{n}] = -\mathfrak{Im}[\ln(Z[\mathbf{n}])], \qquad (4.32)$$

where $\mathfrak{Im}[\cdots]$ denotes the imaginary part. We also note that this definition naturally produces a WZ term with an integer level $k = \mathrm{sgn}[M]N_F$, where N_F (a positive integer) is the number of flavors of fermions that couple to $\mathbf{n}$. In particular, it does not seem possible to generate a WZ term with fractional level in this way.

So far, we have presented two different ways of constructing the WZ term for the NLSM field $\mathbf{n}$ on a curved space $\mathcal{M}$. One interesting aspect of considering the NLSM on general spaces $\mathcal{M}$ is that there are certain choices of $\mathcal{M}$ where neither of these constructions works. This can be seen as follows. First, the standard construction of the WZ term requires that there exists a $\mathcal{B}$ such that $\partial\mathcal{B} = \mathcal{M}$. However, there are some manifolds $\mathcal{M}$ which cannot be realized as the boundary of any higher-dimensional manifold. The precise conditions for $\mathcal{M}$ to be a boundary are given by the following theorem (see, for example, Chap. 4 of Ref. [53]).

Theorem (Thom).—If all of the *Stiefel-Whitney numbers* of $\mathcal{M}$ are zero, then $\mathcal{M}$ can be realized as the boundary of some smooth compact manifold $\mathcal{B}$.

In dimensions $D = 1, 2$, and 3, every orientable $\mathcal{M}$ is a boundary. The situation becomes more interesting for $D \geq 4$. In the case that $D \equiv 0 \bmod 4$, it is easy to construct simple examples of orientable manifolds $\mathcal{M}$ which are not a boundary by taking products of $C\mathbb{P}^{2r}$ for positive integer r, for example, $C\mathbb{P}^2$ in $D = 4$ and $C\mathbb{P}^2 \times C\mathbb{P}^2$ and $C\mathbb{P}^4$ in $D = 8$. Orientable manifolds which are not a boundary also exist in dimensions $D \geq 4$ where D is not equivalent to zero modulo four, for example, in $D = 5, 9, 10$, and 11 [53]. Thus, we find that for many values of $D > 3$ there are choices of $\mathcal{M}$ where the standard construction of the WZ term fails.

The second construction of the WZ term, defined using a path integral for fermions on $\mathcal{M}$, can fail if the manifold $\mathcal{M}$ does not admit a spin structure. If $\mathcal{M}$ does not admit a spin structure, then it is not possible to formulate a consistent theory of fermions on $\mathcal{M}$. The technical requirement for the existence of a spin structure on $\mathcal{M}$ is that $w_2 \in H^2(\mathcal{M}, \mathbb{Z}_2)$, the second *Stiefel-Whitney class* of $\mathcal{M}$, must vanish [54]. Note that we assume that $\mathcal{M}$ is orientable, and so we also require that the first Stiefel-Whitney class of $\mathcal{M}$, $w_1 \in H^1(\mathcal{M}, \mathbb{Z}_2)$, is trivial. In fact, a spin structure cannot be defined on an unorientable manifold, so this condition is crucial for the second construction of the WZ term using a path integral over fermions.

In some cases, the first construction can fail, but the second construction works. One example of such a case can be found in $D = 4$ when $\mathcal{M}$ is taken to be the *Kummer surface*. This four-dimensional manifold is not a boundary but does admit a spin structure (see, for example, Chap. XI of Ref. [55]). A particularly interesting example, also in $D = 4$, is the choice $\mathcal{M} = C\mathbb{P}^2$. In this case, *both* constructions fail. Therefore, we find that in general a third construction of the WZ term is needed. This third

construction should not require $\mathcal{M}$ to be a boundary, and it should also not require $\mathcal{M}$ to admit a spin structure. We refer to such a construction as an "intrinsic construction" since it does not require an extension $\mathcal{B}$ of $\mathcal{M}$. It turns out that such a construction does exist. In particular, in Ref. [51], Alvarez explained how to carry out this construction in detail using the language of Čech cohomology. In Appendix D, we give an example of this type of construction in the simple case that $D = 1$.

To summarize, we find that the ground state wave functional of the $O(D+2)$ NLSM at $\theta = 2\pi k$, $k \in \mathbb{Z}$, in the disordered ($f \to \infty$) phase is

$$\Psi_{\theta=2\pi k}[\mathbf{n}] = e^{-ikS_{\mathrm{WZ}}[\mathbf{n}]}, \qquad (4.33)$$

where $S_{\mathrm{WZ}}[\mathbf{n}]$ is a suitably defined WZ term for the NLSM field $\mathbf{n}$. As we discussed above, the specific construction of the WZ term will depend on the particular spatial manifold $\mathcal{M}$, but the WZ term always exists. The ground state wave functional has several important properties. First, it is independent of the metric of space, which shows that the disordered phase of the NLSM at $\theta = 2\pi k$ is a topological phase. Second, it possesses the full $SO(D+2)$ symmetry of the action for the NLSM with a theta term. These two properties taken together imply that $\Psi_{\theta=2\pi k}[\mathbf{n}]$ can be the ground state wave functional of an SPT phase. As in Sec. III, we can also understand the ground state wave functional as arising from a unitary transformation by the operator $\hat{\mathcal{U}}^{(k)}$ from Eq. (3.11), which transforms the Hamiltonian at $\theta = 2\pi k$ into the Hamiltonian at $\theta = 0$ as in Eq. (3.13). The ground state of the theory at $\theta = 2\pi k$ can then be obtained by applying $\hat{\mathcal{U}}^{(k)}$ to the ground state at $\theta = 0$ as in Eq. (3.14).

C. Uniqueness of the ground state and the energy gap at large f and $\theta = 2\pi k$, $k \in \mathbb{Z}$

In this section, we study the spectrum of the NLSM on the curved space $\mathcal{M}$ by using a *triangulation* of the manifold to implement a latticelike regularization of the NLSM Hamiltonian on $\mathcal{M}$. Using this regularization, we demonstrate the uniqueness of the ground state and the existence of an energy gap in the $O(D+2)$ NLSM in the disordered ($f \to \infty$) phase at $\theta = 0$ on $\mathcal{M}$. Since the NLSM Hamiltonian at $\theta = 2\pi k$, $k \in \mathbb{Z}$, is related to the Hamiltonian at $\theta = 0$ by a unitary transformation, it will follow from the results of this section that the ground state of the NLSM in the disordered phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$, on $\mathcal{M}$ is also unique for all k. Thus, this subsection completes our proof of the absence of topological order in the $O(D+2)$ NLSM in the disordered phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$.

We start by recalling a few basic facts about triangulations of smooth manifolds, following the discussion in Secs. 3.2 and 5.3.2 of Ref. [56]. Intuitively, a triangulation of a manifold is an approximation of the manifold by generalized triangles called *simplices*. A 0-simplex is a

point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3-simplex is a tetrahedron, and so on.[8] A simplicial complex K is a set of simplices in $\mathbb{R}^n$ such that (i) all faces of simplices from K belong to K, (ii) the intersection of any two simplices from K is a face for each of them, and (iii) any point that belongs to one of the simplices from K has a neighborhood which intersects only finitely many simplices from K. For any such simplicial complex K, the space $|K|$ is the topological space which is the union of all simplices of K with the topology induced by $\mathbb{R}^n$. Finally, a triangulation of a manifold $\mathcal{M}$ is a homeomorphism $\rho : \mathcal{M} \to |K|$, where K is a simplicial complex in $\mathbb{R}^n$ for some n (with n greater than or equal to the dimension of $\mathcal{M}$). Any smooth closed manifold $\mathcal{M}$ admits a triangulation.

Now, let us pick particular triangulation $(\rho, |K|)$ of $\mathcal{M}$. Let $\alpha = 1, \ldots, N_D$ label the distinct D-simplices in $|K|$, and define $\mathcal{S}_\alpha \subset \mathcal{M}$ to be the inverse image of the D-simplex α under the map ρ (ρ is a homeomorphism, so it is invertible). We will also refer to $\mathcal{S}_\alpha$ as a D-simplex. For our purposes, the key property of the triangulation is that it allows for a decomposition of $\mathcal{M}$ as $\mathcal{M} = \sum_{\alpha=1}^{N_D} \mathcal{S}_\alpha$, where the sum is the composition of oriented D-chains. To set up a "lattice" on $\mathcal{M}$, we can then pick an arbitrary point p_α in each $\mathcal{S}_\alpha$ to be the points of the lattice. In the lattice regularization on flat space, each lattice point was associated with a hypercubic unit cell of volume a^D, where a was the lattice spacing. In our regularization on curved space, each point p_α is associated with a D-simplex $\mathcal{S}_\alpha$, and each such D-simplex has a volume given by

$$\mathrm{Vol}_\alpha = \int_{\mathcal{S}_\alpha} \mathrm{Vol}_{\mathcal{M}}, \qquad (4.34)$$

where $\mathrm{Vol}_{\mathcal{M}}$ is the volume form on $\mathcal{M}$ determined by its Riemannian metric. In a system of local coordinates (U, ϕ), $U \subset \mathcal{M}$, $\phi : U \to \mathbb{R}^D$, one has $\mathrm{Vol}_{\mathcal{M}} = d^D\mathbf{x}\sqrt{G(\mathbf{x})}$, if $\mathbf{x} \in \mathbb{R}^D$ denotes the image of a point $p \in U$ under ϕ.

Using this regularization, integration of a function $f(p)$ over $\mathcal{M}$, weighted by the volume form $\mathrm{Vol}_{\mathcal{M}}$, can be discretized as

$$\int_{p \in \mathcal{M}} \mathrm{Vol}_{\mathcal{M}} f(p) \to \sum_{\alpha=1}^{N_D} \mathrm{Vol}_\alpha f(p_\alpha). \qquad (4.35)$$

In addition, we can define a Dirac delta function $\delta(p, p')$ on $\mathcal{M}$ by

$$\int_{p \in \mathcal{M}} \mathrm{Vol}_{\mathcal{M}} f(p) \delta(p, p') = f(p'). \qquad (4.36)$$

---

[8]The *standard* n-simplex is the region $\Delta^n \subset \mathbb{R}^{n+1}$ defined by $\Delta^n = \{\sum_{i=1}^{n+1} y^i = 1, y^i \geq 0 \, \forall \, i\}$.

Again, in a system of local coordinates (U, ϕ), one has $\delta(p, p') = \frac{1}{\sqrt{G(\mathbf{x})}} \delta^{(D)}(\mathbf{x} - \mathbf{x}')$, where $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$ are the images of p, p' under the map ϕ. Then, $\delta(p, p')$ is discretized as

$$\delta(p, p') \to \frac{1}{\mathrm{Vol}_\alpha} \delta_{\alpha\beta} \qquad (4.37)$$

if $p = p_\alpha$ and $p' = p_\beta$, and $\delta_{\alpha\beta}$ is the ordinary Kronecker delta.

Throughout this subsection, we considered the NLSM in a particular coordinate patch of $\mathcal{M}$ with local coordinates $\mathbf{x}$. However, using the Dirac delta function $\delta(p, p')$ on $\mathcal{M}$, we can also write the commutation relations for the NLSM on $\mathcal{M}$ as

$$[\hat{n}^a(p), \hat{n}^b(p')] = 0 \qquad (4.38\mathrm{a})$$

$$[\hat{n}^a(p), \hat{\pi}_b(p')] = i(\delta^a{}_b - \hat{n}^a(p)\hat{n}_b(p'))\delta(p, p') \qquad (4.38\mathrm{b})$$

$$[\hat{\pi}_a(p), \hat{\pi}_b(p')] = i(\hat{\pi}_a(p)\hat{n}_b(p') - \hat{\pi}_b(p')\hat{n}_a(p))\delta(p, p'). \qquad (4.38\mathrm{c})$$

For the lattice regularization of the NLSM using the triangulation discussed above, we define lattice variables by

$$\hat{n}^a(p_\alpha) = \hat{n}_\alpha^a \qquad (4.39\mathrm{a})$$

$$\hat{\pi}_a(p_\alpha) = \frac{1}{\mathrm{Vol}_\alpha} \hat{\pi}_{a,\alpha} \qquad (4.39\mathrm{b})$$

for all points p_α in the lattice constructed from the triangulation. The lattice variables obey the commutation relations

$$[\hat{n}_\alpha^a, \hat{n}_\beta^b] = 0 \qquad (4.40\mathrm{a})$$

$$[\hat{n}_\alpha^a, \hat{\pi}_{b,\beta}] = i(\delta^a{}_b - \hat{n}_\alpha^a \hat{n}_{b,\beta})\delta_{\alpha\beta} \qquad (4.40\mathrm{b})$$

$$[\hat{\pi}_{a,\alpha}, \hat{\pi}_{b,\beta}] = i(\hat{\pi}_{a,\alpha}\hat{n}_{b,\beta} - \hat{\pi}_{b,\beta}\hat{n}_{a,\alpha})\delta_{\alpha\beta}, \qquad (4.40\mathrm{c})$$

and the regularized Hamiltonian takes the form

$$\hat{H}(\rho, |K|) = \frac{f}{2} \sum_{\alpha=1}^{N_D} \frac{1}{\mathrm{Vol}_\alpha} \hat{\pi}_{a,\alpha}\hat{\pi}^a{}_\alpha. \qquad (4.41)$$

Here, we have written $\hat{H}(\rho, |K|)$ to indicate the dependence of the regularized Hamiltonian on the choice of a triangulation $(\rho, |K|)$ of $\mathcal{M}$.

Just as in the case on flat space in Sec. III, the regularized Hamiltonian breaks up into a sum of decoupled Hamiltonians for an $O(D + 2)$ NLSM in one spacetime

dimension. Therefore, we can again apply the results of Appendix A to deduce that the ground state of the $O(D+2)$ NLSM in the disordered ($f \to \infty$) phase on $\mathcal{M}$ is unique and corresponds to the state where $\hat{\pi}_{a,\alpha}\hat{\pi}_\alpha^a = 0 \, \forall \, \alpha$. Unlike the case of flat space, however, there is now a different energy cost to creating an excitation on the different sites p_α in the lattice. This is due the presence of the factor $(\mathrm{Vol}_\alpha)^{-1}$ in the summand in Eq. (4.41) [compare to Eq. (3.18) on flat space, which just contains an overall factor of a^{-D}]. It follows that the energy gap for the regularized $O(D+2)$ NLSM on $\mathcal{M}$ is given by

$$m(\rho, |K|) = \frac{f}{2} \frac{(D+1)}{\bar{\mathrm{Vol}}(\rho, |K|)}, \qquad (4.42)$$

where we defined $\bar{\mathrm{Vol}}(\rho, |K|)$ to be the volume of the *largest* D-chain $\mathcal{S}_\alpha$ in the triangulation $(\rho, |K|)$ of $\mathcal{M}$,

$$\bar{\mathrm{Vol}}(\rho, |K|) = \max\{\mathrm{Vol}_\alpha | \alpha \in \{1, ..., N_D\}\}. \qquad (4.43)$$

Thus, we have proven the uniqueness of the ground state and the existence of an energy gap in the $O(D+2)$ NLSM in the disordered phase at $\theta = 0$. From the previous subsection, we know that the Hamiltonian for the $O(D+2)$ NLSM at $\theta = 2\pi k$, $k \in \mathbb{Z}$, is related to the Hamiltonian at $\theta = 0$ by a unitary transformation, so the result in this section then implies uniqueness of the ground state and the existence of an energy gap in the $O(D+2)$ NLSM in the disordered phase at $\theta = 2\pi k$, $k \in \mathbb{Z}$, for all k. This result completes the demonstration that the $O(D+2)$ NLSM in the disordered phase at $\theta = 2\pi k$ is a suitable model for SPT phases of bosons.

V. CONCLUSION

In this paper, we performed an explicit study of the $O(D+2)$ NLSM with a theta term, in its disordered phase and with theta angle $\theta = 2\pi k$, $k \in \mathbb{Z}$, on arbitrary smooth, closed, connected, oriented D-dimensional spatial manifolds $\mathcal{M}$. We showed that in this parameter regime the ground state of the NLSM on $\mathcal{M}$ is unique, and there is a finite energy gap to the lowest lying excited states. In addition, we showed that the ground state wave functional of the NLSM on $\mathcal{M}$ is independent of the metric on $\mathcal{M}$ and takes the form of an exponential of a WZ term for the NLSM field $\mathbf{n}$, just like in the case on flat space [26]. These results taken together imply that the $O(D+2)$ NLSM, in the disordered phase with $\theta = 2\pi k$, $k \in \mathbb{Z}$, is a suitable model for an SPT phase of bosons. In particular, our results show that this model *does not* possess topological order. Thus, our work places the NLSM approach to SPT phases of Ref. [14] on solid ground.

We close the paper with some additional comments and suggestions for future work. First, we mention one puzzle associated with the NLSM description of SPT phases. In several recent works [57–60], it was shown that important

information about the classification of SPT phases with time-reversal symmetry $\mathbb{Z}_2^T$ can be extracted from the partition functions of these phases on unorientable Euclidean spacetime manifolds. For example, in two spacetime dimensions, there is a single nontrivial bosonic SPT phase protected only by $\mathbb{Z}_2^T$, and in Ref. [57], it was shown that this SPT phase can be detected by its partition function $Z_{R\mathbb{P}^2} = -1$ on the spacetime $R\mathbb{P}^2$. An NLSM description of this SPT phase is available (see Sec. IV. B of Ref. [14]), but it seems problematic to calculate the partition function of the NLSM on $R\mathbb{P}^2$. Mathematically, the issue is that the theta term for the $O(D+2)$ NLSM involves the pullback to spacetime of the volume form on S^{D+1}, and this pullback does not seem to make sense when the spacetime is not orientable. Therefore, it would be interesting to see if there is some way to make sense of the NLSM description of SPT phases on unorientable spacetime manifolds.

A second possible direction for future work would be to extend the analysis of this paper to the case of the $O(D+2)$ NLSM in the disordered limit when the theta angle θ is an *odd* multiple of π. In Ref. [13], the authors used a qualitative argument to map out the phase diagram of the $O(4)$ NLSM in $D = 2$ spatial dimensions with a theta term and coefficient $\theta = \pi$. They proposed two possible phases for this theory when the coupling constant f is large: (i) a gapless phase realized at some finite but large value of f and (ii) a gapped phase which is realized in the extreme disordered limit of $f \to \infty$. In addition, the authors of Ref. [13] argued that the ground state in the gapped phase should be doubly degenerate (see also Ref. [12] on this point). It would be interesting to investigate the $f \to \infty$ phase directly within the canonical formalism, with the goal of proving that in this limit the spectrum is indeed gapped and that the ground state is doubly degenerate. It would also be interesting to investigate the dependence of the ground state degeneracy on the topology of the spatial manifold. This problem is quite interesting for the following reason. Typically, the boundary theory of an SPT phase is expected to preserve the symmetry of the SPT phase and be gapless, or to spontaneously break the symmetry in some way (which may lead to a gapped boundary theory). However, at the $(2+1)$-dimensional boundary of a $(3+1)$-dimensional SPT phase (and presumably also in higher dimensions), there is a third possibility: the boundary theory can be gapped and symmetric, but it must also possess intrinsic topological order [61,62]. It is likely that the $O(4)$ NLSM, in the disordered phase and at $\theta = \pi$, can describe such a gapped, symmetry-preserving, and topologically ordered surface state of the bosonic topological insulator phase in $3 + 1$ dimensions [61]. Therefore, it would be interesting to give a proof that the $O(4)$ NLSM in this parameter regime really does possess intrinsic topological order. As we discussed in Sec. III, the unitary transformation which removes the theta angle from the Hamiltonian can only be performed when θ is a multiple of 2π, which means that completely

new methods will be needed to solve the problem of the NLSM in the disordered phase at $\theta = \pi$.

ACKNOWLEDGMENTS

APPENDIX A: CANONICAL QUANTIZATION OF THE $O(N)$ NLSM IN $0+1$ DIMENSIONS

In this Appendix, we review the solution of the $O(N)$ NLSM in one spacetime dimension using the commutation relations of Eqs. (2.29) and the Schrödinger representation in Eq. (2.30). We use this solution in Secs. III and IV to compute the energy gap in regularized versions of the $O(D+2)$ NLSM in the disordered ($f \to \infty$) limit and with $\theta = 2\pi k$, $k \in \mathbb{Z}$, on flat and curved space, respectively. The $O(N)$ NLSM in one spacetime dimension is equivalent to the quantum mechanics problem of a free particle in N spatial dimensions but confined to the surface of the unit sphere S^{N-1}. This is a famous problem in the quantization of constrained systems and has been studied by many authors [41–49]. One finds that the quantum mechanical Hamiltonian is proportional to the Laplace-Beltrami operator $\Delta_{LB}^{(N-1)}$ on the sphere S^{N-1}, so that the energy spectrum is given in terms of the eigenvalues of $\Delta_{LB}^{(N-1)}$. However, there is some controversy in the literature about whether an additional constant term, depending only on N, should appear in the quantum Hamiltonian for this problem. This constant term is irrelevant for the application to our discussion in the context of Secs. III and IV, in which we are interested only in the difference between the energy of the ground state and first excited state. Therefore, in this Appendix, we give a straightforward analysis of the $O(N)$ NLSM in one spacetime dimension, without worrying about subtleties (e.g., Weyl ordering of operators to define the quantum Hamiltonian [41,44,46]) which could lead to an extra constant shift in the energy spectrum. Readers interested in the subtleties associated with this constant term should consult the references cited in this paragraph.

1. Hamiltonian, commutation relations, and the energy spectrum

We consider the $O(N)$ NLSM in one spacetime dimension. For general N, this system does not admit a theta term (a theta term is possible at $N = 2 = D+2$ since $D = 0$ here). However, in this Appendix, we are only interested in discussing the quantization of the theory without a theta term, and so we consider the case of a general N with no theta term. Let $\mathbf{n} = (n^1, \ldots, n^N)$ be the NLSM field. Again, the

NLSM field is subject to the constraint $\mathbf{n} \cdot \mathbf{n} = n^a n_a = 1$. The Lagrangian in one spacetime dimension is

$$\mathcal{L} = \frac{1}{2f}(\partial_t n^a)(\partial_t n_a). \tag{A1}$$

The canonical momentum conjugate to n^a is $\pi_a = \frac{\partial \mathcal{L}}{\partial(\partial_t n^a)} = \frac{1}{f}(\partial_t n_a)$, and the Hamiltonian takes the form

$$H = \frac{f}{2}\pi^a \pi_a. \tag{A2}$$

In one spacetime dimension, the analysis of constrained Hamiltonian systems from Sec. II leads to the commutation relations

$$[\hat{n}^a, \hat{n}^b] = 0 \tag{A3a}$$

$$[\hat{n}^a, \hat{\pi}_b] = i\left(\delta^a{}_b - \frac{\hat{n}^a \hat{n}_b}{\hat{r}^2}\right) \tag{A3b}$$

$$[\hat{\pi}_a, \hat{\pi}_b] = \frac{i}{\hat{r}^2}(\hat{\pi}_a \hat{n}_b - \hat{\pi}_b \hat{n}_a), \tag{A3c}$$

where $\hat{r}^2 = \hat{n}^a \hat{n}_a$. The operator $\hat{r}^2$ commutes with all other operators by construction, and so it can be set equal to 1 at this point, exactly as in Sec. II. However, we find it more convenient for the exposition in this Appendix to leave this operator in place and only set it to 1 at the end of the analysis.

The Schrödinger representation used in this paper for the NLSM commutation relations can be adapted to the case where the operator $\hat{r}^2$ is kept in the commutation relations. In this case, the operator $\hat{n}^a$ again acts as multiplication by the coordinate n^a, but the momentum operator $\hat{\pi}_a$ now takes the form

$$\hat{\pi}^a = -i\left(\delta_a{}^b - \frac{n_a n^b}{r^2}\right)\frac{\partial}{\partial n^b}. \tag{A4}$$

In this representation, the quantity $\hat{\pi}^a \hat{\pi}_a$ appearing in the Hamiltonian takes the explicit form

$$\hat{\pi}^a \hat{\pi}_a = -\left\{\left(\delta^{ab} - \frac{n^a n^b}{r^2}\right)\frac{\partial^2}{\partial n^a \partial n^b} - \frac{(N-1)}{r^2}n^a \frac{\partial}{\partial n^a}\right\}, \tag{A5}$$

so that the Hamiltonian operator is

$$\hat{H} = -\frac{f}{2}\left\{\left(\delta^{ab} - \frac{n^a n^b}{r^2}\right)\frac{\partial^2}{\partial n^a \partial n^b} - \frac{(N-1)}{r^2}n^a \frac{\partial}{\partial n^a}\right\}. \tag{A6}$$

To diagonalize the Hamiltonian, we now prove the following statement.

Claim.—In the Schrödinger representation, the squared sum of canonical momenta is related to $\Delta_{LB}^{(N-1)}$, the Laplace-Beltrami operator on the unit sphere S^{N-1}, by the equation

$$\hat{\pi}^a \hat{\pi}_a = -\frac{1}{r^2} \Delta_{LB}^{(N-1)}. \tag{A7}$$

Proof.—Before imposing the NLSM constraint, the components n^a of the NLSM field are coordinates on $\mathbb{R}^N$. The ordinary Laplacian on $\mathbb{R}^N$ is given in terms of the Laplace-Beltrami operator $\Delta_{LB}^{(N-1)}$ on S^{N-1} by (see, e.g., Sec. II.4 of Ref. [63])

$$\frac{\partial^2}{\partial n^a \partial n_a} = \frac{\partial^2}{\partial r^2} + \frac{(N-1)}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\Delta_{LB}^{(N-1)}. \tag{A8}$$

Now, using $\frac{\partial}{\partial r} = \frac{n^a}{r}\frac{\partial}{\partial n^a}$ and the relation

$$\frac{\partial^2}{\partial r^2} = \frac{n^a n^b}{r^2}\frac{\partial^2}{\partial n^a \partial n^b}, \tag{A9}$$

we find that

$$-\hat{\pi}^a\hat{\pi}_a = \left[\frac{\partial^2}{\partial n^a \partial n_a} - \left(\frac{\partial^2}{\partial r^2} + \frac{(N-1)}{r}\frac{\partial}{\partial r}\right)\right]$$
$$= \frac{1}{r^2}\Delta_{LB}^{(N-1)}. \tag{A10}$$

This completes the proof. ∎

As we discussed earlier in this section, since the operator $\hat{r}^2$ commutes with all other operators, the NLSM constraint $\hat{r}^2 = 1$ can be enforced at any time in the correctly quantized theory. At this point, we can then set $r^2 = 1$ to obtain the final form for the Hamiltonian of the $O(N)$ NLSM,

$$\hat{H} = -\frac{f}{2}\Delta_{LB}^{(N-1)}. \tag{A11}$$

We note here that in more careful approaches to the quantization of this model, the Hamiltonian operator takes the form $\hat{H} = \frac{f}{2}(-\Delta_{LB}^{(N-1)} + \bar{E}(N))$, where $\bar{E}(N)$ is a constant shift of the energy depending only on N [although in the literature there is still some disagreement about the correct value of $\bar{E}(N)$]. In the straightforward approach used in this Appendix, this shift is not present.

The spectrum of $\Delta_{LB}^{(N-1)}$, as well as its eigenfunctions and their multiplicities, can be found in standard references on Riemannian geometry, for example, Ref. [63]. Using these standard results, we find that the eigenvalues of $\hat{H}$ are labeled by a positive integer ℓ and given explicitly by

$$E_\ell = \frac{f}{2}\ell(\ell + N - 2), \qquad \ell \in \mathbb{N}, \tag{A12}$$

in agreement with previous results on the spectrum of this model. The ground state of this theory has energy zero [or $\frac{f}{2}\bar{E}(N)$ for a nonzero shift in the energies], and the difference between the energy of the ground state and first excited state is given by

$$m \equiv E_1 - E_0 = \frac{f}{2}(N-1). \tag{A13}$$

To get a complete understanding of the $O(N)$ NLSM in one spacetime dimension, we now give a full analysis of the symmetries of this system.

2. Symmetry analysis

Consider the Lie algebra $so(N)$ of the Lie group $SO(N)$. In the fundamental (i.e., $N \times N$) representation, one possible basis of the Lie algebra consists of the antisymmetric $N \times N$ matrices $\mathbf{E}^{ij}$ which contain a 1 in the (i, j) entry, a -1 in the (j, i) entry, and zero in all other entries. Since $\mathbf{E}^{ij} = -\mathbf{E}^{ji}$, and since i and j must be distinct for this to make sense, we arrive at the correct number $N(N-1)/2$ of generators of $so(N)$. It is also convenient to define the matrices $\mathbf{E}^{ii}$ for any i to be equal to the matrix with all entries equal to zero. The matrix elements of $\mathbf{E}^{ij}$ are

$$(\mathbf{E}^{ij})^a{}_b = \delta^{ia}\delta^j{}_b - \delta^{ja}\delta^i{}_b. \tag{A14}$$

This definition also works when $i = j$ and yields the zero matrix in that case.

The Lagrangian of Eq. (A1) has an $SO(N)$ global symmetry which is reflected in the fact that it is invariant under the transformation $n^a \to R^a{}_b n^b$ for any $O(N)$ matrix R. The infinitesimal form of this transformation is $n^a \to n^a + \delta n^a$ with $\delta n^a = \epsilon(\mathbf{E}^{ij})^a{}_b n^b$, for a small constant ϵ. By making ϵ time dependent, we derive the conserved currents of this model,

$$J_{ab} = \frac{1}{f}(\partial_t n_a n_b - \partial_t n_b n_a). \tag{A15}$$

Since we are in $0 + 1$ dimensions, the conserved charge operators are obtained from this current simply by replacing $\partial_t n_a$ with $\hat{\pi}_a$, so the conserved charge operators are (note that we have chosen a particular operator ordering here)

$$\hat{Q}_{ab} = \hat{\pi}_a \hat{n}_b - \hat{\pi}_b \hat{n}_a. \tag{A16}$$

The commutator of two momenta from Eq. (A3) can be rewritten (at this point, we set $\hat{r}^2 = 1$ in the commutators) in terms of these charge operators as

$$[\hat{\pi}_a, \hat{\pi}_b] = i\hat{Q}_{ab}. \tag{A17}$$

In the Schrödinger representation (with $r^2 = 1$), the charges $\hat{Q}_{ab}$ take the simple form

$$\hat{Q}_{ab} = i\left(n_a \frac{\partial}{\partial n^b} - n_b \frac{\partial}{\partial n^a}\right). \tag{A18}$$

When acting on functions of n^a, these derivative operators obey the Lie algebra of $so(N)$,

$$[\hat{Q}_{ab}, \hat{Q}_{cd}] = \delta_{ac}\hat{Q}_{bd} - \delta_{ad}\hat{Q}_{bc} + \delta_{bd}\hat{Q}_{ac} - \delta_{bc}\hat{Q}_{ad}. \tag{A19}$$

We now show that the Hamiltonian Eq. (A2) of this system is proportional to the quadratic Casimir of $so(N)$. It then follows that the problem of diagonalizing the Hamiltonian of the $O(N)$ NLSM reduces to a study of the representation theory of $so(N)$, which is already well known. The quadratic Casimir of $so(N)$ is given by the sum of the squares of all the generators $\hat{Q}_{ab}$. We know that half of these are redundant since $\hat{Q}_{ab} = -\hat{Q}_{ba}$, but we can exploit this fact and the fact that $\hat{Q}_{aa} = 0$ to write the Casimir as simply

$$\hat{C} = \frac{1}{2}\hat{Q}_{ab}\hat{Q}^{ab}, \tag{A20}$$

where we have summed over all values of a and b with no restrictions. By explicit computation, one can show that $\hat{C} = \hat{\pi}^a\hat{\pi}_a$ [when we set $r^2 = 1$ in Eq. (A5)], and so the Hamiltonian can be rewritten as

$$\hat{H} = \frac{f}{2}\hat{C}. \tag{A21}$$

In this form, one can clearly see the relationship between the Hamiltonian and the $SO(N)$ symmetry of this model.

APPENDIX B: REGULARIZATION OF THE NLSM HAMILTONIAN

In Sec. III, we studied the energy gap of the NLSM in the disordered ($f \to \infty$) limit and at $\theta = 0$ using a lattice regularization. We briefly indicated there that some kind of regularization scheme was necessary to study the excited states of the NLSM, and then we immediately implemented the lattice regularization. In this Appendix, we explain in detail why it is necessary to regularize the NLSM Hamiltonian to study the excited states, and we also discuss an alternative regularization for the theory on flat space which does not use a lattice. We show that this alternative regularization gives results for the energy gap of the theory which are consistent with the result coming from the lattice regularization. Based on this evidence, we expect that any sensible regularization scheme will give a result for the energy gap of the NLSM which agrees with our result

computed using the lattice regularization. In this Appendix, we focus on the NLSM Hamiltonian at $\theta = 0$. As we explained in Sec. III, the NLSM Hamiltonian at $\theta = 2\pi k$, $k \in \mathbb{Z}$, is related to the NLSM Hamiltonian at $\theta = 0$ by a unitary transformation. Therefore, any result on the spectrum of this theory at $\theta = 0$ will also hold for the theory at $\theta = 2\pi k$ for integer k.

We start by explaining why the NLSM Hamiltonian must be regularized before excited states can be constructed. Recall that in the limit of large coupling f the Hamiltonian for the $O(N)$ NLSM in D spatial dimensions takes the form

$$\hat{H} = \frac{f}{2}\int d^D\mathbf{x}\,\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x}), \tag{B1}$$

where $\hat{\pi}_a(\mathbf{x})$ takes the form shown in Eq. (2.30) in the Schrödinger representation used in this paper. The ground state of this Hamiltonian has zero energy and is characterized by the property that it is annihilated by $\hat{\pi}^a(\mathbf{x})$ for each a. Therefore, to construct the ground state, we only have to consider the action of a single operator $\hat{\pi}^a(\mathbf{x})$ on functionals of the NLSM field. On the other hand, to construct excited states, we need to act with the product $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x})$. This operator is not well defined in the NLSM field theory, as we now show.

To see the problem with the operator $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x})$, we look at the action of $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{y})$ on some functional F of the NLSM field. A short calculation shows that

$$\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{y})F = (N-1)\delta^{(D)}(\mathbf{x}-\mathbf{y})n^a(\mathbf{x})\frac{\delta F}{\delta n^a(\mathbf{x})}$$
$$+ n^a(\mathbf{y})n^b(\mathbf{y})\frac{\delta^2 F}{\delta n^a(\mathbf{x})\delta n^b(\mathbf{y})}$$
$$- n_a(\mathbf{x})n^b(\mathbf{x})n^a(\mathbf{y})n^c(\mathbf{y})\frac{\delta^2 F}{\delta n^b(\mathbf{x})\delta n^c(\mathbf{y})}. \tag{B2}$$

We see that the operator $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{y})$ will diverge as $\mathbf{y}$ approaches $\mathbf{x}$ because of the presence of the delta function in the first term of this expression. This contact divergence implies that the product $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x})$ of momentum operators at the same point $\mathbf{x}$ in space is ill defined, and this is the basic reason why some regularization scheme is needed to construct excited states in this field theory.

Since the divergence in the operator $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x})$ is due to the fact that both factors of $\hat{\pi}_a(\mathbf{x})$ are evaluated at the same point, i.e., the problem is associated with short distances, one way to regulate the operator is to discretize space by introducing a lattice. This is exactly the approach we took in Sec. III. However, other regularization schemes are also possible and should give expressions for the energy gap which agree with the answers obtained from the lattice regularization. To show this, we now discuss one alternative regularization scheme, in which we still consider the

theory on a continuous space, but we introduce some nonlocality to regulate the product $\hat{\pi}_a(\mathbf{x})\hat{\pi}^a(\mathbf{x})$. In this regularization scheme, we first rewrite the Hamiltonian as

$$\hat{H} = \frac{f}{2} \int d^D\mathbf{x} d^D\mathbf{y} \hat{\pi}_a(\mathbf{x}) \hat{\pi}^a(\mathbf{y}) \delta^{(D)}(\mathbf{x} - \mathbf{y}) \qquad \text{(B3)}$$

and then replace the delta function $\delta^{(D)}(\mathbf{x} - \mathbf{y})$ with any known regularized expression for a delta function,

$$\delta^{(D)}(\mathbf{x} - \mathbf{y}) \rightarrow \eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y}). \qquad \text{(B4)}$$

Here, $\eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y})$ is some function of $\mathbf{x} - \mathbf{y}$ which has the property that

$$\lim_{\epsilon \to 0} \int d^D\mathbf{x} f(\mathbf{x}) \eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y}) = f(\mathbf{y}), \qquad \text{(B5)}$$

for any test function $f(\mathbf{x})$. The parameter ϵ has units of length, and it is this small parameter which serves as a regulator for the theory in this regularization scheme. We consider a concrete example of a such a function $\eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y})$ later in this Appendix. In terms of this function, the regularized Hamiltonian takes the form

$$\hat{H}(\epsilon) = \frac{f}{2} \int d^D\mathbf{x} d^D\mathbf{y} \hat{\pi}_a(\mathbf{x}) \hat{\pi}^a(\mathbf{y}) \eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y}). \qquad \text{(B6)}$$

This regularization scheme clearly introduces some non-locality into the theory, since the regularized Hamiltonian (with nonzero ϵ) contains terms which involve the fields $\hat{\pi}_a$ at two different points in space.

Within this alternative regularization scheme, we can also compute the energy gap between the ground state of the system and the first excited state. The vacuum state of the NLSM in the large f limit (and at $\theta = 0$) is the constant wave functional $\Psi[\mathbf{n}] = 1$, which transforms in the trivial representation of $SO(N)$. Experience with the lattice regularization of this theory, and intuition about the role of the $SO(N)$ symmetry in this problem, suggests that the lowest-energy states should transform in the vector representation of $SO(N)$. The simplest such states are given by functionals of the form

$$F^a[\mathbf{n}] = \int d^D\mathbf{x} n^a(\mathbf{x}) F(\mathbf{x}), \qquad \text{(B7)}$$

where $F(\mathbf{x})$ is some arbitrary function of space. For example, we can construct a localized excitation by choosing $F(\mathbf{x})$ to be localized in some region of space. Applying the regularized Hamiltonian to this state gives

$$\hat{H}(\epsilon) F^a[\mathbf{n}] = m(\epsilon) F^a[\mathbf{n}], \qquad \text{(B8)}$$

where the energy $m(\epsilon)$ in this regularization is given by

$$m(\epsilon) = \frac{f}{2}(N - 1)\eta_\epsilon^{(D)}(0). \qquad \text{(B9)}$$

At this point, it is instructive to make a particular choice of regularization of the delta function. We choose the Poisson kernel,

$$\eta_\epsilon^{(D)}(\mathbf{x} - \mathbf{y}) = \prod_{j=1}^{D} \frac{1}{\pi} \frac{\epsilon}{\epsilon^2 + (x^j - y^j)^2}, \qquad \text{(B10)}$$

but other choices are also possible (e.g., a heat kernel, etc.). If we also set $N = D + 2$, which is the case of interest in this paper for constructing NLSMs with a theta term in $D + 1$ spacetime dimensions, then we find that the mass gap in this regularization is

$$m(\epsilon) = \frac{f}{2} \frac{D + 1}{(\pi\epsilon)^D}. \qquad \text{(B11)}$$

This answer is clearly consistent with the expression Eq. (3.21) obtained from the lattice regularization, and the two expressions coincide if we choose the lattice spacing a to be related to the parameter ϵ via $a = \pi\epsilon$. Therefore, we expect that any sensible regularization of the NLSM Hamiltonian will give results consistent with those that we derived in Sec. III using the lattice regularization.

APPENDIX C: SYMPLECTIC GEOMETRY APPROACH TO HAMILTONIAN FORMALISM FOR FIELD THEORIES

In this Appendix, we review the symplectic geometry approach to the Hamiltonian dynamics of a classical field theory. This is essentially a "functional" version of what one does in the symplectic geometry approach to a classical dynamical system with finitely many degrees of freedom (a review of the latter for physicists is given in Chap. 11 of Ref. [64]). We apply this formalism in Sec. IV to determine the correct form of the Poisson bracket in the theory of a free scalar field on a D-dimensional curved space $\mathcal{M}$ with some Riemannian metric G_{ij}, and then we use this information to quantize the $O(D + 2)$ NLSM with a theta term on the curved space $\mathcal{M}$. One of the main advantages of the symplectic geometry approach is that it provides a formalism on which one can rely to understand the classical dynamics, and in particular the correct form of the Poisson bracket, in systems which cannot be analyzed by conventional methods more familiar to physicists. The correct form of the Poisson brackets is essential for quantization, and so this method is useful for the proper quantization of an unfamiliar system. This material is standard in the field theory literature. Therefore, in this section, we simply give a summary of this material in the infinite-dimensional setting in exact analogy to the development on a finite-dimensional phase space as found, for example, in

Ref. [64]. In addition, we note here that a very similar infinite-dimensional symplectic geometry approach is used in establishing the equivariant localization formulas for phase space path integrals in quantum mechanics (see, for example, Sec. 4.3 of Ref. [65]).

To start, consider a field theory on the D-dimensional space $\mathcal{M}$, and let $\{\Phi^a(\mathbf{x})\}_{\mathbf{x}\in\mathcal{M}}$ (for some range of the index a) denote the coordinates on the infinite-dimensional phase space for the system under consideration. In the example of a free scalar field $\phi(\mathbf{x})$, we could choose $\Phi^1(\mathbf{x}) = \phi(\mathbf{x})$, $\Phi^2(\mathbf{x}) = \pi(\mathbf{x})$, where $\pi(\mathbf{x})$ is the momentum conjugate to $\phi(\mathbf{x})$, but in general (as in the finite-dimensional case), it is not necessary to have a definite decomposition into "coordinates" and "momenta." In fact, in many cases, it is *impossible* to find a definition of coordinates and momenta which is valid on the entire phase space. We use the notation $\vec{\Phi}(\mathbf{x}) = (\Phi^1(\mathbf{x}), \Phi^2(\mathbf{x}), \ldots)$ to denote the collection of all field variables $\Phi^a(\mathbf{x})$ at the single point $\mathbf{x}$. The functional derivatives $\frac{\delta}{\delta\Phi^a(\mathbf{x})}$ with respect to the phase space coordinates form a basis of the tangent space at a point in this phase space. We also introduce the coordinate differentials $\delta\Phi^a(\mathbf{x})$, which form a basis for the cotangent space at a point in phase space. We have the natural pairing between the basis elements of the tangent and cotangent spaces,

$$\delta\Phi^a(\mathbf{x})\left(\frac{\delta}{\delta\Phi^b(\mathbf{y})}\right) = \delta^a{}_b\delta^{(D)}(\mathbf{x}-\mathbf{y}). \qquad (C1)$$

On phase space, we also introduce an exterior derivative δ which acts on any functional F of the phase space coordinates as

$$\delta F = \int d^D\mathbf{x}\,\frac{\delta F}{\delta\Phi^a(\mathbf{x})}\delta\Phi^a(\mathbf{x}). \qquad (C2)$$

The wedge product of differentials $\delta\Phi^a(\mathbf{x})$ is defined in the usual way by

$$\delta\Phi^a(\mathbf{x})\wedge\delta\Phi^b(\mathbf{y}) = \delta\Phi^a(\mathbf{x})\otimes\delta\Phi^b(\mathbf{y}) - \delta\Phi^b(\mathbf{y})\otimes\delta\Phi^a(\mathbf{x}). \qquad (C3)$$

A general p-form α on phase space has the form

$$\alpha = \frac{1}{p!}\int\left(\prod_{j=1}^{p}d^D\mathbf{x}_j\right)\alpha_{a_1\cdots a_p}[\vec{\Phi}(\mathbf{x}_1), \ldots, \vec{\Phi}(\mathbf{x}_p);$$

$$\mathbf{x}_1, \ldots, \mathbf{x}_p]\delta\Phi^{a_1}(\mathbf{x}_1)\wedge\cdots\wedge\delta\Phi^{a_p}(\mathbf{x}_p), \qquad (C4)$$

where $\alpha_{a_1\cdots a_p}[\vec{\Phi}(\mathbf{x}_1), \ldots, \vec{\Phi}(\mathbf{x}_p); \mathbf{x}_1, \ldots, \mathbf{x}_p]$ are the components of α. The notation is meant to indicate that the components of α can depend on the fields $\Phi^a(\mathbf{x}_j)$ at the coordinates $\mathbf{x}_j$, and they can also depend explicitly on the coordinates $\mathbf{x}_j$. The action of the exterior derivative δ

on p-forms is defined by the usual axioms: (i) $\delta^2 F = 0$ for any functional F on phase space and (ii) $\delta(\alpha\wedge\beta) = \delta\alpha\wedge\beta + (-1)^p\alpha\wedge\delta\beta$ for any p-form α and any q-form β (i.e., δ is an antiderivation). A general vector field $\underline{V}$ on the phase space has the form

$$\underline{V} = \int d^D\mathbf{x}\,V^a[\vec{\Phi}(\mathbf{x});\mathbf{x}]\frac{\delta}{\delta\Phi^a(\mathbf{x})}, \qquad (C5)$$

where $V^a[\vec{\Phi}(\mathbf{x});\mathbf{x}]$ are the components of $\underline{V}$. The *interior multiplication* of a form α by a vector $\underline{V}$, denoted $i_{\underline{V}}\alpha$, is given by

$$i_{\underline{V}}\alpha = \frac{1}{(p-1)!}\int\left(\prod_{j=1}^{p}d^D\mathbf{x}_j\right)V^a\alpha_{aa_2\cdots a_p}\delta\Phi^{a_2}(\mathbf{x}_2)$$

$$\wedge\cdots\wedge\delta\Phi^{a_p}(\mathbf{x}_p), \qquad (C6)$$

where we suppressed the arguments of $\alpha_{aa_2\cdots a_p}$ and V^a for brevity.

After these preliminaries, we are now ready to develop the canonical formalism on this infinite-dimensional phase space in exact analogy to the development on a finite-dimensional phase space (see, for example, Chap. 11 of Ref. [64]). First, we introduce a symplectic form Ω on phase space, the components of which are defined by

$$\Omega = \frac{1}{2}\int d^D\mathbf{x}_1 d^D\mathbf{x}_2\Omega_{ab}[\vec{\Phi}(\mathbf{x}),\vec{\Phi}(\mathbf{y});\mathbf{x},\mathbf{y}]\delta\Phi^a(\mathbf{x})\wedge\delta\Phi^b(\mathbf{y}).$$

$$(C7)$$

As usual, we require that Ω is closed, $\delta\Omega = 0$. Note also that in this infinite-dimensional case the components $\Omega_{ab}[\vec{\Phi}(\mathbf{x}),\vec{\Phi}(\mathbf{y});\mathbf{x},\mathbf{y}]$ of Ω only need to be antisymmetric under the *simultaneous* exchange $a\leftrightarrow b$ and $\mathbf{x}\leftrightarrow\mathbf{y}$. Next, for any functional F on phase space, we define a vector field $\underline{V}_F$ by the relation

$$\delta F = -i_{\underline{V}_F}\Omega. \qquad (C8)$$

The reason for defining these vector fields in this way is that they allow for a coordinate-independent definition of the Poisson bracket of two functionals F_1 and F_2 on the phase space. The Poisson bracket for F_1 and F_2 is given in terms of the corresponding vector fields $\underline{V}_{F_1}$ and $\underline{V}_{F_2}$ by[9]

$$\{F_1, F_2\} = i_{\underline{V}_{F_1}}i_{\underline{V}_{F_2}}\Omega. \qquad (C9)$$

Finally, Hamilton's equations are equivalent to the single equation

---

[9] We use an opposite sign in this equation as compared with Ref. [64].

$$\delta H = -i_{\underline{V}_H}\Omega, \tag{C10}$$

where $\underline{V}_H$ is the vector field of which the components are the time derivatives of phase space coordinates,

$$\underline{V}_H = \int d^D\mathbf{x} \dot{\Phi}^a(\mathbf{x}) \frac{\delta}{\delta\Phi^a(\mathbf{x})}, \tag{C11}$$

and where the dot represents a time derivative, $(\dot{\ }) := \frac{d}{dt}(\ )$.

APPENDIX D: INTRINSIC CONSTRUCTION OF THE WESS-ZUMINO TERM FOR $D=1$

In this Appendix, we explain in detail a simple example of the intrinsic construction of the WZ term which we mentioned in Sec. IV. This intrinsic construction is the method of choice for constructing the WZ term when other standard constructions fail, for example, in the case where the spatial manifold $\mathcal{M}$ does not admit a spin structure and also cannot be realized as the boundary of any higher-dimensional manifold $\mathcal{B}$. The details of the intrinsic construction are, however, more complicated than the more standard constructions. For this reason, we only present the simplest example of the construction, which is the case where $\mathcal{M}$ has dimension $D=1$, but this intrinsic construction is available in all dimensions. The ideas behind this construction date back to work of Wu and Yang in Ref. [66] and were formalized by Alvarez in Ref. [51] using the language of Čech cohomology. The basic idea is to write the WZ term as a sum of integrals over the space $\mathcal{M}$ of forms which are defined only locally in certain coordinate patches on the target manifold $\mathcal{T}$ of the NLSM. This sum of integrals over $\mathcal{M}$ is then supplemented with additional terms which account for the transition functions which are needed to go between coordinate patches on the target manifold. We also mention here that the methods of Ref. [51] were used in Ref. [67] to give an intrinsic definition of the Abelian Chern-Simons term on a three-dimensional manifold.

We now present the intrinsic construction of the WZ term for the case where the spatial manifold $\mathcal{M}$ has dimension $D=1$. Our discussion in this Appendix applies only to the specific case where the target manifold of the NLSM is the sphere S^2. Other two-dimensional target spaces may require more coordinate patches to be covered properly. As we discussed in the main sections of this paper, we assume that $\mathcal{M}$ is closed, oriented, and connected. For the case $D=1$, this implies that $\mathcal{M}$ is diffeomorphic to a circle. We take $x^1 \in [a,b]$ to be the coordinate on this circle, and the NLSM field is taken to obey periodic boundary conditions $\mathbf{n}(a) = \mathbf{n}(b)$. For $D=1$, we have an $O(3)$ NLSM, and for the purpose of constructing the WZ term, it will be convenient to parametrize the field variables n^a, $a=1$, 2, 3, using spherical coordinates Φ and Θ as

$$n^1 = \cos(\Phi)\sin(\Theta) \tag{D1a}$$

$$n^2 = \sin(\Phi)\cos(\Theta) \tag{D1b}$$

$$n^3 = \cos(\Theta). \tag{D1c}$$

In these coordinates, the volume form on S^2 (the target manifold of the NLSM) takes the form

$$\omega_2 = \sin(\Theta)d\Theta \wedge d\Phi. \tag{D2}$$

The sphere S^2 can be covered by two coordinate patches U_N and U_S, defined in spherical coordinates as $U_N = \{(\Phi,\Theta)|\Phi\in[0,2\pi),\Theta\in[0,\pi-\Theta_0)\}$ and $U_S = \{(\Phi,\Theta)|\Phi\in[0,2\pi),\Theta\in(\Theta_0,\pi]\}$, for some fixed (perhaps small) angle Θ_0. The patch U_N contains the north pole but not the south pole, and the patch U_S contains the south pole but not the north pole. On each patch, the volume form can be expressed as a total derivative $\omega_2 = d\vartheta_N$ or $\omega_2 = d\vartheta_S$, with

$$\vartheta_N = (1-\cos(\Theta))d\Phi \tag{D3a}$$

$$\vartheta_S = -(1+\cos(\Theta))d\Phi. \tag{D3b}$$

On the intersection $U_S \cap U_N$ of the two coordinate patches, we have $\vartheta_S - \vartheta_N = d\psi_{SN}$, where (up to an arbitrary constant)

$$\psi_{SN} = -2\Phi. \tag{D4}$$

At this point, we recall that the ordinary construction of the WZ term for the NLSM on $\mathcal{M}$ uses an extended manifold $\mathcal{B}$ with $\partial\mathcal{B} = \mathcal{M}$ and an extension $\tilde{\mathbf{n}}$ of the NLSM field into $\mathcal{B}$ such that $\tilde{\mathbf{n}}|_{\partial\mathcal{B}} = \mathbf{n}$. In this case, the standard construction for the WZ term is

$$S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2}\int_{\mathcal{B}}\tilde{\mathbf{n}}^*\omega_2. \tag{D5}$$

Using the spherical coordinates Θ and Φ, we find that the variation of the WZ term obtained from this standard construction is

$$\delta S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2}\int_a^b dx^1 \sin(\Theta)(\delta\Theta\partial_1\Phi - \delta\Phi\partial_1\Theta), \tag{D6}$$

where $\partial_1 \equiv \frac{\partial}{\partial x^1}$. We now present the intrinsic construction of the WZ term, which yields an expression for $S_{\text{WZ}}[\mathbf{n}]$ involving only integrations over the physical space $\mathcal{M}$ and gives the same formula for the variation with respect to the NLSM field.

The intrinsic construction takes as its starting point a geometric interpretation of the WZ term. The NLSM field $\mathbf{n}$ is a map from $\mathcal{M}$ to S^2, and the image of $\mathcal{M}$ under this map is a closed curve ℓ on S^2. The curve ℓ inherits an

orientation from the orientation of $\mathcal{M}$. On S^2, there exist regions $\mathcal{C}$ and $\mathcal{C}' = S^2 \backslash \mathcal{C}$ such that $\partial \mathcal{C} = \ell$ and $\partial \mathcal{C}' = \bar{\ell}$, where $\bar{\ell}$ is the curve ℓ with the opposite orientation. Using this information, we define the WZ term for the field configuration $\mathbf{n}$ using the *signed* area of the regions $\mathcal{C}$ or $\mathcal{C}'$ as

$$S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2} \text{Area}[\mathcal{C}], \qquad (\text{D7})$$

or

$$S'_{\text{WZ}}[\mathbf{n}] = -\frac{2\pi k}{\mathcal{A}_2} \text{Area}[\mathcal{C}']. \qquad (\text{D8})$$

The minus sign in the second equation is there to keep track of the fact that the boundary of $\mathcal{C}'$ is $\bar{\ell}$, which has the opposite orientation of $\ell = \partial \mathcal{C}$. The reason for defining the WZ term in this way is that with this definition we have

$$e^{iS_{\text{WZ}}[\mathbf{n}]} = e^{i(S_{\text{WZ}}[\mathbf{n}] - S'_{\text{WZ}}[\mathbf{n}])} e^{iS'_{\text{WZ}}[\mathbf{n}]} = e^{i2\pi k} e^{iS'_{\text{WZ}}[\mathbf{n}]}. \qquad (\text{D9})$$

Therefore, we see that the exponential of the WZ term will be independent of the choice of $S_{\text{WZ}}[\mathbf{n}]$ or $S'_{\text{WZ}}[\mathbf{n}]$ as long as the level k of the WZ term is quantized, $k \in \mathbb{Z}$, which is the usual result. In what follows, we work with the first formula, Eq. (D7).

The formula (D7) instructs us to integrate the volume form ω_2 over the region $\mathcal{C} \subset S^2$. In the case where the curve ℓ is contained only in the coordinate patch U_S on S^2, we have $\omega_2 = d\vartheta_S$, and so the WZ term takes the simple form

$$S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2} \int_\ell \vartheta_s = \frac{2\pi k}{\mathcal{A}_2} \int_{\mathcal{M}} \mathbf{n}^* \vartheta_S$$
$$= -\frac{2\pi k}{\mathcal{A}_2} \int_a^b dx^1 (1 + \cos(\Theta)) \partial_1 \Phi, \qquad (\text{D10})$$

where we used the expression for ϑ_S from Eqs. (D3). If instead the curve ℓ is contained only in U_N, then we have a similar expression for $S_{\text{WZ}}[\mathbf{n}]$ with ϑ_S replaced with ϑ_N. Finally, there is the possibility that the curve ℓ crosses through both coordinate patches on S^2. In this case, the expression for $S_{\text{WZ}}[\mathbf{n}]$ obtained from the formula Eq. (D7) is more complicated.

To construct the WZ term in the case where the curve ℓ passes through both coordinate patches on S^2, we do the following. First, we pick two points s_1 and s_2 on the interval $[a, b]$ such that $\mathbf{n}_1 \equiv \mathbf{n}(s_1)$ and $\mathbf{n}_2 \equiv \mathbf{n}(s_2)$ lie on opposite sides of the curve ℓ in the region $U_S \cap U_N$. Next, we divide the curve ℓ into two pieces ℓ_1 and ℓ_2 such that $\ell = \ell_1 + \ell_2$, where the sum is the composition of oriented 1-chains. Finally, we add a third curve ℓ_3 which connects the points $\mathbf{n}_1$ and $\mathbf{n}_2$ by cutting through the region $U_S \cap U_N$, and we choose the orientation of this curve such that it is directed toward $\mathbf{n}_2$. The curve ℓ_3 also divides
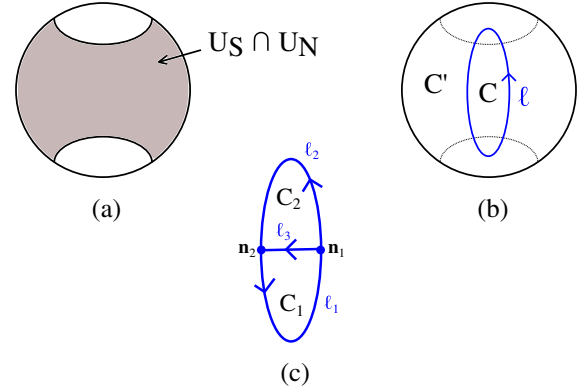

FIG. 1. (a) The shaded region shows the intersection $U_S \cap U_N$ of the two coordinate patches needed to cover the entire sphere S^2. (b) The curve ℓ and the regions $\mathcal{C}$ and $\mathcal{C}'$ of which the union is the entire sphere S^2. The curve ℓ does not lie in a single coordinate patch U_S or U_N, as can be seen from the dotted lines indicating the boundary of the intersection $U_S \cap U_N$. (c) A partition of the region $\mathcal{C}$ into two parts $\mathcal{C}_1$ and $\mathcal{C}_2$ using an additional curve ℓ_3 which starts at the point $\mathbf{n}_1$ and ends at the point $\mathbf{n}_2$. The part $\mathcal{C}_1$ lies entirely in U_S, while the part $\mathcal{C}_2$ lies entirely in U_N.

the region $\mathcal{C}$ into two portions $\mathcal{C}_1$ and $\mathcal{C}_2$. We choose the points s_1 and s_2 and also the curve ℓ_3, so that $\mathcal{C}_1$ lies entirely in U_S and $\mathcal{C}_2$ lies entirely in U_N. This situation is illustrated in Fig. 1. In this case, we can compute Area$[\mathcal{C}]$ as

$$\text{Area}[\mathcal{C}] = \text{Area}[\mathcal{C}_1] + \text{Area}[\mathcal{C}_2]$$
$$= \int_{\ell_1 + \ell_3} \vartheta_S + \int_{\ell_2 - \ell_3} \vartheta_N$$
$$= \int_{\ell_1} \vartheta_S + \int_{\ell_2} \vartheta_N + \int_{\ell_3} (\vartheta_S - \vartheta_N)$$
$$= \int_{\ell_1} \vartheta_S + \int_{\ell_2} \vartheta_N + \psi_{SN}(\mathbf{n}_2) - \psi_{SN}(\mathbf{n}_1), \qquad (\text{D11})$$

where in the last step we used the equation $\vartheta_S - \vartheta_N = d\psi_{SN}$ on $U_S \cap U_N$. The integrals in this expression can be pulled back to $\mathcal{M}$ to give a final expression for the WZ term in the form

$$S_{\text{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2} \left(\int_{\mathbf{n}^{-1}(\ell_1)} \mathbf{n}^* \vartheta_S + \int_{\mathbf{n}^{-1}(\ell_2)} \mathbf{n}^* \vartheta_N \right.$$
$$\left. + \psi_{SN}(\mathbf{n}(s_2)) - \psi_{SN}(\mathbf{n}(s_1)) \right), \qquad (\text{D12})$$

where $\mathbf{n}^{-1}(\ell_1)$ denotes the inverse image of the curve ℓ_1 under the map $\mathbf{n} : \mathcal{M} \to S^2$, and likewise for $\mathbf{n}^{-1}(\ell_2)$.

This form for the WZ term has the advantage that it only involves integrals over the physical space $\mathcal{M}$ and does not require an extended space $\mathcal{B}$ or an extension $\tilde{\mathbf{n}}$ of the field configuration (it also does not require a spin structure
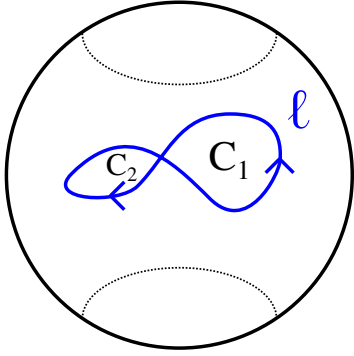
FIG. 2. An example of a curve ℓ with a self-intersection arising from a map $\mathbf{n}: \mathcal{M} \to S^2$ which is not injective. In this case, the WZ term should be computed using the signed area $\mathrm{Area}[\mathcal{C}_1] - \mathrm{Area}[\mathcal{C}_2]$ of the two regions $\mathcal{C}_1$ and $\mathcal{C}_2$ enclosed by the curve ℓ, as shown in Eq. (D13). This is because the curve ℓ goes counterclockwise around the region $\mathcal{C}_1$ but clockwise around the region $\mathcal{C}_2$.

on $\mathcal{M}$). Therefore, we refer to this construction of the WZ term as an *intrinsic* construction. In addition, a short calculation shows that, upon varying this expression with respect to the NLSM field and using the explicit expression for ψ_{SN} from Eq. (D4), the contributions from the points s_1 and s_2 cancel out. The end result for the variation of this form of the WZ term then turns out to be identical to Eq. (D6). Therefore, we have succeeded in our original goal, which was to provide a construction of the WZ term

for the NLSM theory on the manifold $\mathcal{M}$ which does not require a higher-dimensional manifold $\mathcal{B}$ with $\partial \mathcal{B} = \mathcal{M}$. Also, we note here that adding an arbitrary constant to the transition function ψ_{SN}, which still gives a solution to the equation $\vartheta_S - \vartheta_N = d\psi_{SN}$, will not change the final expression (D12) for the WZ term since ψ_{SN} appears in the WZ term in the combination $\psi_{SN}(\mathbf{n}(s_2)) - \psi_{SN}(\mathbf{n}(s_1))$.

Finally, we must discuss what one must do in the case that the curve ℓ on S^2 has self-intersections. The map $\mathbf{n}: \mathcal{M} \to S^2$ is not required to be an *embedding* of $\mathcal{M}$ (i.e., $\mathbf{n}$ is not required to be injective), so in general, the curve ℓ can have self-intersections. In this case, we again define the WZ term using the signed area of the regions enclosed by ℓ. For example, for the situation shown in Fig. 2, we define

$$S_{\mathrm{WZ}}[\mathbf{n}] = \frac{2\pi k}{\mathcal{A}_2}(\mathrm{Area}[\mathcal{C}_1] - \mathrm{Area}[\mathcal{C}_2]). \qquad (\mathrm{D}13)$$

The reason for defining the WZ term in this way is that we ultimately want $S_{\mathrm{WZ}}[\mathbf{n}]$ to reduce to a line integral of ϑ_S or ϑ_N along the curve ℓ, modulo the addition of suitable constant terms in the intersection $U_S \cap U_N$ as discussed above. Since in the example in Fig. 2 the curve ℓ wraps around the regions $\mathcal{C}_1$ and $\mathcal{C}_2$ in opposite directions, we attach opposite signs to the areas of these two regions in the definition of the WZ term so that $S_{\mathrm{WZ}}[\mathbf{n}]$ reduces to a line integral of ϑ_S or ϑ_N along ℓ.

[1] M. Gell-Mann and M. Lévy, Il Nuovo Cimento **16**, 705 (1960).
[2] A. Belavin and A. Polyakov, JETP Lett. **22**, 245 (1975).
[3] A. M. Polyakov, Phys. Lett. **59B**, 79 (1975).
[4] E. Brézin and J. Zinn-Justin, Phys. Rev. B **14**, 3110 (1976).
[5] E. Brezin, J. Zinn-Justin, and J. Le Guillou, Phys. Rev. D **14**, 2615 (1976).
[6] C. J. Hamer, J. B. Kogut, and L. Susskind, Phys. Rev. D **19**, 3091 (1979).
[7] R. P. Feynman, Nucl. Phys. **B188**, 479 (1981).
[8] F. D. M. Haldane, Phys. Lett. **93A**, 464 (1983).
[9] D. H. Friedan, Ann. Phys. (N.Y.) **163**, 318 (1985).
[10] R. Shankar and N. Read, Nucl. Phys. **B336**, 457 (1990).
[11] A. Abanov and P. B. Wiegmann, Nucl. Phys. **B570**, 685 (2000).
[12] T. Senthil and M. P. A. Fisher, Phys. Rev. B **74**, 064405 (2006).
[13] C. Xu and A. W. W. Ludwig, Phys. Rev. Lett. **110**, 200405 (2013).
[14] Z. Bi, A. Rasmussen, K. Slagle, and C. Xu, Phys. Rev. B **91**, 134404 (2015).
[15] Z.-C. Gu and X.-G. Wen, Phys. Rev. B **80**, 155131 (2009).

[16] F. Pollmann, A. M. Turner, E. Berg, and M. Oshikawa, Phys. Rev. B **81**, 064439 (2010).
[17] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Science **338**, 1604 (2012).
[18] M. Levin and Z.-C. Gu, Phys. Rev. B **86**, 115109 (2012).
[19] X. Chen, Z.-C. Gu, Z.-X. Liu, and X.-G. Wen, Phys. Rev. B **87**, 155114 (2013).
[20] T. Senthil, Annu. Rev. Condens. Matter Phys. **6**, 299 (2015).
[21] Y.-Z. You and C. Xu, Phys. Rev. B **90**, 245120 (2014).
[22] Z. Bi, Y.-Z. You, and C. Xu, Phys. Rev. B **90**, 081110 (2014).
[23] Z. Bi, A. Rasmussen, and C. Xu, Phys. Rev. B **89**, 184424 (2014).
[24] C. Xu, Phys. Rev. B **87**, 144421 (2013).
[25] Z. Bi, K. Slagle, and C. Xu, arXiv:1504.04373.
[26] C. Xu and T. Senthil, Phys. Rev. B **87**, 174412 (2013).
[27] D. V. Else and C. Nayak, Phys. Rev. B **90**, 235137 (2014).
[28] Z.-X. Liu and X.-G. Wen, Phys. Rev. Lett. **110**, 067205 (2013).
[29] Y. You and Y.-Z. You, Phys. Rev. B **93**, 195141 (2016).
[30] Y. You, Phys. Rev. B **94**, 195112 (2016).

[31] M. F. Lapa, C.-M. Jian, P. Ye, and T. L. Hughes, Phys. Rev. B **95**, 035149 (2017).

[32] S. Takayoshi, P. Pujol, and A. Tanaka, Phys. Rev. B **94**, 235159 (2016).

[33] E. Fradkin, E. Moreno, and F. A. Schaposnik, Phys. Rev. D **45**, 595 (1992).

[34] E. Corrigan and C. K. Zachos, Phys. Lett. **88A**, 273 (1979).

[35] J. Maharana, Phys. Lett. **128B**, 411 (1983).

[36] M. J. Bowick, D. Karabali, and L. Wijewardhana, Nucl. Phys. **B271**, 417 (1986).

[37] T. Olczyk, P. Panigrahi, and S. Ramaswamy, Z. Phys. C **45**, 653 (1990).

[38] D. Karabali, Int. J. Mod. Phys. A **06**, 1369 (1991).

[39] P. A. Dirac, *Lectures on Quantum Mechanics* (Courier, New York, 2013).

[40] A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976).

[41] H. Kleinert and S. V. Shabanov, Phys. Lett. A **232**, 327 (1997).

[42] B. Podolsky, Phys. Rev. **32**, 812 (1928).

[43] B. S. DeWitt, Rev. Mod. Phys. **29**, 377 (1957).

[44] J. A. Neto and W. Oliveira, Int. J. Mod. Phys. A **14**, 3699 (1999).

[45] C. Neves and C. Wotzasek, J. Phys. A **33**, 6447 (2000).

[46] S.-T. Hong, W. T. Kim, and Y.-J. Park, Mod. Phys. Lett. A **15**, 1915 (2000).

[47] E. Abdalla and R. Banerjee, Braz. J. Phys. **31**, 80 (2001).

[48] A. Scardicchio, Phys. Lett. A **300**, 7 (2002).

[49] S.-T. Hong and K. D. Rothe, Ann. Phys. (N.Y.) **311**, 417 (2004).

[50] E. Witten, Nucl. Phys. **B223**, 422 (1983).

[51] O. Alvarez, Commun. Math. Phys. **100**, 279 (1985).

[52] H. Flanders, *Differential Forms with Applications to the Physical Sciences* (Elsevier, New York, 1963), Vol. 11.

[53] J. Milnor and J. D. Stasheff, *Characteristic Classes (AM-76)* (Princeton University, Princeton, NJ, 2016), Vol. 76.

[54] T. Eguchi, P. B. Gilkey, and A. J. Hanson, Phys. Rep. **66**, 213 (1980).

[55] R. C. Kirby, *The Topology of 4-Manifolds* (Springer, Berlin, 2006), Vol. 1374.

[56] V. V. Prasolov, *Elements of Combinatorial and Differential Topology* (American Mathematical Society, Providence, 2006), Vol. 74.

[57] A. Kapustin, arXiv:1403.1467.

[58] A. Kapustin, R. Thorngren, A. Turzillo, and Z. Wang, J. High Energy Phys. 12 (2015) 01.

[59] C.-T. Hsieh, G. Y. Cho, and S. Ryu, Phys. Rev. B **93**, 075135 (2016).

[60] E. Witten, Rev. Mod. Phys. **88**, 035001 (2016).

[61] A. Vishwanath and T. Senthil, Phys. Rev. X **3**, 011016 (2013).

[62] M. A. Metlitski, C. L. Kane, and M. P. A. Fisher, Phys. Rev. B **88**, 035131 (2013).

[63] I. Chavel, *Eigenvalues in Riemannian Geometry* (Academic, New York, 1984), Vol. 115.

[64] M. Stone and P. Goldbart, *Mathematics for Physics: A Guided Tour for Graduate Students* (Cambridge University Press, Cambridge, England, 2009).

[65] R. J. Szabo, *Equivariant Cohomology and Localization of Path Integrals* (Springer Science & Business Media, Berlin, 2003), Vol. 63.

[66] T. T. Wu and C. N. Yang, Phys. Rev. D **14**, 437 (1976).

[67] A. P. Polychronakos, Nucl. Phys. **B281**, 241 (1987).