

# Decoding perturbation theory using resurgence: Stokes phenomena, new saddle points and Lefschetz thimbles

Aleksey Cherman,<sup>a</sup> Daniele Dorigoni<sup>b</sup> and Mithat Ünsal<sup>c</sup>

<sup>a</sup>*Fine Theoretical Physics Institute, Department of Physics, University of Minnesota,  
Minneapolis, MN, U.S.A.*

<sup>b</sup>*DAMTP, University of Cambridge,  
Wilberforce Road, Cambridge CB3 0WA, U.K.*

<sup>c</sup>*Department of Physics, North Carolina State University,  
Raleigh, NC, 27695 U.S.A.*

*E-mail:* [acherman@umn.edu](mailto:acherman@umn.edu), [d.dorigoni@damtp.cam.ac.uk](mailto:d.dorigoni@damtp.cam.ac.uk),  
[unsal.mithat@gmail.com](mailto:unsal.mithat@gmail.com)

**ABSTRACT:** Resurgence theory implies that the non-perturbative (NP) and perturbative (P) data in a QFT are quantitatively related, and that detailed information about non-perturbative saddle point field configurations of path integrals can be extracted from perturbation theory. Traditionally, only stable NP saddle points are considered in QFT, and homotopy group considerations are used to classify them. However, in many QFTs the relevant homotopy groups are trivial, and even when they are non-trivial they leave many NP saddle points undetected. Resurgence provides a refined classification of NP-saddles, going beyond conventional topological considerations. To demonstrate some of these ideas, we study the  $SU(N)$  principal chiral model (PCM), a two dimensional asymptotically free matrix field theory which has no instantons, because the relevant homotopy group is trivial. Adiabatic continuity is used to reach a weakly coupled regime where NP effects are calculable. We then use resurgence theory to uncover the existence and role of novel ‘fracton’ saddle points, which turn out to be the fractionalized constituents of previously observed unstable ‘uniton’ saddle points. The fractons play a crucial role in the physics of the PCM, and are responsible for the dynamically generated mass gap of the theory. Moreover, we show that the fracton-anti-fracton events are the weak coupling realization of ’t Hooft’s renormalons, and argue that the renormalon ambiguities are systematically cancelled in the semi-classical expansion. Our results motivate the conjecture that the semi-classical expansion of the path integral can be geometrized as a sum over Lefschetz thimbles.

**KEYWORDS:** Solitons Monopoles and Instantons, Nonperturbative Effects, Sigma Models, Renormalization Regularization and Renormalons

ARXIV EPRINT: [1403.1277](https://arxiv.org/abs/1403.1277)

---

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Problems with perturbation theory and the semi-classical expansion	1
1.2	Resurgence theory	2
1.3	Beyond the topological classification of NP saddles	3
1.4	Application: principal chiral model	5
1.5	Outline	6
<b>2</b>	<b>Zero dimensional prototype for resurgence and Lefschetz thimbles</b>	<b>7</b>
2.1	Borel analysis and Stokes phenomena	13
2.2	Stokes automorphism and alien derivative	15
2.3	Reality of resurgent trans-series for real $\lambda$ and BE-summability	16
2.4	Lefschetz thimbles and geometrization of ambiguity cancellation	17
2.5	NP-data in late terms of P-expansion	19
<b>3</b>	<b>Review of the principal chiral model on <math>\mathbb{R}^2</math></b>	<b>19</b>
3.1	Perturbation theory on $\mathbb{R}^2$ and IR-renormalons	22
<b>4</b>	<b>Compactification to <math>\mathbb{R} \times S^1</math> and adiabatic continuity</b>	<b>23</b>
4.1	Large- $L$ expectations	25
4.2	Choosing the right small- $L$ limit	28
4.3	Thermal compactification and non-adiabaticity	30
4.4	Spatial compactification and adiabaticity	31
<b>5</b>	<b>Structure of the perturbative series on small <math>\mathbb{R} \times S_L^1</math></b>	<b>32</b>
5.1	Covering $SU(N)$ with $SU(2)$ s	34
5.2	Large order behavior and Stokes phenomenon	36
5.3	NP-data from the late terms of the P-expansion	40
<b>6</b>	<b>Non-perturbative saddle points</b>	<b>41</b>
6.1	Unitons	42
6.2	$SU(2)$ fractons and KK fractons	44
6.3	$SU(N)$ fractons	47
6.4	Fracton amplitudes	50
<b>7</b>	<b>Resurgence triangle and emergent topological structure</b>	<b>51</b>
7.1	Fracton interactions and $[\mathcal{F}_i \mathcal{F}_i]$ , $[\bar{\mathcal{F}}_i \bar{\mathcal{F}}_i]$ , $[\mathcal{F}_i \bar{\mathcal{F}}_i]$ amplitudes	53
7.2	Reality of resurgent trans-series for real $\lambda$ and BE-summability	60
7.3	Lefschetz thimbles and geometrization of ambiguity cancellation	62
<b>8</b>	<b>Mass gap flow and Borel flow</b>	<b>64</b>

<b>9</b>	<b>Discussion and prospects</b>	<b>68</b>
9.1	Future directions	69
<b>A</b>	<b>Resurgence terminology</b>	<b>70</b>

---

## 1 Introduction

### 1.1 Problems with perturbation theory and the semi-classical expansion

Consider an observable  $\mathcal{O}$  in a quantum theory with a dimensionless coupling  $\lambda$ , defined by an appropriately regularized Euclidean path integral:

$$\langle \mathcal{O}[\lambda] \rangle = Z[\lambda]^{-1} \int dU e^{-S[U;\lambda]} \mathcal{O}, \quad (1.1)$$

where  $Z[\lambda]$  is the partition function. If  $\lambda$  can be kept small,<sup>1</sup> one expects to be able to evaluate the path integral using the saddle-point method, so that (schematically)

$$\langle \mathcal{O}[\lambda] \rangle = \sum_{n=0}^{\infty} p_{0,n} \lambda^n + \sum_c e^{-S_c/\lambda} \sum_{n=0}^{\infty} p_{c,n} \lambda^n. \quad (1.2)$$

Above  $p_{0,n}$  are the perturbative contributions to  $\mathcal{O}$ , and encode an expansion in fluctuations around the trivial “perturbative” saddle-point  $U_0$  of the path integral, which has zero action. There are also contributions from non-perturbative saddle point field configurations  $U_c$ , which have finite actions  $S_c$  measured in units of  $\lambda$ , and contributions from the perturbative fluctuations  $p_{c,n}$  around  $U_c$ .

Eq. (1.2) is traditionally viewed as the semiclassical *approximation* to the original path integral. The reason is that in almost all interesting QFTs, and even in simple quantum mechanics or even simpler ordinary integrals, the perturbative series expansions around both the perturbative saddle  $U_0$  as well as  $U_c$  are actually divergent asymptotic expansions, with  $p_{0,n}, p_{c,n} \sim n!$  [1–3]. The standard way to give a meaning to such perturbative series is via Borel transforms and resummations. After computing the Borel transform of an asymptotic series, and its analytic continuation, one obtains a function with singularities in the ‘Borel plane’. The Borel sum of a perturbative series is defined as the Laplace transform of the analytic continuation of the Borel transform. The issue is that if  $p_{0,n}, p_{c,n} \sim n!$  then there will be singularities on the integration contour in the Borel sum, and the integral — and hence the sum — is not well-defined. Different choices of contour deformations to avoid the singularities give different results for the same physical observable. This is a reflection of the fact that  $\lambda \in \mathbb{R}^+$  is a Stokes line. As a result, the Borel sums of all of the perturbative series appearing in eq. (1.2) are usually not well-defined.

---

<sup>1</sup>In asymptotically free theories,  $\lambda$  grows at low energies. However, its growth can be cut off and  $\lambda$  can be kept naturally small at low energies by e.g. adjoint Higgsing either by an elementary Higgs scalar or Wilson line expectation value in gauge theories, or by turning on appropriate background fields.

Another (much less widely appreciated) fact about the semiclassical expansion is that the amplitudes associated with certain saddle points, for example, correlated instanton-anti-instanton  $[I\bar{I}]$  events, are not well-defined either [4, 5] along the  $\lambda \in \mathbb{R}^+$  Stokes line. That is, in addition to the ambiguities in the sum of the perturbative series, the sum over non-perturbative saddle points also suffers from ambiguities. But if every perturbative series and most of the non-perturbative factors appearing in our expansion are not well-defined, then in what sense, and to what extent, does the semiclassical expansion capture the physics encoded in the original path integral? How do we give a meaning to a saddle point expansion? The standard perspective is that the semiclassical approximation has an inherent ‘fuzziness’ defined by the size of the resummation ambiguities, and eq. (1.2) only approximates the value of the original integral up to semiclassically-incalculable corrections of the order of the ambiguities.

Although the inclusion of the contributions of the NP-saddles seems to make the problems in the semiclassical approximation even worse, we will argue that including the NP saddles is in fact *the solution* to defining our saddle point expansion for  $\langle \mathcal{O}[\lambda] \rangle$  in an ambiguity-free, meaningful way.

## 1.2 Resurgence theory

To see how the program of assigning unambiguous meaning to the semiclassical expansion for  $\langle \mathcal{O}[\lambda] \rangle$  might work, we note that it has been known for some time that there are special cases when an unambiguous meaning can be assigned to eq. (1.2) by carefully including the contributions of the NP saddles. For example, in a double-well or periodic potential problem in quantum mechanics, it is known that the leading ambiguity in perturbation theory is cured by the ambiguity in the  $[I\bar{I}]$  amplitude (and fluctuations around it), and the ambiguity in the perturbation theory around an instanton is cured by the ambiguity in the  $[II\bar{I}]$  amplitude (and fluctuations around it), etc. [4–6], and see also [7, 8].

Such cancellations of ambiguities may seem magical, but in fact underlying the cancellations there is a systematic mathematical framework called resurgence theory, a term coined in a different context by J. Ecalle in early 80s [9]. Applied to QFT, resurgence theory is a generalization of the venerable idea of Borel resummation of the perturbative expansion around the perturbative saddle which systematically incorporates Stokes phenomena [10–14]. As described above, in most interesting quantum mechanical systems and QFTs, Borel resummation does not work (i.e., gives ambiguities) due to singularities in the Borel plane. If the Borel transform of all perturbative series are *endlessly continuable* (i.e. the set of singularities [including poles and branch cut origins] in all Riemann sheets are discrete and there are no natural boundaries), then trans-series of the form of eq. (1.2) can be viewed as expansions of resurgent functions. Ecalle’s work [9] implies that for such trans-series, all would-be ambiguities of the semiclassical representation cancel, see also [15]. The key to these cancellations is that in the trans-series representation eq. (1.2), there are also ambiguities associated with the non-perturbative factors  $e^{-S_c/\lambda}$ , which exactly cancel the leading ambiguities in the perturbation theory, with further (more intricate) relations amongst the various terms in the trans-series leading to the cancellation of ambiguities at higher

orders, in such a way that the trans-series representation is ambiguity-free to all orders.<sup>2</sup> If we conjecture that observables in QFT are resurgent functions, then resurgence theory implies that the expansions around any given saddle-point must contain exact information concerning the expansions around all other saddle-points of the theory. In particular, resurgence implies that the large order terms of *perturbative* series contain encoded precise information about the behavior of the *non-perturbative* saddles and the fluctuations around them. As a suggestive equation, one may call this idea “P-data = NP-data”.

We should emphasize that resurgence suggests a major philosophical shift on the meaning of the semiclassical approximation. If the right hand side of eq. (1.2) can indeed be systematically interpreted in an unambiguous way, then the semiclassical expansion should not be thought of as an approximation. Instead, when viewed as a resurgent trans-series, the saddle point expansion should be viewed as an exact coded *representation* of the observable  $\langle \mathcal{O}(\lambda) \rangle$  in the regime of the QFT which is smoothly connected to the small  $\lambda$  semiclassical limit. In particular, in such a regime, the resurgent transseries gives an arbitrarily accurate representation of observables even for  $\lambda \sim \mathcal{O}(1)$ .

In this work, we take resurgence as our guiding principle, and use it to find new saddles in certain QFTs. We are able to systematically test the predictions of resurgence theory by using the recently developed ideas of adiabatic continuity and weak coupling NP-calculability.

### 1.3 Beyond the topological classification of NP saddles

In the context of QFT, it has recently been proposed to use resurgence theory to provide evidence for a non-perturbative continuum definition in the semi-classical domain [16–19] by invoking the idea of adiabatic continuity [20, 21]. This program provides a new insight into ’t Hooft’s mysterious renormalon problem [3, 22]. In this context, resurgence theory has been applied to e.g. non-Abelian  $SU(N)$  gauge theories on  $\mathbb{R}^3 \times S^1$  and the  $\mathbb{CP}^{N-1}$  non-linear sigma model on  $\mathbb{R} \times S^1$ . In both cases, the theories involved have a non-trivial homotopy group classifying the stable NP saddle points, and consequently, they also have instantons, fractionalized instantons [23–28], and composite configurations made from combinations of correlated instanton and fractionalized instanton events. Using resurgence theory, it has recently been proposed that the ambiguities due to the most severe “infrared renormalon” sources of divergences in these asymptotically-free theories cancel against the contributions of the appropriate neutral bion (fractional instanton-anti-instanton) events with action  $\frac{2}{N}$  in units where the BPST instanton action is normalized to unity, in a semiclassical regime of the theory [16–19]. If it turns out that the cancellations of ambiguities persist to all orders, resurgence theory would yield a systematic non-perturbative semi-classical definition of asymptotically-free theories.

It is important to note that resurgence provides a classification of NP saddles which is more refined than the traditional topological classification of saddle points, based on  $\pi_3[G]$  in 4D gauge theories with gauge group  $G$ , and on  $\pi_2[T]$  in 2D non-linear sigma models with target space  $T$ . If two saddles are in the same conjugacy class in these homotopy groups,

---

<sup>2</sup>See appendix A for a brief overview of Borel summation, trans-series, and resurgence.

then they carry the same topological charge. So topology cannot be used to distinguish them. On the other hand, if these two topologically identical saddle points have different actions, then the non-analyticities in the coupling  $\lambda$  of their contributions to the path integral are different, and hence they are distinguishable using resurgence theory. For example, the perturbative saddle and the instanton saddle by definition constitute two different conjugacy classes according to homotopy, which we will refer to as  $\mathcal{C}_0$  and  $\mathcal{C}_{+1}$  respectively. In presence of a topological  $\theta$ -angle all the elements in  $\mathcal{C}_0$  have actions independent of  $\theta$  while all elements in  $\mathcal{C}_{+1}$  have an imaginary part in their actions proportional to  $(+1) \cdot \theta$ . The elements of these conjugacy classes are

$$\begin{aligned} \mathcal{C}_0 : & \quad \left\{ [0], [\mathcal{I}\bar{\mathcal{I}}], [\mathcal{I}^2\bar{\mathcal{I}}^2], [\mathcal{I}^3\bar{\mathcal{I}}^3], \dots, [\mathcal{I}^n\bar{\mathcal{I}}^n], \dots \right\} \\ \mathcal{C}_{+1} : & \quad \left\{ [\mathcal{I}], [\mathcal{I}^2\bar{\mathcal{I}}], [\mathcal{I}^3\bar{\mathcal{I}}^2], [\mathcal{I}^4\bar{\mathcal{I}}^3], \dots, [\mathcal{I}^{n+1}\bar{\mathcal{I}}^n], \dots \right\}, \end{aligned} \quad (1.3)$$

The elements within each class are not distinguished by homotopy considerations. However, the elements of these conjugacy classes are distinguished according to resurgence theory. This is the motivation of the “resurgence triangle” classification of saddle points discussed in [18, 19], which we explain in detail in section 7.

We should emphasize that all of the NP saddle points appearing in the homotopy conjugacy classes above, except for the perturbative saddle  $[0]$  and the instanton saddle  $[\mathcal{I}]$ , are actually *quasi*-saddle-points. To see what is meant by this, recall that in the semiclassical limit path integrals become dominated by field configurations which come as close as possible to satisfying the equations of motion of the classical action and have finite action. So field configurations that are exact solutions of the equations of motion are of course important in the semiclassical limit, and often they are the only field configurations considered. However, while it is much less widely appreciated, in the semiclassical  $\lambda \ll 1$  limit there are generally also quasi-solutions of the equations of motion, which come parametrically close to satisfying the equations of motion and have finite action. Some prominent examples of such configurations are e.g. *correlated* multi-instanton events in quantum mechanics and QFT, and magnetic bions [29] in gauge theory. We refer to such finite-action quasi-solution field configurations as quasi-saddle-points. While quasi-saddle points are typically not distinguished from exact saddle points by homotopy theory considerations, they are distinguishable using resurgence theory, and are categorized in the resurgence triangle classification of NP saddles. We find that quasi-saddle-points make critical contributions to QFTs in the semiclassical limit.

In this work, we give a more dramatic realization of the idea of the resurgence triangle classification of saddle points. Some interesting QFTs have a trivial homotopy group. Relatedly, they do not possess any known topologically-stable finite-action field configurations like instantons, and hence cannot have fractionalized instantons either! So one might naively think that a semi-classical calculation of observables in such theories would include contributions *only* from the trivial perturbative saddle point.

However, high-order factorial divergences of perturbation theory are ubiquitous and are known to occur even in theories without instantons. If the resurgence formalism is the right way to think about the semiclassical representation of path integrals, it implies that there

must *always* be finite-action configurations that contribute to the path integral whenever the sum of the perturbative series is ambiguous. This must be the case even when homotopy considerations leave no room for contributions from stable instantons or their constituents. This is an even sharper illustration of the point that resurgence theory provides a much more refined classification of the finite-action field configurations that can contribute to path integrals than conventional homotopy-theoretic methods. Understanding how this works in detail is a major focus of this paper. Previous works related to this question include [18, 19, 30] in the context of the  $\mathbb{CP}^{N-1}$  model and [16, 17] in QCD(adj) and deformed YM, which are theories with a non-trivial homotopy group, and [31] in a theory with trivial homotopy group. Indeed, the present paper is a detailed exposition of the results briefly announced in our joint work with G. V. Dunne [31].

We also note a related work [32] in the context of matrix models and topological string theory. Indeed, there is an important body of work applying resurgence theory to matrix models and string theory [33–38], where resurgent analysis is used to find new non-perturbative sectors which must also be taken into account in order to construct full non-perturbative solutions. For a recent review emphasizing resurgence in quantum mechanics and matrix models see [39]. Ideas from resurgence theory have also recently appeared in the context of 4D  $\mathcal{N} = 2$  supersymmetric gauge theories in the body of work triggered by [40, 41], which gave a physical derivation of the Kontsevich-Soibelman (KS) wall-crossing formula [42].

#### 1.4 Application: principal chiral model

A convenient toy model which we will use to examine the above issues is the two dimensional  $SU(N)$  principal chiral model (PCM). Some basic facts and expectations about the dynamics of PCM are summarized in section 3, and section 2.1 of Polyakov’s book [43]. This theory shares many of the most important features of Yang-Mills theory, with the following “similarity list”:

- S1)** Asymptotic freedom;
- S2)** Matrix-like large- $N$  limit (as opposed to a vector-like one), planar dominance;
- S3)** Dynamically generated mass gap;
- S4)** Confinement: free energy  $\mathcal{F}/N^2$  goes to zero in the *low* temperature regime;
- S5)** Deconfinement:  $O(N^2)$  free energy in the *high* temperature regime;
- S6)** Large-order structure of perturbation theory, presence of renormalon singularities.

These features *might* have made the PCM an especially useful toy model for Yang-Mills. In fact, the above similarities are much more pronounced between Yang-Mills theory and the PCM compared to the similarities between YM and the  $\mathbb{CP}^{N-1}$  model. However, historically, the  $\mathbb{CP}^{N-1}$  model has been studied much more intensively. A major reason behind this is the fact that the PCM has some properties which are in stark contrast with Yang-Mills theory. So, we also make a “contrast list”:



- C1)**  $\pi_2[\text{SU}(N)] = 0$ , so the homotopy group which would classify instantons is trivial;
- C2)** Absence of topological charge and topologically stable instantons;
- C3)** Presence of unton saddle (with no proposed quantum interpretation until [31])

Furthermore, to actually be useful, a toy model has to be more friendly than the original theory, QCD, at least in some limit. But unlike other (vector-like) sigma models, such as the  $\mathbb{CP}^{N-1}$  model, for which explicit first-principles analytic solutions are known at large- $N$ , there is no known first-principles large- $N$  analytic solution based on expanding around a large- $N$  saddle point despite heroic efforts by Polyakov [43], section 8.2.<sup>3</sup>

The consequences of the application of resurgence theory to the PCM are rather striking. Resurgence tells us that if the PCM model exist as a quantum theory, item **S6)** in the similarity list implies that all the elements in the contrast list must be consequence of *superficial* reasoning. At a deeper level, the similarity of the large-order growth of perturbation theory in PCM and YM theory makes it *impossible* that the principal chiral model only has a trivial perturbative saddle point. In fact, resurgence theory implies that it must possess a plethora of NP-saddles which is just as rich as in Yang-Mills theory. In this work, we confirm this resurgence theory expectation by explicit calculations.

## 1.5 Outline

The organization of this somewhat lengthy paper is as follows. In section 2, we provide a zero dimensional toy example, related to the 2d theory by dimensional reduction, which exhibits Borel non-summability, Stokes phenomena and the cancellation of ambiguities upon Borel-Ecalle (BE)resummation. In combination with appendix A, we hope that this provides a gentle introduction to some of the methods of resurgence theory. In section 2, we also point out the relation between semiclassical expansions and Lefschetz thimbles, giving a geometric perspective on resurgence. Section 3 summarizes some basic facts and expectations about the dynamics of the PCM on  $\mathbb{R}^2$ . In section 4 we explain the construction of the unique weakly-coupled small- $L$  limit of the PCM on  $\mathbb{R} \times S_L^1$  which is continuously connected to the theory on  $\mathbb{R}^2$ . Our analysis is inspired by the one in [18, 19] for the  $\mathbb{CP}^{N-1}$  model. The weak-coupling parameter turns out to be  $\frac{NLA}{2\pi} \ll 1$ , similarly to deformed YM and QCD(adj) [21] and 3d YM with adjoint matter or twisted boundary conditions [49, 50]. At leading order in  $NLA$ , low-energy observables in the 2D PCM can be described by a

---

<sup>3</sup>Despite not being a first-principles solvable model at large- $N$ , in the sense that one can go directly from the Lagrangian to the physical spectrum and correlation functions without making assumptions about the dynamics, the PCM turns out to be quantum mechanically integrable for any  $N$ , in contrast to the  $\mathbb{CP}^{N-1}$  model, which is not integrable at the quantum level for  $N > 2$ . The integrability of the model, together with a few plausible assumptions, most crucially the existence of a mass gap, leads to a solution for the minimal factorized  $S$ -matrix for the PCM [44, 45] which was discovered a long time ago. For some recent work see e. g. [46–48]. While integrability techniques lead to a solution for the PCM  $S$ -matrix, the relation of the resulting solution for the  $S$ -matrix to the microscopic physics is quite opaque. Moreover, while the assumptions necessary for the derivation of the  $S$ -matrix using integrability — such as the presence of the mass gap — are certainly reasonable, they have yet to be demonstrated from first principles. Finally, the integrability-based approaches do not yield any information about the interpretation of the divergences of renormalized perturbation theory, which are a major focus of our work.



simple one-dimensional effective field theory, which is just quantum mechanics. In section 5 we study large-order perturbative behavior of the weakly-coupled QM limit of the  $SU(N)$  PCM. By using resurgence theory techniques, we identify the non-perturbative ambiguity in the Borel resummation and interpret it as pointing to the presence of new NP-saddles in the problem. In section 6, we show that the model indeed has the predicted non-perturbative saddle points, the unitons and fractons, and describe their properties. In section 7 we show that the amplitudes of correlated fracton-anti-fracton events (which we often refer to as neutral bions) have ambiguous parts on the Stokes line, and these ambiguous parts cancel the renormalon ambiguities of perturbation theory. Thus, we interpret the neutral bion as the semi-classical realization of the infrared renormalon. In section 8 we show that the fractons are responsible for the generation of the mass gap in the bosonic PCM at small- $L$ . We describe a plausible flow of the mass gap as the radius is dialed from small to large- $L$ . Relatedly, we point out that the Borel plane singularities on  $\mathbb{R}^2$  are *twice* as dense as compared to the location of singularities on  $\mathbb{R} \times S^1$ , and argue that there should exist a smooth flow in the location of singularities as the radius is dialed from small to large- $L$ . We refer to this phenomenon as Borel flow. Understanding the exact nature of the Borel flow would amount to solution of the mass gap problem on  $\mathbb{R}^2$ , which is an open problem.

## 2 Zero dimensional prototype for resurgence and Lefschetz thimbles

In this section, we consider a zero dimensional integral using steepest descent methods. as a prototype of the semi-classical approach in path integrals. In fact, the zero dimensional model is related to the 2d QFT by dimensional reduction. Compactifying the 2d QFT on small  $\mathbb{R} \times S^1$  with twisted boundary condition on  $S^1$ , we land on a quantum mechanical problem with periodic potential. Further compactifying the QM problem and going to the small  $S^1 \times S^1$  regime, the integral over the zeroth Kaluza-Klein mode reduces to our 0d prototype, as will become clear in section 4. The perturbative expansion of the resulting finite-dimensional integral already exhibits non-Borel summability, Stokes phenomena and cancellation of ambiguities upon Borel-Ecalle (BE)resummation that also take place in path-integral of PCM, and hence provides a useful playground in which we can show many properties very explicitly.<sup>4</sup>

Consider the zero dimensional partition function  $Z(\lambda)$

$$\begin{aligned} Z(\lambda) &= \int_{-\frac{\pi}{2\sqrt{\lambda}}}^{\frac{\pi}{2\sqrt{\lambda}}} dy e^{-\frac{1}{2\lambda} \sin^2(\sqrt{\lambda}y)} = \frac{1}{\sqrt{\lambda}} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} dx e^{-\frac{1}{2\lambda} \sin^2(x)} \\ &= \frac{\pi e^{-\frac{1}{4\lambda}}}{\sqrt{\lambda}} I_0\left(\frac{1}{4\lambda}\right), \end{aligned} \quad (2.1)$$

where  $I_0$  is the modified Bessel function of the first kind.  $Z(\lambda)$  is an integral of a real function over a real domain on a finite interval  $I = [-\frac{\pi}{2}, \frac{\pi}{2}]$ , hence the result is manifestly

---

<sup>4</sup>The discussion of this section streamlines the analytic continuation ideas of [51, 52] and aims to make the relation to Ecalle's theory of resurgence [9] as simple as possible. The material in this section is already known, however we find it useful to detail it since we use a parallel approach for path integrals later in the paper.

real for real  $\lambda$ . In order to demonstrate the use of some of the resurgence technology that we will use in QFT, we would like to study this integral by using the steepest descent expansion, which is the counter-part of the semi-classical expansion in our QFT example. The fundamental idea of the analysis is that to understand the behavior of the  $Z(\lambda)$  for  $\lambda \in \mathbb{R}^+$  one should understand the behavior of the analytic continuation of  $Z(\lambda)$  when  $\lambda \in \mathbb{C}$ .

Our analysis will proceed as follows:

- 1) Identify all critical points.
- 2) Allow  $\lambda$  to move off  $\mathbb{R}^+$  into  $\mathbb{C}$ , and analytically continue  $Z(\lambda)$  by rewriting the original integration cycle as a sum over steepest descent paths, which are called Lefschetz thimbles in general.
- 3) Develop perturbation theory around the P and NP saddles, and derive the respective asymptotic expansions. This is the counterpart of the semi-classical approximation in QFT.
- 4) Show that the action of the NP saddle governs the growth of late terms in the perturbative series around P-saddle, and that sub-leading corrections to the late terms in the perturbative series around the P-saddle are governed by early terms of the perturbative expansion around the NP-saddle and vice versa.
- 5) Show the cancellation of ambiguities and the reality of the trans-series representation of  $Z(\lambda)$  on the  $\lambda \in \mathbb{R}^+$  Stokes line.

We first view the action as a meromorphic function  $S(z)$ . This leads to a more natural description of the steepest descent method and the semiclassical expansion both in the present zero-dimensional example and in QFT. It is also the natural way to study the properties of partition functions under analytic continuation. In fact, a judicious analysis of the semi-classical expansion urges us to view *all* actions as meromorphic functions of the fields as we will see very explicitly. So we now change perspective on the integration cycle  $I$  as

$$I \subset \mathbb{R} \longrightarrow \Sigma \subset \mathbb{C}, \quad \Sigma = I \quad \text{for } \theta \equiv \arg(\lambda) = 0 \quad (2.2)$$

where  $\Sigma$  has real dimension one for general  $\arg(\lambda)$ .<sup>5</sup> We must now address the question of how the integration cycle in  $Z(\lambda)$  changes once  $\theta \neq 0$ .

There are two non-degenerate critical points, call them  $z_0$  and  $z_1$ , obtained by extremizing the action

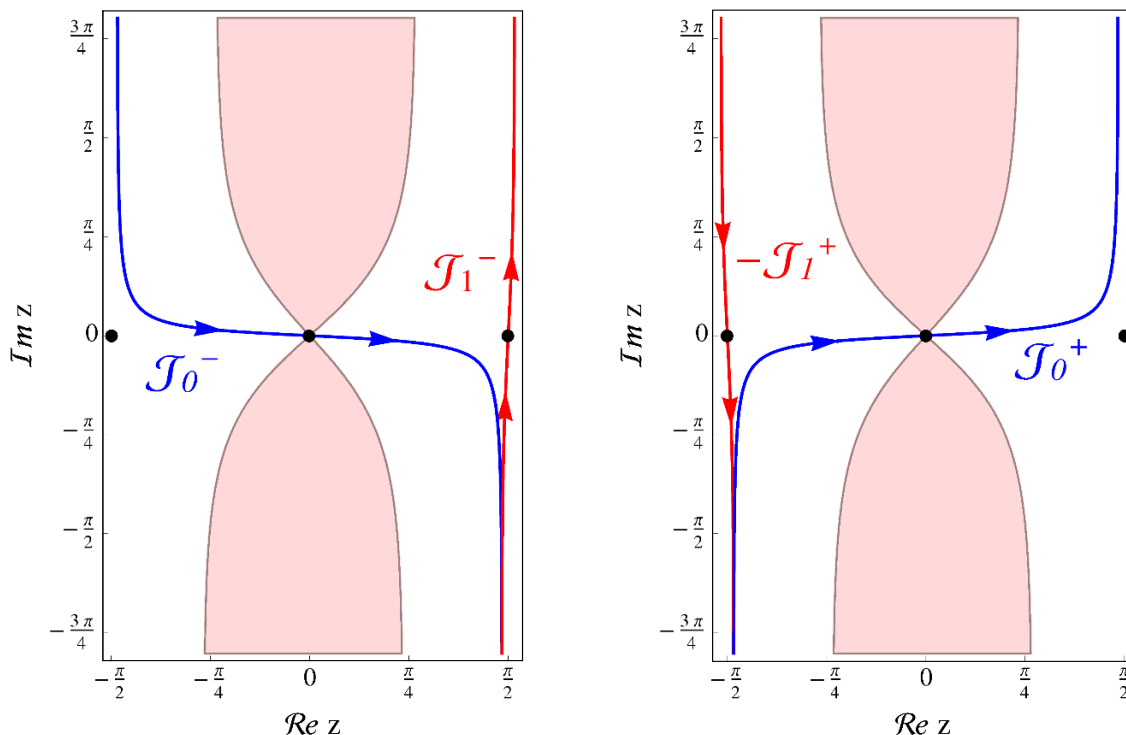
$$\frac{dS}{dz} = 0 \implies \text{critical points: } \{z_0, z_1\} = \left\{0, \frac{\pi}{2}\right\}. \quad (2.3)$$

We call the first one the P-saddle (perturbative vacuum) since it has zero action, and call the latter the NP-saddle since it has a positive action:

$$S(z_0) = 0, \quad S(z_1) = \frac{1}{2\lambda} \quad S_{10} = S_1 - S_0 = \frac{1}{2\lambda}. \quad (2.4)$$

---

<sup>5</sup>More generally, we generalize  $I \subset \mathbb{R}^N \longrightarrow \Sigma \subset \mathbb{C}^N$ , where  $\Sigma$  has real dimension  $N$ .



**Figure 1.** *Left:* Lefschetz thimbles at  $\lambda = e^{i\theta}$  with  $\theta = 0^-$ :  $\mathcal{J}_0 + \mathcal{J}_1$ . *Right:* at  $\theta = 0^+$ .  $\mathcal{J}_0 - \mathcal{J}_1$ . We take  $\theta = \mp 0.1$  to ease visualization.

We have also defined the “relative action”  $S_{10}$  (called the “singulant” by Dingle [10]), which plays an important role in asymptotic analysis.<sup>6</sup>

Associated with each critical point  $z_i$ , there is a unique integration cycle  $\mathcal{J}_i$ , called a Lefschetz thimble or a steepest descent path, along which  $\text{Re } S$  has a downward gradient flow and the phase  $\text{Im } S$  remains stationary. (See [52] for a detailed discussion of the construction using Morse theory.) Indeed, a thimble  $\mathcal{J}_i$  is defined to obey

$$\frac{dz}{dt} = -\overline{\partial_z S(z)} \quad (2.5)$$

where  $t$  is a coordinate along the thimble, with the initial condition  $z(t \rightarrow -\infty) = z_i$ . This definition implies that  $\text{Im } S$  is stationary, meaning that  $\frac{d \text{Im } S}{dt} = 0$ , and

$$\text{Im } S(z)|_{\mathcal{J}_i} = \text{Im } S(z_i), \quad i = 0, 1. \quad (2.6)$$

<sup>6</sup>One might naively think that a singulant is the equivalent of an instanton (which is a non-trivial saddle in the path integral formulation) in quantum mechanics or QFT, since both are nontrivial saddle points. However, in QM, or QFT, there is in general a charge (topological or perhaps emergent, as we will see here) associated with instantons, while the perturbative vacuum is neutral under this charge. Thus, the role that a singulant plays in the large-order behavior of perturbative series in ordinary integrals is actually played by instanton-anti-instanton  $[\mathcal{I}\bar{\mathcal{I}}]$  saddles in QM and/or the weak coupling realization of IR-renormalons in QFT. So, in passing from  $d = 0$  examples to  $d \geq 1$ , the mapping is roughly singulant  $\longleftrightarrow [\mathcal{I}\bar{\mathcal{I}}]$  saddle.

For each critical point  $z_i$  one can also define paths with upward gradient flow  $\mathcal{K}_i$ , and it can be shown that there is a one-to-one correspondence

$$z_i \leftrightarrow \mathcal{J}_i, \mathcal{K}_i \quad (2.7)$$

between the critical points and Lefschetz thimbles. The set of the Lefschetz thimbles may be seen as forming a linearly independent and complete basis of integration cycles for integrals of  $e^{-S(z)}$ . In general, the contours of integration deform smoothly as  $\arg(\lambda)$  is varied, and pass through only the associated saddle. Exactly at the Stokes lines, these contours also pass through a subset of other saddles. Lefschetz thimbles are the natural geometric surfaces (lines in our example) which can be used to describe the analytic continuation of  $Z(\lambda)$  to complex  $\lambda$ .

The Lefschetz thimbles  $\mathcal{J}_i$  are generally unbounded, even when the original integration cycle is bounded, as illustrated in figure 1. Therefore, we must address the issue of the convergence of the integration over a thimble. In doing so, we divide the complex  $z$ -plane into “good” and “bad” regions [52]. A good region corresponds to  $\text{Re}(S(z)) > 0$  such that  $e^{-S(z)} \rightarrow 0$  as  $|z| \rightarrow \infty$ . A bad region is the complement, one in which  $\text{Re}(S(z)) < 0$ . An admissible contour, for which the integral is finite by construction, is the one which connects two good regions. In figure 1, the white regions are good and the red regions are bad. The  $\mathcal{J}_i$  cycles start and end in the good regions, while the  $\mathcal{K}_i$  cycles, which are not shown in the figure, start and end in the bad regions.

A general integration cycle  $\Sigma(\theta)$  on which the integral converges can be written as a sum over the critical point cycles:

$$\Sigma(\theta) = \sum_{i=0,1} n_i \mathcal{J}_i, \quad n_i \in \mathbb{Z}, \quad (2.8)$$

The coefficients  $n_i$  are piece-wise constant, but have jumps when  $\theta$  crosses Stokes lines. To see an illustration of this, note that our original integration cycle  $I = [-\frac{\pi}{2}, \frac{\pi}{2}]$  can actually be written in *two* different ways, depending on how one approaches to  $\theta = \arg(\lambda) = 0$  Stokes line:

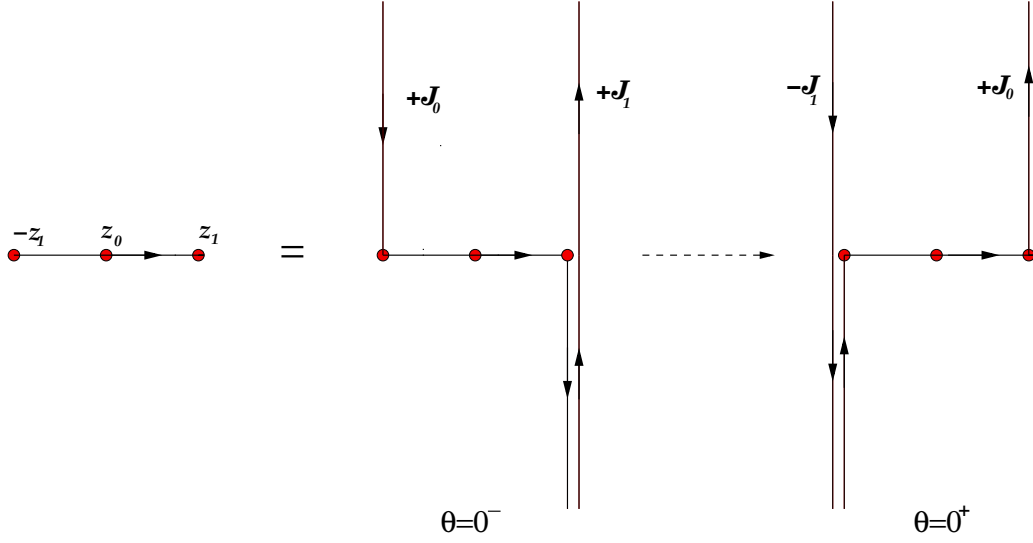
$$I = \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \longrightarrow \Sigma = \begin{cases} \mathcal{J}_0(0^-) + \mathcal{J}_1(0^-) \\ \mathcal{J}_0(0^+) - \mathcal{J}_1(0^+) \end{cases} \quad (2.9)$$

This is illustrated in figure 3.

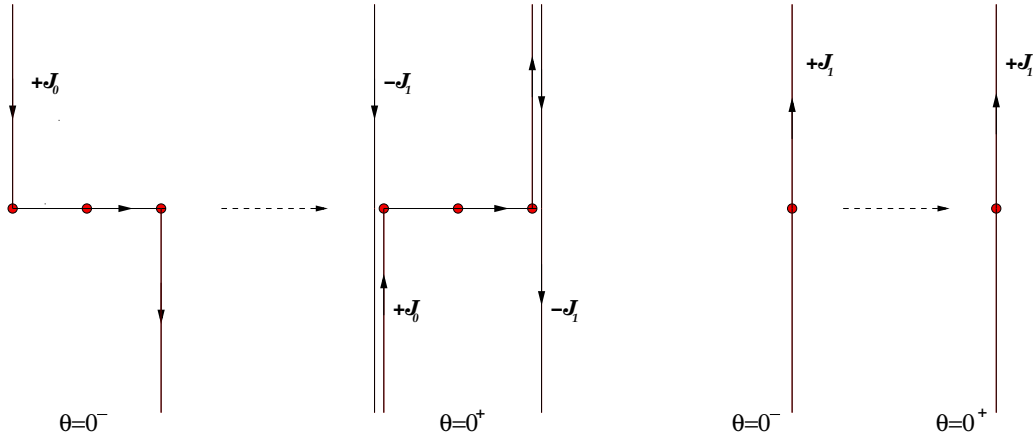
The relation between the two choices of the integration cycles, the notion of cancellation of ambiguities and BE-summability is discussed in section 2.4. Note that despite the fact that  $I$  is a finite interval in  $\mathbb{R}$ , the critical point cycles  $\mathcal{J}_0$  and  $\mathcal{J}_1$  are infinite in  $\mathbb{C}$ . Moreover, while  $\mathcal{J}_0(0^-)$  and  $\mathcal{J}_0(0^+)$  coincide on the real axis, their “tails” in the imaginary direction have a relative sign flip. There is also a flip of the sign of the coefficient of  $\mathcal{J}_1$  at  $\theta = 0$  (and  $\mathcal{J}_0$  at  $\theta = \pi$ ). These sign flips are realizations of the Stokes phenomenon [10]. They are responsible for the cancellation of the imaginary “tail” contributions to integrals running over either of the cycles in eq. (2.9),<sup>7</sup> so that the value of the integral for  $\theta = 0^\pm$

---

<sup>7</sup>Recall that  $z_1$  coincides with  $-z_1$  due to the periodicity of the action.



**Figure 2.** The original integration cycle as a linear combination of Lefschetz thimbles at  $\theta = 0^-$  and  $\theta = 0^+$ .  $\theta = 0$  is a Stokes line.



**Figure 3.** Stokes phenomenon (wall-crossing) at  $\theta = 0$ :  $\mathcal{J}_0 \rightarrow \mathcal{J}_0 - 2\mathcal{J}_1$ , while  $\mathcal{J}_1 \rightarrow \mathcal{J}_1$ . There is also a Stokes line at  $\theta = \pi$  where  $\mathcal{J}_1$  jumps and  $\mathcal{J}_0$  does not.

coincides with its value on the real integration cycle  $\Sigma(\theta = 0) = I$ . This is an elementary but useful perspective on ambiguity cancellation, which is realized in figure 2 geometrically.

More generally, note that as  $\arg(\lambda)$  changes, the cycles  $\mathcal{J}_i$  deform in a smooth manner, except for Stokes lines, where they undergo jumps. The two Stokes lines are at  $\theta = 0$  and  $\theta = \pi$ , and the respective jumps are:

$$\begin{aligned} \theta = 0 : \quad & \mathcal{J}_1 \longrightarrow \mathcal{J}_1, & \mathcal{J}_0 \longrightarrow \mathcal{J}_0 - 2\mathcal{J}_1, \\ \theta = \pi : \quad & \mathcal{J}_1 \longrightarrow \mathcal{J}_1 + 2\mathcal{J}_0, & \mathcal{J}_0 \longrightarrow \mathcal{J}_0. \end{aligned} \quad (2.10)$$

To see where these relations come from, note for example that when  $\theta = 0^-$ ,  $\mathcal{J}_0$  goes from one good domain to another good domain by passing through saddle  $z_0$  and being arbitrarily close to saddle  $-z_1$ . However, at  $\theta = 0^+$ , there is no single path which can achieve

this. To start and end in the same places when  $\theta = 0^+$  as when  $\theta = 0^-$ , one first needs to take the  $-\mathcal{J}_1$  thimble, then the  $\mathcal{J}_0$  thimble, then again go along  $-\mathcal{J}_1$ . So  $\mathcal{J}_0 \rightarrow \mathcal{J}_0 - 2\mathcal{J}_1$ , as is illustrated in figure 3. Similar jumps in contours can be found in monodromy problems associated with certain Picard-Fuchs equations [53].

In summary, the expressions for the analytic continuation of the integration cycle from  $\arg(\lambda) = 0$  to arbitrary  $\arg(\lambda)$  must take into account the Stokes phenomena eq. (2.10) and it depends on the Stokes chamber:

$$\Sigma(\theta) = \begin{cases} +\mathcal{J}_0(\theta) + \mathcal{J}_1(\theta) & -\pi < \theta < 0, \\ +\mathcal{J}_0(\theta) - \mathcal{J}_1(\theta) & 0 < \theta < \pi, \\ -\mathcal{J}_0(\theta) - \mathcal{J}_1(\theta) & \pi < \theta < 2\pi. \end{cases} \quad (2.11)$$

The full thimbles structure continues periodically as we increase (decrease)  $\theta$ , with period  $4\pi$ . This illustrates the reason that Lefschetz thimbles provide the natural basis for semiclassical expansions and the analytic continuation of Feynman path integral.

**Perturbative expansions around P and NP saddles.** So far we have made no approximations in our analysis. If one could evaluate the integrals along the cycles eq. (2.11) exactly, one would obtain an exact result for  $Z(\lambda)$ . Usually, however, this is not possible, and the best one can do is evaluate the integrals using perturbative series. Hence we now find the perturbative expansion around each of the two saddles. The formal asymptotic expansion around the  $z_0$  P-saddle is given by

$$\mathfrak{Z}_0(\lambda) \equiv e^{-\frac{S_0}{\lambda}} \Phi_0(\lambda) = e^{-\frac{S_0}{\lambda}} \sum_{n=0}^{\infty} a_n \lambda^n \equiv \sqrt{2\pi} \sum_{n=0}^{\infty} \frac{\Gamma\left(n + \frac{1}{2}\right)^2}{n! \Gamma\left(\frac{1}{2}\right)^2} (2\lambda)^n. \quad (2.12)$$

This is a non-alternating asymptotic series for  $\theta = 0$ . The late terms grow as  $a_n \sim \frac{n!}{(S_{10})^n}$ . The series is non-Borel-summable in the  $\theta = 0$  direction, but it is Borel-summable in the  $\theta = \pi$  direction. This formal series is a perturbative representation of the contribution of the integral along the  $\mathcal{J}_0$  thimble.

The perturbative expansion around the NP-saddle (including the NP-factor) is given by

$$\mathfrak{Z}_1(\lambda) = e^{-\frac{S_1}{\lambda}} \Phi_1(\lambda) = e^{-\frac{S_1}{\lambda}} \sum_{n=0}^{\infty} (-1)^n a_n \lambda^n \equiv e^{-\frac{1}{2\lambda}} \Phi_1(\lambda). \quad (2.13)$$

This series is an alternating asymptotic series for  $\theta = 0$ . The late terms grow as  $(-1)^n \frac{n!}{(S_{10})^n}$ , and the series is Borel-summable in the  $\theta = 0$  direction. On the other hand, it is not Borel-summable in the  $\theta = \pi$  direction. This formal series is a perturbative representation of the contribution of the integral along the  $\mathcal{J}_1$  thimble.

The semiclassical expansion for  $Z(\lambda)$  can be written as a two term trans-series

$$\mathfrak{Z}(\lambda, \sigma_0, \sigma_1) = \sigma_0 \mathfrak{Z}_0(\lambda) + \sigma_1 \mathfrak{Z}_1(\lambda) = \sigma_0 \Phi_0(\lambda) + \sigma_1 e^{-\frac{1}{2\lambda}} \Phi_1(\lambda), \quad (2.14)$$

where  $\sigma_i$  are trans-series parameters. Here the  $\sigma_i$  parameters have exactly the same role as the  $n_i$  coefficients of the Lefschetz thimbles in eq. (2.8). The trans-series is an algebraic

representation of the geometric information in eq. (2.8). The analytic continuation of our original integral to complex  $\lambda$  involves different linear combinations of  $\mathfrak{Z}_0(\lambda)$  and  $\mathfrak{Z}_1(\lambda)$  in different Stokes wedges. The value of trans-series representation  $\mathfrak{Z}(\lambda, \sigma_0, \sigma_1)$  of  $Z(\lambda)$  is that because it is an algebraic representation, it is well-suited for direct calculations using standard perturbative methods, while the value of the geometric Lefschetz thimble representation of  $Z(\lambda)$  is its ease of visualization.

## 2.1 Borel analysis and Stokes phenomena

The Borel transforms of the formal series in eq. (2.12) and eq. (2.13) are given by

$$\hat{\Phi}_0(t) \equiv B[\Phi_0](t) = \sum_{n=0}^{\infty} a_n \frac{t^n}{n!} = \sqrt{2\pi} \sum_{n=0}^{\infty} \frac{\Gamma(n + \frac{1}{2})^2}{\Gamma(n+1)\Gamma(\frac{1}{2})^2} \frac{(2t)^n}{n!}, \quad (2.15)$$

$$\hat{\Phi}_1(t) \equiv B[\Phi_1](t) = \sum_{n=0}^{\infty} (-1)^n a_n \frac{t^n}{n!}, \quad (2.16)$$

and since the coefficients  $a_n$  grow like  $n!$  both  $\hat{\Phi}_0(t)$  and  $\hat{\Phi}_1(t)$  define two germs of analytic functions at  $t = 0$ . In this simple example we have the luxury of having closed-form expressions for all of the terms of the Borel transforms of the perturbative series, and indeed there are also closed-form expressions for the analytic continuations of these germs into  $\mathbb{C}_s$  in terms of hypergeometric functions:

$$B[\Phi_0](t) = \sqrt{2\pi} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1; 2t\right), \quad B[\Phi_1](t) = \sqrt{2\pi} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1; -2t\right) \quad (2.17)$$

with singularity (branch point) at  $t = 1/2$  for  $B[\Phi_0](t)$  and at  $t = -1/2$  for  $B[\Phi_1](t)$ . In more complicated examples, one might only have closed-form expressions for the low-order and high-order terms in perturbative series. However, this is still enough to determine the positions of the singularities of the Borel transform, since they are determined by the asymptotic behavior of the high-order terms of the perturbative series.

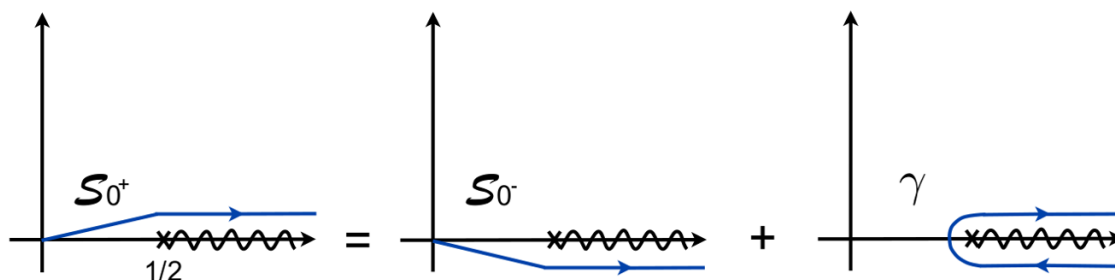
The sectorial (directional) Borel resummations for  $\Phi_i$ ,  $i = 0, 1$  are given by

$$\mathcal{S}_\theta \Phi_i(\lambda) = \frac{1}{\lambda} \int_0^{e^{i\theta}\infty} dt e^{-t/\lambda} B[\Phi_i](t). \quad (2.18)$$

These are well-defined holomorphic functions of  $\lambda$  in  $\text{Re}(e^{i\theta}/\lambda) > 0$ , where  $\theta$  is now parametrizing a generic direction in the complex  $t$ -plane. For real coupling,  $\arg \lambda = 0$ , we cannot directly work with  $\mathcal{S}_0$  due to the presence along the line of integration of a singularity of the Borel transform  $B[\Phi_0](t)$ . This is associated with the Stokes phenomenon in the complex  $\lambda$ -plane: the series eq. (2.12) becomes non-alternating when  $\theta = 0$  and hence  $\Phi_0(\lambda)$  is not Borel summable when  $\theta = 0$ .

$\Phi_0(\lambda)$  is, however, *right and left Borel-summable*. For the right summation one integrates along a contour which avoids the singularity in such a way that the singularity remains to the right of the contour ( $\theta = 0^+$  in eq. (2.18)), and analogously, for left summation, the singularity stays on the left of the contour ( $\theta = 0^-$  in eq. (2.18)).





**Figure 4.** The right Borel resummation can be rewritten as the sum of the left Borel resummation plus the contribution coming from the Hankel contour  $\gamma$ , coming from  $t \rightarrow +\infty$ , circling around the branch cut starting at  $t = 1/2$  and going back to  $+\infty$ .

As illustrated in figure 4, the difference of the right and left Borel resummation can be written as an integral over the Hankel contour  $\gamma$  which starts at  $\infty$  below the imaginary axis, then circles the singular point at  $t = 1/2$ , and then goes back to  $\infty$  above the imaginary axis:

$$\begin{aligned}
 (\mathcal{S}_{0+} - \mathcal{S}_{0-})\Phi_0(\lambda) &= \frac{\sqrt{2\pi}}{\lambda} \int_{\gamma} dt e^{-t/\lambda} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1; 2t\right) \\
 &= \frac{\sqrt{2\pi}}{\lambda} \int_{1/2}^{\infty} dt e^{-t/\lambda} \left[ {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1, 2t + i\varepsilon\right) - {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1, 2t - i\varepsilon\right) \right] \\
 &= \frac{\sqrt{2\pi}}{\lambda} \int_{1/2}^{\infty} dt e^{-t/\lambda} 2i {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1, 1 - 2t\right) \\
 &= 2i\sqrt{2\pi} e^{-1/(2\lambda)} \frac{1}{\lambda} \int_0^{\infty} dt e^{-t/\lambda} {}_2F_1\left(\frac{1}{2}, \frac{1}{2}, 1, -2t\right) \\
 &= 2ie^{-1/(2\lambda)} \mathcal{S}_0\Phi_1(\lambda). \tag{2.19}
 \end{aligned}$$

To obtain third line in eq. (2.19) we used the known discontinuity property of the hypergeometric function [54]:

$${}_2F_1\left(a, b, c \middle| t + i\varepsilon\right) - {}_2F_1\left(a, b, c \middle| t - i\varepsilon\right) = \frac{2\pi i \Gamma(c)}{\Gamma(a)\Gamma(b)} {}_2F_1\left(c - a, c - b, 1 \middle| 1 - t\right), \tag{2.20}$$

valid for  $a + b = c$ , and the last line is the (unambiguous) Borel resummation of the  $\Phi_1(\lambda)$  series in the  $\theta = 0$  direction. If we only had access to the asymptotic expressions for the high-order behavior of the coefficients of the perturbative series, in the last line we would have obtained a perturbative series expression for  $\mathcal{S}_0\Phi_1(\lambda)$ , rather than a closed-form expression for  $\mathcal{S}_0\Phi_1(\lambda)$  itself. The factor of  $2i$  on the right-hand-side of eq. (2.26) is called a Stokes constant (or “analytic invariant” in Ecalle’s terminology):

$$s = 2i. \tag{2.21}$$

Stokes constants are non-vanishing only at singular points on Stokes line and are zero otherwise.

In fact, we can obtain the same result by changing the integration variable from the “field variable”  $z$  to the “action variable”  $u = S(z)$ . This puts the integral into the form of a Borel sum.<sup>8</sup> Then the integrals over the  $\mathcal{J}_0(0^\pm)$  cycles are given by

$$\begin{aligned} \frac{1}{\sqrt{\lambda}} \int_{\mathcal{J}_0(0^\mp)} e^{-\frac{1}{2\lambda} \sin^2(z)} &= \frac{2}{\sqrt{\lambda}} \int_0^{1/2} du \frac{e^{-u/\lambda}}{\sqrt{2u(1-2u)}} \mp \frac{2i}{\sqrt{\lambda}} \int_{1/2}^\infty du \frac{e^{-u/\lambda}}{\sqrt{2u(2u-1)}} \\ &= \frac{2}{\sqrt{\lambda}} \int_0^{1/2} du \frac{e^{-u/\lambda}}{\sqrt{2u(1-2u)}} \mp i e^{-\frac{1}{2\lambda}} \frac{2}{\sqrt{\lambda}} \int_0^\infty du \frac{e^{-u/\lambda}}{\sqrt{(2u+1)2u}} \\ &= \text{Re} \mathcal{S}_0 \Phi_0 \mp i e^{-\frac{1}{2\lambda}} \mathcal{S}_0 \Phi_1 \end{aligned} \quad (2.22)$$

where  $\text{Re} \mathcal{S}_0 \Phi_0$  is unambiguous.<sup>9</sup> More importantly, in the integral over the Lefschetz thimble associated with the P-saddle,  $\mathcal{J}_0(0^\mp)$ , we immediately see the imprint of the NP physics! Moreover, we see all of the data associated with the NP saddle, both its NP weight and the perturbative fluctuations around it: it is encoded in the imaginary part of the integral along  $\mathcal{J}_0(0^\mp)$ .

Clearly, the imaginary part is ambiguous for  $\theta$  exactly zero, since the result depends on how one approaches the Stokes line. This is a reflection of the non-Borel-summability of the perturbative series. In the geometric perspective we are following here, the flip in the imaginary part of the integral over  $\mathcal{J}_0(0^\mp)$  is due to the flip of the infinite “tail” of the integration cycle that takes place when crossing the Stokes line  $\theta = 0$ . See figure 2 and figure 3. This is the geometric realization of non-Borel-summability. Happily, this is not the whole story, because the integration over the interval  $I$  is actually the linear combination of thimbles as seen in eq. (2.9). We will come back to the story of how including the contribution from  $\mathcal{J}_1$  cures this problem in section 2.4. First, however, it is useful to understand how this happens from the algebraic trans-series point of view.

## 2.2 Stokes automorphism and alien derivative

To understand how ambiguity cancellation works in the trans-series representation, it is useful to introduce the notions of Stokes automorphisms and alien derivatives from resurgency theory. To keep the presentation more streamlined, appendix A summarizes some of the results and definitions that we are going to use in what follows.

<sup>8</sup>Note that in this case the Borel transform one sees inside the Borel-sum expression are associated with sending  $\sum a_n \lambda^n \rightarrow \sum_n \frac{a_n}{(n+1/2)!} t^n$ , rather than  $\sum a_n \lambda^n \rightarrow \sum_n \frac{a_n}{n!} t^n$  as we had written above. This highlights the point that the Borel transform of a given power series is not really unique due to the freedom to divide  $a_n$  by  $(n + \alpha)!$  for arbitrary fixed  $\alpha$  when defining the transform. However, there is also a corresponding freedom, depending on  $\alpha$ , in the definition of the Borel sum. One can easily show that the value of the Borel sum is independent of  $\alpha$ .

<sup>9</sup>The integral that we identify with  $\text{Re} \mathcal{S}_0 \Phi_0$  is dominated by  $u \lesssim \lambda$  in the small  $\lambda$  regime. The procedure to obtain the perturbative expansion  $\Phi_0$  from this expression involves two steps. First, one should extend the integration domain to  $[0, \infty)$ . Next, one Taylor expands  $\frac{1}{\sqrt{1-2u}}$  around the origin. Performing the integral will give the divergent asymptotic expansion  $\Phi_0$ . The reason for the divergence is the use of the Taylor expansion beyond its radius of convergence. One can also obtain the (convergent) strong coupling expansion from the integral representation in the  $\lambda \gg 1$  regime, by expanding the exponential into a power series in  $\frac{1}{\lambda}$  and performing order by order integration.

Distinct sectorial solutions on two different sides of a Stokes line are “connected” through the Stokes automorphism,  $\underline{\mathfrak{S}}_\theta$ :

$$\mathcal{S}_{\theta+} = \mathcal{S}_{\theta-} \circ \underline{\mathfrak{S}}_\theta \equiv \mathcal{S}_{\theta-} \circ (1 - \text{Disc}_{\theta-}), \quad (2.23)$$

where  $\text{Disc}_{\theta-}$  denotes is the discontinuity arising on crossing the Stokes line, thus,

$$\mathcal{S}_{\theta+} - \mathcal{S}_{\theta-} = -\mathcal{S}_{\theta-} \circ \text{Disc}_{\theta-}. \quad (2.24)$$

In our example eq. (2.19), the difference of the right and left summations in the  $\theta = 0$ -direction of  $\Phi_0(\lambda)$  is an exponentially small imaginary term given by

$$\begin{aligned} \mathcal{S}_{0+}\Phi_0(\lambda) - \mathcal{S}_{0-}\Phi_0(\lambda) &= 2ie^{-1/(2\lambda)}\mathcal{S}_0\Phi_1(\lambda) = -\mathcal{S}_0 \circ \text{Disc}_0\Phi_0(\lambda), \\ \Rightarrow \text{Disc}_0\Phi_0(\lambda) &= -2ie^{-1/(2\lambda)}\Phi_1(\lambda). \end{aligned} \quad (2.25)$$

The Stokes automorphisms connecting different sectorial Borel sums are non-trivial only at Stokes lines. The Stokes lines are  $\theta = 0$  for  $\Phi_0$  and  $\theta = \pi$  for  $\Phi_1$ , and the resulting Stokes automorphisms are

$$\begin{aligned} \underline{\mathfrak{S}}_0\Phi_0(\lambda) &= \Phi_0(\lambda) + 2ie^{-\frac{1}{2\lambda}}\Phi_1(\lambda), & \underline{\mathfrak{S}}_0\Phi_1(\lambda) &= \Phi_1(\lambda), \\ \underline{\mathfrak{S}}_\pi\Phi_1(\lambda) &= \Phi_1(\lambda) + 2ie^{+\frac{1}{2\lambda}}\Phi_0(\lambda), & \underline{\mathfrak{S}}_\pi\Phi_0(\lambda) &= \Phi_0(\lambda). \end{aligned} \quad (2.26)$$

These equations encode the beautiful structure of resurgence: the P-saddle carries complete information about the NP saddle which is decoded using its Stokes automorphism in the  $\theta = 0$  direction. At the same time, the NP-saddle carries complete information about the P-saddle, which is decoded by its own Stokes automorphism, but in the  $\theta = \pi$  direction.

The set of all formal series appearing in our problem form a closed algebra under the action of the singularity derivative, also called alien derivative, which acts as

$$\begin{aligned} \Delta_{+\frac{1}{2}}\Phi_0 &= 2i\Phi_1, & \Delta_{+\frac{1}{2}}\Phi_1 &= 0, \\ \Delta_{-\frac{1}{2}}\Phi_1 &= 2i\Phi_0, & \Delta_{-\frac{1}{2}}\Phi_0 &= 0. \end{aligned} \quad (2.27)$$

The first one of these relations means that the action of the singularity derivative at  $\omega = +\frac{1}{2}$  on  $\Phi_0$  is just  $\Phi_1$  times the Stokes constant eq. (2.21). The action of the singularity derivative at any other point on  $\Phi_0$  just gives zero, because the Borel transform of  $\Phi_0$  does not have any other singularities, i.e.,  $\Delta_\omega\Phi_0 = 0$  for  $\omega \neq \frac{1}{2}$ .

### 2.3 Reality of resurgent trans-series for real $\lambda$ and BE-summability

As stated earlier, for  $\arg(\lambda) = 0$ , the partition function defined in eq. (2.1) is manifestly real. On the other hand, the formal first sum  $\Phi_0(\lambda)$  in eq. (2.29) is non-Borel summable in the singular direction  $\theta = 0$ , and hence it has an ambiguity, which is of order  $ie^{-\frac{1}{2\lambda}}$ . On the other hand, for  $\Phi_1$ , the singular direction is  $\theta = \pi$ , and hence it can be Borel resummed in the  $\theta = 0$  direction.

For example, for  $\theta = 0^-$  (approaching the real line either from below),

$$\Phi_0(\lambda) + ie^{-\frac{1}{2\lambda}}\Phi_1(\lambda) \xrightarrow{\text{BE-summation}} \mathcal{S}_{0-}\Phi_0 + ie^{-\frac{1}{2\lambda}}\mathcal{S}_{0-}\Phi_1$$

$$\begin{aligned}
 &= (\text{Re}\mathcal{S}_0\Phi_0 + i\text{Im}\mathcal{S}_0-\Phi_0) + ie^{-\frac{1}{2\lambda}}\mathcal{S}_0\Phi_1 \\
 &= \text{Re}\mathcal{S}_0\Phi_0 + i\left(\text{Im}\mathcal{S}_0-\Phi_0 + e^{-\frac{1}{2\lambda}}\mathcal{S}_0\Phi_1\right) \\
 &= \text{Re}\mathcal{S}_0\Phi_0
 \end{aligned} \tag{2.28}$$

and we have similar cancellation for  $\theta = 0^+$ , approaching the real line from above. The non-Borel summability of the perturbative expansion  $\Phi_0$  leads to two-fold purely imaginary ambiguity. But exactly at the Stokes line, the integration path is also two-fold ambiguous,  $\mathcal{J}_0 \pm \mathcal{J}_1$ , for  $\theta = 0^\mp$ . This maps to a two-fold ambiguity of the coefficient of the NP-term in the trans-series. The observable  $Z(\lambda)$  is the combination of the two contributions, and the ambiguities cancel in the appropriate combinations in any of the Stokes chambers, leading to the real (physical) result on positive real axis in coupling constant plane. This is an example of median resummation and Borel-Ecalle summability.<sup>10</sup>

Consequently, the trans-series expansions for the analytic continuation of the partition function eq. (2.1) in different Stokes chambers are given by

$$Z(\lambda) = \begin{cases} \mathfrak{Z}_0(\lambda) + i\mathfrak{Z}_1(\lambda) = \Phi_0(\lambda) + ie^{-\frac{1}{2\lambda}}\Phi_1(\lambda) & -\pi < \theta < 0, \\ \mathfrak{Z}_0(\lambda) - i\mathfrak{Z}_1(\lambda) = \Phi_0(\lambda) - ie^{-\frac{1}{2\lambda}}\Phi_1(\lambda) & 0 < \theta < \pi, \end{cases} \tag{2.29}$$

with a Stokes jump at  $\arg(\lambda) = 0$ . Approaching the  $\lambda \in \mathbb{R}^+$  line from above or below, we observe that the real solution for a trans-series (2.14) is given by

$$\mathfrak{Z}_{\mathbb{R}}(\lambda, 1, 0) = \mathcal{S}_{0-}\mathfrak{Z}\left(z, 1, +\frac{1}{2}s\right) = \mathcal{S}_{0+}\mathfrak{Z}\left(z, 1, -\frac{1}{2}s\right), \tag{2.30}$$

where  $s = 2i$  is once again the Stokes constant. This is a very simple example of the fact that cancellation of nonperturbative ambiguities leads to median resummation of trans-series [34]. This also seems to be valid for non-linear systems with infinitely many Borel plane singularities [15]. We comment on the generalization of this formula to QFT in section 7.

## 2.4 Lefschetz thimbles and geometrization of ambiguity cancellation

We now return to the picture offered by the Lefschetz thimble decomposition of the integration cycle to see the geometrization of ambiguity cancellation on the Stokes line  $\arg(\lambda) = 0$  which we have already seen in the algebraic trans-series representation eq. (2.28). The integral over the P-thimble  $\mathcal{J}_0$  at  $\theta = 0^-$  and at  $\theta = 0^+$  can be written schematically as

$$\begin{aligned}
 \int_{\mathcal{J}_0(0^-)} &= \int_{-\frac{\pi}{2}+i\infty}^{-\frac{\pi}{2}} + \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} + \int_{+\frac{\pi}{2}}^{+\frac{\pi}{2}-i\infty} \xrightarrow[\text{periodicity}]{} \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} + \int_{+\frac{\pi}{2}+i\infty}^{+\frac{\pi}{2}-i\infty} = Z - \int_{\mathcal{J}_1(0)} \\
 \int_{\mathcal{J}_0(0^+)} &= \int_{-\frac{\pi}{2}-i\infty}^{-\frac{\pi}{2}} + \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} + \int_{+\frac{\pi}{2}}^{+\frac{\pi}{2}+i\infty} \xrightarrow[\text{periodicity}]{} \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} + \int_{+\frac{\pi}{2}-i\infty}^{+\frac{\pi}{2}+i\infty} = Z + \int_{\mathcal{J}_1(0)}
 \end{aligned} \tag{2.31}$$

<sup>10</sup>In the QM and QFT examples, the cancellation mechanism of ambiguities is essentially the same, but unlike the 0d example where the NP-term completely disappears, there are (infinitely many) real non-perturbative contributions that also contribute to observables along with the perturbative contribution.

where in the second step, we cut the segment  $[-\frac{\pi}{2} + i\infty, -\frac{\pi}{2}]$  and glued it to  $[\frac{\pi}{2} + i\infty, \frac{\pi}{2}]$ . Because of the  $\pi$  periodicity of the integrand, the integral remains unchanged.  $Z$  is the original real valued partition function eq. (2.1), and  $\int_{\mathcal{J}_1(0)}$  is purely imaginary and is of order  $e^{-\frac{1}{2\lambda}}$ . This formula makes it manifest that the integral over the P-thimble  $\mathcal{J}_0$  has an imaginary part and is *not* equivalent to the original partition function  $Z$ .

For  $\theta = 0^-$  the addition of  $+\int_{\mathcal{J}_1(0)}$  kills the (undesired) imaginary part of the  $\mathcal{J}_0$  integral, and the combination  $\mathcal{J}_0(0^-) + \mathcal{J}_1(0^-)$  is the linear combination of thimbles associated with  $Z$  at  $\theta = 0^-$ , namely

$$\frac{1}{\sqrt{\lambda}} \int_{\mathcal{J}_0(0^-) + \mathcal{J}_1(0^-)} e^{-\frac{1}{2\lambda} \sin^2(z)} = Z. \quad (2.32)$$

Upon the Stokes jump at  $\theta = 0$   $\mathcal{J}_1 \rightarrow \mathcal{J}_1$ ,  $\mathcal{J}_0 \rightarrow \mathcal{J}_0 - 2\mathcal{J}_1$  given in eq. (2.10), we obtain

$$\frac{1}{\sqrt{\lambda}} \int_{\mathcal{J}_0(0^+) - \mathcal{J}_1(0^+)} e^{-\frac{1}{2\lambda} \sin^2(z)} = Z, \quad (2.33)$$

where the combination of  $\mathcal{J}_0(0^+), \mathcal{J}_1(0^+)$  is simply the unique linear combination which cancels the imaginary part exactly at  $\theta = 0^+$ . It is also important to note that except for  $\arg(\lambda) = 0$ , (and  $\arg(\lambda) = \pi$ ), there is never an exact cancellation between the contribution of the two saddles, and both saddles contribute! In fact, at the anti-Stokes lines  $\theta = \pm\frac{\pi}{2}$ , the modulus of the two contributions is the same, and there is an exchange of dominance.

Thus, the ambiguity in the imaginary part of the integration  $\int_{\mathcal{J}_0(0^\mp)}$  is cancelled exactly by the ambiguity in the prefactor of the  $\int_{\mathcal{J}_1(0)}$  integral. This is a simple geometric realization of the cancellation of ambiguities on the Stokes line. Stated another way, one can observe that on approaching the Stokes line from above  $\theta = 0^+$  or from below  $\theta = 0^-$ , the “amplitude” associated with the NP-saddle  $[z_1]_{\theta=0^\pm}$  is given by a two-fold ambiguous result:

$$[\rho_1]_{\theta=0^\pm} = \pm i e^{-\frac{1}{2\lambda}} \mathcal{S}_0 \Phi_1. \quad (2.34)$$

This is the counter-part of the ambiguous structure of instanton-anti-instanton-type amplitudes in QM and QFT examples, where the associated amplitude (which may naively be expected to be real) actually possess an unambiguous real part and a two-fold ambiguous imaginary part:

$$[\mathcal{I}\bar{\mathcal{I}}]_{\theta=0^\pm} \sim e^{-2S_I} \pm i\pi e^{-2S_I}. \quad (2.35)$$

This is taking place in QFT for the same reason as in ordinary integration. Of course, in semi-classically calculable regimes of QFTs and in QM, there are infinitely many saddle points and Lefschetz thimbles. The thimbles are infinite dimensional algebraic varieties and when the theory is regularized on a finite lattice, they become finite dimensional algebraic varieties. The saddle points are the perturbative vacuum and also instantons, bi-instantons, topologically neutral instanton molecules, etc, with appropriate terminological modifications in theories without the topology to support instantons, as we explore in what follows. Let  $\rho_0$  denote the perturbative vacuum and  $\rho_n$  denote various NP-sectors, with action  $S_n$ , that can communicate with the P-sector according to the structure of graded resurgence triangle explained in e.g. section 7. The evidence gathered so far suggests that

the integration over  $\mathcal{J}_n, n \neq 0$  yields both real and imaginary parts in QFT, while, in the 0d example in this section,  $\mathcal{J}_1$  yields only a purely imaginary contribution for  $\arg(\lambda) = 0$ . The cancellation of the imaginary parts in path integral examples is essentially the same as for ordinary integrals. However, in QFT there are also real unambiguous contributions to observables from NP-saddles.

## 2.5 NP-data in late terms of P-expansion

Resurgence and the Stokes automorphism allow one to extract the structure of late terms in both P and NP sectors. The asymptotic large order behavior of the perturbative expansion can be deduced by using Cauchy's theorem. Taking  $z = \frac{1}{\lambda}$ , we can write

$$F(z) = \frac{1}{2\pi i} \sum_a \int_0^{e^{i\theta_a}\infty} d\omega \frac{\text{Disc}_{\theta_a} F(\omega)}{\omega - z} + \frac{1}{2\pi i} \oint_{C^\infty} \frac{F(\omega)}{\omega - z}, \quad (2.36)$$

where summation over  $a$  is over the singular directions in Borel plane, or Stokes lines in the physical coupling plane (for simplicity we omitted the contributions coming from simple poles), and  $C^\infty$  is a closed loop at infinity.

In the present problem, the singular directions are  $\theta = 0, \pi$ . For example, consider the large order behavior of  $\Phi_0$ . The only discontinuity of  $\Phi_0$  is in the  $\theta = 0$  direction and using eq. (2.26) and eq. (2.36), we obtain

$$a_n^{(0)} \sim \frac{s}{2\pi i} \frac{\Gamma(n)}{(S_{10})^n} \left[ a_0^{(1)} + a_1^{(1)} \frac{S_{10}}{(n-1)} + a_2^{(1)} \frac{(S_{10})^2}{(n-1)(n-2)} + \dots \right]. \quad (2.37)$$

So the leading large order behavior of the asymptotic expansion around the P-saddle is determined by the relative action with respect to the NP-saddle. The corrections to the leading behavior are governed by the early terms in the perturbative expansion around the NP-saddle. This is a very explicit realization of the idea of resurgence stated just after eq. (2.26): the information in the series expansion around the NP-saddle surges up, in a disguised form, in the expansion around the P-saddle and vice versa [9].

## 3 Review of the principal chiral model on $\mathbb{R}^2$

This section briefly summarizes some basic aspects of the principal chiral model (PCM) with and without fermions. The bosonic PCM in  $d = 2$  dimensions is an asymptotically free matrix field theory. The classical action is given by

$$S_v = \frac{1}{2g^2} \int_M dt dx \text{Tr} \partial_\mu U \partial^\mu U^\dagger, \quad \frac{1}{g^2} = \frac{N}{\lambda} \quad (3.1)$$

where  $U(t, x) \in \text{SU}(N)$ ,  $M$  is a two-dimensional manifold with  $\mu$  running over  $t, x$ , and  $\lambda = g^2 N$  is a dimensionless coupling constant which must be held fixed when taking the large- $N$  limit. This action is invariant under the global symmetry group  $\text{SU}(N)_L \times \text{SU}(N)_R$  acting as  $U \rightarrow g_L U g_R^\dagger$ , with  $g_L \in \text{SU}(N)_L$  and  $g_R \in \text{SU}(N)_R$ .

Classically, the theory is scale invariant, and the non-linear wave solutions for the  $U$ -field propagate at the speed of light. This means the classical theory has  $N^2 - 1$  gapless

degrees of freedom. This is similar to classical Yang-Mills theory, which also has  $N^2 - 1$  massless (gapless) gluons.

In the quantum theory, the situation is believed to be both qualitatively and quantitatively different. In particular, a macroscopic observer in a hypothetical  $\mathbb{R}^{1,1}$  universe would not see the  $N^2 - 1$  non-linear  $U$ -field waves, just as we do not see non-linear Yang-Mills waves. Indeed, numerical lattice simulations indicate that the theory is gapped, see e.g. [55]. The mass gap is expected to be of the order of the dynamically generated strong scale  $\Lambda$

$$\Lambda^{\beta_0} = \mu^{\beta_0} e^{-\frac{4\pi}{g^2(\mu)}}, \quad \beta_0 = N \quad (3.2)$$

where  $\mu$  is a cut-off scale, and  $\beta_0$  is the leading coefficient of renormalization group beta-function.

This model is interesting because it possesses a matrix-like large- $N$  limit which is dominated by planar diagrams when  $N \rightarrow \infty$  with  $\lambda$  fixed, just like Yang-Mills theory. However, it has only received scant attention as a useful toy model for Yang-Mills. One of the primary reasons behind this is an apparent dissimilarity to Yang-Mills: the PCM does not have instantons while Yang-Mills theory does. We will come to the conclusion that this difference from Yang-Mills is only superficial, because under suitable conditions, there is an infinite class of the NP-saddles in PCM model as well.

As it happens, some NP saddles in the PCM, the unitons, have been discovered some time ago by Uhlenbeck [56]. Unitons are harmonic maps from  $S^2$  to  $SU(N)$ .<sup>11</sup> Unlike instantons, there is no homotopy argument for the stability of a uniton. However, (and sounding almost contrary to the previous statement), the uniton action is quantized in units of  $\frac{8\pi}{g^2}$  [57–61].<sup>12</sup> Thus the uniton amplitude is parametrically of the form:

$$\mathcal{U} \sim e^{-\frac{8\pi}{g^2(\mu)}} \quad \text{and} \quad \Lambda^{2\beta_0} = \mu^{2\beta_0} \mathcal{U}. \quad (3.3)$$

The structure of the moduli of a uniton, its zero and quasi-zero modes are so far not completely understood.

The bosonic principal chiral matrix theory is expected to possess the following properties:

- 1) Mass gap
- 2) Confinement: free energy  $\mathcal{F}/N^2$  approaching zero in the *low* temperature regime.
- 3) Deconfinement:  $O(N^2)$  free energy in the *high* temperature regime.

<sup>11</sup>Here,  $S^2$  should be viewed as one point compactification of  $\mathbb{R}^2$  by including a point at infinity.

<sup>12</sup>We are grateful to N. S. Manton for suggesting an elegant way to think about the quantization of the uniton actions. The key observation is that while  $\pi_2[SU(N)] = 0$ ,  $\pi_3[SU(N)] = \mathbb{Z}$ . The relevance of  $\pi_3$  comes from considering the homotopy groups of the *infinite-dimensional* manifold of field configurations  $\mathcal{M} = \text{Maps} : S^2 \rightarrow SU(N), \pi_n[\mathcal{M}] = \pi_{n+2}(SU(N))$ . This perspective suggests that the vanishing of  $\pi_2[SU(N)]$  is just the statement that there is only one connected component in  $\mathcal{M}$  [62], so the unitons must lie in the same conjugacy class as the vacuum, on the other hand the non-triviality of  $\pi_3[SU(N)]$  means that there are non-contractible one-cycles in  $\mathcal{M}$ , which appear to be realized by unitons, with the integrality of  $\pi_3[SU(N)]$  related to the integrality of the uniton action.



A few remarks are in order about these properties: **1)** In the integrability studies of the PCM where the model is viewed as solvable, the existence of the mass gap is an *assumption*. This assumption is something that we would like to derive in our framework. **2)** We refer to the low temperature regime as a confined regime because there the microscopic  $\mathcal{O}(N^2)$  degrees of freedom are not present in the physical Hilbert space. **3)** We refer to the high temperature regime as deconfined because of the liberation of the  $\mathcal{O}(N^2)$  microscopic degrees of freedom.

**Fermions.** It will also be useful to add Lie algebra valued Majorana fermions to the bosonic Lagrangian, since that makes it easier to draw analogies with QCD with fermions in adjoint representation. The action eq. (3.1) is replaced by

$$S_f = S_b + \frac{1}{2g^2} \int_M d^2x \operatorname{Tr} \left[ i\bar{\psi}_i \gamma^\mu (\partial_\mu \psi^i + \frac{1}{2}[U^\dagger \partial_\mu U, \psi^i]) - \frac{1}{16} \{\bar{\psi}_i, \gamma_5 \psi^i\} \{\bar{\psi}_j, \gamma_5 \psi^j\} \right], \quad (3.4)$$

where  $\psi^i$ , with  $i, \dots, N_f$ , are  $\mathfrak{su}(N)$  valued Majorana fermions and  $\gamma_5$  is the chirality matrix in  $2d$ . The fermions are two-index matrix valued fields. Under  $SU(N)_L \times SU(N)_R$  they transform as  $\psi^i \rightarrow g_R \psi^i g_R^\dagger$ , and despite the asymmetric-looking form of the fermion transformations the theory still has an  $SU(N)_L \times SU(N)_R$  symmetry for any  $N_f$ .

**$N_f = 1$  or  $\mathcal{N} = (1, 1)$ .** The model with  $N_f = 1$  has  $\mathcal{N} = (1, 1)$  supersymmetry, the minimal non-chiral supersymmetry in 2D, see e. g. [63, 64]. On top of the three properties of the bosonic model mentioned above, this fermionic model is also believed to possess the property

- 4) Discrete chiral symmetry breaking and two isolated vacua (for any  $N$ ).

The  $\mathcal{N} = (1, 1)$  model has a  $\mathbb{Z}_2$  discrete chiral symmetry:

$$\mathbb{Z}_2 : \psi \rightarrow \gamma_5 \psi, \quad \text{or} \quad \mathbb{Z}_2 : \operatorname{Tr} \bar{\psi} \psi \rightarrow -\operatorname{Tr} \bar{\psi} \psi \quad (3.5)$$

which is believed to be dynamically broken by the formation of fermion bilinear condensate,  $\langle \frac{1}{N} \operatorname{Tr} \bar{\psi} \psi \rangle = \pm \Lambda$ .

The existence of the two isolated vacua can also be backed up by the following independent argument, for the case of the  $\mathcal{N} = (1, 1)$   $SU(2)$  PCM. The group manifold for the  $SU(2)$  PCM is three-sphere  $S^3$ , same as the  $O(4)$  sigma model. The supersymmetric index for the  $O(4)$  model (or generally for  $O(N)$  models) was calculated a long time ago and it is equal to  $I_S = 2$  [65]. This is indeed compatible with the discrete chiral symmetry breaking and the existence of two isolated vacua. We claim that the supersymmetric Witten index for the PCM model for arbitrary  $N$  must also be equal to two:

$$I_S(SU(N)) = 2. \quad (3.6)$$

**$N_f > 1$ .** The PCM model with  $N_f$  Weyl Majorana fermions has an  $SU(N_f)$  continuous global symmetry and  $\mathbb{Z}_2$  discrete chiral symmetry. A continuous global symmetry in a finite  $N$  theory cannot be broken on  $\mathbb{R}^2$  due to the Mermin-Wagner-Coleman theorem. We

expect that for sufficiently low  $N_f$  the discrete chiral symmetry should be broken, and at large- $N$  the  $SU(N_f)$  global symmetry may be broken as well.

The inclusion of an arbitrary number of fermion flavors does not modify the one loop beta function given in eq. (3.2). Unlike in four dimensions, two-dimensional non-linear sigma models remain asymptotically free even at arbitrarily large- $N_f$ . Whether these theories are confining or exhibit IR conformal behavior should depend on the number of fermionic flavors,  $N_f$ . The value  $N_f^*$  at which these theories move from gapped (for  $U$ -fluctuations) behavior to gapless behavior is currently unknown. We expect that for  $N_f < N_f^*$ , the discrete chiral symmetry should be broken and these theories possess two isolated vacua.

### 3.1 Perturbation theory on $\mathbb{R}^2$ and IR-renormalons

The structure of perturbation theory in PCM is similar to other asymptotically free matrix field theory. On general grounds, we expect two types of factorial divergences in the perturbative series associated with the P saddle point. These are

1. Combinatorial  $n!$  growth in the number of Feynman diagrams
2. Phase space  $n!$  contribution coming from the integration of high and low momenta from a fixed class of diagrams. These contributions are referred to as the UV and IR renormalons, respectively.

Usually, in theories with instantons, the combinatorial growth is associated with instanton-anti-instanton  $[\mathcal{I}\bar{\mathcal{I}}]$  pairs. That is, the ambiguity of Borel resummation of perturbation theory is cancelled against the ambiguity in the  $[\mathcal{I}\bar{\mathcal{I}}]$ -amplitude. However, in the PCM, there are no instantons to begin with. In [31], we argued that the uniton saddle ought to substitute the  $[\mathcal{I}\bar{\mathcal{I}}]$ -saddle and its ambiguity, yielding an ambiguity of order  $\pm i e^{-\frac{8\pi}{g^2(Q)}}$ .

On  $\mathbb{R}^2$ , Fateev, Kazakov, and Wiegmann [66, 67] used integrability techniques to show that there exists a much larger ambiguity. It is guessed there that this is related to the leading renormalon ambiguity:

$$\text{Leading IR renormalon on } \mathbb{R}^2 : \pm i e^{-\frac{8\pi}{g^2(Q)N}} \sim \pm i \frac{\Lambda^2}{Q^2}, \quad (3.7)$$

where  $Q$  is a large Euclidean momentum.

This is a sensible guess in the light of the operator product expansion (OPE). In the PCM, the leading non-trivial condensate is the dimension two operator  $\text{Tr } \partial_\mu U \partial^\mu U^\dagger$ . Since  $\arg \lambda = 0$  is a Stokes line in PCM, the expectation value of generic operators that can mix with perturbation theory are two-fold ambiguous [68]. The condensate evaluated at  $\theta = 0^\pm$  must give

$$\left\langle \frac{1}{N} \text{Tr } \partial_\mu U \partial^\mu U^\dagger \right\rangle_{\theta=0^\pm} Q^{-2} \sim c_1 \frac{\Lambda^2}{Q^2} \pm i c_2 \frac{\Lambda^2}{Q^2} \quad (3.8)$$

where  $c_1$  and  $c_2$  are pure numbers. Of course, one also expects more-suppressed IR renormalon ambiguities of the order  $\Lambda^{2m}/Q^{2m}$ ,  $m > 1$ .

On  $\mathbb{R}^2$ , there is no semi-classical interpretation for IR renormalons. However, at large  $Q^2$  there is a useful hierarchy of non-perturbative scales associated with the IR renormalons

$$\underbrace{e^{-\frac{8\pi}{g^2(Q)N}}}_{1^{\text{st}} \text{ IR-renormalon}} \gg \underbrace{e^{-\frac{16\pi}{g^2(Q)N}}}_{2^{\text{nd}} \text{ IR-renormalon}} \gg \dots \gg \underbrace{e^{-\frac{8\pi}{g^2(Q)}}}_{1\text{-uniton}} \gg \dots$$

or equivalently  $\frac{\Lambda^2}{Q^2} \gg \frac{\Lambda^4}{Q^4} \gg \dots \gg \frac{\Lambda^{2N}}{Q^{2N}} \gg \dots$  (3.9)

which is exploited in the OPE approach. It is worth emphasizing that the leading IR renormalon ambiguity is parametrically much larger than the uniton ambiguity at large Euclidean momentum  $Q$ . It is actually the  $N^{\text{th}}$  root of it.

In theories with fermions, if  $N_f < N_f^*$ , we expect only minor and relatively unimportant changes in the Borel plane structure. If  $N_f > N_f^*$ , one does not expect IR-renormalons to exist, since the coupling does not diverge in the infrared. This drastic change at  $N_f^*$  is consistent with a speculation by 't Hooft, in the context of gauge theories, that IR-renormalons are connected with the mass gap and confinement [22].

#### 4 Compactification to $\mathbb{R} \times S^1$ and adiabatic continuity

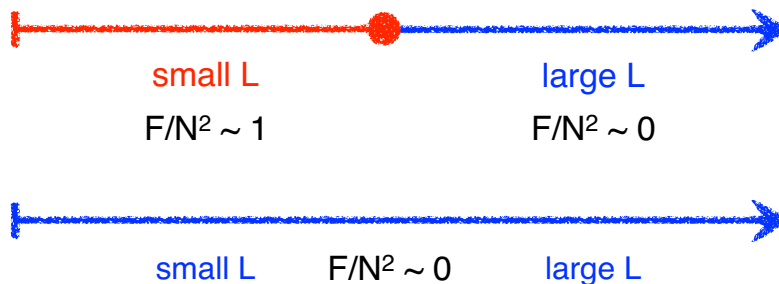
Usually, one can only hope to get analytic insights into the non-perturbative physics of a field theory when it is weakly coupled. This is definitely not the case for the PCM on  $\mathbb{R}^2$ , where the IR physics is strongly coupled in terms of  $\lambda$ . When the PCM is compactified on  $\mathbb{R} \times S^1$ , however, at small enough circle size  $L$  the coupling will become set at the scale  $1/L$ , and the theory is guaranteed to become weakly coupled thanks to asymptotic freedom.

In recent years it has become clear that in QCD-like gauge theories on  $\mathbb{R}^3 \times S^1$ , the appearance of a weak-coupling regime at small- $L$  can occur in two dramatically different ways:

- i) If the  $S^1$  circle is thermal, as  $L$  is dialed from large to small, the gauge theory goes through a phase transition, or a rapid cross-over, in such a way that the physics in the small- $L$  theory is not adiabatically connected to the physics in the large- $L$  decompactification limit.
- ii) If the  $S^1$  circle is spatial in theories with e.g. massless adjoint fermions, or if a center-stabilizing deformation is used, then as  $L$  is dialed from large to small, one finds that the small- $L$  limit of the gauge theory is adiabatically connected to the large- $L$  regime without any phase transition or rapid cross-over.

In gauge theory, thermal Yang-Mills theory and thermal  $\mathcal{N} = 1$  Super-Yang-Mills (SYM) yield examples of the first type of limit, as is well known from thermal field theory [69]. In this case, the long distance<sup>13</sup> physics is strongly coupled and incalculable. Some examples of the second type are deformed Yang-Mills and spatially compactified  $\mathcal{N} = 1$  SYM [21, 29]. In this case, the long distance physics remains weakly coupled and is non-perturbatively calculable.

<sup>13</sup>Here by “long distance” we mean physics dominated by contributions from distance  $\ell \gg (g_3^2 N)^{-1} = 1/(g_4^2 N T)$ .



**Figure 5.** *Top:* with a thermal compactification on  $\mathbb{R} \times S^1_\beta$ , there is a rapid-crossover from an  $F/N^2 \sim O(1)$  (deconfined) behavior of the free energy to  $F/N^2 = 0$  confined regime, which becomes a genuine phase transition at  $N = \infty$ . Even at finite- $N$ , however, the quantitative behavior of the theory changes dramatically between these two regimes, despite the fact that there is no sharp phase transition in a finite volume. *Bottom:* by using spatial compactification, we find a unique small- $L$  limit in which “free energy” scales as  $F/N^2 = 0$ . The behavior of the theory does not change dramatically from small- $L$  to large- $L$ . This is the idea of *adiabatic continuity*. Since the small- $L$  theory is weakly coupled thanks to asymptotic freedom, it is NP-calculable, and the knowledge gained therein is continuously connected to the physics of decompactified theory on  $\mathbb{R}^2$ .

Our goal in this section is to discuss the realization of both of these classes of small- $L$  limits in the PCM by a careful analysis of the compactification procedure, as depicted in figure 5. The theory defined by eq. (3.1) has a global  $SU(N)_L \times SU(N)_R$  symmetry acting as  $U \rightarrow g_L U g_R^\dagger$ . Defining the theory on an Euclidean base space manifold  $\mathbb{R} \times S^1$  requires a choice of boundary conditions on the circle. We will consider a family of compactified theories labeled by a choice of boundary conditions

$$U(t, x + L) = e^{iH_L} U(t, x) e^{-iH_R}, \quad (4.1)$$

$$\psi_i(t, x + L) = (\pm) e^{iH_R} \psi_i(t, x) e^{-iH_R}, \quad (4.2)$$

where  $H_L, H_R \in \mathfrak{su}(N)$ . We refer to the  $(-)$  boundary conditions for the fermions as thermal, since they reduce to the standard anti-periodic thermal boundary conditions when  $H_R = H_L = 0$ , and we refer to  $(+)$  boundary conditions as spatial, since they become purely periodic BCs when  $H_R = H_L = 0$ . For generic  $H_L, H_R$ , these are non-thermal compactifications. For the supersymmetric  $\mathcal{N} = (1, 1)$  theory, the spatial boundary conditions respect supersymmetry.

The question is which theory, in this family, is in the same ‘phase’ at small- $L$  as the theory on  $\mathbb{R}^2$ .<sup>14</sup> With a thermal compactification, there is a rapid cross-over/phase transition at finite/large  $N$  as  $L$  is dialed from large to small values. Our goal is to construct a special small- $L$  limit in which this does not happen, so that the changes between large- $L$  and small- $L$  are ‘adiabatic’. One of the two observables we will use in our construction will

<sup>14</sup>We are working with a two-dimensional theory, for which the Coleman-Mermin-Wagner theorem implies the lack of any local order parameters which could distinguish distinct phases at finite  $N$ . Nevertheless, there can still be rapid cross-overs in the physical properties as a function of  $L$  at finite  $N$  (which become a sharp phase transition at  $N = \infty$ ) and as a result small- $L$  physics can be very different from large- $L$  physics.

sharpen into a proper order parameter in the large- $N$  limit, where the Coleman-Mermin-Wagner theorem no longer forbids such a notion.

To proceed with the analysis we find it convenient to switch variables to  $\tilde{U}, \tilde{\psi}$ :

$$\tilde{U} = e^{-iH_L \frac{x}{L}} U e^{iH_R \frac{x}{L}}, \quad (4.3)$$

$$\tilde{\psi} = e^{-iH_R \frac{x}{L}} \psi e^{iH_R \frac{x}{L}}, \quad (4.4)$$

so that  $\tilde{U}, \tilde{\psi}$  are periodic on  $S^1$ . In terms of  $\tilde{U}, \tilde{\psi}$ , the action becomes

$$\begin{aligned} S[H_L, H_R] &= \frac{N}{2\lambda} \int_{\mathbb{R} \times S^1} dt dx \operatorname{Tr} D_\mu \tilde{U} D^\mu \tilde{U}^\dagger \\ &+ \frac{N}{2\lambda} \int_{\mathbb{R} \times S^1} dt dx \operatorname{Tr} \left[ i \tilde{\bar{\psi}}_i \gamma^\mu \left( D_\mu \tilde{\psi}^i + \frac{1}{2} [\tilde{U}^\dagger D_\mu \tilde{U}, \tilde{\psi}^i] \right) - \frac{1}{16} \{ \tilde{\bar{\psi}}_i, \gamma_5 \tilde{\psi}^i \} \{ \tilde{\bar{\psi}}_j, \gamma_5 \tilde{\psi}^j \} \right], \end{aligned} \quad (4.5)$$

where

$$\begin{aligned} D_\mu \tilde{U} &= \partial_\mu \tilde{U} - i \frac{\delta_{\mu,x}}{L} (H_L \tilde{U} - \tilde{U} H_R) = \partial_\mu \tilde{U} - i \delta_{\mu,x} ([H_V, \tilde{U}] + \{H_A, \tilde{U}\}), \\ D_\mu \tilde{\psi}^i &= \partial_\mu \tilde{\psi}^i - i \frac{\delta_{\mu,x}}{L} [H_R, \tilde{\psi}^i], \end{aligned} \quad (4.6)$$

and  $H_{V,A} = \frac{1}{2L}(H_L \pm H_R)$ . We can interpret  $\frac{1}{L}H_L$  and  $\frac{1}{L}H_R$  as background gauge fields for the global symmetry group  $SU(N)_L \times SU(N)_R$ . The symmetry of the original action eq. (3.1) under  $U \rightarrow g_L U g_R^\dagger$  becomes a symmetry of eq. (4.5) under  $\tilde{U} \rightarrow g_L \tilde{U} g_R^\dagger$  together with  $H_L \rightarrow g_L H_L g_L^\dagger$  and  $H_R \rightarrow g_R H_R g_R^\dagger$ , which are effectively global ‘gauge’ transformations. A theory with the twisted boundary conditions (BCs) of eq. (4.1) can be equivalently viewed as a theory with periodic BCs for  $\tilde{U}, \tilde{\psi}$  with the constant background gauge fields  $\frac{1}{L}H_{L/R}$ .<sup>15</sup> From here onward we will work with the latter picture, and drop the tilde on  $U, \psi$  for simplicity.

#### 4.1 Large- $L$ expectations

To figure out the right choice for  $H_V, H_A$ , we have to decide what properties of the large- $L$  theory we want to capture in the small- $L$  limit. From our perspective, the most essential features of the bosonic PCM on  $\mathbb{R}^2$  and large  $S^1 \times \mathbb{R}$  are the existence of a mass gap and the vanishing of  $F/N^2$  in the large- $N$  limit.

We want to find a small- $L$  limit which allows both of these expected properties to persist. To translate these heuristic expectations into sharp conditions on the theory, we

<sup>15</sup>Note that these constant background fields  $H_{L,R}$  are *not* standard chemical potentials for the conserved currents

$$J_\mu^L = i U^\dagger \partial_\mu U, \quad (4.7)$$

$$J_\mu^R = i \partial_\mu U U^\dagger. \quad (4.8)$$

A chemical potential in Minkowski space enters the action as a constant background *time* component of a gauge field. In Euclidean space, a chemical potentials enters the action as a constant imaginary Euclidean-time component of a background gauge field. In contrast, we interpret the background gauge fields  $H_V, H_A$  as the *space* components of the gauge fields associated to the  $SU(N)_{L,R}$  symmetries, and they are real in *both* Minkowski and Euclidean space.

first observe that the existence of mass gap implies that when  $L \rightarrow \infty$ , the dependence on the boundary conditions must vanish, since in a theory with a mass gap we expect finite-volume effects to vanish as  $e^{-\Delta L}$ , where  $\Delta$  is the mass gap. Let  $Z(L; H_L, H_R)$  be the partition function of the theory with the background fields turned on, and let us define as  $\mathcal{F}(L; H_L, H_R) = \mathcal{V}^{-1} \log Z(L; H_L, H_R)$  as the “twist free energy”, where  $\mathcal{V}^{-1}$  is the volume of the space-time manifold. Then it is clear that in the large- $L$  confined phase, the free energy must be independent of the boundary conditions at leading order in  $N$ . At small- $L$ , it is not possible to have complete independence of the theory from  $L$ .<sup>16</sup> The highest degree of independence from  $L$  one can demand at small- $L$  is

$$\frac{\partial \mathcal{F}}{\partial H_i} = \langle J_x^i \rangle_{H_V, H_A} = 0 \quad i = V, A \quad (4.9)$$

where  $J_\mu^{V/A} = \frac{1}{2}(J_\mu^L \pm J_\mu^R)$ , and the subscript on  $\langle \cdot \rangle_{H_V, H_A}$  is a reminder that the expectation value is to be taken with background fields turned on. Heuristically, this means that we demand that changing boundary conditions should not result in persistent currents in the compactified direction. Equation (4.9) automatically holds in the decompactification limit in the large- $L$  theory, but it becomes a non-trivial constraint at small- $L$ .

Next, to have any chance that the small- $L$  theory is confined according to the count of the degrees of freedom contributing to the physical Hilbert space, the “twist free energy”, i.e, the free energy of the system normalized to  $N^2$  at a given value of the the boundary conditions must zero. For instance, with purely ‘thermal’ boundary conditions, with  $H_V = H_A = 0$ , when  $L$  can be interpreted as an inverse temperature  $1/T$ , at small- $L$  the theory becomes weakly coupled, but one expects the theory to be ‘deconfined’, with all  $N^2 - 1$  components of the matrix  $U$  liberated and contributing equally to free energy density, so that  $\mathcal{F}/N^2 \sim \mathcal{O}(1)$ . (We show this explicitly in section 4.3.) This is in sharp contrast to what happens in the low temperature decompactification limit, where we expect the theory to be in a ‘confining’ phase, in the sense that  $\lim_{N \rightarrow \infty} \mathcal{F}/N^2 = 0$ . That this is the case can be seen from lattice simulations (see e.g. [55]) as well as from the exact solution for the spectrum using integrability [44, 45, 66, 67].

Indeed, as shown in [66, 67, 70], the spectrum consists of  $N - 1$  single-particle modes<sup>17</sup> with masses  $m_k$

$$m_k = m \frac{\sin(\frac{\pi k}{N})}{\sin(\pi/N)}, \quad k = 1, \dots, N - 1 \quad (4.10)$$

where  $m = \Lambda$  is the strong scale of the theory. The states at level- $k$  transform as a bi-product of  $k$ -index antisymmetric irreducible representations with degeneracies  $d_k$ :

$$V_k \otimes V_k \in \text{SU}(N)_L \times \text{SU}(N)_R, \quad d_k = [\dim(V_k)]^2 = (N!/[k!(N-k)!])^2 \quad (4.11)$$

where  $V_k$  is  $k$ -index anti-symmetric representation. To compute the free energy, we need to consider multi-particle configurations. This is subtle because the single-particle modes have non-trivial interactions between each other: the S matrix is not trivial. At low

<sup>16</sup>If it were so, the theory would still be strongly coupled at small- $L$ .

<sup>17</sup>We are grateful to V. Kazakov for discussion related to this issue.

temperatures, the dominant contributions of the interactions will be due to exchange of the lightest massive mode, and their size is controlled by the factor

$$c = N^2 e^{-m/T}. \quad (4.12)$$

The factor of  $N^2 = d_1$  comes from the degeneracy of the lightest states, while  $e^{-m/T}$  is the associated Boltzmann factor. For the interactions to be negligible, we must have  $c \ll 1$ , which requires

$$T \ll m/(2 \log N). \quad (4.13)$$

Note that this scale depends on  $N$ , in contrast to the situation in confining gauge theories. The reason for the difference is that in confining gauge theories the single-particle states are all  $SU(N)$  gauge singlets, while in  $\sigma$ -models like the PCM the spectrum is charged under  $SU(N)_L \times SU(N)_R$ .

When the condition  $c \ll 1$  is met, we are dealing with a dilute gas, and the free energy can be written as

$$\mathcal{F} = -T \sum_{k=1}^{N-1} d_k \int \frac{dp}{2\pi} \log \left( 1 - e^{-\omega_k/T} \right) \quad (4.14)$$

where  $\omega_k^2 = p^2 + m_k^2$ , which reduces to

$$\mathcal{F} \rightarrow N^2 T^2 \left( \frac{m}{2\pi T} \right)^{1/2} e^{-m/T} \sim \frac{c m^2}{(\log N)^{3/2}} \quad (4.15)$$

So as claimed above,

$$\lim_{N \rightarrow \infty} \mathcal{F}/N^2 = 0 \quad (4.16)$$

for small  $T$ .

At large- $N$ , the value of  $\frac{1}{N^2} \mathcal{F}$  becomes a bona-fide order parameter: it is zero in the decompactification limit, but may become order one at small- $L$ , depending on the values of  $H_V, H_A$ . Hence we will demand that  $\mathcal{F}/N^2$  (or more precisely the natural dimensionless quantities  $L^2 \mathcal{F}$  or  $\Lambda^{-2} \mathcal{F}$  associated to  $\mathcal{F}$ ) must go to zero at large  $N$  for any  $L$ . This expectation is believed to be automatically met in the decompactification limit based on previous studies using integrability [44, 45] or lattice Monte Carlo simulations, see e.g. [55]. But demanding  $L^2 \mathcal{F}/N^2 \rightarrow 0$  poses a non-trivial constraint at small- $L$ .

To summarize, our *adiabaticity conditions* are

1. We demand an insensitivity to changes in boundary conditions, eq. (4.9).
2. We demand the free energy normalized to  $N^2$  goes to zero at large  $N$ .
3. For the supersymmetric  $N_f = 1$  PCM, we impose susy preserving boundary conditions. For the non-susy  $N_f > 1$  case we impose the same supersymmetric boundary conditions for all the fermions.<sup>18</sup>

We emphasize that we are *not* assuming that the theory is gapped at small- $L$ . That is a

---

<sup>18</sup>While it is clear that this is a sensible demand for the  $N_f = 1$  theory, the reason to demand this for general  $N_f \geq 1$  is more subtle. The reason we do so is that we expect the large- $N$  PCM with  $N_f$  ‘adjoint’ fermions to have an emergent fermionic symmetry at large- $N$  due to arguments similar to the ones recently given for QCD[Adj] in [71].



dynamical question about the theory, which should be settled by a calculation at small- $L$ . What makes the construction interesting is that the unique small- $L$  theory selected by the conditions above *does* turn out to have a non-perturbatively-generated mass gap! This is consistent with the notion that the small- $L$  limit we construct really is adiabatic.

## 4.2 Choosing the right small- $L$ limit

Our goal for the rest of this section is to see which choice of  $H_V, H_A$  results in a  $Z(L; H_L, H_R)$  which reproduces the large- $L$  expectations encoded in eq. (4.9) and  $L^2 \mathcal{F}/N^2 \rightarrow 0$  even at small- $L$ . When  $L\Lambda$  is small enough, the theory will become weakly coupled in terms of  $U$ . The precise meaning of ‘small enough’ is subtle, and we address it below. If  $U$  is parametrized as

$$U = e^{iW}, \quad W \in \mathfrak{su}(N) \quad (4.17)$$

we expect that at weak coupling the dominant contribution to  $Z$  will come from fluctuations which are quadratic in  $W$ .

We first determine the form of  $H_A$  consistent with eq. (4.9). To do this, we perform a global axial ‘gauge’ transformation using  $g \in \mathrm{SU}(N)_A$ , so that  $H_A \rightarrow g H_A g^\dagger$  is diagonalized and lies within the Cartan subalgebra of  $\mathfrak{su}(N)$   $H_A = a_\alpha t^\alpha$ , with  $\{t^\alpha\}$  being the Cartan generators normalized as  $\mathrm{Tr}(t^\alpha t^\beta) = \delta^{\alpha\beta}/2$ . The small  $W$  (i.e., perturbative) action then becomes

$$S = \frac{N}{2\lambda} \int_{\mathbb{R} \times S^1} dt dx \left( -2 \sum_{\alpha=1}^{N-1} (a_\alpha)^2 + \mathcal{O}(W^2) \right). \quad (4.18)$$

This constitutes a *tree-level* potential for the eigenvalues of  $H_A$ , which is extremized when  $H_A = 0$ . Hence to satisfy eq. (4.9), we must set<sup>19</sup>  $H_A = 0$ . So long as the theory is weakly coupled, quantum effects cannot change the extremum of the potential energy as a function of  $H_A$ , whether or not there are fermions in the theory. The reason is the presence of the *non-vanishing* tree-level potential above. Hence from here onward, we will set  $H_A = 0$ .

Now we can focus on working out the effective potential for  $H_V$ . When  $H_A = 0$  we can use a global vectorial ‘gauge’ transformation to diagonalize  $H_V$ , so that it lies in the Cartan subalgebra. The energy is independent of  $H_V$  at tree-level, so now we must do a one-loop analysis to compute the first non-trivial contributions to the twist free energy.

Before presenting the result for the twist free energies with the two classes (thermal and spatial) of BCs we are considering, it is useful to introduce a Wilson loop operator associated to  $H_V$ , which will allow us to write the one loop twist free energy in a more illuminating form:

$$\Omega = e^{i \oint dx_2 H_V} = e^{i L H_V} \quad \text{where} \quad (H_V)_{jk} = 2\pi \mu^j L^{-1} \delta_{jk}, \quad \mathrm{Tr} H_V = 0. \quad (4.19)$$

Note that the eigenvalues of  $\Omega$  are  $\mathrm{SU}(N)_V$  gauge-invariant, and  $\mathrm{Tr} \Omega$  transforms non-trivially under the  $\mathbb{Z}_N$  center symmetry of  $\mathrm{SU}(N)_V$  acting as  $\Omega \rightarrow \omega \Omega, \omega \in \mathbb{Z}_N$ . The

<sup>19</sup>Note that the partition function actually depends on the conjugacy classes of  $H_L, H_R$ :  $Z[H_L, H_R] = Z[g_L H_L g_L^\dagger, g_R H_R g_R^\dagger]$  with  $g_{L/R} \in \mathrm{SU}(N)_{L/R}$ . Hence once can trade a purely axial background  $H_V = 0, H_A \neq 0$  (as in [66, 67]) for a purely vectorial one, provided one modifies eq. (4.17) accordingly. We thank V. Kazakov and Z. Bajnok for pointing this out to us.

expression for the twist free energy of the PCM in the presence of the background gauge field  $H_V$  can now be rewritten in terms of the  $\Omega$ :<sup>20</sup>

$$\Omega = \begin{pmatrix} e^{2\pi i \mu_1} & 0 & \dots & 0 \\ 0 & e^{2\pi i \mu_2} & \dots & 0 \\ \vdots & & & \\ 0 & 0 & \dots & e^{2\pi i \mu_N} \end{pmatrix}, \quad 0 \leq \mu_1 \leq \mu_2 \leq \dots \leq \mu_N < 1. \quad (4.20)$$

One can now calculate the potential for the background holonomy  $\Omega$  by integrating out the weakly coupled KK-modes at one-loop level. The result is

$$L^2 V_-(\Omega) = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n^2} (-1 + (-1)^n N_f) (|\text{Tr } \Omega^n|^2 - 1) \quad (\text{thermal}), \quad (4.21)$$

$$L^2 V_+(\Omega) = (N_f - 1) \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{1}{n^2} (|\text{Tr } \Omega^n|^2 - 1) \quad (\text{spatial}). \quad (4.22)$$

An intuitive way to derive eq. (4.22) is as follows. (We only detail the derivation for the spatial case; the thermal analysis is very similar.) At sufficiently small- $L$ , the Kaluza-Klein tower of modes in the principal chiral model can be viewed as a collection of simple harmonic bosonic and fermionic oscillators. The strategy of the derivation follows Lüscher and van Baal [72, 73]. However, the idea of adiabatic continuity, which is instrumental for our purpose, did not appear in these earlier works.

In spatial compactification, i.e., with periodic boundary conditions for fermions, the KK modes of bosons and fermions are degenerate at the classical level. (This may or may not be lifted quantum mechanically, depending on the theory.) There are  $(N^2 - 1)$  physical bosonic fluctuations and  $N_f(N^2 - 1)$  fermionic fluctuations for each Kaluza-Klein level  $k \in \mathbb{Z}$ . The vacuum energy density associated with the background eq. (4.20) is the sum of ground state energies of corresponding bosonic and fermionic harmonic oscillators, and it is equal to:

$$\begin{aligned} L\mathcal{E}[\mu_{ij}] &= \sum_{\text{bosons}} \frac{1}{2} \omega^b - \sum_{\text{fermions}} \frac{1}{2} \omega^f \\ &= (1 - N_f) \frac{1}{2L} \sum_{i,j} \sum_{k \in \mathbb{Z}} |\mu_i - \mu_j + 2\pi k|. \end{aligned} \quad (4.23)$$

Since the vacuum energy density eq. (4.23) is periodic in  $\mu_{ij} \equiv \mu_i - \mu_j$  with period  $2\pi$ , it can be Fourier transformed,

$$\mathcal{E}[\mu_{ij}] = \sum_{ij} \sum_{n \in \mathbb{Z} \setminus \{0\}} P_n e^{in\mu_{ij}}, \quad (4.24)$$

which is actually the Poisson resummation of the original formula. Here,  $P_n = \frac{(1-N_f)}{2L^2} I_n$ . The advantage of this form is that  $\sum_{ij} e^{in\mu_{ij}} = |\text{Tr } \Omega^n|^2$ , and we can express the result in

---

<sup>20</sup>We use the  $S_N$  Weyl symmetry of  $\mathfrak{su}(N)_V$  to arrange the eigenvalues of  $\Omega$  in a canonical order in the case where they are not coincident.

terms of the background holonomy. The summation over KK modes  $k \in \mathbb{Z}$  thus turns into a summation over the winding number of the line operator. The prefactor is:

$$\begin{aligned}
 I_n &= \frac{1}{2\pi} \int_0^{2\pi} d\mu \sum_{k \in \mathbb{Z}} |\mu + 2\pi \vec{k}| e^{in\mu} \\
 &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\mu |\mu| e^{in\mu} = \frac{1}{2\pi} \frac{1}{\Gamma(-\frac{1}{2})} \int d\mu \int_0^{\infty} \frac{d\alpha}{\alpha^{3/2}} e^{-\alpha\mu^2} e^{in\mu} \\
 &= \frac{1}{2\pi} \frac{1}{\Gamma(-\frac{1}{2})} \int_0^{\infty} \frac{d\alpha}{\alpha^{3/2}} \left(\frac{\pi}{\alpha}\right)^{1/2} e^{-\frac{n^2}{4\alpha}} = -\frac{1}{\pi} \frac{1}{n^2},
 \end{aligned} \tag{4.25}$$

resulting in vacuum energy density

$$\mathcal{E}[\Omega] = V_+[\Omega] = \frac{1}{2\pi L^2} \sum_{n \in \mathbb{Z} \setminus \{0\}} (N_f - 1) \frac{|\text{Tr } \Omega^n|^2}{n^2}, \tag{4.26}$$

which is identical to eq. (4.22).

Our adiabaticity conditions now tell us that we must find the extrema of the twist free energy, and compute the large- $N$  scaling of  $\mathcal{Z} = L^2 V$  at the extremum. Note that the expressions for the twist free energy is very similar to the effective potential for the eigenvalues of Polyakov loops in  $\text{SU}(N)$  gauge theories on  $\mathbb{R}^3 \times S^1$  [16, 74] with  $N_f$  adjoint Weyl fermions. The minimization problem for the potential is same as the one in QCD(Adj) on  $\mathbb{R}^3 \times S^1$ .<sup>21</sup> It turns out that there are two extrema that we must deal with in general.

### 4.3 Thermal compactification and non-adiabaticity

There are  $N$  ‘thermal’ extrema of the thermal or spatial twist free energy at

$$\Omega^{\text{thermal}} = e^{i\frac{2\pi k}{N}} \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}, \tag{4.27}$$

where  $k$  labels the center-position of the lump of eigenvalues. For these extrema  $H_V = 0$ , and if we use anti-periodic BCs for fermions this is precisely the standard Euclidean thermal compactification. Note that  $\text{Tr } \Omega^{\text{thermal}}$  transforms non-trivially under all non-trivial elements in the center subgroup  $\mathbb{Z}_N \subset \text{SU}(N)_V$ .

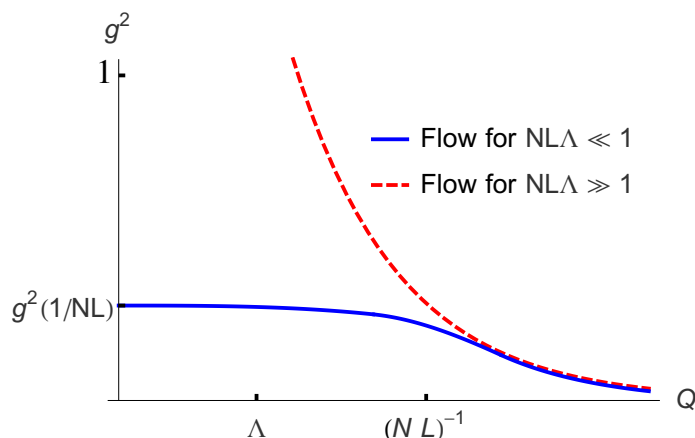
In the thermal case, the free energy density is

$$\mathcal{F} = V_-[\Omega = 1] = (N^2 - 1) \frac{\pi}{6} \left(1 + \frac{N_f}{2}\right) T^2. \tag{4.28}$$

This gives a nice check of our calculations, because this is precisely the expected Stefan-Boltzmann law:  $(N^2 - 1)$  is the number of bosonic degrees of freedom and  $\frac{\pi}{6} T^2$  is the free energy per boson, while  $(N^2 - 1)N_f$  is the number of fermionic degrees of freedom, with

---

<sup>21</sup>This analysis can easily be generalized to PCM with other classical Lie groups such as  $\text{SO}(N)$ ,  $\text{Sp}(N)$ , ... along the same lines as [16].



**Figure 6.** A sketch of the flow of the coupling constant  $g^2$  in the two regimes  $\frac{NL\Lambda}{2\pi} \ll 1$  and  $\frac{NL\Lambda}{2\pi} \gg 1$  in the principal chiral model with a  $\mathbb{Z}_N$ -symmetric background holonomy. If  $N$  is large with  $L$  fixed, large- $N$  volume independence holds in the  $NL\Lambda \gg 1$  regime. The physics of the theory on  $\mathbb{R} \times S^1$  exactly coincides with the physics of the theory on  $\mathbb{R}^2$ , up to  $1/N$  corrections, and the theory is strongly coupled at large distances and not tractable using semiclassical methods. On the other hand, if  $NL\Lambda \ll 1$ , the physics of the theory is continuously connected to the physics on  $\mathbb{R}^2$ , but volume independence does not hold. Here the coupling constant flow freezes at some small value, and the theory is semi-classically tractable.

$\frac{\pi}{12}T^2$  being the free energy per fermion. The factor of two difference follows from statistics, Bose-Einstein versus Fermi-Dirac.

This is the deconfined regime shown in figure 5 where the  $O(N^2)$  degrees of freedom are liberated, not adiabatically connected to the large- $L$  confined regime.

#### 4.4 Spatial compactification and adiabaticity

The only other extremum of the spatial twist free energy for  $N_f = 0, N_f > 1$  is

$$\Omega^{\text{spatial}} = e^{i\frac{\pi}{N}\nu} \begin{pmatrix} 1 & & & \\ & e^{i\frac{2\pi}{N}} & & \\ & & \ddots & \\ & & & e^{i\frac{2\pi(N-1)}{N}} \end{pmatrix}, \quad \nu = 0, 1 \text{ for } N = \text{odd, even.} \quad (4.29)$$

This extremum of the twist free energy is *unique*, and  $\text{Tr } \Omega$  is neutral under the  $\mathbb{Z}_N$  center symmetry  $\mathbb{Z}_N \subset \text{SU}(N)_V$ , with  $\text{Tr } \Omega^n$  vanishing for  $n \bmod N \neq 0$ . This is in sharp contrast with the thermal holonomy in eq. (4.27), for which  $\text{Tr } \Omega^n \neq 0$  for all  $n$ .

For  $N_f = 1$ , the theory has  $\mathcal{N} = (1, 1)$  supersymmetry. The fermion boundary conditions associated with  $V^{(+)}$  also respect supersymmetry. Consequently, the vanishing of the one-loop contribution actually extends to all orders in perturbation theory. At the non-perturbative level, we expect to find non-vanishing contributions to  $V^{(+)}$ , and for the background in eq. (4.29) to be a non-trivial extremum, as is the case in 4D  $\mathcal{N} = 1$  SYM [75] and the 2D  $\mathcal{N} = (2, 2)$   $\mathbb{CP}^{N-1}$  model. However, we leave an explicit verification of this for the PCM to future work.

At the unique center-symmetric extremum characterized by eq. (4.29), the twist free energy is

$$V_+(\Omega^{\text{spatial}}) = \frac{(N_f - 1)}{\pi L^2} \times \frac{\pi^2}{6}. \quad (4.30)$$

Note that it remains  $O(N^0)$  at arbitrarily small- $L$ , as opposed to the thermal case, where the associated free energy is  $O(N^2)$ , as anticipated in figure 5. Hence  $L^2\mathcal{F}/N^2 \rightarrow 0$ , and we refer to this center-symmetric small- $L$  regime as confined. We also observe that *only* the theory associated to  $V_+$  with  $\Omega$  satisfying eq. (4.29) satisfies *all* of our conditions of section 4.1 for continuity between small- $L$  and large- $L$  regimes. Hence we have found a *unique* choice of boundary conditions for the principal chiral model for which one can expect that the physics at small- $L$  should be smoothly connected to physics at large- $L$ . We conjecture that with the  $\mathbb{Z}_N$ -symmetric background holonomy, the PCM on  $\mathbb{R} \times S_L^1$  has an adiabatic small- $L$  limit. If  $L$  is small enough, the theory remains weakly coupled at long distances, as illustrated in figure 6. This is the counter-part of the adjoint-Higgs/Wilson line branch in gauge theories.

## 5 Structure of the perturbative series on small $\mathbb{R} \times S_L^1$

Now that we have devised a weak-coupling limit on  $\mathbb{R} \times S^1$  adiabatically connected to the theory on  $\mathbb{R}^2$ , we can expect long-distance observables to be calculable using semiclassical methods, with a trans-series representation of the form

$$\langle \mathcal{O}[\lambda] \rangle = \sum_{n=0}^{\infty} p_{0,n} \lambda^n + \sum_c e^{-S_c/\lambda} \sum_{n=0}^{\infty} \sum_{k=0}^n p_{c,n,k} \lambda^n [\log(\lambda^{-1})]^k, \quad (5.1)$$

where  $\lambda = \lambda(1/NL)$ . The reason for the inclusion of the  $\log \lambda^{-1}$  terms will become apparent in section 7.

To keep the discussion more focused, from here onward we will discuss the structure of the perturbative expansion and semiclassical field configurations PCM with  $N_f = 0$ , and leave generalizations to  $N_f > 0$  to future work.

As we mentioned in section 3.1 the leading singularity, along the positive real axis of the Borel plane, for the  $\mathbb{R}^2$  theory is the IR-renormalon singularity and it is located at  $1/N$  of the uniton action. On  $\mathbb{R}^2$ , the theory is strongly interacting, and there is no semi-classical interpretation of this singularity. Once we put the theory on  $\mathbb{R} \times S^1$  and go to the adiabatic small-circle limit described in section 4, the Borel singularity positions should move in a smooth way relative to their locations in the  $\mathbb{R}^2$  limit.<sup>22</sup> Since the adiabatic small- $L$  limit we constructed in section 4 is weakly coupled, however, we should be able to see renormalon ambiguities in the large-order behavior of perturbation theory in  $\lambda(1/NL)$ , and reproduce at least the factor of  $N$  in the expression above. We now demonstrate that this is indeed the case.

---

<sup>22</sup>With a thermal compactification, one can show that the renormalon singularities simply disappear, consistent with our claim that the high temperature limit is not smoothly connected to the theory on  $\mathbb{R}^2$ .

The quickest route to see the semi-classical realization of the IR renormalons involves working out the perturbative series describing the contributions to the ground state energy  $E(\lambda)$  from fluctuations around the  $P$  saddle:

$$\mathcal{E}(\lambda) = E(\lambda)\xi^{-1} = \sum_{n=0}^{\infty} p_n \lambda^n, \quad (5.2)$$

where we have chosen to write the ground state energy in the natural units of the problem, which turn out to be  $\xi = 2\pi/(NL)$ . The ground state energy will receive contributions from modes with momenta  $Q > 1/(NL)$ , as well as modes with  $Q < 1/(NL)$ .

The effective coupling constant for the UV modes with  $Q > 1/(NL)$  is  $\lambda(Q)$ , and runs logarithmically, as sketched in figure 6. Such modes will produce factorially-growing and *sign-alternating* contributions to the perturbative coefficients  $p_n$ . More precisely, the sign-alternating part of the perturbative series is

$$p_n g^{2n} \sim (-1)^n \frac{n!}{\left(\frac{8\pi}{g^2 N}\right)^n}. \quad (5.3)$$

In the phase space integration over high momenta, the dominant contribution arises from the UV scale

$$\textbf{UV-renormalon support: } k_{\text{UV}}^* \sim Q e^n \gg \Lambda \iff \ell_{\text{UV}}^* \sim \frac{1}{k_{\text{UV}}^*} \sim L N e^{-n} \ll \Lambda^{-1} \quad (5.4)$$

where  $Q$  is some large external momentum. These are associated with singularities on the negative real axis in the Borel plane, and are called the UV renormalons [3]. The UV renormalons can not be affected by compactification, because  $\ell_{\text{UV}}^* \sim L N e^{-n} \ll L N$ . In other words, UV renormalons probe only the high-momentum behavior of a theory and do not care if the space is compactified or not. So it is natural to see them appear unchanged in the compactified theory. Their presence is not related to the particular regularization used to compute the phase space integrals over high momenta. Given these considerations, it is not very surprising to find that the UV renormalons can be recovered in the latticized theory [76–78].

The contributions from modes with  $Q < 1/(NL)$  are a different and more subtle matter. As illustrated in figure 6, the relevant coupling for these modes is effectively  $\lambda(1/(NL))$ , which ceases to run at energies lower than  $1/(LN)$ . *Naively*, the absence of running would mean that the IR-renormalons should disappear. For example, if we take the PCM on  $\mathbb{R}^2$  with  $N_f > N_f^*$ , where  $N_f^*$  is the value of  $N_f$  above which the theory is conformal in the IR, the coupling stops running in the IR as well, and this was our reason for asserting in section 3.1 that the IR-renormalons must disappear when  $N_f > N_f^*$ . So why is it that the freezing of the coupling at a small value in the infrared, which leads to the absence of IR-renormalon singularities when it happens on  $\mathbb{R}^2$ , does not lead to the same effect on  $\mathbb{R} \times S^1$ ? Are we not running into a contradiction?

In fact, this is precisely at the heart of our adiabatic continuity idea. First, on  $\mathbb{R}^2$ , the dominant contribution to the renormalon singularities comes from the scale

$$\textbf{IR-renormalon support: } k_{\text{IR}}^* \sim Q e^{-n} \ll \Lambda \iff \ell_{\text{IR}}^* \sim \frac{1}{k_{\text{IR}}^*} \sim Q^{-1} e^n \gg \Lambda^{-1} \quad (5.5)$$

which is exactly the scale where the theory becomes strongly coupled and perturbation theory becomes unreliable. Studying the Borel resummation of perturbation theory, as briefly summarized in section 3.1, one finds the leading-order ambiguity  $\pm i\Lambda^2 Q^{-2} \sim \pm i e^{-\frac{8\pi}{g^2(Q)N}}$  for processes involving a large external momentum  $Q$ . Note that this is of order  $[e^{-\frac{8\pi}{g^2(Q)}}]^{1/N}$  where  $e^{-\frac{8\pi}{g^2(Q)}} \sim \Lambda^{2N} Q^{-2N}$  is the uniton amplitude at momentum scale  $Q$ .

Because of the strong coupling problem, there is no semi-classical interpretation for the IR renormalons on  $\mathbb{R}^2$ . And indeed, until very recently it was widely believed that there is *never* a semi-classical interpretation for IR renormalons. However, the more useful/refined question to pose is actually the following.

- Start with an asymptotically free theory on  $\mathbb{R}^d$  which has IR renormalon singularities. Compactify on  $\mathbb{R}^{d-1} \times S^1$  and dial radius of  $S^1$  to a small value. What happens to the renormalon singularities if we are able to work with a compactification which adiabatically connects the theory on  $\mathbb{R}^d$  to a regime where the long distance dynamics becomes weakly coupled?<sup>23</sup>

Below, we show that in the adiabatic small- $L$  limit of the PCM, we can describe the long-distance physics using a small- $L$  effective field theory, which is a quantum mechanical system. This theory has ordinary 1d instantons whose actions are  $\frac{1}{N}$  that of the unitons on  $\mathbb{R}^2$ . We refer to these 1d instantons as fractons, since they are the fractionalized constituents of the uniton saddle from a 2d point of view.

We then show that the IR renormalon singularities on  $\mathbb{R}^2$  which arise from phase space integration and which are located at  $\sim \frac{S_{\text{uniton}}}{N}$  transmute into semi-classical correlated fracton-anti-fracton saddles on small  $\mathbb{R} \times S_L^1$  which are also located at  $\sim \frac{S_{\text{uniton}}}{N}$ . The crucial factor here is the appearance of  $\frac{1}{N}$ , which is the characteristic of renormalons, but the exact location of the singularities will generally change mildly between the semi-classical and non-semi-classical regimes. This is what is guaranteed to hold via adiabaticity. So we have reached the same conclusion in PCM as was found in deformed YM theory, QCD(adj), and the  $\mathbb{CP}^{N-1}$  model [16–18], but in a theory which has no topologically-stable stable points!

## 5.1 Covering $SU(N)$ with $SU(2)$ s

Before discussing the construction of the leading term in the small- $L$  effective field theory, we make some remarks on the organization of perturbation theory more generally. From this point onward, our discussion will only focus on the purely bosonic model 3.1; further comments on the fermionic extension of the PCM can be found in section 9.

The most common way to set up a perturbative calculation would be to write  $U = e^{iW}$ ,  $W \in \mathfrak{su}(N)$ , expand the action in powers of  $W$ , and then do perturbation theory in terms of  $W$ . Indeed, this was how we organized our calculation in section 4. From here onward, however, we find it advantageous to organize our perturbative calculations in a dif-

<sup>23</sup>It is conceivable that there may be alternative setups to compactification in which an analogous question can be posed.



ferent way. We will examine the contributions to perturbative observables from field fluctuations within  $SU(2)$  subgroups of  $SU(N)$ , and then sum over the various  $SU(2)$  subgroups.<sup>24</sup>

Consider the embedding of  $SU(2)$  into  $SU(N)$  as two by two diagonal blocks:

$$U_j^{(1)} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \begin{pmatrix} U_{j,j} & U_{j,j+1} \\ U_{j+1,j} & U_{j+1,j+1} \end{pmatrix} & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \begin{pmatrix} z_1 & iz_2 \\ i\bar{z}_2 & \bar{z}_1 \end{pmatrix} & \\ & & & \ddots \\ & & & & 1 \end{pmatrix}, \quad (5.6)$$

and

$$U_N^{(1)} = \begin{pmatrix} U_{1,1} & & U_{1,N} \\ & 1 & \\ & & \ddots \\ & & & 1 \\ U_{N,1} & & & U_{N,N} \end{pmatrix} = \begin{pmatrix} \bar{z}_1 & & & i\bar{z}_2 \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ iz_2 & & & z_1 \end{pmatrix}, \quad (5.7)$$

where  $z_1, z_2 \in \mathbb{C}$ , and  $|z_1|^2 + |z_2|^2 = 1$ .

One can associate the affine root system of the associated  $\mathfrak{su}(N)$  Lie algebra with each of these embeddings,  $U_j^{(1)}, j = 1, \dots, N$ . Each root gets associated with the corresponding  $SU(2)$  subgroup. The affine root system consists of the  $N-1$  simple roots  $\alpha_i, i = 1, \dots, N-1$  along with the affine root  $\alpha_N$ :

$$\alpha_j = \begin{pmatrix} 0_1 & 0_2 & \dots & 1_j & -1_{j+1} & \dots & 0_N \end{pmatrix}, \quad \alpha_N = \begin{pmatrix} -1_1 & 0_2 & \dots & 0 & 0 & \dots & 1_N \end{pmatrix}. \quad (5.8)$$

While the  $N$  distinguished  $SU(2)$  subgroups above will be the most important for our analysis, note that the generic  $SU(2)$  subgroup looks like

$$U_j^{(k)} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \begin{pmatrix} U_{j,j} & & U_{j,j+k} \\ & 1 & \\ & & \ddots & \\ & & & 1 \\ U_{j+k,j} & & & U_{j+k,j+k} \end{pmatrix} & \\ & & & \ddots \\ & & & & 1 \end{pmatrix} = \begin{pmatrix} 1 & & & \\ & \ddots & & \\ & & \begin{pmatrix} z_1 & & iz_2 \\ & 1 & \\ & & \ddots & \\ & & & 1 \\ i\bar{z}_2 & & & \bar{z}_1 \end{pmatrix} & \\ & & & \ddots \\ & & & & 1 \end{pmatrix}. \quad (5.9)$$

Each one of these  $SU(2)$  subgroups can be associated with sums of roots in the affine Lie algebra which are themselves roots.

---

<sup>24</sup>This is the process by which gauge configurations are generated in lattice Monte Carlo simulations for  $SU(N)$  Yang-Mills theory [79], and is known to cover the whole  $SU(N)$  manifold.

To write down the actions describing the fluctuations within the  $U_i^{(k)}$   $SU(2)$  sub manifolds, we need to choose a parametrization of  $SU(2) \simeq S^3$ . We use the Hopf coordinate parametrization:

$$U \equiv \begin{pmatrix} z_1 & iz_2 \\ i\bar{z}_2 & \bar{z}_1 \end{pmatrix} \quad \text{where} \quad \begin{aligned} z_1 &= r_1 e^{i\phi_1} = \cos \theta e^{i\phi_1}, \\ z_2 &= r_2 e^{i\phi_2} = \sin \theta e^{i\phi_2}, \end{aligned} \quad (5.10)$$

and we take the angular variables to have the ranges  $\theta \in [0, \pi]$ ,  $\phi_1 \in [0, \pi]$ ,  $\phi_2 \in [0, 2\pi]$ . For fixed  $\theta$ , the variables  $\phi_1, \phi_2$  parametrize a torus. At the degeneration points,  $\theta = 0, \pi/2, \pi$ , the torus shrinks to a circle. With this parametrization, for each one of our  $SU(2)$  embeddings, using the bosonic part of eq. (4.5), we find that the fluctuations are described by

$$S_i^{(k)} = \frac{1}{g^2} \int_{\mathbb{R} \times S^1} dt dx \left[ (\partial_\mu \theta)^2 + \cos^2 \theta (\partial_\mu \phi_1)^2 + \sin^2 \theta (\partial_\mu \phi_2 + \xi_{i,k} \delta_{\mu,x})^2 \right], \quad (5.11)$$

where

$$\xi_{i,k} = \frac{2\pi\mu_{i,i+k}}{L} = \frac{2\pi(\mu_{i+k} - \mu_i)}{L}. \quad (5.12)$$

For the  $\mathbb{Z}_N$ -symmetric background,  $\xi_{i,k} = 2\pi k/(NL)$  for all  $i = 1, \dots, N$  (using affine roots when needed).<sup>25</sup>

## 5.2 Large order behavior and Stokes phenomenon

To efficiently deduce the large-order behavior of the perturbative contributions to  $\mathcal{E}$  when  $NL\Lambda/2\pi \ll 1$ , we will use a small- $L$  effective field theory. The IR properties of the theory for small enough  $L$  can be calculated from a theory where one integrates out modes with energies larger than the scale  $(NL)^{-1}$ , up to a subtlety which we comment on below. Therefore, the effective field theory is *forgetful* of UV-renormalon singularities, which can only be extracted from the microscopic theory. The leading term of the action of the small- $L$  effective field theory is

$$S = \frac{L}{2g^2} \int dt \text{Tr} \left( \partial_t U \partial_t U^\dagger + [H_V, U][H_V, U^\dagger] \right) \quad (5.13)$$

$$\rightarrow \frac{L}{g^2} \int dt \left[ \dot{\theta}^2 + \cos^2 \theta \dot{\phi}_1^2 + \sin^2 \theta \dot{\phi}_2^2 + \xi_{i,k}^2 \sin^2 \theta \right], \quad (5.14)$$

where  $\xi_{i,k}$  is the same as eq. (5.12). The expression in the second line comes from a restriction to a generic  $SU(2)$  subgroup of  $SU(N)$ , given in eq. (5.9). This leading term of the small- $L$  EFT is a 0 + 1 dimensional field theory, which is just quantum mechanics.

An important subtlety with the derivation of eq. (5.14) is that it is not quite true that all KK-momentum carrying states decouple at low energies. The issue is that states that carry non-zero *winding number* can contribute to the low-energy dynamics on the same

<sup>25</sup>This action is identical to the one of the  $\mathbb{CP}^1$  model given in [18], except for an overall factor of 2, if one forgets about the  $\phi_1$  coordinate and sets  $k = 1$ . This follows from viewing  $S^3$  through the lens of the Hopf fibration  $S^1 \rightarrow S^3 \rightarrow S^2 \sim \mathbb{CP}^1$  where  $S^3$  is the total space and  $S^1$  is fibered over  $S^2$ .

footing as states that carry zero winding number. To see this, note that if we look at the contribution from  $U_i^{(N-1)}$  and allow  $\phi_2$  to carry  $-1$  units of winding number, we find

$$S = \frac{L}{g^2} \int dt \left[ \dot{\theta}^2 + \cos^2 \theta \dot{\phi}_1^2 + \sin^2 \theta \dot{\phi}_2^2 + \left( \frac{-2\pi}{L} + \xi(N-1) \right)^2 \sin^2 \theta \right] \quad (5.15)$$

$$= \frac{L}{g^2} \int dt \left[ \dot{\theta}^2 + \cos^2 \theta \dot{\phi}_1^2 + \sin^2 \theta \dot{\phi}_2^2 + \xi^2 \sin^2 \theta \right], \quad (5.16)$$

where  $\xi = 2\pi/NL$ . This contribution is in fact associated precisely with the affine root of the  $\mathfrak{su}(N)$  Lie algebra, and its relevance is due to the compactness of the  $\mathbb{Z}_N$ -symmetric background holonomy. So there are contributions to the low-energy physics from some configurations with winding number  $-1$  which are of the same magnitude as contributions from field configurations that carry winding number  $0$ . This is an illustration of the fact that the adiabatically-compactified theory “remembers” its two-dimensional nature even when  $L$  is very small. This subtlety will also be important in the analysis of non-perturbative saddle points in the next section.

To understand the high-order behavior of perturbation theory in  $g^2[1/NL]$  using the leading order part of the EFT action — that is, leading order in an  $NL\Lambda/2\pi \ll 1$  expansion — it turns out to be useful to temporarily move back to Minkowski space and use Hamiltonian methods, because it lets us use some known results from the literature. At small- $L$  the Hamiltonian associated with eq. (5.14) is

$$H = \frac{g^2}{4L} P_\theta^2 + \frac{L\xi_k^2}{g^2} \sin^2 \theta + \frac{g^2}{4L \sin^2 \theta} P_{\phi_1}^2 + \frac{g^2}{4L \cos^2 \theta} P_{\phi_2}^2. \quad (5.17)$$

We emphasize that this Hamiltonian describes the contributions to the energies of the states from the  $N$   $SU(2)$  subgroups  $U_j^{(1)}$ ,  $j = 1, \dots, N$ . In general, the way we have organized perturbation theory, we must sum over the contributions from all  $SU(2)$  subgroups to compute the energies of states in the full  $SU(N)$  PCM. Fortunately, we will see that the  $U_j^{(1)}$  subgroups make the dominant contribution to e.g. the ground state energy at small- $L$ , which simplifies our analysis. Also, to avoid confusion, we note that with the center-symmetric background gauge fields turned on, each of the  $U_j^{(1)}$  subgroups will make the same contribution to  $\mathcal{E}$ . If the background fields were to be slightly perturbed away from the  $\mathbb{Z}_N$  symmetric points this would no longer be the case; the contributions are in principle distinguishable.

Having made the point about the importance of taking into account the various  $SU(2)$  subgroups, we present the subsequent calculations with  $\xi \equiv \xi_1 = 2\pi/(NL)$  to lessen the notational clutter. The dependence on  $\xi_k$  always follows from the replacement  $\xi \rightarrow \xi_k$ .

The potential for  $\theta$  has minima at  $\theta = 0, \pi$ , and the action has a discrete symmetry  $P$  which acts as  $P : \theta \rightarrow \pi - \theta$ . The states of the system will have well-defined eigenvalues  $\pm 1$  under  $P$ . Moreover, since  $\phi_1, \phi_2$  are cyclic compact coordinates, associated to quantized conserved charges, a state with the quantum numbers of  $(\pm, n_1, n_2)$  will have an energy of order

$$E_{\pm, n_1, n_2} \sim g^2 \xi (n_1^2 + n_2^2) \sim \frac{g^2}{NL} (n_1^2 + n_2^2). \quad (5.18)$$

We expect the ground state and first excited states to have the quantum numbers  $(+, 0, 0)$ , and  $(-, 0, 0)$ , with an excitation energy non-perturbatively small compared to the scale  $1/NL$ .

To compute the energy of the ground state as a function of  $g^2$ , we use the Born-Oppenheimer (BO) approximation familiar from atomic physics. The Born-Oppenheimer approximation states that the lowest excitation energies of a coupled quantum system can be extracted by solving the Schrödinger equation for the ‘light’ degrees of freedom (in our case, the dynamics of  $\theta$  field) with the ‘heavy’ degrees of freedom (the dynamics of the  $\phi_1, \phi_2$  fields) frozen to their energy-minimizing values. Essentially, one is exploiting a separation in energy scales between the two sets of degrees of freedom to solve for the behavior of the heavy variables classically, and then treating the light variables quantum-mechanically. In our case, we will see that the separation in energy scales between states with  $n_1, n_2 \neq 0$  and  $n_1 = n_2 = 0$  is exponential in  $g^2$ , so the use of the Born-Oppenheimer is parametrically well-justified when  $g^2 \ll 1$ , as is the case when  $NL$  is small compared to  $\Lambda^{-1}$ . So in the Bohr-Oppenheimer approximation, we only need to compute the ground state energy in the  $n_1 = n_2 = 0$  sector of the theory, described by the Hamiltonian

$$H = \frac{g^2}{4L} P_\theta^2 + \frac{L\xi^2}{g^2} \sin^2 \theta. \quad (5.19)$$

The Schrödinger equation can be written as

$$\left[ \frac{d^2}{d\theta^2} + p + \frac{2\xi^2}{g^2} \cos(2g\theta) \right] \psi = 0, \quad p = 4E - \frac{2\xi^2}{g^2}, \quad \theta \in [0, \pi], \quad (5.20)$$

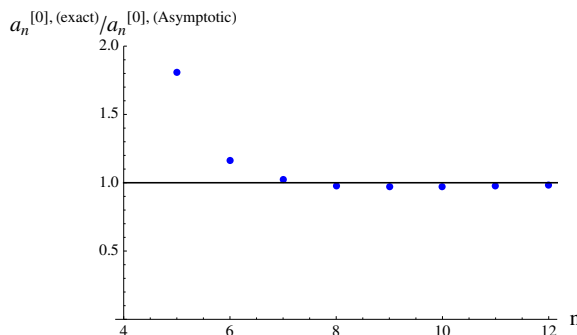
where we have set  $L = 1$ . It is convenient to define a dimensionless Hamiltonian  $\tilde{H} = H\xi^{-1}$ , and bring the kinetic term into canonical form by the change of variables  $\theta = \sqrt{\frac{g^2}{2\xi}} \tilde{\theta}$ . Then the Hamiltonian reads

$$\tilde{H} = -\frac{1}{2} \frac{d^2}{d\tilde{\theta}^2} + \frac{\xi}{2g^2} \cos \left( \sqrt{\frac{2g^2}{\xi}} \tilde{\theta} \right). \quad (5.21)$$

This form is same as the Hamiltonian in Zinn-Justin’s text, section 41.2 [80] with the identification  $g_{ZJ} = \frac{2g^2}{\xi}$ .

The energy of the ground state now follows from the  $P = +$  solution, and the large-order behavior of this solution was already determined by Stone and Reeve [81] using the methods developed by Bender and Wu [82, 83]. From [81] we see that the contributions of each of the  $U_{(i)}^1$  in the  $SU(2)$  subgroups to the perturbative series for the ground state energy,  $\mathcal{E}_i$ , behave as

$$\begin{aligned} \mathcal{E}_i(g) &= E_i(g)\xi^{-1} \\ &= \frac{1}{2} - \frac{1}{2} \left( \frac{g^2}{8\xi} \right) - \frac{1}{2} \left( \frac{g^2}{8\xi} \right)^2 - \frac{3}{2} \left( \frac{g^2}{8\xi} \right)^3 - \frac{53}{8} \left( \frac{g^2}{8\xi} \right)^4 - \frac{297}{8} \left( \frac{g^2}{8\xi} \right)^5 - \frac{3961}{16} \left( \frac{g^2}{8\xi} \right)^6 - \frac{60727}{32} \left( \frac{g^2}{8\xi} \right)^7 \\ &\quad - \frac{2095501}{128} \left( \frac{g^2}{8\xi} \right)^8 - \frac{20057205}{128} \left( \frac{g^2}{8\xi} \right)^9 - \frac{421644859}{256} \left( \frac{g^2}{8\xi} \right)^{10} - \frac{4835954237}{256} \left( \frac{g^2}{8\xi} \right)^{11} + \dots \end{aligned} \quad (5.22)$$



**Figure 7.** The ratio of the exact value of the pre-factors in perturbation theory to the asymptotic formula,  $\frac{a_n^{[0], \text{exact}}}{a_n^{[0], \text{asymptotic}}}$ .

with an asymptotic form given by

$$\mathcal{E}_i(g) = \sum_{n=0}^{\infty} a_n^{[0]} g^{2n}, \quad a_n^{[0]} \sim -\frac{2}{\pi} \left( \frac{1}{8\xi} \right)^n n! \left( 1 - \frac{5}{2n} - \frac{13}{8n^2} + \mathcal{O}(n^{-3}) \right). \quad (5.23)$$

We have confirmed the correctness of the asymptotic form (which includes sub-leading corrections to the leading asymptotic) by numerical analysis. See figure 7 for an illustration of the rapid convergence of the series coefficients to their asymptotic large-order form, truncated at order  $\mathcal{O}(n^{-3})$ , given by  $a_n^{[0], \text{asymptotic}} = \frac{2}{\pi} \left( \frac{1}{8\xi} \right)^n n! \left( 1 - \frac{5}{2n} - \frac{13}{8n^2} \right)$ .

The expansion parameter of perturbation theory is  $g^2/(8\xi) = g^2 N/(16\pi)$ . The perturbative series is non-alternating and has factorially-growing coefficients, of the form  $n! [g^2/(8\xi)]^n$ . This means that the perturbative expansion is along a Stokes line, and its resummation is ambiguous. The analytic continuation of the Borel transform for the leading order  $n!$  divergence is given by

$$B\mathcal{E}_i(t) = -\frac{2}{\pi} \frac{1}{1 - \frac{t}{8\xi}}. \quad (5.24)$$

Hence the perturbative series will not be Borel summable along  $\mathbb{R}^+$  in Borel plane, due to the (leading) singularity located at  $t_* = 8\xi = \frac{16\pi}{N}$ . If the integration contour defining the Borel-resummation of the series is deformed to pass either above or below the real axis near this singularity, so that it is avoided, one can obtain a finite result. However, the result of the resummation will depend on the choice of contour, leading to a two-fold ambiguity. An equivalent way to think about this, which we find particularly useful in working with the NP saddles, is to analytically continue  $g^2 \rightarrow g^2 e^{i\epsilon}$ ,  $\epsilon \in (-\pi, \pi)$ . Then so long as  $\epsilon \neq 0$ , the Borel sum converges, but the analytic continuation back to  $\epsilon = 0$  depends on whether  $\epsilon$  approaches 0 from above the real axis or below the real axis. This matches what one sees with the contour deformations.

Using either of these equivalent approaches we can define the lateral eq. (2.18) Borel resummation as in the simple example discussed in section 2.1. In particular, the right  $\mathcal{S}_{0+}\mathcal{E}_i$  and left  $\mathcal{S}_{0-}\mathcal{E}_i$  Borel resummations yield

$$\mathcal{S}_{0\pm}\mathcal{E}_i = \frac{1}{g^2} \int_{C^\pm} dt e^{-t/g^2} B\mathcal{E}_i(t) = P \frac{1}{g^2} \int dt e^{-t/g^2} B\mathcal{E}_i(t) \mp i \frac{16\xi}{g^2} e^{-\frac{8\xi}{g^2}}$$

$$= \text{Re } \mathcal{S}_0 \mathcal{E}_i \mp i \frac{32\pi}{g^2 N} e^{-\frac{16\pi}{g^2 N}}, \quad (5.25)$$

where  $P$  stands for the (unique) Cauchy principal value.

**Stokes jumps.** We can now determine the action of the Stokes automorphism:

$$(\mathcal{S}_{0+} - \mathcal{S}_{0-})\mathcal{E} = -i \frac{32\xi}{g^2} e^{-\frac{8\xi}{g^2}} = -i \frac{64\pi}{g^2 N} e^{-\frac{16\pi}{g^2 N}}. \quad (5.26)$$

Note that this expression is valid for each of the individual  $\text{SU}(2)$ 's embedded into  $\text{SU}(N)$  as in eq. (5.6). There are  $N$  such embeddings associated with the  $N$  roots of the affine Lie algebra in a center-symmetric background for which  $\xi = \mu_{i+1} - \mu_i = \frac{2\pi}{N}$ . So there are  $N$  leading-order ambiguities in the Borel resummation of perturbative series describing the fluctuations within the  $U_i^{(1)}$  subgroups. These  $N$  ambiguities are identical when the background holonomy is  $\mathbb{Z}_N$ -symmetric, so that the overall leading ambiguity in the  $\text{SU}(N)$  theory is just  $N$  times the expression above.<sup>26</sup>

Of course, there are also ambiguities coming from other embeddings of  $\text{SU}(2)$  into  $\text{SU}(N)$  such as the one shown in eq. (5.9) which are associated with roots which can be written as positive linear combination of the simple roots. In this case, with the  $U_j^{(k)}$  embeddings in the center-symmetric background, the leading non-perturbative ambiguity is of order

$$\mp i \frac{16\xi k}{g^2} e^{-\frac{8\xi k}{g^2}}, \quad k \geq 2. \quad (5.27)$$

As advertised, this is exponentially small compared to the leading ambiguities coming from fluctuations living within  $U_i^{(1)}$ .

### 5.3 NP-data from the late terms of the P-expansion

Energy eigenstates in a real periodic potential must be real. Yet we have just seen that the perturbative series generates a result which has an imaginary part upon Borel resummation. Furthermore, as seen in eq. (5.25), the imaginary part is ambiguous. A non-perturbative completion of this result which gets rid of both of these problems is the resurgent trans-series. See [15] for a very explicit discussion. In fact, the NP-completion of the perturbative results, even in the case of simplest single parameter trans-series, leads to infinitely many NP-saddles.

If we assume that the semiclassical representation of observables in the PCM is resurgent, the appearance of the ambiguity in perturbation theory described above implies that there must exist some finite-action field configuration with action  $S = \frac{16\pi}{Ng^2}$  whose amplitude should be ambiguous in just the right way to cancel this leading ambiguity. Another way to say this is that resurgence theory *predicts* the existence of such non-perturbative saddle-points. However, as we have mentioned before, naive topological considerations seem to leave no room for stable finite-action solutions in the PCM. In the next section, we resolve this puzzle by showing the existence of such saddles, which in general are *not* classified according to the topological structure of the microscopic theory.

---

<sup>26</sup>The ambiguities would become split if the  $\mathbb{Z}_N$  symmetry were to be slightly broken.

In the low energy effective field theory for the PCM on small  $\mathbb{R} \times S^1$ , there is a sense in which topology is *emergent*. Indeed, the periodic potential eq. (5.19) has one-dimensional instantons, which can be classified according to topology in quantum mechanics. However, in QM and QFT, it is well-known that instantons cannot mix with perturbation theory. The first NP-saddles which can mix with perturbation theory and fix its ambiguity are of the instanton-anti-instanton form. Indeed, we should expect a left/right  $[\mathcal{F}\bar{\mathcal{F}}]_{\pm} = \text{Re}[\mathcal{F}\bar{\mathcal{F}}] \pm ie^{-\frac{16\pi}{Ng^2}}$  correlated amplitude for some event with amplitude  $\mathcal{F}$ : this is the only way the leading ambiguities of perturbation theory can be cancelled in the trans-series. So resurgence theory and perturbation theory in the semi-classically calculable regime of the 2d PCM compactified down to  $\mathbb{R} \times S^1$  predicts that there must exist a NP saddle with amplitude

$$\mathcal{F} \sim e^{-\frac{4\xi}{g^2}} \sim e^{-\frac{8\pi}{g^2 N}}. \quad (5.28)$$

Note that the action of this NP saddle in the compactified theory is  $1/N$  of the uniton action. Explicit field configurations with such actions are discussed in detail in the following section.

## 6 Non-perturbative saddle points

We now begin our exploration of the non-perturbative features of the PCM at small  $S^1$ , with the boundary conditions/background gauge field corresponding to eq. (4.29). As we have mentioned in the introduction, there are no stable topological defects in the theory on  $\mathbb{R}^2$ . However, on  $\mathbb{R} \times S^1$  with a  $\mathbb{Z}_N$ -symmetric background, the situation is different. The background holonomy amounts to a potential on the  $SU(N)$  target space, effectively modifying it to  $U(1)^{N-1}$ , the maximal torus of  $SU(N)$ , at energies low compared to  $1/(NL)$ . This is the counterpart of adjoint Higgsing in gauge theories via a Wilson line, see e.g. [29, 75].

As we briefly mentioned in section 5.2, this potential has isolated minima on the group manifold. Consequently, there are tunneling events between these minima, which should admit a semiclassical Euclidean description as *stable* instanton-like field configurations when  $L$  is small enough and the theory is weakly coupled. This happens despite the fact that the microscopic theory has a trivial  $\pi_2$  homotopy group and hence has no instantons. This is the sense in which there is an emergent topological classification in the low energy effective field theory.

We will give a careful classification of these new saddle points of the PCM path integral. Moreover, we will see that these small- $L$  field configurations are the constituents of unitons, with the minimal uniton fractionalizing into  $N$  constituents at small- $L$ . For this reason, we refer to these small- $L$  stable saddle points as *fractons*. This terminology was originally introduced in [23] in the multi-flavor Schwinger model for fractionalized instantons. The fractons play a critical role in the non-perturbative physics of the PCM on small  $S^1$ , and are responsible for the mass gap of the theory. Moreover, they lead us to the resolution of the deep puzzle about the interpretation of renormalon ambiguities in the PCM.



## 6.1 Unitons

We start with a description of the unitons, which are finite-action solutions of the second-order Euclidean equations of motion [56]. The reason that one should expect such solutions is that the  $\mathbb{CP}^{N-1}$  manifold can be embedded as a totally geodesic submanifold into  $SU(N)$  [84]. This means that solutions of the  $\mathbb{CP}^{N-1}$  field equations can be lifted to solutions of the  $SU(N)$  field equations. However, while finite-action solutions on  $\mathbb{CP}^{N-1}$  carry a topological charge, and have no negative modes while staying within the  $\mathbb{CP}^{N-1}$  manifold, once the solution can evolve in the full  $SU(N)$  target space, there is no longer any conserved topological charge. Hence unitons are known to be “unstable”, which just means that the fluctuation operator around a uniton has negative modes [85].

It turns out that in general, not all of the solutions of the PCM are obtained directly from the  $\mathbb{CP}^{N-1}$  embedding, and in [56] Uhlenbeck gave an exhaustive classifications of all possible solutions to the PCM equation of motion. However, our interest here will be in what happens to the minimal-action uniton, which *can* be obtained from a  $\mathbb{CP}^{N-1}$  embedding, and has the action

$$S_{\text{uniton}} = \frac{8\pi}{g^2} = N \times \frac{8\pi}{\lambda}. \quad (6.1)$$

This minimal-action uniton solution on  $\mathbb{R} \times S^1$  with the boundary conditions determined by eq. (4.29) is associated with the minimal-action instanton of  $\mathbb{CP}^{N-1}$ , and takes the form

$$U_{\text{uniton}} = e^{i\pi/N} (1 - 2\mathbb{P}), \quad \mathbb{P}_{ij} = \frac{v_i v_j^\dagger}{v^\dagger \cdot v}, \quad (6.2)$$

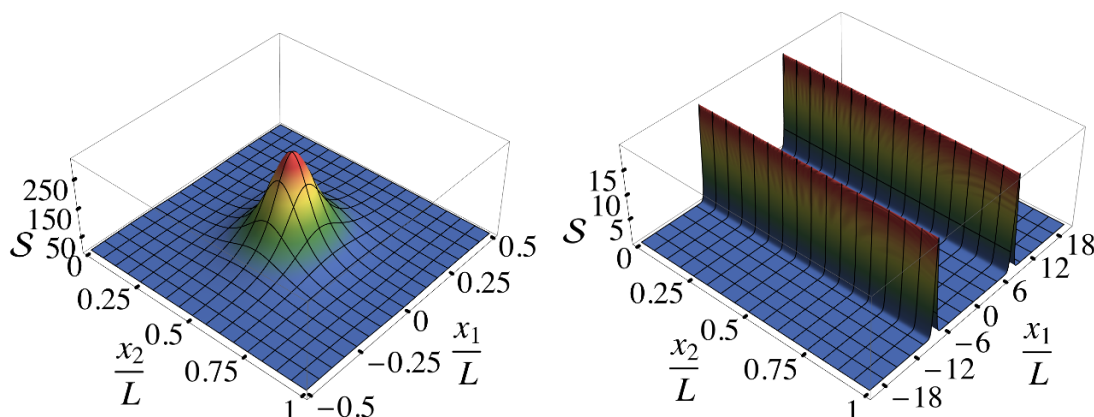
where  $\mathbb{P}$  is a projector,  $\mathbb{P}^2 = \mathbb{P}$ . Here, we defined  $v = \Omega(z) \cdot v_I$

$$v_I = \begin{pmatrix} 1 \\ \lambda_N \\ \lambda_{N-1} \\ \vdots \\ \lambda_3 \\ \lambda_2 + e^{2\pi z} \lambda_1 \end{pmatrix} \quad (6.3)$$

and

$$\Omega(z) = e^{i\frac{\pi}{N}\nu} \begin{pmatrix} 1 & & & \\ & e^{\frac{2\pi}{N}z} & & \\ & & \ddots & \\ & & & e^{\frac{2\pi(N-1)z}{N}} \end{pmatrix}, \quad \nu = 0, 1 \text{ for } N = \text{odd, even}, \quad (6.4)$$

and  $z = (x_1 + ix_2)/L$ , with  $x_{1,2}$  coordinates on Euclidean  $\mathbb{R} \times S^1$ , while  $\lambda_i \in \mathbb{C}$  are moduli of the solution. We note that  $v(z)$  are precisely the twisted instanton configurations discussed in the context of the  $\mathbb{CP}^{N-1}$  model in [26, 27] and [18, 19], with the correct twisted periodicity in  $x_2$ .



**Figure 8.** Action densities for the SU(2) unitons: left: small unition. Right: large unition fractionalized to two fractons. We used the following values for the moduli:  $\lambda_1 = \lambda_2 = 1$  (left) and  $\lambda_1 = \lambda_2 = e^{-25}$  (right).

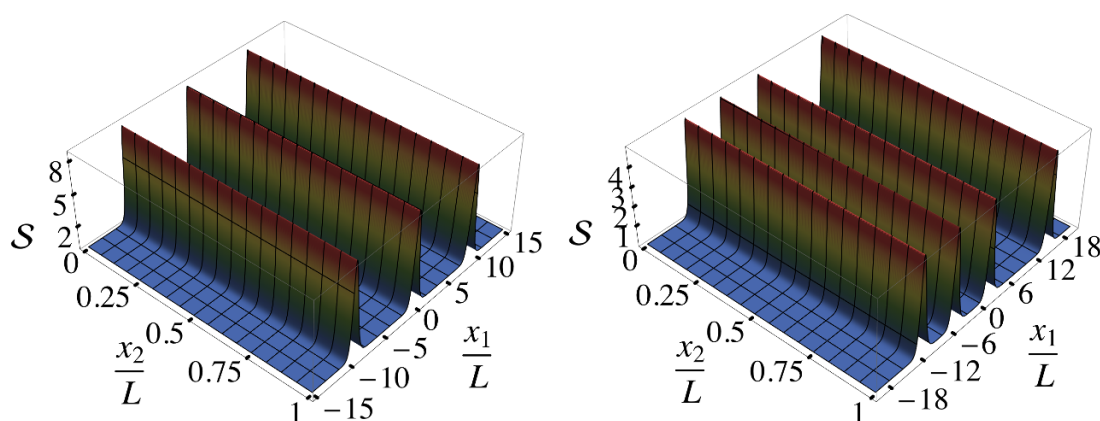
Once we work on  $\mathbb{R} \times S^1$  with a center-symmetric background holonomy, we expect the emergence of some stable instanton-like field configurations, the fractons. The fastest way to see the emergence of the fractons is to look at some plots of the action densities associated to these solutions as a function of the moduli  $\lambda_i$  shown above.<sup>27</sup> It turns out that ‘small’ unitons, which is to say ones whose characteristic size is small compared to  $L$ , resemble the profile of a unition on  $\mathbb{R}^2$ . An example of such a small-union configuration is shown on the left in figure 8, where the minimal unition looks like a single lump of Euclidean action density,  $\mathcal{L}_E$ , centered near  $x_1 = 0, x_2 = L/2$ . On the other hand, unitons which are large compared to  $L$  tend to fractionalize, in the background eq. (4.29), into multiple lumps, the fractons. The locations of the different fractons are controlled by the  $\lambda_i$ , with an example shown on the right of figure 8. The number of lumps is at most  $N$ .<sup>28</sup> More examples of fractionalization are shown for the SU(3) and SU(4) cases in figure 9 on the left and right respectively. As we change the moduli  $\lambda_i$ , it is possible to see in figure 10 how the fractons merge together, returning to the single lump configuration for the small unition shown on the left of figure 8.

From the plots we see that in the fractionalized limit, the lumps are constant along the compact direction  $x_2$ , and seem to approach a vacuum configuration in between each other. This strongly suggests that the PCM should have isolated fracton solutions<sup>29</sup> for

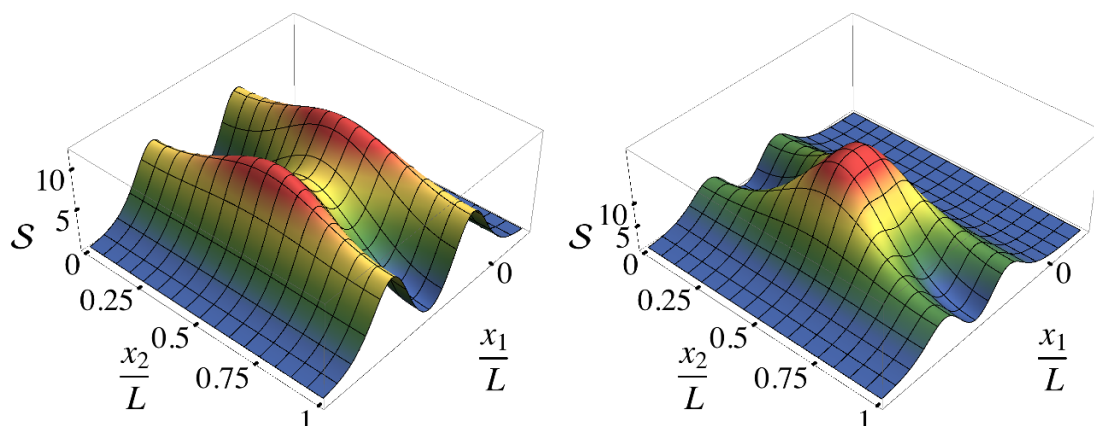
<sup>27</sup>The unition moduli  $\lambda_i$  that we show in the text are inherited from the moduli of the  $\mathbb{CP}^{N-1}$  instanton, but they are not the only moduli of the unition solution, since there are also moduli associated with the embedding of  $\mathbb{CP}^{N-1}$  into  $SU(N)$ . The full structure of the zero and quasi-zero modes of a unition is not known.

<sup>28</sup>These type of plots, in the context of fractionalization of BPST instantons in the background of non-trivial holonomy, goes back to the work of van Baal et al. [25] in QCD-like theories. For other works emphasizing similar effects in various QFTs, see e.g. [23, 24, 26–28, 86–90]. The amusing feature illustrated by the PCM is that the idea of fractionalization works even in the absence of topologically stable instantons! A more direct route to study the constituents of instanton appears in [21, 29, 75] by studying the theory in a weakly coupled calculable regime, see e.g. [91–93].

<sup>29</sup>In [94] it was shown that a direct extrapolation of classical field configurations from small- $L$  to large- $L$ , using only the leading term in the small- $L$  EFT action, is quite subtle for non-BPS solutions. On the other hand, we already know that the PCM, in the decompactified  $\mathbb{R}^2$  regime, becomes strongly coupled and one cannot recover the full quantum theory starting from classical field configurations. Indeed, even before thinking about the strongly-coupled  $L \rightarrow \infty$  limit, one can see that the small- $L$  EFT action itself gets corrections as  $L$  is increased. In this paper, we confine our attention to the weakly-coupled small- $L$  limit, in which the subtleties explored in [94] are not relevant.



**Figure 9.** Action densities for the large SU(3) and SU(4) unitons. They split to three and four fractons, respectively. We used the following values for the moduli: SU(3) uniton  $\lambda_1 = \lambda_2 = e^{-19}, \lambda_3 = 0$  (left), SU(4) uniton with  $\lambda_1 = \lambda_2 = e^{-18}, \lambda_3 = e^8, \lambda_4 = e^5$  (right).



**Figure 10.** Action densities for intermediate size SU(2) unitons. The fractons merge together and return to the single lump uniton configuration. We used the following values for the moduli:  $\lambda_1 = e^1, \lambda_2 = e^{-4}$  (left), and  $\lambda_1 = e^2, \lambda_2 = e^{-4}$  (right).

small- $L$ , and this will be explicitly verified in the next section.

We also note that in the thermal compactification, at high temperatures, where the trivial holonomy of eq. (4.27) must be used, the fractionalization of unitons does not occur.

## 6.2 SU(2) fractons and KK fractons

It is instructive to start our search for the fracton solutions by considering the SU(2) PCM, in which case the group manifold is  $S^3$ . The discussion of the fractons becomes most transparent if we parametrize the group element using Hopf coordinates as we did in section 5, see eq. (5.10). The round metric on the 3-sphere in the Hopf coordinates is given by

$$ds^2 = d\theta^2 + \cos^2 \theta d\phi_1^2 + \sin^2 \theta d\phi_2^2, \quad (6.5)$$

and this fixes the kinetic term. From the action eq. (4.5), we see that a non-trivial potential term is induced on the  $S^3$  manifold due to the background field determined by eq. (4.29). The associated Euclidean Lagrangian on  $\mathbb{R} \times S^1$  is given by

$$S = \frac{1}{g^2} \int_{\mathbb{R} \times S^1} dx_1 dx_2 \left[ (\partial_\mu \theta)^2 + \cos^2 \theta (\partial_\mu \phi_1)^2 + \sin^2 \theta (\partial_\mu \phi_2 + \xi \delta_{\mu, x_2})^2 \right], \quad (6.6)$$

where  $\xi = 2\pi/NL = \pi/L$ .

Given the fact that unitons which are large compared to  $L$  split into  $N$  configurations which appear to have flat profiles along the  $S^1$ , it is tempting to start by setting all Kaluza-Klein (KK) momenta in eq. (6.6) to zero, and thus obtain an effective one-dimensional theory. The usual reasoning behind such an approach is that field configurations carrying  $n$  units of KK momentum have an energy density  $n^2/L^2$ , so for studying physics on length scales large compared to  $L$  it is sufficient to focus on states with  $n = 0$ , since these are parametrically lighter. This would certainly be correct if  $\xi$  were set to zero. However, as we already saw in the perturbative context, this approach is too naive for the PCM in the  $\mathbb{Z}_N$ -symmetric small- $L$  limit, since the small- $L$  theory actually ‘remembers’ that it is microscopically a *two-dimensional* theory, and neglecting all configurations with non-zero winding number (and hence KK momentum) is not quite correct. It will turn out to be important to consider finite-action field configurations also carrying non-zero winding number for  $\phi_2$ .

**SU(2) fractons.** Let us start by looking for instanton configurations carrying *zero* units of KK momentum in the compact direction  $x_2$  for  $\theta, \phi_1$  and  $\phi_2$ . By setting  $\partial_{x_2} = 0$ , the Euclidean action for the low-lying modes reduces to

$$S = \frac{L}{g^2} \int_{-\infty}^{+\infty} dx_1 \left[ \dot{\theta}^2 + \cos^2 \theta \dot{\phi}_1^2 + \sin^2 \theta \dot{\phi}_2^2 + \xi^2 \sin^2 \theta \right], \quad (6.7)$$

where dotted quantities are derived with respect to  $x_1$ . As expected, the KK reduction lands us on a model with a non-trivial potential on the target space  $\mathcal{T} = S^3$ , due to presence of the  $\mathbb{Z}_2$ -symmetric background holonomy.

The Euclidean equations of motion associated with this action are

$$\ddot{\theta} - \frac{1}{2} \sin 2\theta [(\dot{\phi}_2)^2 - (\dot{\phi}_1)^2 + \xi^2] = 0, \quad (6.8)$$

$$\partial_{x_1} \left( \cos^2 \theta \dot{\phi}_1 \right) = 0, \quad (6.9)$$

$$\partial_{x_1} \left( \sin^2 \theta \dot{\phi}_2 \right) = 0. \quad (6.10)$$

Setting  $\phi_{1,2}$  constant in time solves the second and third equations while the first one reduces to the equation for a kink-instanton — the fracton from the 2d viewpoint — in one dimension:

$$\ddot{\theta} - \frac{\xi^2}{2} \sin 2\theta = 0. \quad (6.11)$$

The existence of the *stable* fracton events is the crucial difference compared to  $\mathbb{R}^2$ , or to thermal compactification  $\mathbb{R} \times S^1_\beta$  associated with the trivial background holonomy.

We can find the action of the instanton configuration by focusing only on the  $\theta$  field, and using Bogomolny's method to rewrite the action in the simple form

$$S = \frac{L}{g^2} \int_{-\infty}^{+\infty} dx_1 \left[ (\dot{\theta} \mp \xi \sin \theta)^2 \pm 2\xi \dot{\theta} \sin \theta \right] \geq \frac{2L\xi}{g^2} \int_0^\pi d\theta \sin \theta \geq \frac{4L\xi}{g^2} \quad (6.12)$$

$$\Rightarrow S_F = \frac{4\pi}{g^2} = \frac{S_{\text{uniton}}}{2}, \quad (6.13)$$

where we substituted the  $\mathbb{Z}_2$  symmetric value for  $\xi = \pi/L$ . Note that the action of a fracton  $S_F$  is precisely  $1/2$  ( $1/N$  in the  $SU(N)$  case) times the action of a uniton eq. (6.1), supporting our claim that the fractons are constituents of unitons.

The equality in eq. (6.13) holds whenever the fracton/anti-fracton satisfy the first order BPS equation

$$\dot{\theta}(x_1) - \xi \sin \theta = 0, \quad (6.14)$$

$$\dot{\bar{\theta}}(x_1) + \xi \sin \bar{\theta} = 0. \quad (6.15)$$

The solutions to these equations take the form

$$\theta(x_1; x_1^{(0)}) = 2 \operatorname{arcCot} \left[ e^{-\xi(x_1 - x_1^{(0)})} \right], \quad (6.16)$$

$$\bar{\theta}(x_1; x_1^{(0)}) = \pi - 2 \operatorname{arcCot} \left[ e^{-\xi(x_1 - x_1^{(0)})} \right], \quad (6.17)$$

where  $x_1^{(0)}$  is a position modulus. The fracton obeys the boundary conditions  $\theta(x_1 \rightarrow -\infty) = 0$ ,  $\theta(x_1 \rightarrow +\infty) = \pi$ , while the anti-fracton goes from  $\pi$  to 0 and these solutions have precisely the BPS-saturated (from the small- $L$  effective field theory point of view) action

$$S_{\text{fracton}} = \frac{4\pi}{g^2} = \frac{S_{\text{uniton}}}{2}. \quad (6.18)$$

**KK fractons, and unitons from fractons.** As we have mentioned above, discarding all KK non-zero modes misses some critical information, which is important even when  $L$  is small. In particular, consider doing the dimensional reduction to 1D with the *periodic* field  $\phi_2 \in [0, 2\pi]$  winding  $n$  times as we move along the  $S^1$  (i.e.  $\phi_2 = 2\pi n x/L$ ), rather than just being set to a constant, as we did above. We will shortly see that configurations with  $n = -1$  have a distinguished role. Let us assume that

$$\phi_2(x_1, x_2) = \phi_2(x_1) + \frac{2\pi n x_2}{L}. \quad (6.19)$$

This means that  $\phi_2$  carries some KK momentum, by way of a non-trivial winding number  $n$ . The reduced action becomes

$$S = \frac{L}{g^2} \int_{-\infty}^{+\infty} dx_1 \left[ \dot{\theta}^2 + \cos^2 \theta \dot{\phi}_1^2 + \sin^2 \theta \dot{\phi}_2^2 + \left( \frac{2\pi n}{L} + \xi \right)^2 \sin^2 \theta \right]. \quad (6.20)$$

The same arguments as before then show that this action gives rise to stable instanton configurations, which we will refer to as KK-fractons.<sup>30</sup> These configurations have action

$$S^{(n)} = \frac{4|2\pi n + \xi L|}{g^2}. \quad (6.21)$$

Now observe that when  $\xi$  takes on the  $\mathbb{Z}_2$  symmetric value  $\xi = \pi/L$  and  $n = -1$ , this becomes

$$S_{\text{KK fracton}} = \frac{4\pi}{g^2} = S_F = \frac{S_{\text{uniton}}}{2}. \quad (6.22)$$

The action for a generic KK fracton with  $n \neq -1$ , will be higher than  $S_{\text{fracton}}$ , and hence their contribution will be suppressed in the semiclassical expansion. But fractons and the  $n = -1$  KK fractons enter the semiclassical expansion on the *same* footing!

Despite having the same action as the fractons, it is important to emphasize that the  $n = -1$  KK fractons and  $n = 0$  fractons are *distinct* field configurations. Indeed, the explicit solutions for the field configurations of the  $n = -1$  KK fractons are

$$\theta(x_1; x_1^{(0)})_{\text{KK}} = 2 \operatorname{arcCot} \left[ e^{-(\xi - 2\pi/L)(x_1 - x_1^{(0)})} \right], \quad (6.23)$$

$$\bar{\theta}(x_1; x_1^{(0)})_{\text{KK}} = \pi - 2 \operatorname{arcCot} \left[ e^{-(\xi - 2\pi/L)(x_1 - x_1^{(0)})} \right]. \quad (6.24)$$

These KK fractons behave as  $\theta(x_1 \rightarrow -\infty)_{\text{KK}} \rightarrow \pi$  and  $\theta(x_1 \rightarrow +\infty)_{\text{KK}} \rightarrow 0$ , while  $\bar{\theta}(x_1 \rightarrow -\infty)_{\text{KK}} \rightarrow 0$  and  $\bar{\theta}(x_1 \rightarrow +\infty)_{\text{KK}} \rightarrow \pi$ , in contrast to the standard fractons.

From the solution for the minimal fractionalized  $\text{SU}(2)$  uniton (6.2), depicted in figure 8, we can indeed read the exact form of the various component fields  $\theta, \phi_{1,2}$ . The  $\theta$  component can be seen as the gluing of a fracton at  $x_1 = x_1^{(0)}$ , followed by a KK-fracton at some  $x_1 > x_1^{(0)}$ , so that the combined field configurations goes from  $\theta = 0$  at  $x_1 = -\infty$  back to  $\theta = 0$  at  $x_1 = +\infty$ . To see why the second object is a KK-fracton rather than an anti-fracton one can study the field  $\phi_2$  in the uniton solution on  $\mathbb{R} \times S^1$  with the  $\mathbb{Z}_N$ -symmetric background holonomy. It can be shown that  $\phi_2$  interpolates between a configuration with zero winding in the  $x_2$  direction at  $x_1 = -\infty$  and a configuration with  $-1$  winding in the  $x_2$  direction at  $x_1 = +\infty$ :

$$\phi_2(x_1, x_2)|_{\text{uniton}} = \frac{-\pi x_2}{L} - \arctan \left[ \tanh \left( \frac{\pi x_1}{L} \right) \tan \left( \frac{\pi x_2}{L} \right) \right], \quad (6.25)$$

where for simplicity we fixed the moduli  $\lambda_1 = \lambda_2 = 1$ . Hence we see the precise sense in which the fractons are the constituents of unitons.

### 6.3 $\text{SU}(N)$ fractons

We now discuss the generalization of the  $\text{SU}(2)$  fractons to the  $\text{SU}(N)$  PCM. This can be done by embedding the  $\text{SU}(2)$  fractons into the  $\text{SU}(N)$  model, by using eq. (5.9). This construction permits us to classify all the fracton events with action

$$S^{(k)} = k \times S_F, \quad k \in \mathbb{Z}_+, \quad (6.26)$$

---

<sup>30</sup>This construction is analogous to the one done for gauge theories on  $\mathbb{R}^3 \times S^1$ , where the resulting configurations are known as KK-monopoles [24, 25].

where

$$S_F = \frac{8\pi}{g^2 N} = \frac{S_{\text{uniton}}}{N} \quad (6.27)$$

is the action of the minimal fracton, and  $k$  may be seen as “level” of the associated action.

**Naive approach.** Consider the root system of the Lie algebra  $\mathfrak{su}(N)$ . Denote the simple root system as

$$\Delta_1 = \{\alpha_1, \alpha_2, \dots, \alpha_{N-1}\}. \quad (6.28)$$

We can use  $\Delta_1$  to build the complete root system  $A_{N-1}$ . For each root  $\alpha$  (simple or not), there is an associated  $\mathfrak{su}(2)$  sub-algebra. All fracton embeddings associated with simple roots are minimal action, level-1 events, with action:

$$\mathcal{F}_{\alpha_i}^{(1)} \equiv \mathcal{F}_i : \quad S_{i,1} = \frac{8\pi\mu_{i+1,i}}{g^2} = \frac{4\xi}{g^2} = \frac{8\pi}{g^2 N} = \frac{S_{\text{uniton}}}{N} \quad i \in [1, N-1] \quad (6.29)$$

where we set  $\mu_i$  to their center-symmetric values. There are  $N-1$  simple roots, and hence there are  $N-1$  minimal action fracton events.

The positive roots which can be written as a sum of just two adjacent simple roots correspond to level-2 fractons. There are  $N-2$  such positive roots. The actions of these events are (in the center-symmetric background)

$$\mathcal{F}_{\alpha_i + \alpha_{i+1}}^{(2)} : \quad S_{i,2} = \frac{8\pi\mu_{i+2,i}}{g^2} = 2 \times \frac{8\pi}{g^2 N} = 2 \times \frac{S_{\text{uniton}}}{N} \quad i \in [1, N-2]. \quad (6.30)$$

Similarly positive roots which can be written as a sum of just three (adjacent) simple roots correspond to level-3 fractons and so on. The general construction is that the root space  $A_{N-1}$  can be split into simple roots and roots that can be written as  $k$ -linear combination of simple roots:

$$\begin{aligned} k=1 : & \quad (N-1) \text{ positive roots : } \Delta_1^+ \equiv \{\alpha_1, \alpha_2, \dots, \alpha_{N-1}\}, \\ k=2 : & \quad (N-2) \text{ positive roots : } \Delta_2^+ \equiv \{\alpha_1 + \alpha_2, \alpha_2 + \alpha_3, \dots, \alpha_{N-2} + \alpha_{N-1}\} \\ k=3 : & \quad (N-3) \text{ positive roots : } \Delta_3^+ \equiv \{\alpha_1 + \alpha_2 + \alpha_3, \dots, \alpha_{N-3} + \alpha_{N-2} + \alpha_{N-1}\} \\ \dots & \quad \dots \quad \dots \\ k=N-2 : & \quad 2 \text{ positive roots : } \Delta_{N-2}^+ \equiv \{\alpha_1 + \alpha_2 + \dots + \alpha_{N-2}, \alpha_2 + \alpha_3 + \dots + \alpha_{N-1}\} \\ k=N-1 : & \quad 1 \text{ positive root : } \Delta_{N-1}^+ \equiv \{\alpha_1 + \alpha_2 + \dots + \alpha_{N-2} + \alpha_{N-1}\} \end{aligned} \quad (6.31)$$

We can similarly define the negative roots,  $\Delta_j^- = -\Delta_j$ . Obviously, the root space  $A_{N-1}$  can be decomposed as

$$A_{N-1} = \bigoplus_{k=1}^{N-1} (\Delta_j^+ + \Delta_j^-). \quad (6.32)$$

This root space  $A_{N-1}$  contains  $N^2 - N$  roots, half positive and half negative, and together with the Cartan subalgebra, whose dimension is  $N-1$ , they form the complete set of generators of  $\text{SU}(N)$ .

As is well-known, there is an  $\mathfrak{su}(2)$  sub-algebra of  $\mathfrak{su}(N)$  generated by  $E_\alpha, E_{-\alpha}, \alpha \cdot \mathbf{H}$  where  $E_{\pm\alpha}$  are raising and lowering operators, and  $\mathbf{H}$  are Cartan sub-algebra matrices.



We can embed an  $\mathfrak{su}(2)$  fracton into  $\mathfrak{su}(N)$  by using any root  $\alpha$ . Clearly, we have fracton embeddings associated with  $\alpha$ ,

$$\mathcal{F}_\alpha^{(k)} : \quad \alpha \in \Delta_k^+, \quad S_{i,k} = \frac{8\pi(\mu_{i+k} - \mu_i)}{g^2} = \frac{8\pi k}{g^2 N} = k \times S_F = \frac{k S_{\text{uniton}}}{N}. \quad (6.33)$$

According to this rationale, there are  $N - 1$  fractons at level-1,  $N - 2$  fractons at level-2, so on and so forth, and only one fracton at level  $N - 1$ . Clearly, the densities of events of level-1, level-2, etc are also hierarchical, and obeys

$$e^{-S_F} \gg e^{-2S_F} \gg \dots \gg e^{-(N-1)S_F} \quad (6.34)$$

**More carefully, accounting for the compactness of holonomy.** The previous discussion does not take into account the fact that the background holonomy is compact. For example, one might naively think that events in  $\Delta_{N-1}^+$  are very rare tunneling paths, exponentially suppressed with respect to the minimal fractons, because their actions are  $S_{1,N-1} = (N-1) \times S_{i,1}$ . Indeed, this is largely correct. However, if we choose, for  $k = N-1$ , a configuration with winding number  $n = -1$  for  $\phi_2$ , as was done in the  $SU(2)$  PCM in section 6.2, and then apply the Kaluza-Klein-reduction to the action afterwards, we obtain

$$S = \frac{1}{g^2} \int_{\mathbb{R}} \left[ (\partial_t \theta)^2 + \left( \frac{2\pi}{L} (-1 + \mu_N - \mu_1) \right)^2 \sin^2 \theta \right]. \quad (6.35)$$

This happens because  $\mu_i - \mu_j$  behaves as a fractional momentum in the compact direction, and combining  $(\mu_N - \mu_1)$  with  $-1$  unit of winding number results in a *twisted KK-reduction*. In the full 2D QFT, this tunneling path is exactly on the same footing with the elementary fractons corresponding to  $\Delta_1^+$ , and its action coincides with them (6.29) in the center-symmetric background of eq. (4.29):

$$S_{\text{twisted}} = \frac{8\pi |-1 + (\mu_N - \mu_1)|}{g^2} = \frac{8\pi}{g^2 N} = S_{i,1} = \frac{S_{\text{uniton}}}{N}. \quad (6.36)$$

We refer to this instanton-type event as the KK fracton, denoted  $\mathcal{F}_N$ . It is associated with the affine root of the  $\mathfrak{su}(N)$  Lie algebra, defined by

$$\alpha_N = - \sum_{i=1}^{N-1} \alpha_i, \quad (6.37)$$

which is itself a negative root. The  $\mathcal{F}_N$  fracton is the counter-part of the so called KK-monopole-instanton (also called affine or twisted instanton) in gauge theories on  $\mathbb{R}^3 \times S^1$  [24, 25].

Similarly, by inserting  $-1$  units of winding number into the two events living in  $k = N-2$ , we can turn them into tunneling events with action  $\frac{2}{N} S_{\text{uniton}}$ . This pattern continues

for all levels:

$$\begin{aligned}
 k = 1 : \quad & \Delta_1^{\text{aff}} \equiv \Delta_1^+ + \Delta_{N-1}^- \equiv \{\alpha_i\} \\
 k = 2 : \quad & \Delta_2^{\text{aff}} \equiv \Delta_2^+ + \Delta_{N-2}^- \equiv \{\alpha_i + \alpha_{i+1}\} \\
 k = 3 : \quad & \Delta_3^{\text{aff}} \equiv \Delta_3^+ + \Delta_{N-3}^- \equiv \{\alpha_i + \alpha_{i+1} + \alpha_{i+2}\} \\
 \dots & \\
 k = N - 2 : \quad & \Delta_{N-2}^{\text{aff}} \equiv \Delta_{N-2}^+ + \Delta_2^- \equiv \{\alpha_1 + \alpha_2 + \dots + \alpha_{i-1} + \alpha_{i+2} \dots + \alpha_N\} \\
 k = N - 1 : \quad & \Delta_{N-1}^{\text{aff}} \equiv \Delta_{N-1}^+ + \Delta_1^- \equiv \{\alpha_1 + \alpha_2 + \dots + \alpha_{i-1} + \alpha_{i+1} \dots + \alpha_N\}
 \end{aligned} \tag{6.38}$$

where  $i \in [1, N]$ . In particular, we learn that at level  $k$ , there are *always*  $N$  NP-saddles associated with the roots  $\Delta_k^{\text{aff}}$ .<sup>31</sup> This is actually a consequence of the cyclic  $\mathbb{Z}_N$  symmetry of the background holonomy. All  $N$ -saddles in  $\Delta_k^{\text{aff}}$  have actions  $S^{(k)} = \frac{8\pi}{g^2 N} \times k = \frac{S_{\text{uniton}}}{N} \times k$ .

The existence of the KK fracton events is due to the compactness of the  $\text{SU}(N)$  target space and of the background holonomy  $\Omega$  in eq. (4.20). Equivalently, although the long distance effective field theory on small  $\mathbb{R} \times S_L^1$  may appear one-dimensional, the underlying microscopic theory is two-dimensional, and this is encoded in the structure of the effective field theory. As we saw above both in the perturbative and non-perturbative contexts, the details of dimensional reduction at small- $L$  are quite subtle in the context of theories with adiabatic small- $L$  limits.

The fractons associated with  $\Delta_1^{\text{aff}}$  are the leading (minimal action) topological configurations in the PCM. In the bosonic model, this will be of crucial importance in the determination of the mass gap. Furthermore, certain correlated fracton-anti-fracton events (neutral bions) will be crucial in resolving the renormalon ambiguity. Both of these features are similar to the story seen in the  $\mathbb{CP}^{N-1}$  model and in YM theory [16–19].

#### 6.4 Fracton amplitudes

The fracton amplitude is the same as the 1-d instanton amplitude associated with the quantum mechanical system in eq. (5.19). Within the Born-Oppenheimer approximation in the low energy effective theory, we find there is effectively only one bosonic zero mode, the position modulus of the fracton. Using dimensionless units introduced in eq. (5.21), the amplitude of the fracton event, is given by

$$\mathcal{F}_i = J_{\tau_0} e^{-S_i} \left[ \frac{\det' M}{\det M_0} \right]^{-\frac{1}{2}} \tag{6.40}$$

<sup>31</sup>The advantage of  $\Delta_k^{\text{aff}}$  with respect to  $\Delta_k^+$  is that it makes the fact that the KK-fractons are on the same footing with the regular fractons manifest, along with the  $\mathbb{Z}_N$  cyclicity. Its main disadvantage is that one loses the notion of positivity associated with level- $k$ . Recall that all roots living in  $\Delta_k^+$  were positive. Since

$$\Delta_k^{\text{aff}} = \Delta_k^+ + \Delta_{N-k}^- , \tag{6.39}$$

and all roots (involving the affine root in some way) living in  $\Delta_{N-k}^-$  are negative,  $\Delta_k^{\text{aff}}$  is not comprised of positive roots only. Unfortunately, one cannot have both properties at once. The concept of positive roots will play some role when we discuss correlated fracton events, and hence, we work with both representation.

where  $M = \frac{\delta^2 S}{\delta\theta_c(t_1)\delta\theta_c(t_1)} = \left[ -\left(\frac{d}{dt_1}\right)^2 - V''(\theta_c(t_1)) \right] \delta(t_1 - t_2)$  is the quadratic fluctuation operator in the background of the fracton,  $\det' M$  indicates that the zero mode is dropped in evaluation of the determinant, and  $\det M_0$  is the determinant of the quadratic fluctuation operator in the vacuum background, i.e.  $M_0 = [(d/dt_1)^2 - \text{const.}] \delta(t_1 - t_2)$ . The determinant in eq. (6.40) can be evaluated in multiple different ways [39, 80], for example, via the Gelfand-Yaglom method [95]. The Jacobian associated with the zero mode is

$$J_{\tau_0} = \sqrt{\frac{S_i}{2\pi}}. \quad (6.41)$$

The result of this calculation is given in section 41.2 of [80], and is equal to  $\mathcal{I} = \frac{4}{\sqrt{\pi g_{ZJ}}} e^{-8/g_{ZJ}}$ . The conversion to our notation, as already noted after eq. (5.21), involves the relation  $g_{ZJ} = \frac{2g^2}{\xi}$ , and we can write the minimal fracton amplitude ( $i = 1, \dots, N$ ) in the center-symmetric background as

$$\mathcal{F}_i = \sqrt{\frac{2S_i}{\pi}} e^{-S_i} = \sqrt{\frac{8\xi}{\pi g^2}} e^{-\frac{4\xi}{g^2}} = \sqrt{\frac{16}{g^2 N}} e^{-\frac{8\pi}{g^2 N}}. \quad (6.42)$$

There are perturbative fluctuations around the fracton, and the amplitude incorporating those fluctuations is given by

$$[\mathcal{F}] \times \Phi_{\mathcal{F}}(g^2) = \sqrt{\frac{16}{g^2 N}} e^{-\frac{8\pi}{g^2 N}} \Phi_{\mathcal{F}}(g^2), \quad \Phi_{\mathcal{F}}(g^2) \equiv \sum_{n=0}^{\infty} a_n^{[\mathcal{F}]} g^{2n}. \quad (6.43)$$

It should be noted that there is no non-perturbative ambiguity in the NP part of the amplitude, given by  $\sqrt{\frac{16}{g^2 N}} e^{-\frac{8\pi}{g^2 N}}$ . On the other hand we expect  $\Phi_{\mathcal{F}}(g^2)$  to be a non-Borel summable asymptotic series, similarly to perturbative expansion around the P-saddle eq. (5.22).

## 7 Resurgence triangle and emergent topological structure

Now that we have found the fractons, the minimal-action non-perturbative field configurations at weak coupling, we are in a position to see how the renormalon ambiguities in the Borel resummation of the perturbative series, such as eq. (5.25) are cancelled. We will show that they cancel against corresponding ambiguities in the contributions from the non-perturbative sector, leaving well-defined results order by order in an expansion in  $e^{-S_F} = e^{-8\pi/(g^2 N)}$ . The phenomenon we are describing is an example of a semiclassical expansion which is not Borel summable, but *is* Borel-Ecalle summable. This is a generalization of eq. (2.28) for ordinary integrals.

We first emphasize again that there is no non-perturbative ambiguity in the fracton amplitude  $[\mathcal{F}_i]$  itself.<sup>32</sup> Although there is no topological charge in the microscopic theory, there is an emergent topological structure in the low energy effective theory, and the fractons are

---

<sup>32</sup>Of course, the perturbative fluctuations  $\Phi_{\mathcal{F}}(g^2)$  around it are still non-Borel summable and ambiguous, but that is a different story.

1d-instantons from that point of view, and hence they cannot mix with perturbation theory. In theories with a topological  $\Theta$  angle it is always true that events with non-zero instanton number cannot cure the ambiguity of perturbation theory [18] due to the  $\Theta$  dependence of instanton amplitudes. In our case, despite the absence of a microscopic topological argument, at small- $L$  we find that fractons cannot mix with perturbation theory either.

Second, it is now worthwhile to re-inspect perturbation theory around the perturbative saddle, and rewrite eq. (5.22) and eq. (5.23) as expansions in

$$\frac{1}{2S_F} = \frac{g^2}{8\xi} = \frac{g^2 N}{16\pi}. \quad (7.1)$$

Then

$$\begin{aligned} \mathcal{E}_i(S_F) = & \frac{1}{2} - \frac{1}{2} \left( \frac{1}{2S_F} \right) - \frac{1}{2} \left( \frac{1}{2S_F} \right)^2 - \frac{3}{2} \left( \frac{1}{2S_F} \right)^3 - \frac{53}{8} \left( \frac{1}{2S_F} \right)^4 - \frac{297}{8} \left( \frac{1}{2S_F} \right)^5 - \frac{3961}{16} \left( \frac{1}{2S_F} \right)^6 - \dots \\ & - \frac{2}{\pi} \left( \frac{1}{2S_F} \right)^n n! \left( 1 + O\left(\frac{1}{n}\right) \right). \end{aligned} \quad (7.2)$$

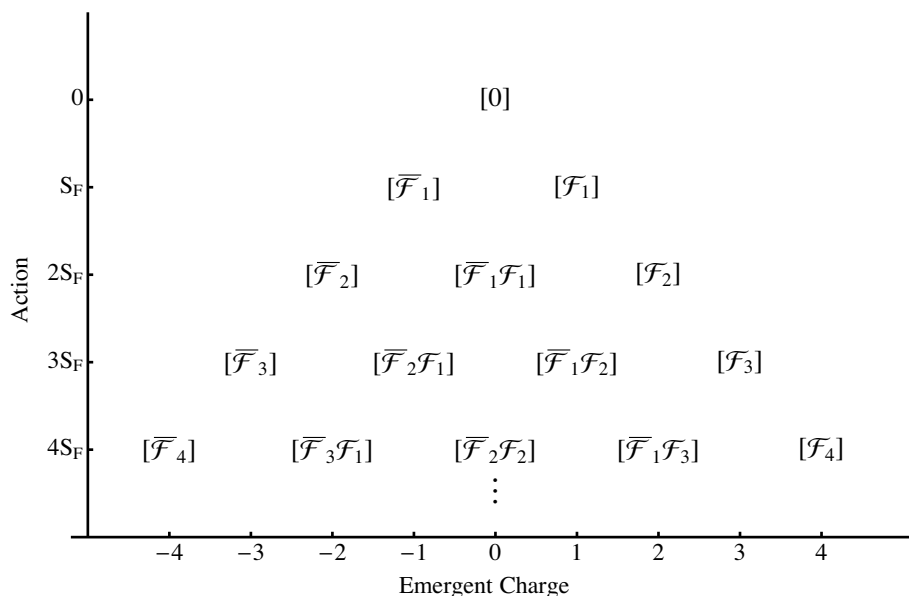
This emphasizes that the perturbative expansion parameter is the inverse of *two* times the fracton action. As already mentioned in footnote 6, this is one of the major differences between ordinary integrals and path integrals. The leading ambiguity of perturbation theory is given by the right/left Borel resummation eq. (5.25). Expressing these left/right resummations in terms of the action,

$$\mathcal{S}_{0\pm} \mathcal{E}_i = \text{Re } \mathcal{S}_0 \mathcal{E} \mp 2\pi i [\mathcal{F}_i][\bar{\mathcal{F}}_i] = \text{Re } \mathcal{S}_0 \mathcal{E} \mp i(4S_F) e^{-2S_F} = \text{Re } \mathcal{S}_0 \mathcal{E} \mp i \frac{32\pi}{g^2 N} e^{-\frac{16\pi}{g^2 N}} \quad (7.3)$$

makes it clear that the ambiguities can only be canceled by contributions which live in the fracton-anti-fracton sector. We show that this is the case in section 7.2. The first step in seeing how this works is to understand the interactions of fractons and anti-fractons, which will be our task in section 7.1.

Our main idea about emergent topology and resurgence is encoded in the resurgence triangle [18] classification of saddle points, which is illustrated for the PCM on  $\mathbb{R} \times S^1$  in figure 11 in a schematic form, which does not take into account the “ramification” phenomenon we will describe later in this section. The entries in the resurgence triangle are  $[\bar{\mathcal{F}}_{\bar{n}} \mathcal{F}_n]$ , which denote *correlated* events involving  $n$ -fracton and  $\bar{n}$  anti-fractons. The vertical axis is the action of the events, while the horizontal axis tracks the “charge”  $q$  of the events, defined as  $q = n - \bar{n}$ . Then the perturbative vacuum is the “primary” or “level zero” for the  $q = 0$  sector,  $[\mathcal{F}_1]$  is the “level zero” event for the  $q = 1$  sector,  $[\mathcal{F}_2]$  is the level zero event for the  $q = 2$  sector, and so on. The cancellation of ambiguities happens between saddles in a given column, and different columns cannot mix in the cancellation of their respective ambiguities. Note that unlike theories in which different columns are classified according to different homotopy classes of the microscopic theory, in the present example where there is no microscopic topological classification, the classification of columns takes place according to the topology in the low energy effective theory, which is *emergent*.

It should be emphasized that the resurgence triangle and the variety of NP-saddles for the PCM are just as rich as  $\mathbb{CP}^{N-1}$  model [18] or deformed Yang-Mills theory or



**Figure 11.** Schematic version of the resurgence triangle classification of saddle points induced by resurgence theory for the PCM on small  $\mathbb{R} \times S^1$ . Class  $[\mathcal{F}_2]$  includes both single uncorrelated events with charge  $\alpha \in \Delta_2^{\text{aff}}$  in eq. (6.38) as well as correlated two-events such as  $[\mathcal{F}_i\mathcal{F}_i]$  with charge  $2\alpha_i$  which itself is not a root. In particular, (7.24) provides a more realistic Lie algebraic ramified description of this cartoon, and (7.24) provides the quantitative distinction between the amplitudes of these two types of events.

QCD(adj) [16]. In other words, the fact that the homotopy group  $\pi_2$  is trivial in the microscopic PCM is only a superficial difference from Yang-Mills theory. At the deeper level of insight allowed by resurgence theory, it appears that the set of NP saddles in the PCM is in fact in *one-to-one* correspondence to the set of NP saddles in Yang-Mills theory, because in both cases the set of saddles are determined by the properties of the underlying  $\mathfrak{su}(N)$  Lie algebra!

### 7.1 Fracton interactions and $[\mathcal{F}_i\mathcal{F}_i]$ , $[\bar{\mathcal{F}}_i\bar{\mathcal{F}}_i]$ , $[\mathcal{F}_i\bar{\mathcal{F}}_i]$ amplitudes

To understand how the weak coupling realization of the renormalon ambiguity explained around eq. (5.25) and eq. (7.3) is cured, we must understand the structure of the semi-classical expansion at second and higher order. As we have described above, at first order in the semi-classical expansion, one has the minimal-action fractons, but their contributions ( $\sim e^{-S_F}$ ) are **1**) exponentially *larger* than the renormalon ambiguity ( $\sim e^{-2S_F}$ ) and **2**) are unambiguous. Hence the minimal-action fractons contributing at first order in the semi-classical expansion cannot contribute to resolving the issues of perturbation theory, which can only be cured at second order as shown by eq. (7.3).

At second order in the semi-classical expansion, class  $[\mathcal{F}_2]$ , we encounter two-class of events: *correlated* fracton events, such as  $[\mathcal{F}_i\mathcal{F}_i]$ ,  $[\bar{\mathcal{F}}_i\bar{\mathcal{F}}_i]$ ,  $[\mathcal{F}_i\bar{\mathcal{F}}_i]$  with charges  $2\alpha_i, 0, -2\alpha_i$ , respectively (note that these are not roots, hence these are viewed as correlated events) as well as *uncorrelated* single fracton events associated with roots  $\alpha \in \Delta_2^{\text{aff}}$  in eq. (6.38).

(Around (7.24), we provide more quantitative distinction between the amplitudes of these two types of events.) The uncorrelated events correspond to field configurations which are exact solutions of the equations of motion of the theory. The correlated events have finite actions which are parametrically close to those of the uncorrelated events at this order in the  $\lambda \ll 1$  limit, and the associated field configurations are quasi-solutions to the equations of motion rather than exact solutions, meaning that the equations of motion are satisfied with parametrically good accuracy when  $\lambda \ll 1$ . The correlated events are *quasi*-saddle-points, and make critically important contributions to the path integral in the semiclassical limit.

According to the emergent topological structure associated with the resurgence triangle of figure 11,  $[\mathcal{F}_i \mathcal{F}_i]$ ,  $[\bar{\mathcal{F}}_i \bar{\mathcal{F}}_i]$  are the leading saddles in their classes, and cannot cure the ambiguity of perturbation theory either. It is the fracton-anti-fracton configurations  $[\mathcal{F}_i \bar{\mathcal{F}}_i]$ ,  $i = 1, 2, \dots, N$  (which we refer to as “neutral bions” due to the universality of such configurations across a wide range of asymptotically-free theories) which can mix with perturbation theory. They are the first sub-leading saddles in the column associated with the perturbative saddle.

Below, we demonstrate that the non-perturbative amplitudes (without incorporating the perturbative fluctuations) associated with the correlated events are given by:

$$\begin{aligned}
 [\mathcal{F}_j \mathcal{F}_j] &= [\bar{\mathcal{F}}_j \bar{\mathcal{F}}_j] = (-\log(4S_F) - \gamma) [\mathcal{F}_j] [\mathcal{F}_j] \\
 &= (-\log(4S_F) - \gamma) \frac{2S_F}{\pi} e^{-2S_F} \\
 &= \left( -\log \left[ \frac{32\pi}{g^2 N} \right] - \gamma \right) \frac{16}{g^2 N} e^{-\frac{16\pi}{g^2 N}}, \tag{7.4}
 \end{aligned}$$

$$\begin{aligned}
 [\mathcal{F}_j \bar{\mathcal{F}}_j]_{\pm} &= [\bar{\mathcal{F}}_j \mathcal{F}_j]_{\pm} = (-\log(4S_F) - \gamma) [\mathcal{F}_j] [\bar{\mathcal{F}}_j] \pm i\pi [\mathcal{F}_j] [\bar{\mathcal{F}}_j] \\
 &= (-\log(4S_F) - \gamma) \frac{2S_F}{\pi} e^{-2S_F} \pm i\pi \times \frac{2S_F}{\pi} e^{-2S_F} \\
 &= \left( -\log \left[ \frac{32\pi}{g^2 N} \right] - \gamma \right) \frac{16}{g^2 N} e^{-\frac{16\pi}{g^2 N}} \pm i \frac{16\pi}{g^2 N} e^{-\frac{16\pi}{g^2 N}}, \tag{7.5}
 \end{aligned}$$

where  $\gamma$  is the Euler-Mascheroni constant. Each of these representations is useful for slightly complementary reasons. Note that the correlated  $[\mathcal{F}_j \mathcal{F}_j]$  event amplitude is *not* the same quantity as  $[\mathcal{F}_j] [\mathcal{F}_j]$ , and these are genuinely different (isolated, non-degenerate) saddles. In fact, in the semi-classical evaluation of partition function, we must sum over both types of saddles independently. For example, up to second order in NP trans-series expansion, we can write a grand canonical ensemble of all defects entering up to second order in the resurgence triangle:

$$\begin{aligned}
 e^{-E\beta} &\sim e^{-\frac{\omega}{2}(1+O(g))\beta} \prod_{\mathcal{T}} \left( \sum_{n_{\mathcal{T}}} \frac{(\beta \mathcal{T})^{n_{\mathcal{T}}}}{n_{\mathcal{T}}!} \right) \\
 &= e^{-\frac{\omega}{2}(1+O(g))\beta} \left( \sum_{n_{\mathcal{F}}} \frac{(\beta \mathcal{F})^{n_{\mathcal{F}}}}{n_{\mathcal{F}}!} \right) \left( \sum_{n_{\bar{\mathcal{F}}}} \frac{(\beta \bar{\mathcal{F}})^{n_{\bar{\mathcal{F}}}}}{n_{\bar{\mathcal{F}}}!} \right) \left( \sum_{n_{[\mathcal{F}\mathcal{F}]}} \frac{(\beta [\mathcal{F}\mathcal{F}])^{n_{[\mathcal{F}\mathcal{F}]}}}{n_{[\mathcal{F}\mathcal{F}]!} \right) \dots \\
 &= e^{-\left(\frac{\omega}{2}(1+O(g)) - \mathcal{F} - \bar{\mathcal{F}} - [\mathcal{F}\mathcal{F}] - [\bar{\mathcal{F}}\bar{\mathcal{F}}] - [\mathcal{F}\bar{\mathcal{F}}] + \dots\right)\beta}, \tag{7.6}
 \end{aligned}$$

where we denoted with  $\omega/2$  the energy of the ground state in the free case  $g = 0$ . A pictorial description of this generalized grand canonical ensemble can be found in [8]. It is also implicit in the Zinn-Justin's exact quantization formula which incorporates all orders in non-perturbative effects [80] via a resurgent trans-series.

**Computation of  $[\mathcal{F}_j \bar{\mathcal{F}}_j]$  and  $[\mathcal{F}_j \mathcal{F}_j]$  and amplitude.** Amplitudes for correlated two-events in the  $\mathbb{CP}^{N-1}$  model on  $\mathbb{R} \times S^1$  with  $\mathbb{Z}_N$ -twisted boundary conditions were calculated in [18], and as we have remarked there is a striking degree of similarity between the NP saddles of the  $\mathbb{CP}^{N-1}$  model and the NP saddles of the PCM associated with the affine simple roots of  $\mathfrak{su}(N)$ . Moreover, it turns out that at small- $L$  it is precisely (the interactions of) these  $N$  minimum-action saddle points of the PCM which are responsible for the cancellation of renormalon ambiguities. Hence our discussion below will amount to a review of the argument of [18], with a few additional clarifying comments. We suppress the index  $i$  in what follows to lessen the clutter, since we are looking for the correlated amplitudes for  $[\mathcal{F}_\alpha \bar{\mathcal{F}}_\alpha]$  and  $[\mathcal{F}_\alpha \mathcal{F}_\alpha]$  for any  $\alpha \in A_{N-1}$ . Indeed,  $\alpha$  may be chosen to be  $\alpha_i \in \Delta_1^{\text{aff}}$  in the classification of eq. (6.38), without loss of generality.

Consider an  $SU(2)$  subgroup embedded into  $SU(N)$  PCM, and let  $[\mathcal{F}\mathcal{F}]$  and  $[\mathcal{F}\bar{\mathcal{F}}]$  configurations denote correlated events therein. Then consider the ansatz for e.g.  $[\mathcal{F}\bar{\mathcal{F}}]$ :

$$\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau) = \theta_{\mathcal{F}}(t_0 + \tau/2) + \theta_{\bar{\mathcal{F}}}(t_0 - \tau/2). \quad (7.7)$$

Here,  $t_0$  is the center of action coordinate of the pair, while  $\tau$  is the separation of the constituent fractons. While  $\theta_{\mathcal{F}}$  and  $\theta_{\bar{\mathcal{F}}}$  are individually solutions to the equations of motion, we do *not* expect  $\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau)$  to be a solution, because the equations of motion are nonlinear. However, when  $\tau \gg \xi$ , we expect  $\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau)$  to become a quasi-solution to the equations of motion, in the sense that the equations of motion are satisfied up to terms exponentially small in  $\xi\tau$ . In the far-separated regime, we expect the action of  $\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau)$  to approach  $2S_F$ . So we expect  $\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau)$  to be a quasi-saddle-point of the theory.

Fluctuations around exact saddle points are either zero modes, which cost zero action, or perturbative modes, which have an action cost of order  $\xi$ . Fluctuations around quasi-saddle-points are more subtle. In addition to zero modes and perturbative fluctuation modes, there are also quasi-zero modes, which have an action cost which is exponentially small in  $\xi$ . For  $\theta_{\mathcal{F}\bar{\mathcal{F}}}(t_0; \tau)$  it turns out that  $t_0$  is an exact zero mode, while  $\tau$  is a quasi-zero mode (with an ‘energy’ parametrically separated from the perturbative fluctuations). The contribution of  $[\mathcal{F}\bar{\mathcal{F}}]$  events to the path integral is given by

$$[\mathcal{F}\bar{\mathcal{F}}] = [\mathcal{F}][\bar{\mathcal{F}}] \int d\Omega d\tau e^{-V(\tau)}, \quad (7.8)$$

where  $[\mathcal{F}][\bar{\mathcal{F}}]$  are the product of uncorrelated amplitudes. The explicit integration over  $\tau$ , the separation between the constituent fractons, appears because it is a quasi zero mode, due to the presence of the “interaction potential” between the fractons. The integration over  $d\Omega$  is an instruction to integrate over the 1d “solid angle”, which simply counts for the two different (distinguishable) orderings of the events namely,  $[\mathcal{F}\bar{\mathcal{F}}]$  and  $[\bar{\mathcal{F}}\mathcal{F}]$ .

One may evaluate the interactions between two fracton events separated by a distance  $\tau$  by directly computing the action on the ansatz in eq. (7.7). We find

$$S_{2\text{-event}} = 2S_{1\text{-event}} + V(\tau), \quad (7.9)$$

where  $V(\tau)$  characterize the interaction and are given by:

$$\begin{aligned} V(\tau) &= +\frac{4L\xi}{g^2} \frac{1}{\sinh^2(\frac{\xi\tau}{2})} \left(1 - \frac{\xi\tau}{\sinh(\xi\tau)}\right) \xrightarrow{\tau \gg \xi^{-1}} +\frac{16L\xi}{g^2} e^{-\xi\tau} \quad \text{for } [\mathcal{F}\mathcal{F}], \\ V(\tau) &= -\frac{4L\xi}{g^2} \frac{1}{\cosh^2(\frac{\xi\tau}{2})} \left(1 + \frac{\xi\tau}{\sinh(\xi\tau)}\right) \xrightarrow{\tau \gg \xi^{-1}} -\frac{16L\xi}{g^2} e^{-\xi\tau} \quad \text{for } [\mathcal{F}\bar{\mathcal{F}}]. \end{aligned} \quad (7.10)$$

In the latter expressions, we assumed the dilute fracton regime, where the separation  $\tau$  between fractons is much larger than fracton size  $r_{\mathcal{F}} \sim \xi^{-1}$ . Note that  $V(\tau)$  is exponentially small in  $\xi\tau$ , as advertised. In Euclidean space, where we map the proliferation of fractons into a dilute classical gas, these results can be interpreted as *repulsive* interactions between widely separated fracton-fracton events, and *attractive* interactions between widely separated fracton anti-fracton events.

According to the structure of the resurgence triangle of figure 11, the integral over the quasi-zero modes for  $[\mathcal{F}\mathcal{F}]$  should not yield an imaginary ambiguous part, while the one for  $[\mathcal{F}\bar{\mathcal{F}}]$  should have an ambiguity. The quasi-zero mode integrals are of the form

$$I_1(g^2) = \int d(\xi\tau) \left[ e^{-\frac{16L\xi}{g^2} e^{-\xi\tau}} - 1 \right] \quad \text{for } [\mathcal{F}\mathcal{F}], \quad (7.11)$$

$$I_2(g^2) = \int d(\xi\tau) \left[ e^{+\frac{16L\xi}{g^2} e^{-\xi\tau}} - 1 \right] \quad \text{for } [\mathcal{F}\bar{\mathcal{F}}], \quad (7.12)$$

where  $(-1)$  subtracts off the uncorrelated fracton-fracton events [4, 18, 19], and arises from the semi-classical expansion of the partition function automatically. In fact, not subtracting this factor would amount to double-counting the uncorrelated fracton events.

For the  $I_1(g^2)$  integral, using an integration by parts, we obtain

$$I_1(g^2) = \frac{16L\xi}{g^2} \int d(\xi\tau) \tau e^{-\left(\frac{16L\xi}{g^2} e^{-\xi\tau} + \xi\tau\right)} = \left( -\log \left[ \frac{g^2}{16L\xi} \right] - \gamma \right), \quad (7.13)$$

with  $\gamma$  once again the Euler-Mascheroni constant. This has its main support at the length scale  $\tau_*$ :

$$\tau_* = \frac{1}{\xi} \log \frac{16L\xi}{g^2} \gg \frac{1}{\xi}. \quad (7.14)$$

The integrand dies off for  $\tau \lesssim \tau_*$  because of the repulsion, and it dies off for  $\tau \gtrsim \tau_*$  because of the subtraction of uncorrelated events. Thus, we identify  $\tau_*$  as the characteristic size of the correlated 2-events. Since  $\tau_*\xi \gg 1$ , the correlated event is a *quasi-solution* to the equations of motion. Moreover, note that  $\tau_*$  is parametrically larger than the fracton size,



but parametrically smaller than the inter-fracton separation, which is in turn much smaller than the typical separation between 2-events. So we have a hierarchy of scales:

$$\begin{array}{ccccccc} r_{\mathcal{F}} & \ll & r_{[\mathcal{F}\mathcal{F}]} & \ll & d_{\mathcal{F}} & \ll & d_{[\mathcal{F}\mathcal{F}]} \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \frac{1}{\xi} & \ll & \frac{1}{\xi} \log(4S_F) & \ll & \frac{1}{\xi} e^{+S_F} & \ll & \frac{1}{\xi} e^{2S_F} \end{array} \quad (7.15)$$

This hierarchy of scales means that the use of semi-classical methods for 1-events *and* 2-events is simultaneously justified.

Now consider the  $I_2(g^2)$  integral. The integral is dominated at separations where we cannot meaningfully talk about isolated fracton-anti-fracton events. This is neither a bug nor an accident. Rather, it is an important feature! Recall that the  $\arg(g^2) > 0$  line is a Stokes line, along which the Borel resummation of the perturbative series is ambiguous. The left or right Borel resummations are not ambiguous, but there is a Stokes jump associated with the crossing of the Stokes line. The  $[\mathcal{F}\bar{\mathcal{F}}]$  configuration can mix with perturbation theory, and if perturbation theory is ambiguous with an ambiguity at order  $e^{-2S_F}$ , it is conceivable that configurations that can mix with perturbation theory may also have ambiguities of the same order (but not larger).

In fact, rather than trying to compute the  $[\mathcal{F}\bar{\mathcal{F}}]$  directly on a Stokes line, we should instead calculate the right and left  $[\mathcal{F}\bar{\mathcal{F}}]_{\pm}$  amplitudes. The simplest way to do this is to take  $g^2 \rightarrow -g^2$ , and then observe that  $I_2(-g^2) = I_1(g^2)$ , an integral already performed. The analytic continuation back to  $+g^2$  from  $-g^2$  through the complex  $g^2$  plane is two-fold ambiguous, with the result depending on whether we approach the positive real axis from above or from below. This method of evaluating the  $[\mathcal{F}\bar{\mathcal{F}}]$  amplitude is called the Bogomolny-Zinn-Justin (BZJ) prescription [4, 5].<sup>33</sup> Following the BZJ method, we find

$$I_{2,\pm}(g^2) = \left( -\log \left[ \frac{16\xi}{g^2} \right] - \gamma \pm i\pi \right). \quad (7.16)$$

So the correlated  $[\mathcal{F}\bar{\mathcal{F}}]_{\pm}$  events have an imaginary ambiguity.

**Remark on analytic continuation.** The reader may feel concerned by the following aspect of the previous derivation. It naively looks like we are taking  $g^2 \rightarrow -g^2$  for  $\mathcal{F}\bar{\mathcal{F}}$ , evaluating the QZM integration there, and then we take  $-g^2 \rightarrow +g^2$  again either clock-wise or anti-clock-wise, producing a two-fold ambiguous result. But, naively, we are not performing any analytic continuation while calculating amplitudes for  $\mathcal{F}\mathcal{F}$  events. Does this mean that we are treating the theory inconsistently, by treating one sector differently from the other?

As a matter of fact, we can (and should) move off the  $\arg(g^2) = 0$  line, the Stokes line, both for  $\mathcal{F}\bar{\mathcal{F}}$  as well as  $\mathcal{F}\mathcal{F}$ . The point is that we can do so by just taking  $g^2 \rightarrow g^2 e^{\pm i\epsilon}$ , where  $0 < \epsilon < \pi$ , evaluate the integral, and then come back to the  $\arg(g^2) = 0$  line by taking  $\epsilon \rightarrow 0$ . Then, we find that

$$[\mathcal{F}\bar{\mathcal{F}}]_{+} - [\mathcal{F}\bar{\mathcal{F}}]_{-} = 2i\pi \times \frac{2S_F}{\pi} e^{-2S_F},$$

---

<sup>33</sup>The BZJ prescription is very general, and is also relevant for theories with massless fermions, for some important early works on this application see [96, 97].

$$[\mathcal{FF}]_+ - [\mathcal{FF}]_- = 0, \quad (7.17)$$

so that the  $[\mathcal{F}\bar{\mathcal{F}}]$  amplitude has a Stokes jump while the  $[\mathcal{FF}]$  does not!

In fact, this is how things must work out if the resurgence triangle of figure 11 is the consistent characterization of the semi-classical regime of the principle chiral model or other QFTs studied so far [16, 18, 31]. As we will demonstrate in the next section, the ambiguity associated with  $[\mathcal{F}\bar{\mathcal{F}}]_{\pm}$  events cancels the ambiguity associated with the non-Borel-summability of the perturbative vacuum [0] which was calculated in eqs. (5.25), (7.3). According to the emergent topological structure,  $[\mathcal{FF}]$  is the lowest action configuration (the “level zero” or “primary”) in the sector with “charge” +2. If  $[\mathcal{FF}]$  were to have an ambiguity in its NP-part, that would imply that there must exist another configuration with lower action, and charge +2, which is impossible by construction.

**SU(N).** We now discuss the generalization of this analysis to  $SU(N)$ , where the minimal action fractons are labeled by the roots  $\alpha \in \Delta_1^{\text{aff}}$  of the  $\mathfrak{su}(N)$  algebra. Based on some general arguments we expect that the interaction potential between the fractons are given by

$$V^{(ij)}(\tau) = \frac{8L\xi}{g^2}(\alpha_i \cdot \alpha_j)e^{-\xi\tau}, \quad (7.18)$$

where  $\alpha_i \cdot \alpha_j = 2\delta_{i,j} - \delta_{i,j+1} - \delta_{i,j-1}$ , which reduces to eq. (7.10) for  $i = j$ .

The classification of 1- and 2-events in the bosonic  $SU(N)$  PCM on  $\mathbb{R} \times S^1$  is quite analogous to the classification of tunneling events in deformed Yang-Mills on  $\mathbb{R}^3 \times S^1$  and  $\mathbb{CP}^{N-1}$  on  $\mathbb{R} \times S^1$ . There are tunneling 1-events in field space associated with the change of the field by  $\alpha \in A_{N-1}$ , where  $\alpha$  is any element of root space. These are uncorrelated single events which can be embedded as exact solution into  $SU(N)$  PCM. At level-2 and above, there are a number of subtle issues which we address below.

- **Neutral bions:** these are correlated  $[\mathcal{F}_i\bar{\mathcal{F}}_i]_{\pm} = [\mathcal{B}_{ii}]_{\pm}$  events, a tunneling occurring in an  $SU(2)$  subgroup associated with root  $\alpha_i$  followed by an anti-tunneling associated with  $-\alpha_i$ . In the parametrization of eq. (4.17), we have

$$W_0 \rightarrow W_0 + \pi H_{\alpha_i} - \pi H_{\alpha_i}. \quad (7.19)$$

These exist for all positive entries of the extended Cartan matrix  $\hat{A}_{ii}$ . For  $\hat{A}_{ii} > 0$ , the interaction between the constituents is attractive. Since neutral bions have the quantum numbers of the perturbative vacuum, they can (and do) play a role in the cancellation of the ambiguities of perturbation theory. Furthermore, since  $\alpha_i - \alpha_i = 0$ , there is no single uncorrelated event associated with neutral bion. By its very nature, it is a correlated event. It also generates a non-perturbative contribution for the background holonomy potential (4.20) on top of the one-loop perturbative potential (4.22), and may play a role in the deconfinement phase transition (rapid crossover, for finite  $N$ ) in PCM.

- **Charged bions:** these are correlated  $[\mathcal{F}_i\bar{\mathcal{F}}_j] = \mathcal{B}_{ij}$  events that exist for all negative entries of the extended Cartan matrix,  $\hat{A}_{ij}$ , thus  $j = i \pm 1$ , which can be described as

$$W_0 \rightarrow W_0 + \pi H_{\alpha_i} - \pi H_{\alpha_j}$$

$$\underbrace{\longrightarrow}_{\text{e.g., } j=i+1} W_0 + \pi \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & \begin{pmatrix} 1_{i,i} & 0 & 0 \\ 0 & -2_{i+1,i+1} & 0 \\ 0 & 0 & 1_{i+2,i+2} \end{pmatrix} & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} . \quad (7.20)$$

Since  $\hat{A}_{i,i\pm 1} < 0$ , the interaction between the constituents is repulsive, and there is no ambiguity associated with these two-events. Furthermore, since  $\alpha_i - \alpha_{i\pm 1}$  is not a simple root, there is no single uncorrelated event associated with charged bion. Like the neutral bion, by its very nature, the charged bions are fundamentally correlated events. In theories with fermions, they play the leading role in the generation of the mass gap.

- **Higher action elementary fractons:** for all roots of the Lie algebra of  $\mathfrak{su}(N)$ , an exact solution can be embedded into an  $SU(2)$  subgroup of  $SU(N)$ . In particular this includes the roots  $\alpha_i + \alpha_{i+1} \in \Delta_2^{\text{aff}}$  associated with the tunneling event

$$W_0 \rightarrow W_0 + \pi H_{\alpha_i + \alpha_{i+1}} \longrightarrow W_0 + \pi \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & \begin{pmatrix} 1_{i,i} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1_{i+2,i+2} \end{pmatrix} & & \\ & & & \ddots & \\ & & & & 0 \end{pmatrix} . \quad (7.21)$$

Since this tunneling event is associated with a root, unlike the neutral and charged bions events, it should be considered as an elementary event, even though it carries two units of action  $2S_F$ . There is no ambiguity in the NP-part of the elementary events, while P-fluctuations around them will always have ambiguities.

- **Higher action composite fracton-fracton pairs:** consider two roots  $\alpha$  and  $\beta$  whose sum  $\alpha + \beta$  is not a root itself, for example,  $2\alpha_i$ . In all such cases, the tunneling event must be seen as composite. This is because one cannot embed a simple  $SU(2)$  fracton associated with the sum  $\alpha + \beta$  into  $SU(N)$ . For *all* such correlated events, the interaction between the constituents is *always* repulsive, and so there is no ambiguity associated with the NP-part of such events:

$$W_0 \rightarrow W_0 + \pi H_\alpha + \pi H_\beta . \quad (7.22)$$

For example, for  $SU(N)$ ,  $N \geq 3$ ,  $\mathcal{F}_{\alpha_1 + \alpha_2}$  is a single event with action  $2S_F$  while  $[\mathcal{F}_{\alpha_1} \mathcal{F}_{\alpha_1}]$  is a correlated event with action  $2S_F$ . This is because  $\alpha_1 + \alpha_2$  is a root, while  $2\alpha_1$  is not a root.

With these remarks in mind, we observe that the rows of the resurgence triangle of figure 11 get “ramified” according to a Lie algebraic structure. For example, the first three rows of the resurgence triangle in figure 11 should really have been written as:

$$\begin{aligned}
 & [0] \\
 & [\bar{\mathcal{F}}_{\alpha_i}] \qquad \qquad \qquad [\mathcal{F}_{\alpha_i}] \qquad \qquad \qquad (7.23) \\
 & \{[\bar{\mathcal{F}}_{\alpha_i}^2], [\bar{\mathcal{F}}_{\alpha_i+\alpha_{i+1}}]\} \qquad \{[\mathcal{F}_{\alpha_i}\bar{\mathcal{F}}_{\alpha_i}]_{\pm}, [\mathcal{F}_{\alpha_i}\bar{\mathcal{F}}_{\alpha_{i\pm 1}}]\} \qquad \{[\mathcal{F}_{\alpha_i}^2], \{[\mathcal{F}_{\alpha_i+\alpha_{i+1}}]\}\}
 \end{aligned}$$

This carries rather refined information about the structure of the semi-classical expansion. There are essentially *six* type of events at order  $e^{-2S_F}$  in the semiclassical expansion. This is the same order as the leading ambiguity in perturbation theory. Only *one* class out of the six, the neutral bion, participates in fixing the ambiguity of perturbation theory around the perturbative vacuum. The other classes of events have other roles in the rich inner life of the theory.

Despite the fact that all of these events appear at the same order in the semi-classical expansion, ( $e^{-2S_F}$ ), their amplitudes<sup>34</sup> differ in crucial ways:

$$\begin{aligned}
 [\mathcal{F}_{\alpha_i}^2] &= (-\log(4S_F) - \gamma) \frac{2S_F}{\pi} e^{-2S_F}, \\
 [\mathcal{F}_{\alpha_i+\alpha_{i+1}}] &= \sqrt{\frac{4S_F}{\pi}} e^{-2S_F}, \\
 [\mathcal{F}_{\alpha_i}\bar{\mathcal{F}}_{\alpha_i}]_{\pm} &= (-\log(4S_F) - \gamma \pm i\pi) \frac{2S_F}{\pi} e^{-2S_F}, \\
 [\mathcal{F}_{\alpha_i}\bar{\mathcal{F}}_{\alpha_{i\pm 1}}] &= (-\log(2S_F) - \gamma) \frac{2S_F}{\pi} e^{-2S_F}. \qquad (7.24)
 \end{aligned}$$

Even the two events which carry “charge” 2 under the emergent topological structure,  $\{[\mathcal{F}_{\alpha_i}^2], [\mathcal{F}_{\alpha_i+\alpha_{i+1}}]\}$  differ. The reason for this, as explained above, is that  $2\alpha_i$  is not a root, and consequently the corresponding event is a correlated one. On the other hand,  $\alpha_i + \alpha_{i+1}$  is a root, and is associated with a single tunneling within the corresponding  $SU(2)$  subgroup of  $SU(N)$ .

The expressions given in eq. (7.24) do not include perturbative fluctuations around the NP saddles. Incorporating the perturbative fluctuations around the 2-events, we get

$$[\mathcal{F}\bar{\mathcal{F}}]_{\pm} \times \Phi_{[\mathcal{F}\bar{\mathcal{F}}]} \equiv \left( -\log \left[ \frac{32\pi}{g^2 N} \right] - \gamma \pm i\pi \right) \frac{16}{g^2 N} e^{-\frac{16\pi}{g^2 N}} \times \sum_{n=0}^{\infty} a_n^{[\mathcal{F}\bar{\mathcal{F}}]} g^{2n}, \qquad (7.25)$$

for the  $[\mathcal{F}\bar{\mathcal{F}}]_{\pm}$  event, and similar expression for other events.

## 7.2 Reality of resurgent trans-series for real $\lambda$ and BE-summability

The principal chiral model is a matrix field theory with a real action and a stable ground state. Hence its partition function must be real and unambiguous, similarly to our toy

<sup>34</sup>As one can see, most of these amplitudes contain logarithms. Similar logarithms also appear in the case of resonances in resurgent trans-series [15, 37, 38].

example eq. (2.1). In the preceding sections, however, we have calculated the leading perturbative and non-perturbative contributions to the the partition function, and have found that:

1. Perturbation theory is non-Borel resummable on the  $\arg(g^2) = 0$  Stokes line, meaning that the Borel sum of the perturbative series has a two-fold ambiguous imaginary part.
2. At second order in semi-classical expansion, we have calculated the neutral bion amplitude (via the BZJ prescription) and showed that it also has a two-fold ambiguous imaginary part.

Each of these ambiguities would by themselves be disastrous, indicating that the theory is not well-defined. The framework of resurgence suggests the resolution. Perturbation theory by itself is indeed not well-defined. The semi-classical expansion by itself is not well-defined either. However, neither is a direct physical observable, only their sum is. In fact, the ambiguities at order  $e^{-2S_F}$  cancel exactly to yield a result which is ambiguity free up to order  $e^{-4S_F}$ . Resurgence is the statement that these cancellations repeat order by order in the resurgent trans-series expansion for every physical observable.

Using eq. (7.3) and eq. (5.26), the right/left Borel resummation of the perturbative series can now be written in terms of fracton amplitudes as:

$$\mathcal{S}_{0\pm}\mathcal{E} = \text{Re } \mathcal{S}_0\mathcal{E} \pm \frac{s}{2}[\mathcal{F}][\bar{\mathcal{F}}], \quad s = -4\pi i, \quad (7.26)$$

where  $s = -4\pi i$  is the purely imaginary Stokes constant (analytic invariant) of the problem. Similarly, using eq. (7.4), the right/left neutral bion amplitudes are given by

$$[\mathcal{F}\bar{\mathcal{F}}]_{\pm} + [\bar{\mathcal{F}}\mathcal{F}]_{\pm} = 2(-\log(4S_F) - \gamma)[\mathcal{F}][\bar{\mathcal{F}}] \mp \frac{s}{2}[\mathcal{F}][\bar{\mathcal{F}}]. \quad (7.27)$$

Note that the *same* Stokes constant appears in the imaginary ambiguous part of the neutral bion amplitude.

The ambiguities in both quantities are a manifestation of the fact that we are performing an expansion on a Stokes line. Consequently, the perturbative series is non-Borel summable, and exhibits a Stokes jump, which is *mirrored* by the jump in the neutral bion amplitude, leading to the cancellation of ambiguities:

$$\text{Im} [\mathcal{S}_{0\pm}\mathcal{E} + ([\mathcal{F}\bar{\mathcal{F}}]_{0\pm} + [\bar{\mathcal{F}}\mathcal{F}]_{0\pm})] = 0. \quad (7.28)$$

This is the counterpart of the cancellation of ambiguities we saw in  $d = 0$  example eq. (2.28), and it is an explicit realization of the median resummation [9, 15] and BE-summability.

The sum is ambiguity free up to higher order effects, and the non-canceling terms are of the form

$$\mathcal{S}_{0\pm}\mathcal{E} + [\mathcal{F}\bar{\mathcal{F}}]_{\pm} + [\bar{\mathcal{F}}\mathcal{F}]_{\pm} = \text{Re } \mathcal{S}_0\mathcal{E} + 2 \left( -\log \left[ \frac{32\pi}{g^2 N} \right] - \gamma \right) \frac{16}{g^2 N} e^{-\frac{16\pi}{g^2 N}}. \quad (7.29)$$

Physically, this quantity is the average of the ground state and first excited state. The difference of the first excited state and the ground state is the mass gap, and will be discussed in the next section.

Despite the fact that we have only shown eq. (7.28) to be true at order  $O(e^{-2S_F})$ , resurgence actually implies that

$$\text{Im} \left( \mathcal{S}_{0\pm} \Phi_0 + 2[\mathcal{F}\overline{\mathcal{F}}]_{0\pm} \times \mathcal{S}_{0\pm} \Phi_{[\mathcal{F}\overline{\mathcal{F}}]} \right) = 0 \text{ up to } O(e^{-4S_F}). \quad (7.30)$$

This expression can be inverted, using eq. (2.36), to obtain the non-alternating part of the late terms in the perturbative expansion around the perturbative vacuum:

$$a_n^{[0]} \sim -\frac{2}{\pi} \frac{\Gamma(n+1)}{(2S_F)^n} \left[ a_0^{[\mathcal{F}\overline{\mathcal{F}}]} + a_1^{[\mathcal{F}\overline{\mathcal{F}}]} \frac{(2S_F)}{n} + a_2^{[\mathcal{F}\overline{\mathcal{F}}]} \frac{(2S_F)^2}{n(n-1)} + \dots \right] + \dots \quad (7.31)$$

which indeed agrees with eq. (5.23). Note that this expression is very much in the same spirit as our zero dimensional example eq. (2.37).

Finally, although we do not attempt to derive it here (it is beyond the scope of our present work), we believe that all of the formal series appearing in our problem form an infinite dimensional algebra, the resurgence algebra, closed under the action of the singularity (“alien”) derivative. For example, we expect to have the relations

$$\Delta_{2S_F} \Phi_0 = s_1 \Phi_{[\mathcal{F}\overline{\mathcal{F}}]}, \quad (7.32)$$

$$\Delta_{4S_F} \Phi_0 = s_2 \Phi_{[\mathcal{F}^2\overline{\mathcal{F}}^2]}, \quad (7.33)$$

$$\Delta_{-2S_F} \Phi_{[\mathcal{F}\overline{\mathcal{F}}]} = \tilde{s}_1 \Phi_0, \quad (7.34)$$

...

where  $s_i, \tilde{s}_1$  are Stokes constants and  $\Delta$  are Ecalle’s alien derivatives. The first one of these relations means that the action of the alien derivative on  $\Phi_0$  at  $2S_F$  yields the formal perturbative series around the  $[\mathcal{F}\overline{\mathcal{F}}]$  saddle point. The second relation means that the action of the alien derivative on  $\Phi_0$  at  $4S_F$  yields the formal perturbative series describing fluctuations around the  $[\mathcal{F}^2\overline{\mathcal{F}}^2]$  saddle. The third relation means that the action of the alien derivative on  $\Phi_{[\mathcal{F}\overline{\mathcal{F}}]}$  at  $-2S_F$  should yield the formal series around the perturbative saddle. It would be very interesting to derive the entire resurgence algebra for QM or QFT.

### 7.3 Lefschetz thimbles and geometrization of ambiguity cancellation

In this section, we briefly sketch our conjecture of the geometric reasons for the ambiguity cancellation, in connection with semi-classics and Lefschetz thimbles.

Our results strongly suggest that the geometric explanation of the cancellations of the imaginary parts is very similar in the  $d = 0$  and  $d = 1, 2$  examples. It is tied up with the steepest descent (or semi-classical) expansion. The lesson of resurgence theory is that whenever we consider a semi-classical expansion, we should in fact always work with a *complexified* version of the path integral. That this is the case becomes clear when one appreciates the nature of our approach in the preceding sections, which involved analytically continuing  $g$ . So to make sense of the ambiguities and use resurgence theory, we already had to work with a complexified version of the path integral.

More specifically, in our example we first have to generalize

$$U(x) \in \text{SU}(N) \longrightarrow Z(x) \in \text{SL}(N, \mathbb{C}), \quad (7.35)$$

where  $\text{SL}(N, \mathbb{C})$  is the complex special linear group. In a lattice version of the model, the infinite-dimensional path integral is regularized into a finite dimensional integral,<sup>35</sup> and the connection of resurgence theory to the Lefschetz thimble decompositions of integration cycles (which was explained in detail in [52, 100]) becomes particularly clear. Let the two dimensional lattice have  $L_1 L_2$  sites. Then the original partition function is an integral over  $[\text{SU}(N)]^{L_1 L_2}$ , an  $(N^2 - 1)L_1 L_2$  dimensional space. When we complexify, the dimension of space is doubled,

$$\dim([\text{SL}(N, \mathbb{C})]^{L_1 L_2}) = 2(N^2 - 1)L_1 L_2. \quad (7.36)$$

But since our goal is to find an analytic continuation of the original integral, we must pick special integration cycles within the complexified field space. Hence the integration runs over a sub-manifold  $\Sigma$  which is again  $(N^2 - 1)L_1 L_2$  dimensional, same as original purely real cycle.

For each saddle in the discretized theory, there exists a unique Lefschetz thimble attached to it,  $\mathcal{J}_{[\text{saddle}]}(\theta)$ , whose structure depends on the phase  $\arg(g^2) = \theta$ . In fact, in the finite dimensional case,  $\mathcal{J}_{[\text{saddle}]}(\theta)$  lives in  $[\text{SL}(N, \mathbb{C})]^{L_1 L_2}$  and its dimension is half of the complex space. The analytic continuation of the original integration cycle may be expressed as a linear combination of Lefschetz thimbles:

$$\Sigma(\theta) = \sum_{i \in \text{saddles}} n_i \mathcal{J}_{[i]}(\theta), \quad (7.37)$$

and  $n_i$  are piece-wise constant (in Stokes wedges) but jump at the Stokes lines.

For example, consider the Lefschetz thimble attached to perturbative vacuum,  $\mathcal{J}_{[0]}(\theta)$  at  $\theta = 0^+$  and  $\theta = 0^-$ . The integration cycle must have a dramatic change (upon crossing a Stokes line) which does not alter the real part of the integration, but leads to a jump in the imaginary part. Figure 3 provides a cartoon of this phenomenon for an ordinary integral. In fact, figure 3 is related to our present problem via a dimensional reduction, in the one-site limit of the lattice model  $L_1 = L_2 = 1$  with a twisted reduction in the  $L_2$  direction. Figure 3 shows that the real part of the cycle must remain unaltered upon a Stokes jump at  $\theta = 0$ , while the “tail” in imaginary direction is reversed. We believe that a generalization of these phenomena to the continuum limit, where the integrals become infinite dimensional, is operative in the path integrals of quantum mechanics and field theory,<sup>36</sup> although the explicit construction of the Lefschetz thimbles in the infinite dimensional case may be quite subtle [100, 102] and may require a generalization of the techniques explained in [52, 100].

We expect infinitely many thimbles associated with the infinite number of saddle points of the path integral in the semi-classical domain. In fact, we can make the thimble version of our resurgence triangle (in the unramified notation to avoid clutter):

$$\mathcal{J}_{[0]}$$

$$\mathcal{J}_{[\mathcal{F}_1]} \quad \mathcal{J}_{[\mathcal{F}_1]}$$

---

<sup>35</sup>Note that a deformation of the domain of integration to a complexified extension of the path integral, and subsequently integrating over Lefschetz thimbles where the phase is stationary, might be of extreme importance in theories affected by the sign problem [98, 99].

<sup>36</sup>See also [101].

$$\begin{array}{cccc}
\mathcal{J}_{[\bar{\mathcal{F}}_2]} & \mathcal{J}_{[\bar{\mathcal{F}}_1\mathcal{F}_1]} & \mathcal{J}_{[\mathcal{F}_2]} & \\
\mathcal{J}_{[\bar{\mathcal{F}}_3]} & \mathcal{J}_{[\bar{\mathcal{F}}_2\mathcal{F}_1]} & \mathcal{J}_{[\bar{\mathcal{F}}_1\mathcal{F}_2]} & \mathcal{J}_{[\mathcal{F}_3]} \\
& \vdots & & 
\end{array} \tag{7.38}$$

The combination of the solvable  $d = 0$  dimensionally reduced model and the non-Borel summability of the perturbative series eq. (7.3) tells us that the integral over just the P-thimble  $\mathcal{J}_{[0]}(\theta)$  will have a pathological  $\theta = 0$  limit, resulting in a two-fold ambiguous result, depending on the direction of the approach to  $\theta = 0$  in the complex  $g^2$  plane. We have already seen that this pathology can be fixed by integration over other NP-thimbles. The integration over the fracton-anti-fracton thimble  $\mathcal{J}_{[\mathcal{F}\bar{\mathcal{F}}]}(0^\pm)$  contributes at the same order as the ambiguity of the P-thimble on the  $\theta = 0$  direction. In fact, we expect

$$\int_{\mathcal{J}_{[0]}(0^-) + \mathcal{J}_{[\mathcal{F}\bar{\mathcal{F}}]}(0^-)} DZ(x_1, x_2) e^{-S[Z(x_1, x_2)]} \tag{7.39}$$

to be ambiguity free at order  $e^{-2S_F}$ , but to have some ambiguities at order  $e^{-4S_F}$ , which are cancelled thanks to the fact that the full integration contour includes thimbles which pass through the appropriate higher-action NP saddle points, and so on. It would be very interesting to understand the structure of all the thimbles and Stokes phenomena in this problem. We believe that this would provide a geometric understanding of the intricate relations between P and NP data which leads to the cancellation of ambiguities in resurgent trans-series, as described in e.g. [15].

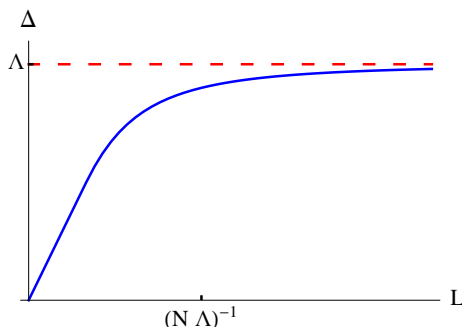
## 8 Mass gap flow and Borel flow

In one of the first works on renormalons, 't Hooft speculated that they may be connected to the mass gap and confinement in gauge theories [22]. Using resurgence, we find a refinement and a confirmation of this idea in a semi-classical regime continuously connected to the strongly coupled regime of gauge theories and non-linear sigma models.

In every semi-classically calculable example studied so far, it turns out that the mass gap is due to *half* a renormalon in the semi-classical domain [16–19, 31]. In deformed Yang-Mills on small  $\mathbb{R}^3 \times S^1$ , the mass gap is due to monopole-instantons  $\mathcal{M}$  [21], while  $[\mathcal{M}\bar{\mathcal{M}}]_\pm$  yields the leading semi-classical realization of the renormalons [16, 17]. In  $\mathcal{N}=1$  SYM and QCD(adj) on small  $\mathbb{R}^3 \times S^1$ , the mass gap is due to magnetic bions  $\mathcal{B}$  [29], while the leading semi-classical realization of the renormalon is the neutral bion  $[\mathcal{B}\bar{\mathcal{B}}]_\pm$  [16, 17], etc.

In the SU(2) PCM, the evaluation of the mass gap in the small- $L$  regime follows very closely the calculation of the mass gap in the  $\mathbb{CP}^1$  model. In section 7.2, we already argued that the weak coupling realization of the renormalon is again the neutral bion,  $[\mathcal{F}\bar{\mathcal{F}}]_\pm$ . Below, we show that the mass gap at leading order in semi-classical expansion in the small  $\mathbb{R} \times S^1$  regime is due to fractons  $\mathcal{F}$ . This is also true for the SU( $N$ ) model in the  $\frac{NL\Lambda}{2\pi} \ll 1$  small circle limit adiabatically connected to  $\mathbb{R}^2$ .





**Figure 12.** *Mass gap flow:* a sketch of the simplest behavior for mass gap the mass gap as a function of  $L\Lambda$ , interpolating between the known behaviors at small and large  $L\Lambda$ . In the small  $\mathbb{R}^1 \times S^1$  regime, corresponding to  $\frac{L\Lambda}{2\pi} \ll 1$  is semi-classically calculable. At leading order, it is a one fracton effect. On  $\mathbb{R}^2$  or large  $\mathbb{R}^1 \times S^1$ , a reliable analytical method which can address the mass gap question is at present unknown and this figure is a sketch of the expected flow. This paper collects evidence to support the conjecture that the small  $\mathbb{R}^1 \times S^1$  with  $\mathbb{Z}_N$ -twisted boundary conditions is adiabatically connected to the theory on  $\mathbb{R}^2$ .

The mass gap is defined as the energy required to excite the system from the ground state  $E^{(0)}$  to the first excited state  $E^{(1)}$ . As discussed in section 5, in the small- $L$  regime, we can work with the Hamiltonian eq. (5.17) which describes the dynamics of a small- $L$  EFT with zero KK-momentum, or of a small- $L$  EFT with  $-1$  units of winding number on the  $S^1$ . In the Born-Oppenheimer approximation, we can further focus on eq. (5.19), since the states which carry non-zero  $P_{\phi_1}$  and  $P_{\phi_2}$  momentum acquire a gap of order  $g^2/L$ , while (as we show below) the low lying states of eq. (5.17) are split by a non-perturbatively small amount, justifying the use of the Born-Oppenheimer approximation.

In the  $SU(2)$  PCM the ground state is two-fold degenerate to all orders in perturbation theory. This degeneracy is lifted by non-perturbative fracton effects. The two lowest lying eigenstates described by eq. (5.17) have the quantum numbers  $|\pm, n_1 = 0, n_2 = 0\rangle$ , where  $\pm$  are the eigenvalues of the parity operator  $P$  which acts on the polar coordinate  $\theta$  in the Hopf parametrization as  $\theta \rightarrow \pi - \theta$ . At leading non-perturbative order in the semiclassical expansion the  $SU(2)$  PCM mass gap is then given by

$$\Delta = (E_- - E_+) = \xi(2\mathcal{F}_1 + 2\mathcal{F}_2) = \frac{4\pi}{L} \sqrt{\frac{16}{g^2 N}} e^{-\frac{4\pi}{g^2}} \sim \Lambda(\Lambda L). \quad (8.1)$$

For  $SU(N)$ , parametrically,<sup>37</sup> the mass gap is of order

$$\Delta = (E_- - E_+) \sim \frac{2\pi}{LN} \sqrt{\frac{16}{g^2 N}} e^{-\frac{8\pi}{g^2 N}} \sim \frac{1}{LN} e^{-S_{\text{uniton}}/N} \sim \Lambda(\Lambda LN), \quad (8.2)$$

<sup>37</sup>At leading order in the  $SU(N)$  PCM, the mass gap gets contributions from the  $N$  minimal-action fractons which describe tunneling between the ground state and the  $N$  directions in field space parametrized by of the  $N$  affine simple roots. To compute an exact expression for the mass gap in the  $SU(N)$  case, one must diagonalize the resulting “tunneling matrix” and compute its smallest eigenvalue. The expression we show in eq. (8.2) is the parametric form of the result which this calculation would give.

in the weak coupling semi-classical regime  $\frac{NLA}{2\pi} \ll 1$ .<sup>38</sup> Note that the mass gap at small- $L$  is  $N$ -independent in the large  $N$  limit, just as it is on  $\mathbb{R}^2$ , because when  $NLA \ll 1$  and  $N \gg 1$ ,  $L$  scales as  $L \sim 1/N$ . This lends further support for the claim that our small- $L$  limit is adiabatic. At the boundary of the region of validity of the semi-classical regime  $NLA \sim 1$  where we can no longer rely on semiclassics, we observe that the mass gap acquires a strong scale value  $m_g \sim \Lambda$ . In the strong coupling regime,  $NLA \gg 1$ , we expect the mass gap to be independent of the size of the circle. In fact, the onset of  $\mathbb{R}^2$  behavior at the compactification scale  $(\Lambda/N)^{-1}$  rather than  $\Lambda^{-1}$  is a hallmark of large  $N$  volume independence, which can be shown to apply to the theory we are working with. Therefore, provided we are given a value  $m_0$  for the gap at some  $NL_0\Lambda \gg 1$ , as  $N$  is varied  $m_0$  can only change by order  $O(1/N^2)$  corrections in the  $SU(N)$  model. In other words, we expect the mass gap to plateau and remain fixed in this regime as shown in figure 12, which shows a sketched picture of the expected form of the mass gap as a function of  $L$ .

The connection with renormalons should now be clear. The field configurations  $\mathcal{F}_i$  that give rise to the mass gap at order  $e^{-S_F}$  then produce the leading renormalon singularities at the next order of the semi-classical expansion  $e^{-2S_F}$ . So the mass gap is tied to “half” of a renormalon. This is a concrete realization of ’t Hooft’s idea [22].

**Borel flow.** The idea of Borel flow is a more abstract version of the mass gap flow. Borel flow is tied up with all non-perturbative observables in the problem. The IR-singularities in the Borel plane on the small  $S_L^1 \times \mathbb{R}$  regime and on  $\mathbb{R}^2$  are located at

$$\begin{aligned} t_m^{S_L^1 \times \mathbb{R}} &= \frac{16\pi}{N} m, & m = 1, 2, \dots \\ t_m^{\mathbb{R}^2} &= \frac{8\pi}{N} m, & m = 1, 2, \dots \end{aligned} \quad (8.3)$$

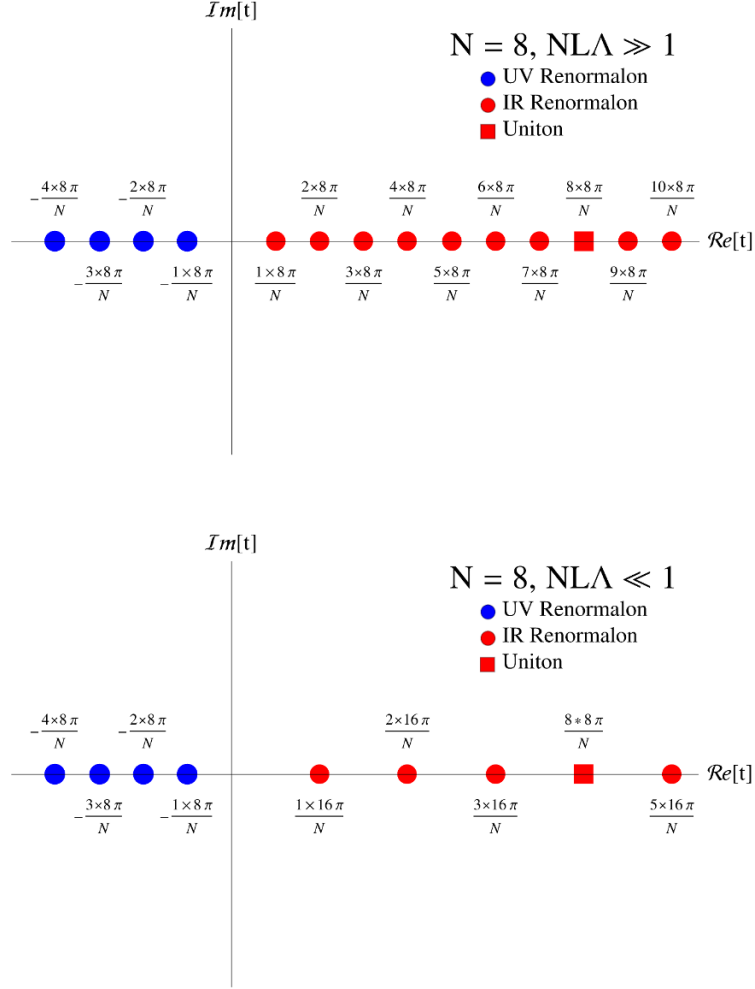
while the location of the UV-renormalon singularities remain unchanged no matter the value of  $L$ . See figure 13. The most dominant singularities ( $m = 1$ ) lead to ambiguities of order

$$\begin{aligned} S_L^1 \times \mathbb{R} : & \quad \pm i e^{-\frac{16\pi}{g^2(1/LN)N}} \sim \pm i(\Lambda LN)^4, \\ \mathbb{R}^2 : & \quad \pm i e^{-\frac{8\pi}{g^2(Q)N}} \sim \pm i(\Lambda^2 Q^{-2}), \end{aligned} \quad (8.4)$$

where at small  $S_L^1 \times \mathbb{R}$ , the ’t Hooft coupling is evaluated at distance  $LN$ , while on  $\mathbb{R}^2$ , it is determined at a high (Euclidean) momentum scale  $Q$  (entering through an OPE with an external momentum insertion  $Q$ .) The crucial point is that  $Q \gg \Lambda$  and  $\frac{1}{LN} \gg \Lambda$  so that the coupling is weak at the scale of  $Q$  and  $\frac{1}{LN}$ , and in both cases, this gives a control parameter over the small NP-induced term. On  $S_L^1 \times \mathbb{R}$  these ambiguities are cancelled respectively by the ambiguity in neutral bion events shown in eq. (7.5), while on  $\mathbb{R}^2$  they are cancelled by the ambiguity in the condensate eq. (3.8).

---

<sup>38</sup>The appearance of the non-perturbative factor in the mass gap is the major difference with respect to thermal compactification. In thermal theory at small  $S_\beta^1 \times \mathbb{R}$ , usual KK-reduction works, and the gap is given  $\Delta_{\text{gap}}^{\text{thermal}} \sim \frac{g^2}{L}$ . The thermal low energy theory does not remember its two dimensional origin, in contrast to the adiabatic small- $L$  limit we have constructed.



**Figure 13.** *Borel flow:* (Top) The expected structure of the Borel plane on  $\mathbb{R}^2$ . (Bottom) The Borel plane structure on small  $\mathbb{R}^1 \times S^1$ . The small- $L$  effective field theory (in which small- $L$  physics is integrated out) does not capture the UV renormalon structure by construction, but does capture the IR-renormalon singularities of the small- $L$  theory.

In the semiclassical regime, the mass gap is generated by half a renormalon, i.e., a fracton. The dimensionless mass gap is  $m_g L N$ , and takes the form

$$S_L^1 \times \mathbb{R} : \quad m_g L N \sim (\Lambda^2 L^2 N^2) \longrightarrow m_g \sim \Lambda(\Lambda L N), \quad (8.5)$$

demonstrating explicitly the relation between the mass gap and the leading renormalon. Compare this with the first line of eq. (8.4). If we accept the behavior we have seen in the semi-classical regime as a rough guide to the behavior we should expect in the strongly coupled domain, we would deduce that the dimensionless mass gap (now measured in units of some external large momentum  $Q$ ) behaves as

$$\mathbb{R}^2 : \quad m_g Q^{-1} \sim (\Lambda Q^{-1}) \longrightarrow m_g \sim \Lambda, \quad (8.6)$$

which is a sensible result on  $\mathbb{R}^2$ . Thus, we are tempted to sharpen 't Hooft speculation. In asymptotically free non-linear sigma models (including for instance  $\mathbb{CP}^{N-1}$ ,  $O(N)$ , Grass-

mannian, and principal chiral model-type matrix field theories) there may exist a quantitative relation between the mass gap of the theory and the location of the first renormalon singularity for all values of  $L$ .

As illustrated in figure 13 the IR-singularities in the Borel plane are twice as dense on  $\mathbb{R}^2$  with respect to  $\mathbb{R} \times S_L^1$ . The crucial point for adiabatic continuity is the fact that in both regimes the singularities are spaced by units of  $\sim \frac{1}{N}$ . Our framework strongly suggests that as we dial the radius from small to large, the singularities must exhibit a flow towards the origin rendering them twice as dense. The same phenomenon should also take place in deformed Yang-Mills, in which the dilution factor between the weak coupling regime and strong coupling regime is  $\frac{11}{3}$  [17].

Clearly, the flow of the singularities in the Borel plane as the radius is dialed, i.e, the Borel flow, and the flow of the mass gap as the theory is dialed from a weak coupling to strong coupling are intimately related. They are very likely manifestations of the same underlying dynamics. We believe that developing a thorough understanding of these flow equations would constitute a major step towards the solution of the mass gap problem in a large variety of asymptotically free theories.

## 9 Discussion and prospects

We have employed resurgence and adiabatic continuity to give a classification of the P- and NP- saddles in the principal chiral model. Due to insights from various techniques, such as lattice Monte Carlo calculations and integrability, these theories were believed to have highly non-trivial similarities to Yang-Mills theory. But the theory had no known NP-saddles, except for the uniton saddle discussed mostly in the mathematics literature [56], whose role in the quantum version of the PCM never became clear. So, from a semi-classical point of view, the only relevant saddle seemed to be the perturbative vacuum. This appeared to be a dramatic difference from Yang-Mills theory.

In this work we have constructed an infinite class of NP-saddles such as the fractons, as well as the zoo of correlated events, which turn out to play a crucial role in the dynamics of the PCM. The fractons lead to the generation of the mass gap of the theory, while neutral bions (correlated fracton-anti-fracton events) lead to the semi-classical realization of the IR renormalons in the weak coupling calculable regime  $\frac{N\Lambda}{2\pi} \ll 1$ . Our analysis is inspired by the recent treatment of the 2D  $\mathbb{CP}^{N-1}$  model in [18, 19], and has many parallels with the analysis of gauge theories on  $\mathbb{R}^3 \times S^1$  initiated in [20, 21, 29], and recently revisited in the context of resurgence in [16, 17].

In the present work, in the small- $L$  regime, we were able to demonstrate the existence of fractons, whose action is  $S_F = S_{\text{uniton}}/N$ , by three independent methods:

- Large order analysis of the small- $L$  perturbative series describing fluctuations around the perturbative saddle point implies, via resurgence, that the non-perturbative completion of the problem involves saddles with action  $2S_F$ . Moreover, the notion of emergent topology encoded in the resurgence triangle leads to the conclusions that there must also be other saddles, with action  $S_F$ , which are precisely the fractons.

- The effective field theory obtained via adiabatic continuity, which is just a particular quantum mechanical theory, allows a simple study of the NP-saddles. The NP saddles with the smallest action have actions  $S_F$ , and they come in  $N$  different types.
- When the  $\mathbb{Z}_N$  symmetric background holonomy is turned on we have seen that the uniton splits into  $N$  lumps, each of which carries an action  $S_F$ , as nicely shown by the plots in figure 9.

This last phenomenon is morally similar to the splitting of calorons (periodic instantons) into  $N$ -monopole instantons in gauge theories [24, 25], but it is again worth emphasizing that the PCM does not have any instantons.

### 9.1 Future directions

There are a large number of interesting possible extensions of the study we have performed here. A few of them are:

1. **WZW term and sign problem:** addition of a WZW-term to the action modifies the IR dynamics on  $\mathbb{R}^2$  and introduces a sign problem if one were to try to attack the system using Monte Carlo simulations. It would be interesting to study this system on the calculable small- $L$  regime.
2. **Borel and mass gap flows:** an understanding of the non-perturbative dynamics on  $\mathbb{R}^2$  can perhaps be achieved by summing the resurgent trans-series at small- $L$ , and extrapolating the sum to large- $L$ . Less ambitiously, one may develop an understanding of an infinitesimal version of Borel flow from a more detailed analysis of the construction of the action of the small- $L$  effective field theory, by exploring the effects of the higher-derivative terms in the action induced by integrating out high-momentum modes.
3. **PCM with fermions:** the PCM model, like other non-linear sigma models, is asymptotically free for any number of fermion flavors (unlike QCD-like theories on  $\mathbb{R}^4$ .) It would be interesting to perform a detailed investigation of the impact of fermions on the dynamics, and to determine the boundary between the confining and IR-conformal regimes.
4. **Large- $N$  reduced model:** we can reduce the theory with  $N_f \geq 1$  Majorana fermions to a one-site lattice theory, by imposing 't Hooft twisted boundary conditions. Let  $\Phi$  represent the bosonic/fermionic degrees of freedom  $U, \psi_i$  with  $\mathbf{e}_1, \mathbf{e}_2$  the two lattice vectors and then reduce the matrix model to one-site by imposing

$$\Phi(\mathbf{x} + \mathbf{e}_1) = \Omega_1 \Phi(\mathbf{x}) \Omega_1^\dagger, \quad \Phi(\mathbf{x} + \mathbf{e}_2) = \Omega_2 \Phi(\mathbf{x}) \Omega_2^\dagger, \quad \Omega_1 \Omega_2 = e^{i \frac{2\pi}{N}} \Omega_2 \Omega_1, \quad (9.1)$$

where the last condition ensures compatibility of the fields at  $\mathbf{x} \sim \mathbf{x} + \mathbf{e}_1 + \mathbf{e}_2$ . It would be useful to examine the dynamics of the associated reduced large- $N$  model.

5. **Renormalons in the large- $N$  reduced model:** in the single-site reduction which is enabled by large- $N$  volume independence, the space-time volume of the theory on  $\mathbb{R}^2$  is mapped to the matrix size ( $N = \infty$ ) of the 1-site matrix model. Within planar perturbation theory there is an exact mapping between summation over spacetime momenta, and summation over the adjoint  $SU(N)$  indices. This implies that there must exist a matrix field interpretation for both IR and UV renormalons. It would be interesting to understand this in detail.

## Acknowledgments

We are very grateful to Ines Aniceto, Gokce Basar, Gerald Dunne, David Gross, Daniel Harlow, Vladimir Kazakov, Peter Koroteev, Sungjay Lee, Jonathan Maltz, Nick Manton, Soo-Jong Rey, Ricardo Schiappa, Misha Shifman, Edward Witten, Kenny Wong and Larry Yaffe for stimulating and enjoyable discussions. We are also particularly grateful to Gerald Dunne for collaboration on related work. M.Ü. thanks KITP, Kobayashi-Maskawa Institute at Nagoya University and Seoul National University for hospitality. He also thanks the participants of the “New Methods in Nonperturbative Quantum Field Theory” program for fruitful discussions. D. D. is grateful for the support of European Research Council Advanced Grant No. 247252, Properties and Applications of the Gauge/Gravity Correspondence.

## A Resurgence terminology

In this appendix we briefly discuss some of the mathematics behind asymptotic series and resurgence methods. For further details, see e.g. [9, 13, 39, 103].

As we already mentioned at the beginning of our paper, most perturbative series appearing in physics are not convergent. When we compute a generic physical observable, by means of perturbative expansion, this takes the form

$$f(g) = \sum_{n=0}^{\infty} a_n g^n, \quad (\text{A.1})$$

where  $g$  is the coupling constant. When the constant term  $a_0$  in (A.1) vanishes the asymptotic series is called a *small* power series. As noted long time ago by Dyson and Lipatov [1, 2], a generic feature of quantum field theory is the factorial growth  $n!$  of the coefficients  $a_n$  (i.e. combinatorics of Feynman diagrams or phase space integration of UV renormalons), which effectively makes the series eq. (A.1) divergent for all non-zero  $g$ . It turns out that series usually found in physics are only asymptotic series, meaning that the difference between the function  $f(g)$  and the partial sum tends to zero as

$$\lim_{g \rightarrow 0} g^{-N} |f(g) - \sum_{n=0}^N a_n g^n| = 0, \quad (\text{A.2})$$

this for all  $N$ . Clearly this does not imply any convergence, and a naive finite-order partial sum may differ enormously from the actual function  $f$ .

We now define an important unitary subalgebra of the algebra of formal power series with coefficients in  $\mathbb{C}$ ,  $\mathbb{C}[g]$ . Expansions of the form of eq. (A.1) are examples of *Gevrey order 1* formal power series with coefficients  $a_n$ , satisfying the property that  $|a_n|/n!$  is growing at most as a geometric series. Some Gevrey-1 series can be assigned a meaningful sum by the method of Borel summation. To define the Borel sum of a Gevrey-1 series, we first insert a factor of “1” into the series using the well-known formula

$$\frac{1}{n!} \int_0^\infty dt t^n e^{-t} = 1. \quad (\text{A.3})$$

If we commute the integral with the sum, we get the Borel sum  $\mathcal{S}[f](g)$  of the original series

$$\mathcal{S}[f](g) = a_0 + \int_0^\infty dt e^{-t/g} B[f](t). \quad (\text{A.4})$$

where the Borel operator  $B$  takes the formal power series  $f(g)$  and gives

$$B[f](t) = \sum_{n=1}^{\infty} \frac{a_n}{(n-1)!} t^{n-1}, \quad (\text{A.5})$$

called the Borel transform of  $f$ . Note that since we assumed the coefficients  $a_n$  to be of Gevrey-1 type, the Borel transform  $B[f]$  has a finite radius of convergence and eq. (A.5) defines an analytic function in the origin  $t = 0$ . By a change of variables in the integral one can see that the expansion for  $g \sim 0$  of eq. (A.5) leads to our initial expansion eq. (A.1). The Borel sum, represented by the operator  $\mathcal{S}$ , of the asymptotic series (A.1) is then simply the Laplace transform of the analytic continuation of the Borel transform  $B[f]$ . This leads to a well defined expression for  $\mathcal{S}[f](g)$  as an analytic function in the half-plane  $\text{Re}(g) > 0$ , as shown schematically in figure 14.

As an example of the use of this machinery, consider the following formal series

$$E(g) = \sum_{n=0}^{\infty} (-1)^n n! g^{n+1}. \quad (\text{A.6})$$

This is a formal solution to Euler’s equation<sup>39</sup>

$$g^2 E'(g) + E(g) = g. \quad (\text{A.7})$$

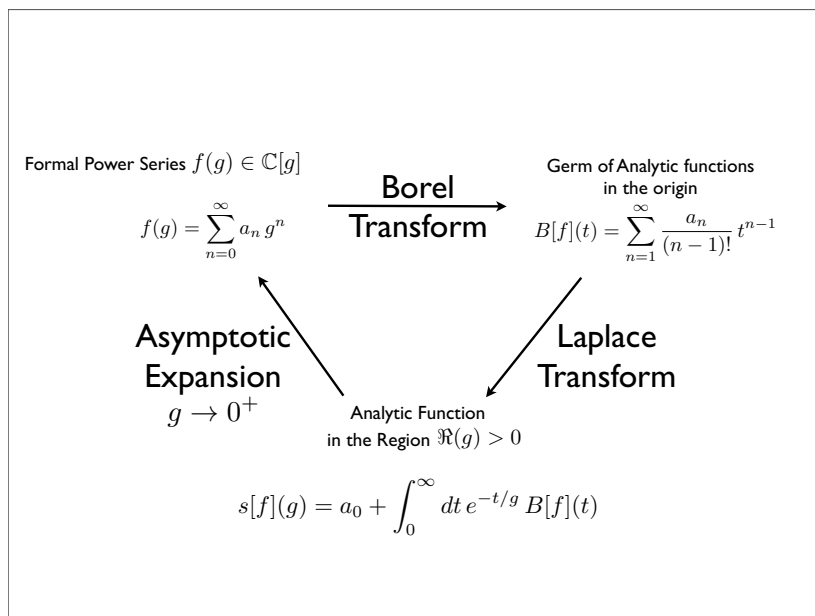
The formal series diverges for all  $g \neq 0$ . However, the coefficients  $a_n = (-1)^{n-1}(n-1)!$  alternate in sign, leading to a well defined and unique Borel sum. The Borel transform can be obtained from the definition eq. (A.5)

$$B[E](t) = \frac{1}{1+t}. \quad (\text{A.8})$$

Next we compute the Laplace transform of  $B[E](t)$  to obtain an analytic function in the half-plane  $\text{Re}(g) > 0$ , which solves Euler’s equation (A.7).

---

<sup>39</sup>Note that usually Euler’s equation is written in terms of  $x = 1/g$ .



**Figure 14.** Schematic representation of the Borel resummation procedure.

Consider now a simple modification of the formal series eq. (A.6):

$$F(g) = \sum_{n=0}^{\infty} n! g^{n+1}. \quad (\text{A.9})$$

This is a solution of the ordinary differential equation

$$g^2 F'(g) - F(g) = -g. \quad (\text{A.10})$$

The coefficients  $a_n = (n-1)!$  are now non-alternating in sign, and the Borel transform of  $F$  is

$$B[F](t) = \frac{1}{1-t}. \quad (\text{A.11})$$

Due to the non-alternating nature of the coefficients  $a_n$  we obtain a two-fold ambiguity in the Laplace transform, since the integration contour  $t \in [0, \infty)$  is a Stokes line, that is a singular direction since it contains a pole for  $t = 1$ . If we allow the contour of integration to move into the complex  $t$  plane, also called Borel plane, we can avoid the singularity by either passing above it,  $\arg(t) > 0$ , or below it,  $\arg(t) < 0$ . Hence we can define two Laplace transforms of  $B[F]$ ,  $\mathcal{S}_+[F]$  and  $\mathcal{S}_-[F]$ , obtained by integrating on the two different contours. The functions  $\mathcal{S}_+[F]$  and  $\mathcal{S}_-[F]$  give two possible analytic continuations of the original formal series eq. (A.9). One can show that their difference is related to the residue around  $t = 1$ , and is given by

$$\mathcal{S}_+[F](g) - \mathcal{S}_-[F](g) = 2\pi i e^{-1/g}. \quad (\text{A.12})$$

This exponential term is the hallmark of non-perturbative physics, and the example just presented is just a simplified version of what normally happens in generic asymptotically-free quantum field theories. As a result, the Borel sum of the formal perturbative power



series expansion (A.1), is ambiguous when the integration line is a Stokes line containing poles (or more generically branch cuts) of the Borel transform.

For a generic asymptotic series  $f$  of Gevrey type 1 we will require its Borel transform to have only a “few” singularities in the Borel plane. More precisely we require the germ of analytic functions  $B[f]$  to be *endlessly continuable* on  $\mathbb{C}$ , meaning that for all  $L > 0$  there exists only a finite set  $\Omega_L(B[f]) \subset \mathbb{C}$ , called the set of  $L$ -accessible singularities, such that  $B[f]$  has an analytic continuation along every path whose length is less than  $L$ , while avoiding the set  $\Omega_L(B[f])$ . This definition is slightly stronger than the original definition given by Ecalle. Both our previous examples obtained from the  $E$  and  $F$  series satisfy this requirement. A working hypothesis, consistent with all results available so far, is that in fact the perturbative series arising from physical QFTs also satisfy this requirement, which is necessary for the technology of resurgence theory to be useful.

We will also say that  $B[f]$  has only *simple singularities* if for all paths  $\gamma$  ending at a singular point  $t_*$ , the analytic continuation  $B[f]_\gamma$  of the germ  $B[f]$ , along the path  $\gamma$ , in a neighborhood of  $t_*$  takes the form

$$B[f]_\gamma(t) = \frac{a_\gamma}{2\pi i(t-t_*)} + b_\gamma(t-t_*) \frac{\log(t-t_*)}{2\pi i} + h_\gamma(t-t_*), \quad (\text{A.13})$$

where  $a_\gamma \in \mathbb{C}$ , while  $b_\gamma$  and  $h_\gamma$  are some analytic germs around the origin. The germs  $b_\gamma$  and  $h_\gamma$  are themselves endlessly continuable functions with simple singularities. With the concepts just introduced, we define the formal power series  $f$  in (A.1) to be a *simple resurgent function* if it is of Gevrey order 1 and its Borel transform is an endlessly continuable function with only simple singularities. The set of simple resurgent functions is actually a subalgebra of  $\mathbb{C}[g]$  denoted by  ${}^+\mathcal{R}(1)$  as proven by Ecalle [9].

It is useful at this point to introduce the concept of *directional Borel summation*. Given a simple resurgent function  $f$ , we can compute its Borel transform  $B[f]$  eq. (A.5). Then from the germ  $B[f]$  we can compute the directional Laplace transform

$$\mathcal{S}_\theta[f](g) = a_0 + \int_0^{e^{i\theta}\infty} dt e^{-t/g} B[f](t), \quad (\text{A.14})$$

where the contour of integration is the line, lying in the complex Borel plane, starting from the origin,  $t = 0$ , and going to infinity in the direction  $\arg(t) = \theta$ . This integral is convergent in the half-plane defined by  $P_\theta = \{g \in \mathbb{C} \text{ s.t. } \arg(t/g) > 0\}$ . If there are no singularities of  $B[f](t)$  in the region  $P_\theta \cap P_\phi$ , it follows by Cauchy’s theorem that  $\mathcal{S}_\theta[f]$  and  $\mathcal{S}_\phi[f]$  will coincide on  $P_\theta \cap P_\phi$ , so they are analytic continuations of one another. Furthermore they all have the same asymptotic expansion given by eq. (A.1). A direction  $\theta$  which contains singularities for  $B[f](t)$  is called a Stokes line. Thus we define lateral Borel sums by considering  $\mathcal{S}_{\theta+}[f]$  and  $\mathcal{S}_{\theta-}[f]$ , where we avoid the singularity by slightly deforming the contour of integration  $\arg(t) = \theta + \epsilon$  or  $\arg(t) = \theta - \epsilon$ .

If we take a singular direction  $\theta$ , and we assume for simplicity that along this direction the Borel transform  $B[f]$  of the simple resurgent function  $f$  has only one singularity at  $t = t_*$  of the form eq. (A.13), then we can compute

$$\mathcal{S}_{\theta+}[f](g) - \mathcal{S}_{\theta-}[f](g) = \int_{C_\theta} dt e^{-t/g} B[f](t), \quad (\text{A.15})$$

where the contour  $C_\theta$  comes from infinity in the direction  $\arg(t) = \theta$ , turns around the singular point  $t_\star$  clockwise, and then goes back to infinity once again in the direction  $\theta$ , as displayed in figure 4. Given the form of the singularity eq. (A.13), one can show that

$$(\mathcal{S}_{\theta+} - \mathcal{S}_{\theta-})[f](g) = -e^{-t_\star/g} a_\theta - e^{-t_\star/g} \mathcal{S}_\theta \circ B^{-1}[b_\theta](g), \quad (\text{A.16})$$

where  $B^{-1}$  is the inverse of the Borel transform. As we can see from this last equation, the ambiguity in the Borel sum  $\mathcal{S}_\theta[f]$  is related to the presence of infinitesimal terms  $e^{-t_\star/g}(1 + O(g))$ . These terms cannot be captured by our perturbative series eq. (A.1). On a Stokes line it is essential to take into account non-perturbative contributions in order to be able to assign a well-defined meaning to the sum of the perturbative series. Hence we have to replace our asymptotic series eq. (A.1) with a *trans-series* of the schematic form

$$f_{TS}(g) = \sum_{n=0}^{\infty} a_n g^n + \sum_c e^{-t_c/g} \sum_{n=0}^{\infty} a_{c,n}. \quad (\text{A.17})$$

This is precisely what we would expect from an observable computed using a saddle-point method, see eq. (1.2). The  $e^{-t_c/g}$  factors are non-analytic for  $g \rightarrow 0$ , so they have to be treated as objects external to the algebra of simple resurgent functions. We can see the algebra of transseries (A.17) as an extension of the algebra of formal power series (A.1) with a grading induced by the new objects  $e^{-t_c/g}$ .

As shown above, when the direction  $\theta$  is a singular one, the Borel summation jumps as we cross this *Stokes line*, and the full discontinuity across this direction plays a crucial role in linking perturbative and non-perturbative terms. The *Stokes automorphism*  $\underline{\mathfrak{S}}_\theta$  is defined by

$$\mathcal{S}_{\theta+} = \mathcal{S}_{\theta-} \circ \underline{\mathfrak{S}}_\theta = \mathcal{S}_{\theta-} \circ (\text{Id} - \text{Disc}_\theta), \quad (\text{A.18})$$

$$\mathcal{S}_{\theta+} - \mathcal{S}_{\theta-} = -\mathcal{S}_{\theta-} \circ \text{Disc}_\theta, \quad (\text{A.19})$$

where  $\text{Disc}_\theta$  encodes the full discontinuity across  $\theta$ .

When the Stokes automorphism in a particular direction  $\theta$  acts as the identity operator, it means that the Borel transform of  $f(g)$  has no singularities along the  $\theta$  direction and is given by a Borel-summable power series. Across a Stokes line  $\mathfrak{S}_\theta$  is non-trivial and it encodes the jump between the two lateral resummations. Passing from a standard asymptotic series eq. (A.1) to a trans-series eq. (A.17) with the inclusion of non-analytic (non-perturbative) terms of the form  $e^{-t/g}$  is crucial. While these terms are exponentially suppressed for  $g \sim 0$  compared to terms of the form  $a_n g^n$ , when sitting on a Stokes line these terms are critical for making observables well-defined, and they must be taken into account.

By a contour deformation it is possible to show that the difference between the  $\theta^+$  and  $\theta^-$  deformation is nothing but a sum over Hankel contours, and the discontinuity of  $\mathcal{S}$  across  $\theta$  is arising as an infinite sum of contribution coming from each one of the singular points. The logarithm of the Stokes automorphism defines the *alien derivative*

$$\underline{\mathfrak{S}}_\theta = \exp \left( \sum_{t_\star \in \Gamma_\theta} e^{-t_\star/g} \Delta_{t_\star} \right), \quad (\text{A.20})$$

where we denoted with  $\Gamma_\theta$  the set of singular points of the Borel transform along the  $\theta$  direction. It is possible to show that this operator is a real derivation acting on the space of simple resurgent functions [13]. When the Borel transform of  $f$  has only one simple singularity eq. (A.13) at  $t_\star$  in the direction  $\theta = \arg(t_\star)$ , the alien derivative takes the simpler form

$$\Delta_t f(g) = 0, \quad t \neq t_\star, \quad (\text{A.21})$$

$$\Delta_{t_\star} f(g) = a_\gamma + B^{-1}[b_\gamma](g), \quad (\text{A.22})$$

where the path  $\gamma$  is the line emanating from the origin in the direction  $\theta = \arg(t_\star)$ . A more general definition is possible when there are multiple singular points along the chosen Stokes line but we will not need it for the present work [13].

In the particular case in which  $t_\star$  is the only simple singularity for the asymptotic series eq. (A.1) along  $\theta$ , we can rewrite the Stokes automorphism using eq. (A.20)

$$\mathfrak{S}_\theta f(g) = \left( 1 + e^{-t_\star/g} \Delta_{t_\star} + \frac{e^{-2t_\star/g}}{2} \Delta_{t_\star}^2 + \dots \right) f(g). \quad (\text{A.23})$$

When  $B^{-1}[b_\gamma](g)$  has no singularities in this direction, as in the 0 dimensional example in the main text (2.20), the above equation simplifies even further giving

$$\mathfrak{S}_\theta f(g) = f(g) + e^{-t_\star/g} (a_\gamma + B^{-1}[b_\gamma](g)), \quad (\text{A.24})$$

which is just a manifestation of Stokes phenomena written in the language of alien derivatives.

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

## References

- [1] F.J. Dyson, *Divergence of perturbation theory in quantum electrodynamics*, *Phys. Rev.* **85** (1952) 631 [[INSPIRE](#)].
- [2] L.N. Lipatov, *Divergence of the perturbation theory series and the quasiclassical theory*, *Sov. Phys. JETP* **45** (1977) 216 [*Zh. Eksp. Teor. Fiz.* **72** (1977) 411] [[INSPIRE](#)].
- [3] M. Beneke, *Renormalons*, *Phys. Rept.* **317** (1999) 1 [[hep-ph/9807443](#)] [[INSPIRE](#)].
- [4] E.B. Bogomolny, *Calculation of instanton-anti-instanton contributions in quantum mechanics*, *Phys. Lett. B* **91** (1980) 431 [[INSPIRE](#)].
- [5] J. Zinn-Justin, *Multi-instanton contributions in quantum mechanics*, *Nucl. Phys. B* **192** (1981) 125 [[INSPIRE](#)].
- [6] E. Delabaere, H. Dillinger and F. Pham, *Exact semiclassical expansions for one-dimensional quantum oscillators*, *J. Math. Phys.* **38** (1997) 6126.
- [7] G.V. Dunne and M. Ünsal, *Generating nonperturbative physics from perturbation theory*, *Phys. Rev. D* **89** (2014) 041701 [[arXiv:1306.4405](#)] [[INSPIRE](#)].

- [8] G.V. Dunne and M. Ünsal, *Uniform WKB, multi-instantons and resurgent trans-series*, *Phys. Rev. D* **89** (2014) 105009 [[arXiv:1401.5202](#)] [[INSPIRE](#)].
- [9] J. Ecalle, *Les fonctions resurgentes* (in French), vol. I–III, Publ. Math. Orsay, France (1981).
- [10] R.B. Dingle, *Asymptotic expansions: their derivation and interpretation*, Academic Press, U.S.A. (1973).
- [11] M. Berry and C. Howls, *Hyperasymptotics*, *Proc. Roy. Soc. London A* **430** (1990) 653.
- [12] M. Berry and C. Howls, *Hyperasymptotics for integrals with saddles*, *Proc. Roy. Soc. London A* **434** (1991) 657.
- [13] D. Sauzin, *Resurgent functions and splitting problems*, [arXiv:0706.0137](#).
- [14] O. Costin, *Asymptotics and Borel summability*, CRC Press, U.S.A. (2008).
- [15] I. Aniceto and R. Schiappa, *Nonperturbative ambiguities and the reality of resurgent transseries*, *Commun. Math. Phys.* **335** (2015) 183 [[arXiv:1308.1115](#)] [[INSPIRE](#)].
- [16] P.C. Argyres and M. Ünsal, *The semi-classical expansion and resurgence in gauge theories: new perturbative, instanton, bion and renormalon effects*, *JHEP* **08** (2012) 063 [[arXiv:1206.1890](#)] [[INSPIRE](#)].
- [17] P. Argyres and M. Ünsal, *A semiclassical realization of infrared renormalons*, *Phys. Rev. Lett.* **109** (2012) 121601 [[arXiv:1204.1661](#)] [[INSPIRE](#)].
- [18] G.V. Dunne and M. Ünsal, *Resurgence and trans-series in quantum field theory: the  $\mathbb{CP}^{N-1}$  model*, *JHEP* **11** (2012) 170 [[arXiv:1210.2423](#)] [[INSPIRE](#)].
- [19] G.V. Dunne and M. Ünsal, *Continuity and resurgence: towards a continuum definition of the  $\mathbb{CP}^{N-1}$  model*, *Phys. Rev. D* **87** (2013) 025015 [[arXiv:1210.3646](#)] [[INSPIRE](#)].
- [20] M. Shifman and M. Ünsal, *QCD-like theories on  $R_3 \times S_1$ : a smooth journey from small to large  $r(S_1)$  with double-trace deformations*, *Phys. Rev. D* **78** (2008) 065004 [[arXiv:0802.1232](#)] [[INSPIRE](#)].
- [21] M. Ünsal and L.G. Yaffe, *Center-stabilized Yang-Mills theory: confinement and large- $N$  volume independence*, *Phys. Rev. D* **78** (2008) 065035 [[arXiv:0803.0344](#)] [[INSPIRE](#)].
- [22] G. 't Hooft, *Can we make sense out of quantum chromodynamics?*, *Subnucl. Ser.* **15** (1979) 943.
- [23] M.A. Shifman and A.V. Smilga, *Fractons in twisted multiflavor Schwinger model*, *Phys. Rev. D* **50** (1994) 7659 [[hep-th/9407007](#)] [[INSPIRE](#)].
- [24] K.-M. Lee and P. Yi, *Monopoles and instantons on partially compactified D-branes*, *Phys. Rev. D* **56** (1997) 3711 [[hep-th/9702107](#)] [[INSPIRE](#)].
- [25] T.C. Kraan and P. van Baal, *Periodic instantons with nontrivial holonomy*, *Nucl. Phys. B* **533** (1998) 627 [[hep-th/9805168](#)] [[INSPIRE](#)].
- [26] F. Bruckmann, *Instanton constituents in the  $O(3)$  model at finite temperature*, *Phys. Rev. Lett.* **100** (2008) 051602 [[arXiv:0707.0775](#)] [[INSPIRE](#)].
- [27] W. Brendel, F. Bruckmann, L. Janssen, A. Wipf and C. Wozar, *Instanton constituents and fermionic zero modes in twisted  $\mathbb{CP}^n$  models*, *Phys. Lett. B* **676** (2009) 116 [[arXiv:0902.2328](#)] [[INSPIRE](#)].

- [28] D. Harland, *Kinks, chains and loop groups in the  $\mathbb{CP}^n$   $\sigma$ -models*, *J. Math. Phys.* **50** (2009) 122902 [[arXiv:0902.2303](#)] [[INSPIRE](#)].
- [29] M. Ünsal, *Magnetic bion condensation: a new mechanism of confinement and mass gap in four dimensions*, *Phys. Rev. D* **80** (2009) 065001 [[arXiv:0709.3269](#)] [[INSPIRE](#)].
- [30] R. Dabrowski and G.V. Dunne, *Fractionalized non-self-dual solutions in the  $\mathbb{CP}^{N-1}$  model*, *Phys. Rev. D* **88** (2013) 025020 [[arXiv:1306.0921](#)] [[INSPIRE](#)].
- [31] A. Cherman, D. Dorigoni, G.V. Dunne and M. Ünsal, *Resurgence in quantum field theory: nonperturbative effects in the principal chiral model*, *Phys. Rev. Lett.* **112** (2014) 021601 [[arXiv:1308.0127](#)] [[INSPIRE](#)].
- [32] S. Pasquetti and R. Schiappa, *Borel and Stokes nonperturbative phenomena in topological string theory and  $c = 1$  matrix models*, *Annales Henri Poincaré* **11** (2010) 351 [[arXiv:0907.4082](#)] [[INSPIRE](#)].
- [33] M. Mariño, R. Schiappa and M. Weiss, *Nonperturbative effects and the large-order behavior of matrix models and topological strings*, *Commun. Num. Theor. Phys.* **2** (2008) 349 [[arXiv:0711.1954](#)] [[INSPIRE](#)].
- [34] M. Mariño, *Nonperturbative effects and nonperturbative definitions in matrix models and topological strings*, *JHEP* **12** (2008) 114 [[arXiv:0805.3033](#)] [[INSPIRE](#)].
- [35] M. Mariño, R. Schiappa and M. Weiss, *Multi-instantons and multi-cuts*, *J. Math. Phys.* **50** (2009) 052301 [[arXiv:0809.2619](#)] [[INSPIRE](#)].
- [36] S. Garoufalidis, A. Its, A. Kapaev and M. Mariño, *Asymptotics of the instantons of Painlevé I*, *Int. Math. Res. Not.* **2012** (2012) 561 [[arXiv:1002.3634](#)] [[INSPIRE](#)].
- [37] I. Aniceto, R. Schiappa and M. Vonk, *The resurgence of instantons in string theory*, *Commun. Num. Theor. Phys.* **6** (2012) 339 [[arXiv:1106.5922](#)] [[INSPIRE](#)].
- [38] R. Schiappa and R. Vaz, *The resurgence of instantons: multi-cut Stokes phases and the Painlevé II equation*, *Commun. Math. Phys.* **330** (2014) 655 [[arXiv:1302.5138](#)] [[INSPIRE](#)].
- [39] M. Mariño, *Lectures on non-perturbative effects in large- $N$  gauge theories, matrix models and strings*, *Fortsch. Phys.* **62** (2014) 455 [[arXiv:1206.6272](#)] [[INSPIRE](#)].
- [40] D. Gaiotto, G.W. Moore and A. Neitzke, *Four-dimensional wall-crossing via three-dimensional field theory*, *Commun. Math. Phys.* **299** (2010) 163 [[arXiv:0807.4723](#)] [[INSPIRE](#)].
- [41] D. Gaiotto, G.W. Moore and A. Neitzke, *Wall-crossing, Hitchin systems and the WKB approximation*, [arXiv:0907.3987](#) [[INSPIRE](#)].
- [42] M. Kontsevich and Y. Soibelman, *Stability structures, motivic Donaldson-Thomas invariants and cluster transformations*, [arXiv:0811.2435](#) [[INSPIRE](#)].
- [43] A.M. Polyakov, *Gauge fields and strings*, CRC Press, U.S.A. (1987).
- [44] A.M. Polyakov and P.B. Wiegmann, *Theory of non-Abelian Goldstone bosons*, *Phys. Lett. B* **131** (1983) 121 [[INSPIRE](#)].
- [45] P. Wiegmann, *Exact factorized  $S$  matrix of the chiral field in two-dimensions*, *Phys. Lett. B* **142** (1984) 173 [[INSPIRE](#)].
- [46] A.C. Cubero, *Multiparticle form factors of the principal chiral model at large- $N$* , *Phys. Rev. D* **86** (2012) 025025 [[arXiv:1205.2069](#)] [[INSPIRE](#)].

- [47] A.C. Cubero and P. Orland, *Correlation functions of the  $SU(\infty)$  principal chiral model*, *Phys. Rev. D* **88** (2013) 025044 [[arXiv:1306.1930](#)] [[INSPIRE](#)].
- [48] A.C. Cubero and P. Orland, *Dynamical mass reduction in the massive Yang-Mills spectrum in  $1+1$  dimensions*, *Phys. Rev. D* **89** (2014) 085027 [[arXiv:1403.0276](#)] [[INSPIRE](#)].
- [49] A. Armoni, D. Dorigoni and G. Veneziano, *k-string tension from Eguchi-Kawai reduction*, *JHEP* **10** (2011) 086 [[arXiv:1108.6196](#)] [[INSPIRE](#)].
- [50] M. García Pérez, A. González-Arroyo and M. Okawa, *Spatial volume dependence for  $2+1$  dimensional  $SU(N)$  Yang-Mills theory*, *JHEP* **09** (2013) 003 [[arXiv:1307.5254](#)] [[INSPIRE](#)].
- [51] F. Pham, *Vanishing homologies and the  $n$  variable saddlepoint method*, *Proc. Symp. Pure Math.* **2** (1983) 319.
- [52] E. Witten, *Analytic continuation of Chern-Simons theory*, *AMS/IP Stud. Adv. Math.* **50** (2011) 347 [[arXiv:1001.2933](#)] [[INSPIRE](#)].
- [53] T. Gulden, M. Janas, P. Koroteev and A. Kamenev, *Statistical mechanics of Coulomb gases as quantum theory on Riemann surfaces*, *Zh. Eksp. Teor. Fiz.* **144** (2013) 574 [*J. Exp. Theor. Phys.* **117** (2013) 517] [[arXiv:1303.6386](#)] [[INSPIRE](#)].
- [54] F.W. Olver, D.W. Lozier, R.F. Boisvert and C.W. Clark, *NIST handbook of mathematical functions*, Cambridge University Press, U.S.A. (2010).
- [55] P. Rossi and E. Vicari, *Two-dimensional  $SU(N) \times SU(N)$  chiral models on the lattice*, *Phys. Rev. D* **49** (1994) 1621 [Erratum *ibid.* **D 55** (1997) 1698] [[hep-lat/9307014](#)] [[INSPIRE](#)].
- [56] K. Uhlenbeck, *Harmonic maps into Lie groups (classical solutions of the chiral model)*, *J. Diff. Geom.* **30** (1989) 1.
- [57] G. Valli, *On the energy spectrum of harmonic 2-spheres in unitary groups*, *Topology* **27** (1988) 129.
- [58] B. Piette and W.J. Zakrzewski, *Properties of classical solutions of the  $U(N)$  chiral  $\sigma$ -models in two-dimensions*, *Nucl. Phys. B* **300** (1988) 223 [[INSPIRE](#)].
- [59] J.C. Wood, *Explicit construction and parametrization of harmonic two-spheres in the unitary group*, *Proc. London Math. Soc.* **58** (1989) 608.
- [60] R. Ward, *Classical solutions of the chiral model, unitons, and holomorphic vector bundles*, *Commun. Math. Phys.* **128** (1990) 319.
- [61] G.V. Dunne, *Selfdual Chern-Simons theories*, [hep-th/9410065](#) [[INSPIRE](#)].
- [62] N. Manton and P. Sutcliffe, *Topological solitons*, Cambridge University Press, U.S.A. (2004).
- [63] P. Di Vecchia, V.G. Knizhnik, J.L. Petersen and P. Rossi, *A supersymmetric Wess-Zumino Lagrangian in two-dimensions*, *Nucl. Phys. B* **253** (1985) 701 [[INSPIRE](#)].
- [64] J.M. Evans and T.J. Hollowood, *Exact scattering in the  $SU(N)$  supersymmetric principal chiral model*, *Nucl. Phys. B* **493** (1997) 517 [[hep-th/9603190](#)] [[INSPIRE](#)].
- [65] E. Witten, *Constraints on supersymmetry breaking*, *Nucl. Phys. B* **202** (1982) 253 [[INSPIRE](#)].
- [66] V.A. Fateev, P.B. Wiegmann and V.A. Kazakov, *Large- $N$  chiral field in two-dimensions*, *Phys. Rev. Lett.* **73** (1994) 1750 [[INSPIRE](#)].
- [67] V.A. Fateev, V.A. Kazakov and P.B. Wiegmann, *Principal chiral field at large- $N$* , *Nucl. Phys. B* **424** (1994) 505 [[hep-th/9403099](#)] [[INSPIRE](#)].



- [68] F. David, *On the ambiguity of composite operators, IR renormalons and the status of the operator product expansion*, *Nucl. Phys. B* **234** (1984) 237 [INSPIRE].
- [69] D.J. Gross, R.D. Pisarski and L.G. Yaffe, *QCD and instantons at finite temperature*, *Rev. Mod. Phys.* **53** (1981) 43 [INSPIRE].
- [70] E. Ogievetsky, P. Wiegmann and N. Reshetikhin, *The principal chiral field in two-dimensions on classical Lie algebras: the Bethe ansatz solution and factorized theory of scattering*, *Nucl. Phys. B* **280** (1987) 45 [INSPIRE].
- [71] G. Basar, A. Cherman, D. Dorigoni and M. Ünsal, *Volume independence in the large- $N$  limit and an emergent fermionic symmetry*, *Phys. Rev. Lett.* **111** (2013) 121601 [arXiv:1306.2960] [INSPIRE].
- [72] M. Lüscher, *Some analytic results concerning the mass spectrum of Yang-Mills gauge theories on a torus*, *Nucl. Phys. B* **219** (1983) 233 [INSPIRE].
- [73] P. van Baal, *The small volume expansion of gauge theories coupled to massless fermions*, *Nucl. Phys. B* **307** (1988) 274 [Erratum *ibid.* **B 312** (1989) 752] [INSPIRE].
- [74] P. Kovtun, M. Ünsal and L.G. Yaffe, *Volume independence in large- $N_c$  QCD-like gauge theories*, *JHEP* **06** (2007) 019 [hep-th/0702021] [INSPIRE].
- [75] N.M. Davies, T.J. Hollowood and V.V. Khoze, *Monopoles, affine algebras and the gluino condensate*, *J. Math. Phys.* **44** (2003) 3640 [hep-th/0006011] [INSPIRE].
- [76] C. Bauer, G.S. Bali and A. Pineda, *Compelling evidence of renormalons in QCD from high order perturbative expansions*, *Phys. Rev. Lett.* **108** (2012) 242002 [arXiv:1111.3946] [INSPIRE].
- [77] G.S. Bali, C. Bauer, A. Pineda and C. Torrero, *Perturbative expansion of the energy of static sources at large orders in four-dimensional SU(3) gauge theory*, *Phys. Rev. D* **87** (2013) 094517 [arXiv:1303.3279] [INSPIRE].
- [78] G.S. Bali, C. Bauer and A. Pineda, *Perturbative expansion of the plaquette to  $O(\alpha^{35})$  in four-dimensional SU(3) gauge theory*, *Phys. Rev. D* **89** (2014) 054505 [arXiv:1401.7999] [INSPIRE].
- [79] N. Cabibbo and E. Marinari, *A new method for updating SU( $N$ ) matrices in computer simulations of gauge theories*, *Phys. Lett. B* **119** (1982) 387 [INSPIRE].
- [80] J. Zinn-Justin, *Quantum field theory and critical phenomena*, *Int. Ser. Monogr. Phys.* **113** (2002) 1 [INSPIRE].
- [81] M. Stone and J. Reeve, *Late terms in the asymptotic expansion for the energy levels of a periodic potential*, *Phys. Rev. D* **18** (1978) 4746 [INSPIRE].
- [82] C.M. Bender and T.T. Wu, *Anharmonic oscillator*, *Phys. Rev.* **184** (1969) 1231 [INSPIRE].
- [83] C.M. Bender and T.T. WU, *Large order behavior of perturbation theory*, *Phys. Rev. Lett.* **27** (1971) 461 [INSPIRE].
- [84] H. Eichenherr and M. Forger, *More about nonlinear  $\sigma$ -models on symmetric spaces*, *Nucl. Phys. B* **164** (1980) 528 [Erratum *ibid.* **B 282** (1987) 745] [INSPIRE].
- [85] B. Piette, I. Stokoe and W.J. Zakrzewski, *On stability of solutions of the U( $N$ ) chiral model in two-dimensions*, *Z. Phys. C* **37** (1988) 449 [INSPIRE].
- [86] A. Gonzalez-Arroyo and P. Martinez, *Investigating Yang-Mills theory and confinement as a function of the spatial volume*, *Nucl. Phys. B* **459** (1996) 337 [hep-lat/9507001] [INSPIRE].

- [87] P. van Baal, *QCD in a finite volume*, [hep-ph/0008206](#) [[INSPIRE](#)].
- [88] B. Collie and D. Tong, *The partonic nature of instantons*, *JHEP* **08** (2009) 006 [[arXiv:0905.2267](#)] [[INSPIRE](#)].
- [89] A. Parnachev and A.R. Zhitnitsky, *Phase transitions,  $\theta$  behavior and instantons in QCD and its holographic model*, *Phys. Rev. D* **78** (2008) 125002 [[arXiv:0806.1736](#)] [[INSPIRE](#)].
- [90] M.C. Ogilvie, *Phases of gauge theories*, *J. Phys. A* **45** (2012) 483001 [[arXiv:1211.2843](#)] [[INSPIRE](#)].
- [91] E. Thomas and A.R. Zhitnitsky, *Topological susceptibility and contact term in QCD. A toy model*, *Phys. Rev. D* **85** (2012) 044039 [[arXiv:1109.2608](#)] [[INSPIRE](#)].
- [92] H. Nishimura and M.C. Ogilvie, *Confinement in a Higgs model on  $R^3 \times S^1$* , *Phys. Rev. D* **85** (2012) 065021 [[arXiv:1111.6101](#)] [[INSPIRE](#)].
- [93] M. Garcia Perez, A. Gonzalez-Arroyo and A. Sastre, *Adjoint fermion zero-modes for  $SU(N)$  calorons*, *JHEP* **06** (2009) 065 [[arXiv:0905.0645](#)] [[INSPIRE](#)].
- [94] S. Bolognesi and W. Zakrzewski, *Clustering and decomposition for non-BPS solutions of the  $\mathbb{CP}^{N-1}$  models*, *Phys. Rev. D* **89** (2014) 065013 [[arXiv:1310.8247](#)] [[INSPIRE](#)].
- [95] G.V. Dunne, *Functional determinants in quantum field theory*, *J. Phys. A* **41** (2008) 304006 [[arXiv:0711.1178](#)] [[INSPIRE](#)].
- [96] I.I. Balitsky and A.V. Yung, *Instanton molecular vacuum in  $N = 1$  supersymmetric quantum mechanics*, *Nucl. Phys. B* **274** (1986) 475 [[INSPIRE](#)].
- [97] A.V. Yung, *Instanton vacuum in supersymmetric QCD*, *Nucl. Phys. B* **297** (1988) 47 [[INSPIRE](#)].
- [98] AURORASCIENCE collaboration, M. Cristoforetti, F. Di Renzo and L. Scorzato, *New approach to the sign problem in quantum field theories: high density QCD on a Lefschetz thimble*, *Phys. Rev. D* **86** (2012) 074506 [[arXiv:1205.3996](#)] [[INSPIRE](#)].
- [99] G. Aarts, *Lefschetz thimbles and stochastic quantization: complex actions in the complex plane*, *Phys. Rev. D* **88** (2013) 094501 [[arXiv:1308.4811](#)] [[INSPIRE](#)].
- [100] E. Witten, *A new look at the path integral of quantum mechanics*, [arXiv:1009.6032](#) [[INSPIRE](#)].
- [101] D.D. Ferrante, G.S. Guralnik, Z. Guralnik and C. Pehlevan, *Complex path integrals and the space of theories*, [arXiv:1301.4233](#) [[INSPIRE](#)].
- [102] D. Harlow, J. Maltz and E. Witten, *Analytic continuation of Liouville theory*, *JHEP* **12** (2011) 071 [[arXiv:1108.4417](#)] [[INSPIRE](#)].
- [103] E. Delabaere and F. Pham, *Resurgent methods in semi-classical asymptotics*, *Ann. Inst. Henri Poincaré A* **71** (1999) 1.