UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Matrix models and phase transitions in gauge theories

" Documento Definitivo"

Doutoramento em Matemática

Especialidade de Física Matemática e Mecânica dos Meios Contínuos

Leonardo Santilli

Tese orientada por:

Miguel Tierz Parra

Carlos A. Arango Florentino

Documento especialmente elaborado para a obtenção do grau de doutor

2022

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS

Matrix models and phase transitions in gauge theories

Doutoramento em Matemática

Especialidade de Física Matemática e Mecânica dos Meios Contínuos

Leonardo Santilli

Tese orientada por:

Miguel Tierz Parra

Carlos A. Arango Florentino

Júri:

Presidente:

- Doutor Fernando Jorge Inocêncio Ferreira, Professor Catedrático e Presidente do Departamento de Matemática da Universidade de Lisboa;

Vogais:

- Doutor David Pérez García, Profesor titular, Facultad de Ciencias Matemáticas da Universidad Complutense de Madrid, Espanha;
- Doutor Roberto Tateo, Professore, Dipartimento di Fisica da Università degli studi di Torino, Itália;
- Doutor Miguel Tierz Parra, Investigador, Facultad de Ciencias Matemáticas da Universidad Complutense de Madrid, Espanha, (orientador);
- Doutor Gabriel Czerwionka Lopes Cardoso, Professor Catedrático, Instituto Superior Técnico da Universidade de Lisboa;
- Doutora Maria Teresa Faria da Paz Pereira, Professora Associada com Agregação, Faculdade de Ciências da Universidade de Lisboa.

Documento especialmente elaborado para a obtenção do grau de doutor

Tese desenvolvida com o apoio financeiro da Fundação para a Ciência e a Tecnologia (FCT) com bolsa doutoral SFRH/BD/129405/2017

2022

There's only two or three things I know for sure.
Only two or three things. That's right.
Of course it's never the same things,
and I'm never as sure as I'd like to be.

D. Allison

Ho cercato di non affezionarmi alle mie idee
in modo da liberarmene senza troppo dolore
non appena arrivasse qualcuno
in grado di dimostrarmi che erano false.

E. Albinati

Resumo

O objetivo desta tese é o estudo de diferentes aspectos das teorias de gauge definidas em duas, três e cinco dimensões, utilizando modelos matriciais. Mais especificamente, vamos considerar a teoria de Yang–Mills pura num espaço-tempo bidimensional, bem como deformações da mesma, e teorias de gauge supersimétricas em três e cinco dimensões que incluem termos de Yang–Mills e de Chern–Simons, além de acoplamentos supersimétricos a campos de matéria. As ferramentas fornecidas pela teoria das matrizes aleatórias vão-nos permitir explorar quantitativamente uma vasta gama de propiedades e facetas das teorias de gauge, incluindo transições de fase, operadores de linha e as simetrías generalizadas associadas, e integrabilidade.

O foco deste trabalho está no estudo e caraterização de funções de partição e de operadores circulares de Wilson nestes modelos, através da aplicação de resultados de matrizes aleatórias a estas quantidades físicas, cuja apresentação em forma de modelos matricias é obtida por um mecanismo de localização. Vamos seguir duas linhas principais de investigação: o cálculo exato e explícito das quantidades, e a busca de transições de fase quânticas para N grande.

A apresentação dos resultados obtidos está dividida em três partes, de acordo com a seguinte classificação:

I. resultados exatos em teorias de Chern–Simons supersimétricas com matéria,

II. transições de fase quânticas em teorias supersimétricas com matéria, e

III. deformações da teoria de Yang–Mills em duas dimensões.

A Parte I está baseada nos artigos [1, 2]. Nesta parte vamos calcular de forma exata as funções de partição de algumas famílias de teorias tridimensionais de Chern–Simons supersimétricas, usando a integral de Mordell. Vamos também propor uma expansão em caracteres expressada com polinômios de Schur, cujos coeficientes são dados por invariantes topológicos.

Além disto, vamos mostrar uma correspondência entre dois modelos matriciais, um que computa invariantes topológicos na teoria tridimensional de Chern–Simons pura e o outro surgindo de uma teoria quântica de campos não-comutativa em duas dimensões. A correspondência é estendida a modelos de supermatrizes. Neste caso, a teoria supersimétrica ABJ(M) vai substituir a teoria topológica de Chern–Simons.

A Parte II está baseada em [3–7] e constitui o centro da tese. Nesta parte vamos analizar transições de fase quânticas para N grande em teorias de gauge usando os modelos matriciais. Apresentamos um estudo sistemático e a subseqüente classificação das transições de fase em teorias de gauge supersimétricas em esferas de três e cinco dimensões. Para toda teoria que tenha um limite superconforme conhecido, vamos demostrar que as transições são sempre de terceira ordem, enquanto teorias $U(N)$ em cinco dimensões mostram transições de segunda ordem.

São estudadas também várias familias de modelos de matrizes unitarias, para as quais determinamos os diagramas de fases.

Nesta parte, incluimos o estudo de teorias supersimétricas com grupo de gauge de característica pequena acopladas a campos quânticos de matéria, para valores fixos destes campos de matéria,

estabelecendo uma correspondência com o sistema integrável de Calogero–Moser, assim como no limite de um número grande de campos de matéria, estabelecendo a presença de uma transições de fase de segunda ordem.

A Parte III está baseada em [8–10]. Vamos aplicar o método de localização à teoria de Yang–Mills supersimétrica em cinco dimensões, posta numa variedade da forma $\mathbb{S}^3_b \times \Sigma$, onde $\mathbb{S}^3_b$ representa uma esfera topológica com uma deformação da metrica e Σ é uma superfície de Riemann orientada e fechada. Desta forma, introduzimos uma nova deformação de origem geométrica da teoria efetiva de Yang–Mills, já q-deformada, definida sobre Σ.

Prosseguindo no estudo das deformações das teorias de Yang–Mills em duas dimensões, vamos analisar a sua perturbação pelo operador irrelevante $T\overline{T}$ e demostramos que a Abelianização é mantida por esta deformação, embora outras propriedades características sejam perdidas.

Finalmente, mostramos que a transição de fase de Douglas–Kazakov se estende ao caso deformado com $T\overline{T}$. Quando esta deformação e a q-deformação são conjuntamente presentes, os dois efeitos contribuem para a determinação do diagrama de fases. Uma das duas fases desaparece no regime de $T\overline{T}$-deformação forte, enquanto a outra fase desaparece no regime de q-deformação forte.

Outros artigos não incluídos nesta dissertação são [11–14].

Palavras-chave: Teoria quântica de campos, modelos matriciais, transições de fase, teoria de gauge.

Abstract

Aspects of gauge theories in two, three and five dimensions are investigated using matrix models. Specifically, we consider pure Yang–Mills theory and its deformations in two dimensions, and supersymmetric Yang–Mills and Chern–Simons-matter theories in three and five dimensions. The random matrix approach allows us to explore a vast range of features of the gauge theories, including phase transitions, one-form symmetries and integrability.

Partition functions and Wilson loops are studied in these setups by exploiting their matrix model presentation derived by localization. Two main lines of research are pursued: the computation of exact results at fixed N and the quest for quantum phase transitions at large N.

The partition functions of several three-dimensional quiver Chern–Simons-matter theories are computed exactly using Mordell integrals, and we put forward a character expansion in terms of Schur polynomials, with coefficients given by topological invariants. A correspondence between two matrix models is provided as well, one computing topological invariants in pure Chern–Simons theory and the other arising from a two-dimensional, noncommutative scalar field theory. The correspondence is extended to supermatrix models, with ABJ(M) theory replacing topological Chern–Simons theory in this case. Partition functions and Wilson loop expectation values in three-dimensional $\mathcal{N} = 4$ gauge theories are also computed, uncovering a relation with Calogero–Moser integrable systems.

Furthermore, we apply localization to five-dimensional supersymmetric Yang–Mills theory on compact product manifolds $\mathbb{S}^3 \times \Sigma$, where Σ is a closed oriented Riemann surface, and introduce in this way a novel, "squashed" deformation of q-deformed Yang–Mills theory on Σ. Proceeding in the study of deformations of two-dimensional Yang–Mills theory, we analyze their perturbation by the operator $T\overline{T}$ and prove that Abelianization still holds, although other characteristic properties such as factorization of the partition function break down.

The analysis of large N quantum phase transitions in matrix models and gauge theories constitutes the core of the thesis. We present a systematic study and classification of phase transitions for supersymmetric gauge theories on three- and five-dimensional spheres of large radius. The transitions are always third order for gauge theories connected to a known superconformal point, but are second order for generic five-dimensional $U(N)$ theories. Several multi-parameter families of unitary matrix models are also considered and their phase diagrams are established.

Finally, we show how the Douglas–Kazakov transition of two-dimensional Yang–Mills on the sphere extends to its newly derived deformations. When both $T\overline{T}$ and q-deformations are turned on, the two effects compete, and the system has two phases in the most part of the parameter space, but the weak coupling phase is removed in the regime of strong $T\overline{T}$-deformation, whereas the strong coupling phase is removed in the strong q-deformation regime.

Keywords: Quantum field theory, matrix models, phase transitions, gauge theory.

Acknowledgements

Firstly and mostly, I would like to express my sincere gratitude to my advisor Miguel. What I know about being a researcher, I have learnt from him. He never stopped encouraging me, inspiring me and believing in me. Besides, I have been extremely lucky to meet Carlos, who accepted to co-advice me on the run and who has always taken my work and my ideas very seriously.
Many thanks to those experts who listened to my questions, discussed with me and helped me growing as a researcher in any way, Richard J. Szabo, Roberto Tateo and Carlos Nuñez above all. I also thank the Group of Mathematical Physics, especially Jean-Claude Zambrini and Davide Masoero.
Thank you to Carlos Sotillo, David García, Riccardo Conti and Gabriele Degano, for the long discussions and countless lunches.
Finally, I am indebted to my brother and my parents: without their unconditional support, none of my achievements would have been possible.

I gratefully acknowledge the financial support from the Fundação para a Ciência e a Tecnologia (FCT) through the Scholarship SFRH/BD/129405/2017. Part of the work presented was also supported by FCT Project PTDC/MAT-PUR/30234/2017.

Contents

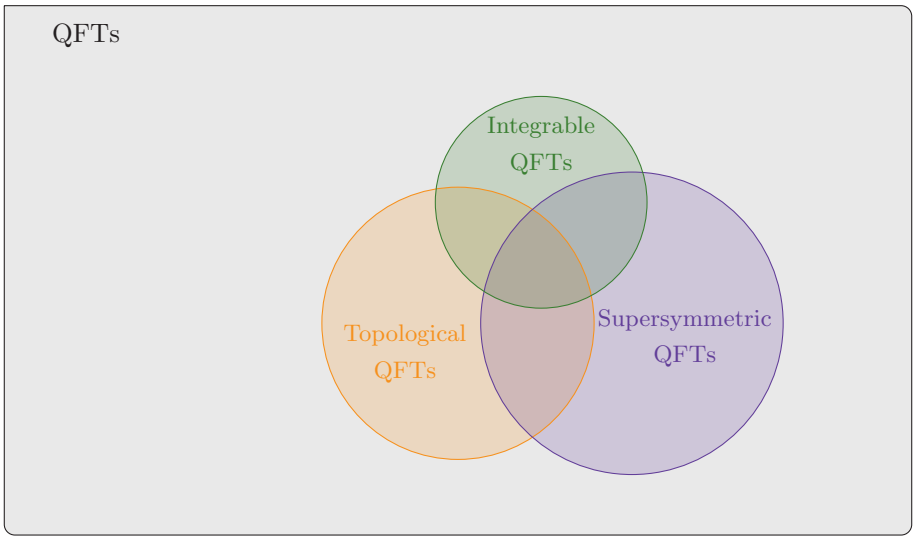
Chapter 1

Introduction

Quantum Field Theory (QFT) is an ubiquitous and terrifically effective formalism which, to the date, represents our best description of physical phenomena at the quantum level. However, most QFTs remain hardly accessible beyond the perturbative regime. Finding exact solutions that compute physical observables in the QFT realm seems therefore a pressing challenge in theoretical and mathematical physics.

The situation is greatly ameliorated when additional assumptions are imposed on a theory. Requiring it to satisfy certain selected properties constrains its degrees of freedom and renders it tractable. One option is to restrict to low enough spacetime dimensions and specific choices of field content. Additional distinguishing properties that characterize theories amenable to analytic study are: topological invariance, supersymmetry and integrability, to name a few. These categories, and even more those theories that lie in their intersection, have received an enormous amount of attention over the years, with the goal of extracting exact, non-perturbative answers and testing general properties that would be untamed in less constrained QFTs.

The present thesis is carved out of such conceptual framework and adds on this vast program.



This dissertation is devoted to the study of gauge theories. The main characters of these theories are the gauge fields, which are connections in a principal bundle over the spacetime. We refer to the mathematically-oriented textbook [15] or to [16] for a comprehensive account. A cornerstone of gauge theories is thus the presence of matrix-valued fields. The large N limit is a powerful tool that yields a firm grasp on the dynamics and phase structure of any theory with matrix degrees of freedom [17, 18], see [19, Ch.8] for an early review.

The importance of the role played by the large N limit is hardly overstated: it impulsed developments spanning from QCD [17, 20] and lattice simulations [21–25], to two-dimensional quantum gravity [26, 27] and the AdS/CFT correspondence [28–31]. The reasons to be interested in large N limits are several. The most obvious one is that, even for phenomenological theories with small N, it gives a first approximation which is exact in the 't Hooft coupling. More precise predictions are then obtained by computing higher order corrections in the $1/N$ expansion [32]. There are many instances, though, in which the physical fields are approximated by finite-rank matrices, and physical predictions are only valid in the large N limit. A prototypical example in this class is two-dimensional quantum gravity [27]. Additionally, the large N limit is instrumental to explore the landscape of critical phenomena [33, 34].

Among all those QFTs that show nuances of solvability, the focus of this thesis is on the so-called cohomological field theories [35]. Topological field theories and theories with extended supersymmetry, for instance, belong to this class.[1] The underlying cohomological structure guarantees that the computation of a great deal of protected physical observables reduces to a matrix model. This statement schematically asserts the content of the localization principle [37, 38].

Random matrix theory [39–41] provides us with the proper mathematical toolkit to evaluate matrix integrals like the ones produced by localization. Entering the domain of random matrices has a two-fold advantage. On one hand, analytic treatment of physical quantities becomes feasible, as random matrix theory oftentimes yields closed-form expressions, or at least explicit multi-parametric expansions, for them. On the other hand, the integral representation is well-suited for a steepest descent analysis at large N.

Throughout this dissertation we pursue both lines. The cohomological field theories we consider are supersymmetric Chern–Simons-matter theories in three and five dimensions, supersymmetric QCD in three dimension, and Yang–Mills theory and its deformations in two dimensions. Moreover, we include the analysis of unitary matrix models which, despite not coming from localization, show intriguing features and serve as toy models for low-dimensional QFTs.

We study matrix models derived from gauge theory and solve them exactly in a long list of examples in three spacetime dimensions. We also put forward a character expansion for a class of supersymmetry-protected observables, the expectation values of half-BPS Wilson loops on the sphere, in three-dimensional Chern–Simons-matter theories, which expresses these quantities as a generating polynomial written in the Schur polynomial basis. The coefficients of these polynomials are given by topological invariants, computed by Chern–Simons quantum field theory [42] and evaluated explicitly.

The random matrix formulation is instrumental in uncovering correspondences between otherwise seemingly unrelated QFTs. Two of the results of this thesis are inserted in this framework. In Chapter 3, a family of spectra of noncommutative field theories in two dimensions is proved to be equivalent with a family of topological invariants in Chern–Simons theory; the relation extends to supermatrix models, with ABJ(M) theory [43, 44] replacing pure Chern–Simons in the supermatrix version. Moreover, a correspondence between certain mass-deformed three-dimensional $\mathcal{N} = 4$ gauge theories and hyperbolic Calogero–Moser integrable systems is established in Chapter 7.

Furthermore, in the central part of this work, we analyze the large N limit of matrix models of different kinds. This will lead us to establish the phase diagram of the corresponding theories, uncovering rich patterns of quantum phase transitions and, along the way, demystifying certain subtle aspects that are often overlooked in the physics literature.

[1] These two families of QFTs overlap, as one may trade a supersymmetric theory for a topological one via a topological twist [36].

One of the central results of this thesis, for what concerns the large N limit, is to question the common assumption that $SU(N)$ gauge theories are well approximated by $U(N)$ theories at large N. We provide explicit and quantitative examples that clarify this issue. Most importantly, the mathematical results are supported by the physical expectations. Concretely, our study of large N three-dimensional Chern–Simons-matter theories in Chapter 5 shows that there is no essential difference between $SU(N)$ and $U(N)$ gauge theories, in agreement with the fact that both gauge theories flow to superconformal points that are easily connected by gauging an Abelian symmetry. In contrast, our study of five-dimensional Chern–Simons-matter theories at large N proves a discrepancy between the $SU(N)$ and the $U(N)$ theories, a reflection of the intuition that the superconformal limits of the latter, if they exist, need not belong to the same universality class of the superconformal limits of the former gauge theories.

Another major achievement of this work is to explore and enhance the landscape of deformations of two-dimensional Yang–Mills theory. We derive a new, one-parameter deformation from reducing topologically twisted super-Yang–Mills in five dimensions on a product manifold $\mathbb{S}^3_b \times \Sigma$, with Σ a closed Riemann surface and $\mathbb{S}^3_b$ a squashed sphere. The geometric parameter b deforms the effective 2d theory on Σ in such a way that the resulting matrix ensemble is a biorthogonal version of the discrete q-ensemble that characterizes q-deformed Yang–Mills.

Finally, we study the consequences of turning on the operator $T\overline{T}$ [45, 46] in the family of two-dimensional Yang–Mills theories. We prove that the cohomological localization of the path integral survives the perturbation by $T\overline{T}$, but introduces non-local interactions, reflected in multi-trace terms in the matrix model. This effect leads to the breakdown of some of the characteristic properties of two-dimensional pure Yang–Mills theories, such as the factorization of the partition function. The phase transition of pure Yang–Mills persists in the $T\overline{T}$-deformed model, but the weak coupling phase is eventually ruled out in the strong deformation regime. When both $T\overline{T}$ and q-deformation are turned on, the two effects compete in determining the phase diagram.

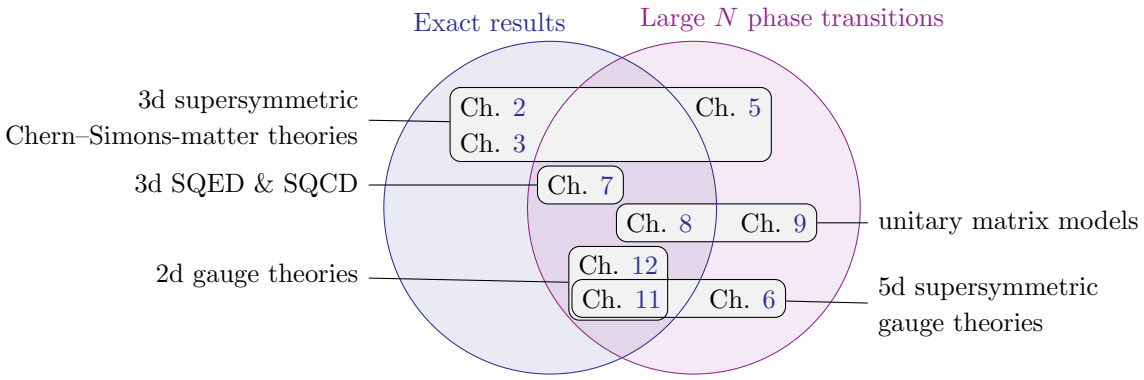
1.1 Organization of the material and summary

This dissertation is organized as follows. The topics covered are grouped in three parts, approximately tracing the three impacts my research had on mathematical physics. Part I investigates exact solutions in three-dimensional Chern–Simons theories. Part II is devoted to the study of phase transitions in gauge theories, derived from the matrix models. Part III introduces and analyzes new deformations of two-dimensional Yang–Mills theory.

A chart of the QFTs considered, the corresponding results and the chapters in which they are presented, is sketched in Figure 1.1.

Part I contains two contributions. Chapter 2 is based on [1]. We study several quiver Chern–Simons-matter theories on the three-sphere, combining the matrix model formulation with a systematic use of Mordell's integral, computing partition functions and checking dualities. We also consider Wilson loops in ABJ(M) theories. Using the Berele–Regev factorization of supersymmetric Schur polynomials [47], we express the expectation value of the Wilson loops in terms of sums of observables of two factorized copies of $U(N)$ pure Chern–Simons theory on the sphere. Then, we use the Cauchy identity to study the partition functions of a number of quiver Chern–Simons-matter models and the result is interpreted as a perturbative expansion in the parameters $t_j = -e^{2\pi m_j}$, where m_j are the masses.

Chapter 3 is based on [2]. The starting point of the analysis is the Langmann–Szabo–Zarembo (LSZ) matrix model, a complex matrix model with a quartic interaction and two external matrices that appears in the study of a scalar field theory on the noncommutative plane [48, 49]. We prove that the LSZ matrix model computes the probability of atypically large fluctuations in the Stieltjes–Wigert matrix model, which is a q-ensemble describing $U(N)$ Chern–Simons theory on



	3d supersymmetric gauge theories	5d supersymmetric gauge theories	2d gauge theories	unitary matrix models
Exact results	Ch. 2,3,7	Ch. 11	Ch. 11,12	Ch. 8
Phase transitions	Ch. 5,7	Ch. 6	Ch. 11,12	Ch. 8,9

Figure 1.1. Organization of the contents.

the three-sphere. The correspondence holds in a generalized sense: depending on the spectra of the two external matrices, the LSZ matrix model either describes probabilities of large fluctuations in the Chern–Simons partition function, in the unknot invariant or in the two-unknot invariant. We extend the result to supermatrix models, and show that a generalized LSZ supermatrix model describes the probability of atypically large fluctuations in the ABJ(M) matrix model.

Part II is the most extensive part of the thesis and consists of six chapters. Chapter 4 is an introduction to the large N method. It mostly reviews well-known results, but rephrased so to enlighten the ensuing discussion. In particular, Section 4.2 compares the large N limit of $SU(N)$ and $U(N)$ theories and may be of independent interest.

Chapters 5-6 are devoted to the study of phase transitions in three- and five-dimensional Chern–Simons-matter theories, respectively. In Chapter 5 we study $\mathcal{N} = 3$ $U(N)$ Chern–Simons theory on $\mathbb{S}^3$ coupled to an arbitrary number of massive hypermultiplet in the fundamental representation. We consider the large N limit and characterize the phase diagram in the large volume regime, uncovering a series of third order phase transitions. The results are checked to satisfy the $\mathcal{F}$-theorem. We then study Wilson loops in the fundamental and antisymmetric representations, with the latter that introduce an additional scaling parameter corresponding to the size of the representation. Additionally, we discuss the large N limit of three-dimensional $\mathcal{N} = 4$ SQCD at fixed radius.

The chapter is partly based on [3]. It contains the salient aspects of [3], but is completely rewritten and enhanced with novel observations and results, that substantiate physical conclusions on the low-energy behaviour of $U(N)$ supersymmetric Chern–Simons-matter theories. The large N analysis of SQCD on $\mathbb{S}^3$ is new and extends previous results in the literature, which used different methods from our saddle point analysis.

Five-dimensional $\mathcal{N} = 1$ theories with gauge group $U(N)$, $SU(N)$, $USp(2N)$ and $SO(N)$ on a large $\mathbb{S}^5$ are studied at large N in Chapter 6, based on [4]. The phase diagram of theories with fundamental hypermultiplets is universal and characterized by third order phase transitions, with the exception of $U(N)$, that shows both second and third order transitions. The phase diagram

of theories with adjoint or (anti-)symmetric hypermultiplets is also determined and found to be universal. Wilson loops in fundamental and antisymmetric representations are analyzed in this limit as well. All the results substantiate the $\mathcal{F}$-theorem.

In Chapter 7, we study supersymmetric Yang–Mills theories on $\mathbb{S}^3$, with massive matter and Fayet–Iliopoulos parameter, showing second order phase transitions in the limit of a large number of flavours. We study both partition functions and Wilson loops. Moreover, two interpretations of the partition function as eigenfunctions of the A_1 and free A_{N-1} hyperbolic Calogero–Moser integrable model are given. This chapter is taken from [5].

Chapters 8-9 are devoted to the analysis of unitary matrix models at large N. Following [6], we analyze two families of unitary matrix models, with weight functions that can be interpreted as characteristic polynomial insertions, in Chapter 8. We find phase transitions of the second and third order, depending on the model. A relationships between the unitary matrix models and continuous random matrix ensembles on the real line, of Cauchy–Romanovski type, is presented and studied both exactly and asymptotically.

Chapter 9 is based on [7]. In it, we study a unitary matrix model with Gross–Witten–Wadia weight function and determinant insertions. After some exact evaluations, we characterize the intricate phase diagram. There are five possible phases: an ungapped phase, two different one-cut gapped phases and two other two-cut gapped phases. The transition from the ungapped phase to any gapped phase is third order, but the transition between any one-cut and any two-cut phase appears to be second order. The physics of tunneling from a metastable vacuum to a stable one and of different releases of instantons is discussed. Wilson loops, β-functions and aspects of chiral symmetry breaking are investigated as well. Furthermore, we study in detail the meromorphic deformation of a general class of unitary matrix models, in which the integration contour is not anchored to the unit circle.

The focus of Part III is on Yang–Mills theories in two dimensions. We find two new solvable deformations of that theory, that we study both exactly and asymptotically. The basics of Yang–Mills theory in two dimensions are reviewed in Chapter 10.

Chapter 11 contains the results from [8]. We revisit the duality between five-dimensional supersymmetric gauge theories and deformations of two-dimensional Yang–Mills theory from a new perspective. We give a unified treatment of supersymmetric gauge theories in three and five dimensions using cohomological localization techniques and the Atiyah–Singer index theorem. We survey various known results in a unified framework and provide simplified derivations of localization formulas, as well as various extensions including the case of irregular Seifert fibrations. We describe the reductions to four-dimensional gauge theories, and give an extensive description of the dual two-dimensional Yang–Mills theory when the three-dimensional part of the geometry is a squashed three-sphere, including its extension to non-zero area, and a detailed analysis of the resulting matrix model. The squashing parameter b yields a further deformation of the usual q-deformation of two-dimensional Yang–Mills theory, which for rational values $b^2 = p/s$ yields a new correspondence with Chern–Simons theory on lens spaces $L(p, s)$.

Finally, in Chapter 12, we derive the $T\overline{T}$-perturbed version of two-dimensional q-deformed Yang-Mills theory on an arbitrary Riemann surface by coupling the unperturbed theory in the first order formalism to Jackiw–Teitelboim gravity. We show that the $T\overline{T}$-deformation results in a breakdown of the connection with a Chern-Simons theory on a Seifert manifold, and of the large N factorization into chiral and anti-chiral sectors. For the $U(N)$ gauge theory on the sphere, we show that the large N phase transition persists, and that it is of third order and induced by instantons. The effect of the $T\overline{T}$-deformation is to decrease the critical value of the 't Hooft coupling, and also to extend the class of line bundles for which the phase transition occurs. The same results are shown to hold for (q, t)-deformed Yang-Mills theory. Chapter 12 is largely based on [10], with the inclusion of an additional section that presents the results of [9].

We include a brief appendix that collects notation and conventions recurrently used throughout the thesis.

Other results not included in the dissertation have been presented in the articles [11–14]. The work [11] is similar in spirit to Chapter 8: we studied unitary matrix models that compute certain quantum amplitudes in a multi-parametric family of spin chains, both exactly and at large N, with special emphasis on the appearance of phase transitions.

In [12] we showed how the use of random matrix theory and q-ensembles provides exact results on the geometry of collections of data, while in [13] we computed a character expansion in terms of Schur polynomials, akin to those in Chapters 2-3, for random matrix reproducing kernels [39].

In [14] we put forward a new tool to study the Coulomb branches of three-dimensional supersymmetric gauge theories. Expanding on the web of correspondences among these Coulomb branches [50, 51], Hanany–Witten brane setups in type IIB string theory [52] and the affine Grassmannian [53], we introduced a combinatorial approach hinged upon Kashiwara crystals [54, 55].

Part I

Exact results in supersymmetric gauge theories

Chapter 2

Exact results and Schur expansions in quiver Chern–Simons-matter theories

2.1 Introduction to the chapter

In the last decade, the combined use of random matrix techniques together with the application of the supersymmetric localization method [56] has produced a wealth of analytical results in the study of supersymmetric gauge theories on compact manifolds, in a number of dimensions [57]. Both finite N properties and large N phenomena such as phase transitions have been elucidated by applying standard matrix model tools.

A very tractable set of theories in this area corresponds to Chern–Simons theories with supersymmetric matter in three dimensions [58] (see [59] for an overview on localization in three dimensions and [60] for an early review of Chern–Simons-matter matrix models). While a large number of results have already been uncovered for these models, we develop here further exact analytical characterizations of such theories, using the matrix model formulation. For this, we will be supplementing the matrix model approach with other analytical tools, such as the consideration of the so-called Mordell integral [61] which, in spite of its deceptively simple appearance, unpacks a wealth of analytical and physical information.

In Section 2.2 we will be presenting the necessary details, not only on Mordell's integral but on the other mathematical tools used. This section will provide physics background as well, while we sketch now the results and methods followed in a more panoramic manner.

In contrast to previous works following a similar approach to the one in the first part of this chapter [62–64], our study will include quiver Chern–Simons-matter theories. In this way, in this first part, contained in Section 2.3, we compute exactly the partition functions of various examples of Chern–Simons-matter theories on the three-sphere, systematically exploiting and interpreting the above mentioned result by Mordell [61]. The theories to be studied will be mostly Abelian quiver models whose computation is nevertheless laborious, but made possible by Mordell's result.

Some of these evaluations are actually duality checks. For example, we compute explicitly the partition function of the $U(1)^3$ theory, which, once particularized to Chern–Simons levels $(k_1, k_2, k_3) = (1, -1, 1)$, becomes the so-called Model III of Jafferis and Yin [65], and we obtain, as expected by duality, the equality with the simpler to compute, and known, partition function of SQED with two fundamental hypermultiplets and no Chern–Simons coupling.

Non-Abelian quiver theories, with the $U(1)_{k_1} \times U(2)_{k_2}$ theory as main example, are also briefly discussed and moreover we present the setup and sufficient conditions to evaluate the partition function of Abelian linear Chern–Simons quivers of arbitrary rank by iterative application of Mordell integrals.

We shall also be studying ABJ theories [44], that are $\mathcal{N} = 6$ $U(N_1)_k \times U(N_2)_{-k}$ Chern–

Simons theories with two bi-fundamental hypermultiplets. As it is well-known, they generalize ABJM theory [43], which is recovered when $N_1 = N_2 =: N$. There are exact computations of observables in ABJM theory when $N = 2$ in [66], and in [67] when mass and Fayet–Iliopoulos deformations are turned on. Besides, the partition function of ABJ theory with arbitrary rank has been evaluated in [68] using the continuation from Chern–Simons theory on the lens space $L(2, 1)$ as introduced in [69]. In [70], the result was confirmed through a direct integral transformation.

Thus, in Subsection 2.3.4, we complement these works by extending this type of analytical evaluations. We will give mass to the bi-fundamental hypermultiplets and add a Fayet–Iliopoulos parameter, and consider the deformed Abelian ABJM theory with Chern–Simons levels k_1 and k_2, reflecting the presence of a Romans mass in the dual gravitational theory [71].

Then, in Section 2.4, we will focus our attention on the vacuum expectation values of correlators of Wilson loops in ABJ(M) theory on $\mathbb{S}^3$. Among the various order loop operators in ABJ(M) theories [72], we will consider $\frac{1}{2}$-BPS Wilson loops [73], whose expectation value is captured by a matrix model that corresponds to the insertion of supersymmetric Schur polynomials in the ABJ(M) matrix model [73, 69].

As a novel consideration in the context of the study of such averages, we distinguish between typical (long) and atypical (short) representations and focus on the former, using the so-called Berele–Regev factorization of supersymmetric Schur polynomials [47] to give expressions in terms of sums of observables of $U(N)$ Chern–Simons theory on $\mathbb{S}^3$. The necessary background is given in the introductory Subsections 2.4.1 and 2.4.2.

As a matter of fact, the case of correlators is often simpler, with this approach, than the study of a single Wilson loop average. For example, by considering the case of two Wilson loops, we shall show that a consequence of the Berele–Regev factorization [47], is that the interacting term of the ABJ two-matrix model cancels out directly, and we immediately obtain the direct, disentangled product of two correlators of pairs of Wilson loops in $U(N)$ Chern–Simons on $\mathbb{S}^3$, each one computed independently, giving quantum dimensions. Furthermore, we will show in Subsection 2.4.5 how this formalism extends to quivers.

Then, in Section 2.5, we discuss a broad class of quiver Chern–Simons theories and our main tool will be the Cauchy identity. Its use entails expanding the matter contribution in a basis of symmetric functions, the Schur basis. As we shall show, these Schur expansions in the matrix model are perturbative evaluations of the observables described by the matrix model representation.

The perturbative meaning of the results has its roots in the nature of the Cauchy identity, reviewed in Subsection 2.2.4. Importantly, the results are not perturbative in the gauge couplings, but in certain other variables playing the role of fugacities for the flavour symmetries. In Subsection 2.5.4 we will combine the Cauchy identity with the Berele–Regev factorization and the results of Section 2.4 to study the expectation value of a single Wilson loop in ABJ theory.

2.2 Physics background and mathematical setup

2.2.1 Chern–Simons theories on $\mathbb{S}^3$

We consider Chern–Simons-matter theories with $\mathcal{N} \geq 3$ supersymmetry in three dimensions. These theories are obtained deforming the action of $\mathcal{N} = 4$ theories of vector and hypermultiplets by Chern–Simons couplings that preserve at least six of the eight supercharges. The resulting theories have a $SU(2)_{\mathrm{R}}$ R-symmetry, but when the microscopic, non-conformal theory is put on $\mathbb{S}^3$, only a maximal torus $U(1)_{\mathrm{R}} \subset SU(2)_{\mathrm{R}}$ is manifest. On a practical level, this amount of supersymmetry guarantees that:

- the Chern–Simons levels are not renormalized, and that

- we can identify the R-charges in the UV, where our computations are performed, with the R-charges in the IR, where the theory is strongly coupled.

There exists a vast literature describing the moduli spaces of vacua of these gauge theories, the most directly relevant for the present work being [74–77]. All the theories we discuss can be engineered in type IIB string theory [78, 79].

We first recall how to write the partition function of a 3d $\mathcal{N} \geq 3$ theory on $\mathbb{S}^3$ [58], which also serves as a presentation of our notation and conventions. The Chern–Simons theories we study have unitary gauge groups of the form

$$U(N_1) \times U(N_2) \times \cdots \times U(N_r).$$

Besides, we mainly consider hypermultiplets in the fundamental representation of a gauge group factor $U(N_p)$, as well as in the bi-fundamental representation of $U(N_p) \times U(N_{p+1})$.

In quiver notation, the number of nodes is r, with the node p corresponding to a gauge factor $U(N_p)$, for $p = 1, \ldots, r$. Unoriented edges between two nodes represent the bi-fundamental hypermultiplets.

The partition function receives the contributions [58]:

$$\begin{aligned}
&\text{vector multiplet at node } p\text{:} && \prod_{1\leq a<b\leq N_p} \left(2\sinh \pi(x_{p,a}-x_{p,b})\right)^2 ,\\
&\text{CS term at node } p\text{:} && \prod_{a=1}^{N_p} e^{i\pi k_p (x_{p,a})^2},\\
&\text{bi-fund. hypers between } p \text{ and } p'\text{:} && \prod_{a=1}^{N_p}\prod_{a'=1}^{N_{p'}} \left(2\cosh \pi(x_{p,a}-x_{p',a'})\right)^{-1} .
\end{aligned}$$

The Chern–Simons levels are k_p, which are required to be integers when $N_p > 1$ but can be rational for an Abelian gauge factor, $N_p = 1$. The isomorphism $U(N) \simeq [U(1) \times SU(N)]/\mathbb{Z}_N$ shows that each node yields an Abelian factor, to which there corresponds a topological global $U(1)_{\text{top}}$ symmetry. Real Fayet–Iliopoulos (FI) parameters ζ_p are introduced as background values of a twisted Abelian vector multiplet for the $U(1)_{\text{top},1} \times \cdots \times U(1)_{\text{top},r}$ symmetry. Furthermore, if we attach $N_{f,p}$ fundamental hypermultiplets to the node p, we can turn on real masses $\vec{m} = (m_{p,j})^{p=1,\ldots,r}_{j=1,\ldots,N_{f,p}}$ as background values of a vector multiplet for the global symmetry $PS[U(N_{f,1}) \times \cdots \times U(N_{f,r})]$ rotating the fundamentals. The tracelessness condition constrains the masses $|\vec{m}| = 0$. The contributions of FI terms and massive hypermultiplets to the partition function are:

$$\begin{aligned}
&\text{FI term at node } p\text{:} && \prod_{a=1}^{N_p} e^{i2\pi\zeta_p x_{p,a}},\\
&\text{fund. hypers at node } p\text{:} && \prod_{j=1}^{N_{f,p}}\prod_{a=1}^{N_p} \left(2\cosh \pi(x_{p,a}+m_{p,j})\right)^{-1} .
\end{aligned}$$

Eventually, we have to integrate over all the $x_{p,a}$. These variables parametrize the Cartan subalgebra of the gauge group,

$$\vec{x} = (x_{p,a})^{p=1,\ldots,r}_{a=1,\ldots,N_p} \in \mathfrak{u}(1)^{N_1} \times \cdots \times \mathfrak{u}(1)^{N_r} \cong \mathbb{R}^{N_1+\cdots+N_r} .$$

If we let r be the radius of $\mathbb{S}^3$, these adimensional variables are $\vec{x} = r\sigma|_{\text{loc}}$, where $\sigma|_{\text{loc}}$ is the value of the real scalar σ in the vector multiplet at the localization locus. The parameters $\vec{m}$ are adimensional as well, $\vec{m} = r\sigma_{\text{b.g.}}$ for $\sigma_{\text{b.g.}}$ the scalar in the background vector multiplet.

ABJ(M) theories and Chern–Simons levels

ABJ(M) theories [43, 44] are $U(N_1) \times U(N_2)$ Chern–Simons theories with $\mathcal{N} = 6$ supersymmetry, which forces the Chern–Simons level to be $(k_1, k_2) = (k, -k)$. In quiver notation, these are extended $\widehat{A}_1$ quiver theories. They have their origin in string/M-theory, and have been conceived as the gauge theory dual to the supergravity solution on $\mathrm{AdS}_4 \times \mathbb{CP}^3$.

A natural question on the gauge theory side is whether there exists a theory with generic levels (k_1, k_2). This point has been addressed in the early days of ABJM theory by Gaiotto and Tomasiello in [71]. It is possible to deform ABJ(M) theories to arbitrary levels deforming the gravity dual solution by a Romans mass, commonly denoted F_0. There are different ways to do so [71], and we will only consider the $\mathcal{N} = 3$ supersymmetric solution. The resulting gauge theory has the same field content of the ABJ(M) theory, but with Chern–Simons levels (k_1, k_2) that obey $k_1 + k_2 = F_0$. For the deformation of other Chern–Simons-matter theories by a Romans mass in the gravity dual, see for example [80, 81].

We remark that, by mass deformations, we will never refer to a Romans mass, and instead always mean the procedure described above to give mass to the hypermultiplets promoting the masses to background scalar fields.

$\frac{1}{2}$-BPS Wilson loops

In $\mathcal{N} \geq 3$ supersymmetric Chern–Simons theories, supersymmetry-preserving Wilson loops in a representation $\mathcal{R}$ of the gauge group wrap a great circle in $\mathbb{S}^3$. Their vacuum expectation value (vev) is computed by localization [58]:

$$\langle W_{\mathcal{R}} \rangle = \left\langle \mathrm{Tr}_{\mathcal{R}} e^{2\pi r \sigma|_{\mathrm{loc}}} \right\rangle .$$

In the formula, $2\pi r$ is the length of the great circle, $\sigma|_{\mathrm{loc}}$ is the value of the real scalar σ at the localization locus as explained above, $\mathrm{Tr}_{\mathcal{R}}$ is the normalized trace in the representation $\mathcal{R}$ and $\langle \cdots \rangle$ means the average of the quantity in the ensemble obtained from localization, which of course depends on the theory under study.

In quiver Chern–Simons theories it is possible to construct Wilson loops charged under a $U(N)$ factor of the gauge group that preserve (at least) two supercharges. For the special case of ABJ theories, however, it is possible to consider Wilson loops in a representation $\mathcal{R}$ of the supergroup $U(N_1|N_2)$ that preserve half of the $\mathcal{N} = 6$ supersymmetry, that is $\frac{1}{2}$-BPS Wilson loops [73, 82].

Unknot invariant in pure Chern-Simons theory

The vev of a Wilson loop in bosonic pure Chern–Simons theory computes the unknot invariant [42]. It was first evaluated with the Chern–Simons matrix model in [83], giving:

$$\langle W_\mu \rangle_{\mathrm{CS}(N)} = (\dim_q \mu)\ q^{-\frac{1}{2}\mathsf{C}_{2;N}(\mu)}. \tag{2.2.1}$$

In the computation in [83] it was shown that the integration of a Schur polynomial in a Stieltjes–Wigert ensemble (and equivalently, in the Chern–Simons matrix model [84]) gives the principal specialization of the Schur polynomial, leading to the expression (2.2.1). This property has been discussed, later on, in a broader sense in [85, 86] (see also [87] for a general discussion of two Schur polynomial insertions).

In (2.2.1) the q-parameter was taken to be real $0 < q = e^{-g} < 1$ and is related to the q-parameter of Chern–Simons theory at level k through the analytic continuation

$$g \mapsto \frac{i2\pi}{k}, \tag{2.2.2}$$

$\dim_q \mu$ is the quantum dimension of the representation μ and $\mathsf{C}_{2;N}(\mu)$ is the quadratic Casimir of $U(N)$ in the representation μ. Knot and link invariants computed in Chern–Simons theory come equipped with a framing [42], and we stress that (2.2.1) is computed not in canonical framing but in the matrix model framing, which is a specific case of Seifert framing.

2.2.2 Mordell integrals

The two integrals we will exploit are [61]:

$$\begin{aligned}\Psi_+(\xi,\lambda;\kappa,\varrho) &:= \int_{-\infty}^{+\infty} dx \frac{e^{i\pi\frac{\kappa}{\varrho}x^2-2\pi x\xi}}{e^{2\pi x}-e^{i2\pi\lambda}} \\ &= \frac{e^{-i\pi\lambda\left(2+2\xi+\frac{\kappa}{\varrho}\lambda\right)}}{e^{i\pi\varrho\left(2\xi+2\frac{\kappa}{\varrho}\lambda-\kappa\right)}-1}\left[-\sqrt{\frac{i\varrho}{\kappa}}\sum_{\alpha=1}^{\kappa} e^{i\pi\frac{\varrho}{\kappa}\left(\xi+\frac{\kappa}{\varrho}\lambda+\alpha\right)^2}+i\sum_{\beta=1}^{\varrho} e^{i\pi\beta\left(2\xi+2\frac{\kappa}{\varrho}\lambda-\frac{\kappa}{\varrho}\beta\right)}\right]\end{aligned} \tag{2.2.3}$$

and

$$\begin{aligned}\Psi_-(\xi,\lambda;\kappa,\varrho) &:= \int_{-\infty}^{+\infty} dx \frac{e^{-i\pi\frac{\kappa}{\varrho}x^2-2\pi x\xi}}{e^{2\pi x}-e^{i2\pi\lambda}} \\ &= \frac{e^{-i\pi\lambda\left(2+2\xi-\frac{\kappa}{\varrho}\lambda\right)}}{e^{i\pi\varrho\left(2\xi-2\frac{\kappa}{\varrho}\lambda-\kappa\right)}-1}\left[\sqrt{-\frac{i\varrho}{\kappa}}\sum_{\alpha=0}^{\kappa-1} e^{-i\pi\frac{\varrho}{\kappa}\left(\xi-\frac{\kappa}{\varrho}\lambda-\alpha\right)^2}+i\sum_{\beta=1}^{\varrho} e^{i\pi\beta\left(2\xi-2\frac{\kappa}{\varrho}\lambda+\frac{\kappa}{\varrho}\beta\right)}\right],\end{aligned} \tag{2.2.4}$$

valid for $\kappa,\varrho\in\mathbb{Z}_{>0}$ and $0<\Re\lambda<1$. Note that the left-hand side only depends on the ratio κ/ϱ. Strictly speaking, these identities only appear in [61, Eq.s (8.1)-(8.2)] for $\lambda=0$, but can be easily extended mimicking the manipulations that lead to [61, Eq. (3.8)]. Since, in doing so, there is a subtlety in the choice of integration contour, we spell the details in Appendix 2.A for completeness.

The interest in the closed form evaluations (2.2.3)-(2.2.4) arose originally in the context of number theory. Related expressions, that in the gauge theory setup correspond to massless hypermultiplets, where first considered by Ramanujan [88].

The building blocks of our solutions will be the integrals

$$\mathcal{I}_k(y,\check{\xi}) := \int_{-\infty}^{+\infty} dx \frac{e^{i\pi k x^2+2\pi x\check{\xi}}}{e^{2\pi x}+e^{2\pi y}}, \tag{2.2.5}$$

for rational k. Comparing with (2.2.3) and (2.2.4), it is clear that

$$\mathcal{I}_k(y,\check{\xi}) = \begin{cases}\Psi_+(\xi=-\check{\xi},\lambda=\frac{1}{2}-iy;\kappa,\varrho) & \text{if } k=+\frac{\kappa}{\varrho};\\ \Psi_-(\xi=-\check{\xi},\lambda=\frac{1}{2}-iy;\kappa,\varrho) & \text{if } k=-\frac{\kappa}{\varrho},\end{cases} \tag{2.2.6}$$

for $\kappa,\varrho\in\mathbb{Z}_{>0}$.

2.2.3 Moments of the log-normal

We introduce the moments of the log-normal distribution, which will appear in our computations. Using a change of variables of the form $X_a=e^{2\pi x_a}$, it was shown in [84] that the partition function of $U(N)$ Chern–Simons theory on $\mathbb{S}^3$, analytically continued to $q=e^{-g}$, $g>0$, is proportional the

partition function of the Stieltjes–Wigert (SW) ensemble. Hence, pure Chern–Simons is solved by polynomials orthogonal with respect to the measure

$$e^{-\frac{1}{2g}(\log X)^2}dX,$$

on $0 < X < \infty$, after the continuation $g \mapsto \frac{i2\pi}{k}$ introduced in (2.2.2). The moments of the log-normal measure are

$$\mu_\alpha(g) = \int_0^{+\infty} \frac{dX}{2\pi X} X^\alpha e^{-\frac{1}{2g}(\log X)^2} = \int_{-\infty}^{\infty} dx e^{-\frac{4\pi}{g}x^2+2\pi\alpha x},$$

defined for $\Re g > 0$ and $\alpha \in \mathbb{Z}$. We immediately find

$$\mu_\alpha(g) = \sqrt{\frac{g}{2\pi}} e^{g\frac{\alpha^2}{2}} = \sqrt{\frac{g}{2\pi}} q^{-\frac{\alpha^2}{2}}.$$

We can collect these moments into a formal generating series:

$$P(z;g) = \sum_{\alpha\in\mathbb{Z}} z^\alpha \mu_\alpha(g). \tag{2.2.7}$$

In the present work, as usual for Chern–Simons theories, we are interested in the analytic continuation (2.2.2), and we write

$$\widetilde{\mu}_\alpha(k) = \sqrt{\frac{i}{k}} q^{-\frac{\alpha^2}{2}} \tag{2.2.8}$$

to denote the moment continued as prescribed in (2.2.2). When q is a κ^{th} root of unity, namely $|k| = \frac{\kappa}{\varrho}$, the generating series (2.2.7) only contains $|\kappa|$ different terms, hence we can resum it:

$$\begin{aligned}\widetilde{P}(z;k) &= \sum_{n\in\mathbb{Z}}\sum_{\alpha=1}^{\kappa} z^{n\kappa+\alpha}\widetilde{\mu}_{\alpha+n\kappa}(k) = \sum_{\alpha=1}^{\kappa} z^\alpha \widetilde{\mu}_\alpha(k)\left[\sum_{n\in\mathbb{Z}} z^{n\kappa}\widetilde{\mu}_{n\kappa}(k)\right] \\ &= \sum_{\alpha=1}^{\kappa} z^\alpha \widetilde{\mu}_\alpha(k).\end{aligned} \tag{2.2.9}$$

This type of resummation is the reason [89] why only integrable $U(N)$ representations contribute to the partition function of Chern–Simons theory on $\mathbb{S}^3$ with q root of unity [42], while all the unitary irreducible representations contribute when q is analytically continued to $q = e^{-g}$.

Looking back at the Mordell integrals (2.2.3)-(2.2.4) we notice that the first of the two sums in $\Psi_\pm$ when $\xi \in \mathbb{Z}$ gives precisely

$$e^{i\pi k\lambda^2}\widetilde{P}\left(e^{i2\pi\lambda\text{sign}(k)};k\right),$$

up to an irrelevant shift in the range of the variable α, now running on $1+\xi,\dots,\kappa+\xi$ in Ψ_+ and on $-\xi,\dots,\kappa-\xi-1$ in Ψ_-. The overall Gaussian coefficient is cancelled by a contribution from the numerator of the overall term in (2.2.3)-(2.2.4). We will see that the fugacity $e^{i2\pi\lambda}$ will play a central role, as further discussed in Sections 2.3.3 and 2.5.

2.2.4 Cauchy identities, Gauss sums and notation

We state here relevant mathematical identities which we will exploit in the text.

Cauchy identity

The Cauchy identity [90, 91]:

$$\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2}\frac{1}{1-X_aY_{\dot{a}}}=\sum_{\nu}\mathfrak{s}_\nu(X_1,\dots,X_{N_1})\mathfrak{s}_\nu(Y_1,\dots,Y_{N_2}) \tag{2.2.10}$$

where $\mathfrak{s}_\nu$ is the Schur polynomial [90, 91] labelled by the Young diagram ν satisfying

$$\text{length}(\nu)\leq\min\{N_1,N_2\}\,.$$

This is a well-known identity, which has become increasingly familiar, and useful, in many contexts, especially in random matrix theory, see e.g. [92, 93].

Note that the equality in (2.2.10) holds analytically if $|X_a|<1$ and $|Y_{\dot{a}}|<1$, and algebraically otherwise [92].

When $N_2=1$, the Cauchy identity reduces to the generating function of the complete homogeneous symmetric polynomials $\mathfrak{h}_\nu$ [90]:

$$\prod_{a=1}^{N}\frac{1}{1-X_aY}=\sum_{\nu=0}^{\infty}Y^\nu\mathfrak{h}_\nu(X_1,\dots,X_N), \tag{2.2.11}$$

where

$$\mathfrak{h}_\nu(X_1,\dots,X_N)=\sum_{1\leq a_1\leq a_2\leq\cdots\leq a_\nu\leq N}X_{a_1}\cdots X_{a_\nu}.$$

Equivalently,

$$\mathfrak{h}_\nu(X_1,\dots,X_N)=\sum_{\mu,|\mu|=\nu}\mathfrak{m}_\mu(x_1,\cdots,x_N),\qquad\forall\nu\in\mathbb{Z}_{\geq0},$$

with $\mathfrak{m}_\mu(x_1,\dots,x_N)=x_1^{\mu_1}\cdots x_N^{\mu_N}$ being the monomials [90] and the sum running over all partitions μ of size $|\mu|=\nu$.

There exists a related identity, known as *dual* Cauchy identity [91]:

$$\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2}(1+X_aY_{\dot{a}})=\sum_{\nu}\mathfrak{s}_{\nu'}(X_1,\dots,X_{N_1})\mathfrak{s}_\nu(Y_1,\dots,Y_{N_2}) \tag{2.2.12}$$

with ν' the conjugate partition, obtained transposing rows and columns of the Young diagram of the partition ν. An important aspect of (2.2.12) is that, differently from (2.2.10), the sum contains a finite number of terms, due to the restriction

$$\text{length}(\nu')=\nu_1\leq N\equiv\min\{N_1,N_2\}\,.$$

Therefore the sum in (2.2.12) only involves partitions whose Young diagrams fit in a $N\times N$ square.

When $N_2=1$, this latter Cauchy identity reduces to the generating function of the elementary symmetric polynomials $\mathfrak{e}_\nu$ [90]:

$$\prod_{a=1}^{N}(1+X_aY)=\sum_{\nu=0}^{N}Y^\nu\mathfrak{e}_\nu(X_1,\dots,X_N), \tag{2.2.13}$$

where

$$\mathfrak{e}_\nu(X_1,\dots,X_N)=\sum_{1\leq a_1<a_2<\cdots<a_\nu\leq N}X_{a_1}\cdots X_{a_\nu}.$$

Gauss sums

The Gauss sum identity:

$$\frac{1}{\sqrt{i\kappa}}\sum_{\alpha=0}^{\kappa-1} e^{\frac{i\pi}{\kappa}\left(\alpha-\ell-\frac{\kappa}{2}\right)^2} = 1, \tag{2.2.14}$$

valid for $\kappa \in \mathbb{Z}_{>0}$ and for every $\ell \in \mathbb{Z}$. This formula will be instrumental to obtain the massless limit of all the computations in Section 2.3.

Remarks on notation

To avoid clutter, whenever possible we will change the notation $x_{p,a}$ for a more suitable one. For example, for $r = 2$, we will write $(x_a, y_{\dot{a}})$ instead of $(x_{1,a}, x_{2,a'})$. Similarly, we will mostly denote the masses simply $\{m_j\}$, when it is clear from the context to which node each one is attached. Besides, throughout the work, we will sometimes switch to exponentiated variables, which we will denote with upper case letters. So, for example, we will use $X_a = e^{2\pi x_a}$, $M_j = e^{2\pi m_j}$ and so on. Moreover, for a given Chern–Simons level k, we define as usual

$$q = \exp\left(-\frac{i2\pi}{k}\right). \tag{2.2.15}$$

The field content of the theories we study is conveniently encoded in A-type Dynkin diagrams or in affine $\widehat{A}$-type Dynkin diagrams. We will interchangeably call the first class A quivers or linear quivers, and the second class $\widehat{A}$ quivers, extended quivers or necklace quivers.

We draw such quivers in the 3d $\mathcal{N} = 4$ quiver notation, so the edges represent hypermultiplets in the bi-fundamental representations and are not directed. The Chern–Simons levels will be implicit in the quiver diagrams.

2.3 Evaluation of partition functions

2.3.1 $U(1)_k$ Chern–Simons theory with a fundamental hypermultiplet

We start our analysis revisiting the simplest Chern–Simons theory that includes matter: $U(1)_k$ Chern–Simons theory with a fundamental hypermultiplet, represented in Figure 2.1.

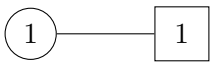


Figure 2.1. $U(1)_k$ theory with $N_f = 1$ fundamental flavour. This is an Abelian A_1 quiver.

The moduli space of vacua of the theory in flat space has been analyzed in [94], with a focus on its S-duality properties. The Chern–Simons term gives a topological mass to the vector multiplet, lifting the Coulomb branch. The moduli space has a non-compact one-dimensional Higgs branch, which is also lifted turning on a real mass deformation. In an Abelian Chern–Simons theory, admitting rational k, S-duality acts as the S-matrix of the $SL(2,\mathbb{Z})$ group on the coupling while exchanging mass and FI terms

$$k \mapsto -\frac{1}{k}, \quad \zeta \mapsto m \quad m \mapsto -\zeta. \tag{2.3.1}$$

The theory with gauge group $U(1)$ and $N_f = 1$ is self-dual under S-duality [94].

The partition function at rational Chern–Simons level k and with mass and FI parameters turned on is

$$\mathcal{Z}_{U(1),1}\left(k,m,\zeta\right)=\int_{-\infty}^{+\infty}dx\frac{e^{i\pi x^2k+i2\pi\zeta x}}{2\cosh\pi(x+m)}. \tag{2.3.2}$$

We do not need to consider both deformations: shifting variables $x'=x+m$ we get

$$\mathcal{Z}_{U(1),1}\left(k,m,\zeta\right)=e^{i\pi km^2+i2\pi m\zeta}\mathcal{Z}_{U(1),1}\left(k,0,\zeta-km\right), \tag{2.3.3}$$

while shifting variables $x'=x+\zeta/k$ we get

$$\mathcal{Z}_{U(1),1}\left(k,m,\zeta\right)=e^{-i\frac{\pi}{k}\zeta^2}\mathcal{Z}_{U(1),1}\left(k,m-\frac{\zeta}{k},0\right). \tag{2.3.4}$$

Therefore, it is sufficient to take one of the two deformations, and the more general result follows immediately. Note how the prefactor suffers a change $k\mapsto-\frac{1}{k}$ when the roles of m and ζ are exchanged, as well as the presence of the additional phase $e^{i2\pi\zeta m}$ in (2.3.3), coupling the FI background twisted vector multiplet to the flavour background vector multiplet.[2]

From the integral representation (2.3.2) the self-duality is easily proven:

$$\begin{aligned}\mathcal{Z}_{U(1),1}\left(k,m,\zeta\right)&=\int_{-\infty}^{+\infty}dx\int_{-\infty}^{+\infty}dy\frac{e^{i\pi x^2k-i2\pi x(y-\zeta)}e^{-i2\pi ym}}{2\cosh(\pi y)}\\&=\sqrt{\frac{i}{k}}e^{-i2\pi m\zeta}\int_{-\infty}^{+\infty}dy\frac{e^{-i\frac{\pi}{k}y^2-i2\pi ym}}{2\cosh\pi(y+\zeta)}=\sqrt{\frac{i}{k}}e^{-i2\pi m\zeta}\mathcal{Z}_{U(1),1}\left(-k^{-1},\zeta,-m\right),\end{aligned}$$

where we have used the fact that $(\cosh\pi x)^{-1}$ is Fourier transformed into itself.

We now use Mordell's formula to evaluate exactly the partition function. Starting with $m\neq0$ and $\zeta=0$ in (2.3.2) we have

$$\mathcal{Z}_{U(1),1}\left(k,m,0\right)=e^{-\pi m}\mathcal{I}_k\left(-m,\frac{1}{2}\right), \tag{2.3.5}$$

given in terms of a Mordell integral. For rational k with $|k|=\frac{\kappa}{\varrho}$, (2.2.6) gives

$$\begin{aligned}\mathcal{Z}_{U(1),1}\left(k>0,m,0\right)=\frac{1}{1-(-1)^{\kappa\varrho-\kappa+\varrho}e^{-2\pi\kappa m}}\Bigg\{&-e^{i\pi k\left(m-\frac{i}{2}\right)^2}\sum_{\beta=1}^{\varrho}\left(-e^{-2\pi km}\right)^{\beta}e^{-i\pi k\beta(\beta-1)}\\&+\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{\kappa}\left(-e^{-2\pi m}\right)^{\alpha-\frac{1}{2}}e^{i\frac{\pi}{k}\left(\alpha-\frac{1}{2}\right)^2}\Bigg\},\end{aligned} \tag{2.3.6}$$

$$\begin{aligned}\mathcal{Z}_{U(1),1}\left(k<0,m,0\right)=\frac{1}{1-(-1)^{\kappa\varrho-\kappa-\varrho}e^{2\pi\kappa m}}\Bigg\{&e^{i\pi k\left(m-\frac{i}{2}\right)^2}\sum_{\beta=1}^{\varrho}\left(-e^{-2\pi km}\right)^{\beta}e^{-i\pi k\beta(\beta-1)}\\&+\sqrt{\frac{i}{|k|}}\sum_{\alpha=0}^{\kappa-1}\left(-e^{2\pi m}\right)^{\alpha-\frac{1}{2}}e^{i\frac{\pi}{k}\left(\alpha-\frac{1}{2}\right)^2}\Bigg\}.\end{aligned} \tag{2.3.7}$$

The factor $e^{-\pi m}$ in (2.3.5) is cancelled against a contribution from the overall factor in the Mordell integrals (2.2.3)-(2.2.4).

[2]The minus sign $m\mapsto-\zeta$ in (2.3.1) comes from our conventions, presented in Subsection 2.2.1. The necessity of that sign can be checked applying S-duality to (2.3.3) and (2.3.4). It is a $\mathbb{Z}_2^C$ twist by charge conjugation that compensates the fact that both ζ and m are in the conjugate representation than the one required by S-duality.

When $k \in \mathbb{Z}$, hence $\varrho = 1$, these latter two expressions reduce to

$$\mathcal{Z}_{U(1),1}\left(k>0,m,0\right) = \frac{e^{i\pi k\left(m^2-\frac{1}{4}\right)+\pi m}}{2\cosh(\pi k m)} + \frac{1}{1+e^{-2\pi k m}}\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{k}\left(-e^{-2\pi m}\right)^{\alpha-\frac{1}{2}} q^{-\frac{1}{2}\left(\alpha-\frac{1}{2}\right)^2},$$

$$\mathcal{Z}_{U(1),1}\left(k<0,m,0\right) = -\frac{e^{i\pi k\left(m^2-\frac{1}{4}\right)+\pi m}}{2\cosh(\pi k m)} + \frac{1}{1+e^{-2\pi k m}}\sqrt{\frac{i}{k}}\sum_{\alpha=0}^{|k|-1}\left(-e^{2\pi m}\right)^{\alpha-\frac{1}{2}} q^{-\frac{1}{2}\left(\alpha-\frac{1}{2}\right)^2}.$$

The result is a real analytic function of m, and is holomorphic in the usual "physical" strip $-\frac{1}{2} < \Im m < \frac{1}{2}$. Note that, using the relation $\lambda = \frac{1}{2} + im$ (see (2.2.6)) between the physical variable m and the variable λ of [61], the result is holomorphic in $0 < \Re\lambda < 1$, in agreement with the proof in Appendix 2.A, based on [61].

Setting instead $m = 0, \zeta \neq 0$ in (2.3.2) we have

$$\mathcal{Z}_{U(1),1}\left(k,0,\zeta\right) = \mathcal{I}_k\left(0,\check{\xi}=\frac{1}{2}+i\zeta\right).$$

The solution is read off from (2.2.6) for any rational k,

$$\begin{aligned}\mathcal{Z}_{U(1),1}\left(k>0,0,\zeta\right) = &-\frac{e^{-i\frac{\pi}{4}k-\pi\zeta}}{1-(-1)^{\kappa-\varrho-\kappa\varrho}e^{2\pi\varrho\zeta}}\sum_{\beta=1}^{\varrho}\left(-e^{2\pi\zeta}\right)^{\beta} e^{-i\pi k\beta(\beta-1)}\\ &-\frac{ie^{-i\frac{\pi}{k}\zeta^2}}{1-(-1)^{\kappa-\varrho-\kappa\varrho}e^{2\pi\varrho\zeta}}\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{\kappa}\left(-e^{2\frac{\pi}{k}\zeta}\right)^{\alpha-\frac{1}{2}} e^{i\frac{\pi}{k}\left(\alpha-\frac{1}{2}\right)^2},\end{aligned} \tag{2.3.8}$$

$$\begin{aligned}\mathcal{Z}_{U(1),1}\left(k<0,0,\zeta\right) = &-\frac{e^{-i\frac{\pi}{4}k-\pi\zeta}}{1-(-1)^{\kappa+\varrho+\kappa\varrho}e^{2\pi\varrho\zeta}}\sum_{\beta=1}^{\varrho}\left(-e^{2\pi\zeta}\right)^{\beta} e^{-i\pi k\beta(\beta-1)}\\ &-\frac{ie^{-i\frac{\pi}{k}\zeta^2}}{1-(-1)^{\kappa+\varrho+\kappa\varrho}e^{2\pi\varrho\zeta}}\sqrt{\frac{i}{|k|}}\sum_{\alpha=1}^{\kappa}\left(-e^{-2\frac{\pi}{k}\zeta}\right)^{\alpha-\frac{1}{2}} e^{i\frac{\pi}{k}\left(\alpha-\frac{1}{2}\right)^2}.\end{aligned} \tag{2.3.9}$$

When $k \in \mathbb{Z}$ it takes the simpler form:

$$\mathcal{Z}_{U(1),1}\left(k>0\right) = \frac{e^{-i\pi\frac{k}{4}}}{2\cosh(\pi\zeta)} - \frac{1}{e^{2\pi\zeta}+1}\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{k}(-1)^{\alpha}q^{-\frac{1}{2}\left(\alpha-\frac{1}{2}-i\zeta\right)^2},$$

$$\mathcal{Z}_{U(1),1}\left(k<0\right) = \frac{e^{-i\pi\frac{k}{4}}}{2\cosh(\pi\zeta)} - \frac{1}{e^{2\pi\zeta}+1}\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{|k|}(-1)^{\alpha}q^{-\frac{1}{2}\left(\alpha-\frac{1}{2}+i\zeta\right)^2},$$

where we recall that $q = e^{-i2\pi/k}$ from (2.2.15). The solution is real analytic in $\zeta \in \mathbb{R}$ and holomorphic in the strip $-\frac{1}{2} < \Im\zeta < \frac{1}{2}$.

We recognize the generating polynomial of the moments of the SW distribution when q is a k^{th} root of unity, $\widetilde{P}(z;k)$, evaluated at $z = -e^{-\text{sign}(k)2\pi m}$ for the theory with only mass term and at $z = -q^{\check{\xi}}$ for the theory with only FI term.

Direct inspection shows that

$$(2.3.6) \;=\; \sqrt{\frac{i}{k}} \times \left[(2.3.9) \text{ with } -\zeta = m \text{ and } \kappa \leftrightarrow \varrho\right],$$

and likewise for (2.3.7) and (2.3.8). This together with the relations (2.3.3)-(2.3.4) gives a full check of the self S-duality of the solution.

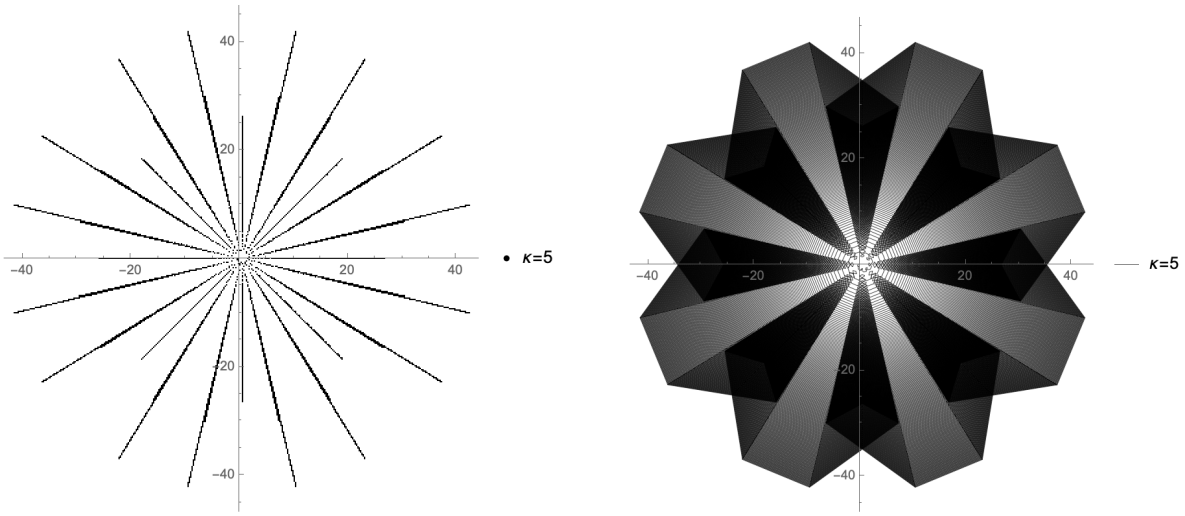


Figure 2.2. Left: Plot of $\mathcal{Z}_{U(1),1}$ with $k = \frac{\kappa}{\varrho}$, $\zeta = 0$, $m = 0.2$ at fixed $\kappa = 5$ and varying $\varrho = 1, \ldots, 10^4$. Right: Same plot, with points obtained from consecutive values of ϱ joined by a segment.

We plot the result (2.3.6) of $\mathcal{Z}_{U(1),1}$ with positive rational k and $\zeta = 0$ in Figure 2.2 and 2.3. Being q a κ^{th} root of unity, at fixed κ and varying ϱ the values of the partition function are placed along rays in $\mathbb{C}$. Increasing κ increases the number of rays.

As the result holds upon complexification of the mass with $|\Im m| < \frac{1}{2}$, it is instructive as well to plot the partition function at fixed κ and increasing ϱ for complex values of m, as we do in Figure 2.4 (for $\kappa = 5$) and Figure 2.5 (for $\kappa = 8$).

2.3.2 Single node quivers

Abelian A_1 theory with two flavours

We consider a $U(1)$ Chern–Simons theory with two massive hypermultiplets in the fundamental representation, see Figure 2.6. The result we present for this theory has been first derived in [62], and we revisit it here as a warm up.

The theory has $SU(2)$ flavour symmetry and the hypermultiplets have masses $(+m, -m)$. The partition function is

$$\begin{aligned}\mathcal{Z}_{U(1),2}(k,m) &= \int_{-\infty}^{\infty} dx \frac{e^{i\pi k x^2}}{4\cosh\pi\,(x-m)\cosh\pi\,(x+m)} \\ &= \frac{1}{2\sinh(2\pi m)} \int_{-\infty}^{\infty} dx e^{i\pi k x^2 + 2\pi x} \left[\frac{1}{e^{2\pi x} + e^{-2\pi m}} - \frac{1}{e^{2\pi x} + e^{2\pi m}}\right] \\ &= \frac{\mathcal{I}_k(-m,1) - \mathcal{I}_k(m,1)}{2\sinh(2\pi m)},\end{aligned}$$

where in the last line we have recognized (2.2.5). The solution in terms of the Mordell integrals (2.2.3) (when $k > 0$) or (2.2.4) (when $k < 0$) holds for any non-zero rational Chern–Simons level. However, the expressions are clearer for $k \in \mathbb{Z}$. Under such assumption, from equation (2.2.6) and simple manipulations, we obtain

$$\mathcal{Z}_{U(1),2}(k,m) = \frac{1}{2\sinh(2\pi m)} \left[-\frac{i e^{i\pi k\left(m^2 - \frac{1}{4}\right)}}{\sinh(\pi k m)} - \frac{\widetilde{P}(-e^{-2\pi m}; k)}{e^{-2\pi|k|m} - 1} + \frac{\widetilde{P}(-e^{2\pi m}; k)}{e^{2\pi|k|m} - 1}\right], \tag{2.3.10}$$

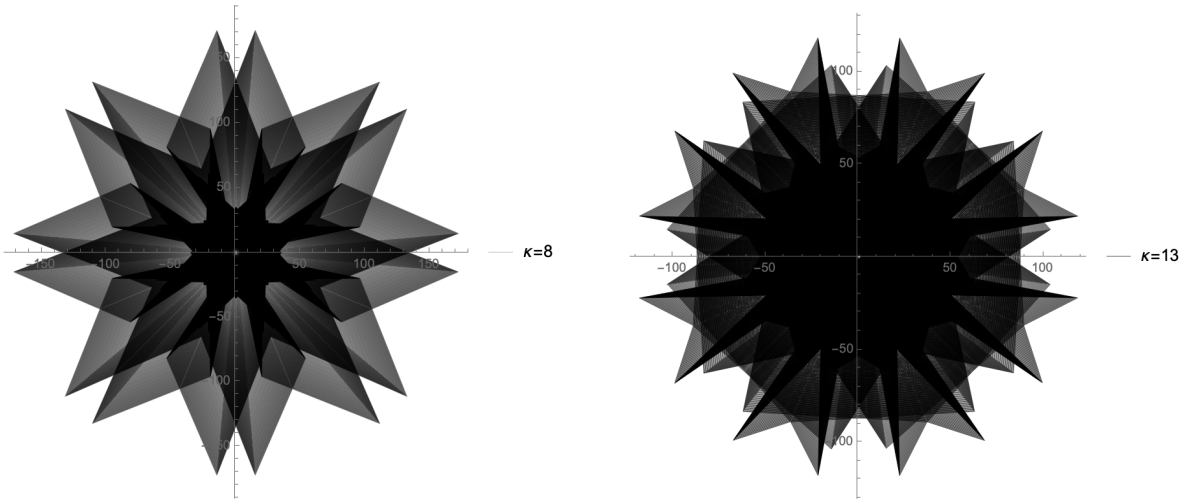


Figure 2.3. Plot of $\mathcal{Z}_{U(1),1}$ with $k = \frac{\kappa}{\varrho}$, $\zeta = 0$, $m = 0.02$ at fixed κ and varying ϱ. The points obtained from consecutive values of ϱ are joined by a segment. Left: $\kappa = 8$, $\varrho = 1, \dots, 2 \times 10^4$. Right: $\kappa = 13$, $\varrho = 1, \dots, 10^4$.

with the polynomial $\widetilde{P}(z;k)$ defined in (2.2.9). We also have shifted the summation range hidden in $\widetilde{P}(z;k)$, so that the sum runs over $\alpha = 0, \dots, k-1$ if $k > 0$ and $\alpha = 1, \dots, |k|$ if $k < 0$.

The masses of the hypermultiplets have played a central role in the derivation, but we can take the massless limit of our final result [62]. Despite each term being divergent, a careful analysis and the application of the Gauss sum identity (2.2.14) show that the result is finite and well defined, and reads

$$\begin{aligned}\mathcal{Z}_{U(1),2}(k>0, m\to 0^+) &= \frac{e^{-i\pi\frac{k}{4}}}{(2\pi m)^2 k}\left[-i + i\frac{1}{\sqrt{ik}}\sum_{\alpha=0}^{k-1} e^{\frac{i\pi}{k}(\alpha+\frac{k}{2})^2}\left(1+2\pi^2 m^2\left(\alpha^2+\frac{k^2}{6}-\alpha k\right)\right)\right] \\ &= \sqrt{\frac{i}{k}}\sum_{\alpha=0}^{k-1}(-1)^\alpha q^{-\frac{\alpha^2}{2}}\left[\frac{1}{k}\left(\alpha-\frac{k}{2}\right)^2-\frac{k}{12}\right],\end{aligned}$$

where to go from the first to the second line we have used (2.2.14). The analogous result when $k < 0$ is derived by the same steps.

The solution of the Mordell integrals $\Psi_\pm$ requires $0 < \Re\lambda < 1$, and we have used $\lambda = \frac{1}{2} \pm im$. Therefore we can complexify the masses in the strip $-\frac{1}{2} < \Im m < \frac{1}{2}$, which is the usual "physical" region in which the integrals from localization do not develop singularities.

Abelian A_1 theory with N_f flavours

The analysis of the Abelian A_1 theory with two flavours is easily generalized to the case of N_f flavours, represented in Figure 2.7. We assume the hypermultiplets have distinct masses, $m_s \neq m_j$ for $s \neq j$, $j = 1, \dots, N_f$, and also turn on a FI parameter $\zeta \in \mathbb{R}$. Using the identity

$$\prod_{j=1}^{N_f}\frac{1}{1+M_j X} = \sum_{j=1}^{N_f}\frac{1}{1+M_j X}\prod_{s\neq j}\frac{1}{1-\frac{M_s}{M_j}}, \tag{2.3.11}$$

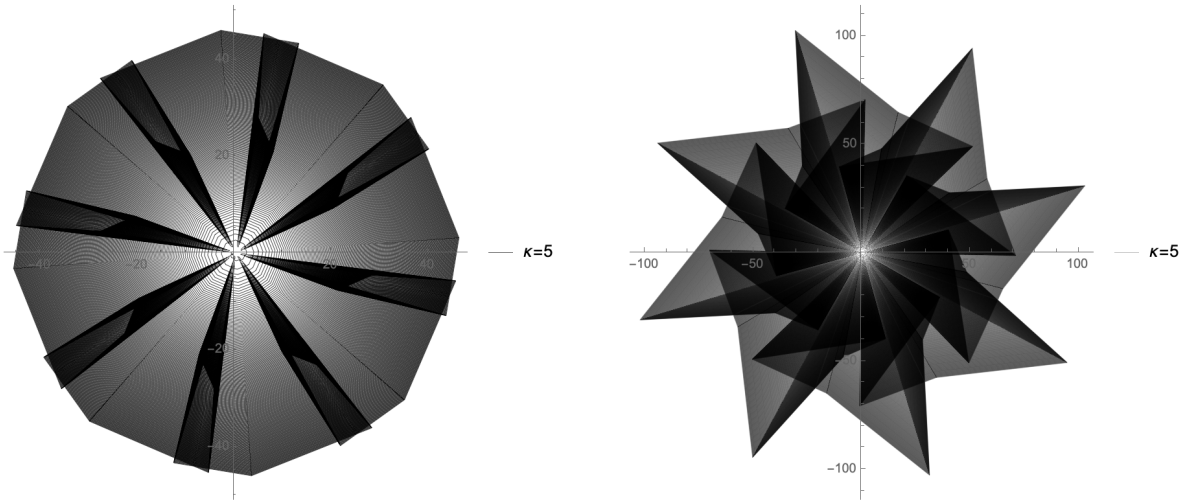


Figure 2.4. Plot of $\mathcal{Z}_{U(1),1}$ with $k = \frac{\kappa}{\varrho}$, at fixed $\kappa = 5$. Left: $m = \sqrt{0.03} + i0.1$ and $\varrho = 1, \ldots, 10^4$. Right: $m = 0.1 + i\sqrt{0.03}$ and $\varrho = 1, \ldots, 2 \times 10^4$.

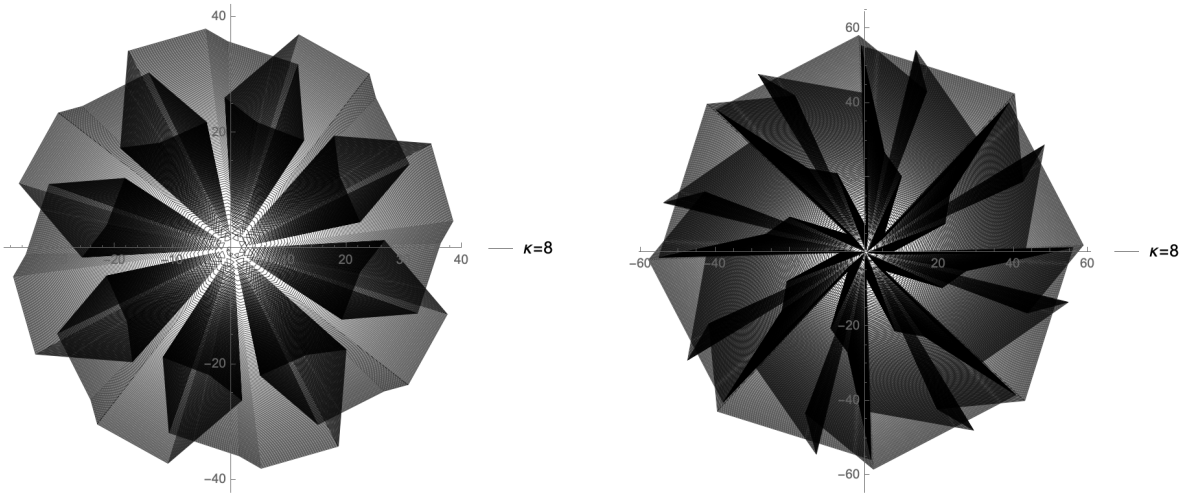


Figure 2.5. Plot of $\mathcal{Z}_{U(1),1}$ with $k = \frac{\kappa}{\varrho}$, at fixed $\kappa = 8$ and varying $\varrho = 1, \ldots, 10^4$. Left: $m = \sqrt{0.03} + i0.1$. Right: $m = 0.1 + i\sqrt{0.03}$.

we can rewrite the partition function of the theory as

$$\begin{aligned}\mathcal{Z}_{U(1),N_f}(k, \vec{m}, \zeta) &= \int dx \frac{e^{i\pi k x^2 + i2\pi\zeta x}}{\prod_{j=1}^{N_f} 2\cosh \pi(x + m_j)} \\ &= \sum_{j=1}^{N_f} \frac{e^{2\pi m_j (N_f - 2)}}{\prod_{s \neq j} (e^{2\pi m_j} - e^{2\pi m_s})} \mathcal{I}_k \left(-m_j, \frac{N_f}{2} + i\zeta \right),\end{aligned}$$

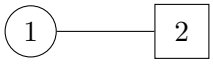


Figure 2.6. Abelian A_1 quiver with $N_f = 2$ fundamental flavours.

Figure 2.7. Abelian A_1 quiver with N_f fundamental flavours.

where we have used $\sum_{j=1}^{N_f} m_j = 0$. From (2.2.6) we obtain an explicit solution in terms of a sum of Mordell integrals for every rational value of the Chern–Simons level k:

$$\begin{aligned}\mathcal{Z}_{U(1),N_f}(k>0,\vec{m},\zeta=0) = \sum_{j=1}^{N_f} \frac{e^{\pi m_j(N_f-2)+i\pi\frac{N_f}{2}}}{\prod_{s\neq j}\left(e^{2\pi m_j}-e^{2\pi m_s}\right)} \frac{1}{1-(-1)^{\varrho\left(N_f+1-\kappa+\frac{\kappa}{\varrho}\right)}e^{-2\pi k m_j}} \\ \times \left[ie^{-i\pi k\left(\frac{1}{2}+im_j\right)^2}\sum_{\beta=1}^{\varrho}(-1)^{\beta N_f}e^{-i\pi k\beta(\beta-1)-2\pi k m_j\beta} \right. \\ \left. -\sqrt{\frac{i}{k}}\sum_{\alpha=1}^{\kappa}\left(-e^{-2\pi m_j}\right)^{\alpha+\frac{N_f}{2}} q^{-\frac{1}{2}\left(\alpha+\frac{N_f}{2}\right)^2} \right],\end{aligned}$$

and

$$\begin{aligned}\mathcal{Z}_{U(1),N_f}(k<0,\vec{m},\zeta=0) = \sum_{j=1}^{N_f} \frac{e^{\pi m_j(N_f-2)+i\pi\frac{N_f}{2}}}{\prod_{s\neq j}\left(e^{2\pi m_j}-e^{2\pi m_s}\right)} \frac{1}{1-(-1)^{\varrho\left(N_f+1-\kappa-\frac{\kappa}{\varrho}\right)}e^{-2\pi k m_j}} \\ \times \left[ie^{-i\pi k\left(\frac{1}{2}+im_j\right)^2}\sum_{\beta=1}^{\varrho}(-1)^{\beta N_f}e^{-i\pi k\beta(\beta-1)-2\pi k m_j\beta} \right. \\ \left. +\sqrt{\frac{i}{k}}\sum_{\alpha=0}^{\kappa-1}\left(-e^{2\pi m_j}\right)^{\alpha-\frac{N_f}{2}} q^{-\frac{1}{2}\left(\alpha-\frac{N_f}{2}\right)^2} \right],\end{aligned}$$

When the number of flavours is even the sums in the last line of each expression become (cf. (2.2.9))

$$\widetilde{P}\left(-e^{-\mathrm{sign}(k)2\pi m_j};k\right),$$

but with the summation range shifted by $-\frac{N_f}{2}$. As we have already pointed out in Subsection 2.2.3, these are polynomials in the variable $e^{i2\pi\lambda\mathrm{sign}(k)}$, hence are holomorphic in $\mathbb{C}\setminus\mathbb{R}_{\geq 0}$.

The effect of reintroducing the FI parameter ζ can be reabsorbed in a change of variable, and the result is the same as above up to a shift of the masses, as in (2.3.4). Besides, the result holds upon complexification of the masses and FI parameters, as long as $\left|\Im m_j - \frac{\Im\zeta}{k}\right| < \frac{1}{2}$.

The solution relied on the assumption of generic masses, but the theory has a well defined confluent limit when two masses become equal. One approach to this case is based on a direct analysis of the cancellations in the formula above. An alternative and especially convenient approach is to interpret the partition function as the average of inverse characteristic polynomials in the Stieltjes–Wigert ensemble, expressing it then as a $N_f \times N_f$ determinant, whose limit is well known to give a Wronskian determinant [95]. We discuss this approach in Subsection 2.5.5. A third approach, valid for all equal masses, was taken in [63].

Wilson loops from Mordell integrals: Abelian A_1 theory

We consider again the Abelian Chern–Simons theory with N_f massive fundamental hypermultiplets and insert a circular Wilson loop in a complex irreducible $U(1)$ representation μ, identified with an integer $\mu \in \mathbb{Z}$. Its expectation value is:

$$\begin{aligned}\langle W_\mu \rangle_{U(1),N_f} &= \frac{1}{\mathcal{Z}_{U(1),N_f}} \int_{-\infty}^{\infty} dx \frac{e^{i\pi k x^2 + 2\pi\mu x}}{\prod_{j=1}^{N_f} 2\cosh \pi(x+m_j)} \\ &= \frac{1}{\mathcal{Z}_{U(1),N_f}} \sum_{j=1}^{N_f} \frac{e^{2\pi m_j (N_f-2)}}{\prod_{s\neq j}(e^{2\pi m_j} - e^{2\pi m_s})} \mathcal{I}_k\left(-m_j, \frac{N_f}{2}+\mu\right),\end{aligned}$$

hence the result can be easily extracted from the above analysis or directly from (2.2.6). When $N_f = 2$ we have the particularly simple relation

$$\langle W_\mu \rangle_{U(1),N_f=2} = \frac{\mathcal{Z}_{U(1),2\mu+2}}{\mathcal{Z}_{U(1),2}}.$$

Non-Abelian A_1 theory with N_f flavours

The next example is the $U(N)$ theory with N_f flavours, as in Figure 2.8. The partition function at $N_f = 2$ and no FI term has been solved in [62], using a change of variables of the form $X_a = e^{2\pi(x_a - c)}$ and writing the resulting expression as a Hankel determinant [84]. The crucial difference from [62, 63] is that we consider generic masses and also allow a FI term. In flat space, this choice lifts the Higgs branch and reduces the moduli space to isolated vacua.

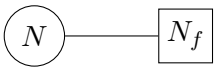


Figure 2.8. Non-Abelian A_1 quiver with N_f fundamental flavours.

We get [62]:

$$\begin{aligned}\mathcal{Z}_{U(N),N_f}(k,\vec{m}) &= \prod_{a=1}^{N} \int_{-\infty}^{\infty} dx_a \frac{e^{i\pi k x_a^2 + i2\pi\zeta x_a} \prod_{b\neq a} 2\sinh\pi(x_b - x_a)}{\prod_{j=1}^{N_f} 2\cosh\pi\,(x_a + m_j)} \\ &= e^{i\pi\frac{N}{k}\left[N\left(N_f-2N\right)-\zeta^2\right]} \det_{1\leq a,b\leq N} \left[e^{\frac{i\pi}{k}\left(N_f-2N\right)(a+b-1)} \int_{-\infty}^{+\infty} dx \frac{e^{i\pi k x^2 + 2\pi x \ell_{ab}}}{\prod_{j=1}^{N_f}\left(1+e^{2\pi(x-m'_j)}\right)} \right]\end{aligned}$$

where we defined for shortness $\ell_{ab} = a + b - 1 - N + \frac{N_f}{2}$ and $m'_j = m_j - \frac{\zeta}{k}$. Using (2.3.11) each entry of the determinant is written as a sum of N_f Mordell integrals:

$$\begin{aligned}\mathcal{Z}_{U(N),N_f}(k,\vec{m}) = e^{i\pi\frac{N}{k}\left[N\left(N_f-2N\right)-\zeta^2\right]} \det_{1\leq a,b\leq N} \Bigg[& e^{\frac{i\pi}{k}\left(N_f-2N\right)(a+b-1)} \\ & \times \sum_{j=1}^{N_f} \frac{e^{2\pi m'_j(N_f-2)}}{\prod_{s\neq j}\left(e^{2\pi m'_j} - e^{2\pi m'_s}\right)} \mathcal{I}_k\left(-m'_j, \ell_{ab}\right) \Bigg]. \qquad (2.3.12)\end{aligned}$$

This results extends [62, 63] to generic deformations, using a different approach than [95]. The massless limit can be taken, exploiting the identity (2.2.14) to see the cancellation of the singularities, cf. [62, Eq. (2.38)], while the limit of coinciding masses is better understood in the formalism of [95].

2.3.3 Lessons so far

Before discussing quiver gauge theories, we pause to analyze the information that can be extracted by the exact solutions in terms of $\mathcal{I}_k\left(-m, \frac{N_f}{2}\right)$, as defined in (2.2.5).

A first observation is that the sums appearing in the right-hand side of the Mordell integrals are all of the form

$$\begin{aligned} &\sim \sum_{\alpha=1}^{k} \left(-e^{-2\pi m}\right)^{\alpha} q^{-\frac{1}{2}\left(\alpha-\frac{N_f}{2}\right)^2} && (k>0), \\ &\sim \sum_{\alpha=0}^{-k-1} \left(-e^{2\pi m}\right)^{\alpha} q^{-\frac{1}{2}\left(\alpha+\frac{N_f}{2}\right)^2} && (k<0). \end{aligned}$$

Here we are considering $\zeta = 0$ for clarity, but the argument goes through in exactly the same way turning on a real FI parameter. The shift in the Gaussian factor in each summand accounts for the shift $k \mapsto k - \frac{N_f}{2}$ from integrating out massive hypermultiplets.

Another important aspect is that Mordell's solution is a holomorphic function of $\lambda = \frac{1}{2} + im$ [61]. We notice that λ is precisely the variable $\frac{\mathfrak{t}}{2} + im$ identified by Jafferis [96] (see also the exhaustive discussion in [97]), with respect to which the partition function on $\mathbb{S}^3$ is holomorphic. Here $\mathfrak{t}$ parametrizes the trial $U(1)_{\mathrm{R}}$ R-charge of the hypermultiplet in the microscopic theory, and in our case is fixed to $\mathfrak{t} = 1$ by the $\mathcal{N} = 3$ extended supersymmetry.

Related to the just mentioned aspect, we stress the role of the numerator in the overall multiplicative term in (2.2.3) and (2.2.4). This term always generates an overall factor $e^{-i\pi k\lambda^2}$, which is a Chern–Simons coupling for the background vector multiplet of the global symmetry, precisely given in terms of the holomorphic variable $\lambda = \frac{1}{2} + im$. On the other hand, a pure $U(1)_k$ Chern–Simons theory coupled to a background vector multiplet generates an effective Chern–Simons term $e^{i\frac{\pi}{k}\lambda^2}$ [98, 97], which emerges from the integrals. We also know that, for $N_f = 1$, the theory must be self-dual under $k \mapsto -\frac{1}{k}$ [94], as we have extensively discussed in Subsection 2.3.1. Therefore, the overall factor derived in [61] is essential to guarantee the invariance of the partition function under the S-duality when $N_f = 1$, or more in general to reproduce the correct Chern–Simons couplings for the background vector multiplets [98, 97].[3]

$U(1)_k$ theory at rational k

As we have learned from the plots in Subsection 2.3.2 and the surrounding discussion, the study of the partition function $\mathcal{Z}_{U(1),N_f} \in \mathbb{C}$ at fixed m as a function of $k = \frac{\kappa}{\varrho}$ uncovers a rich structure when ϱ is increased keeping κ fixed. This observation is compatible with the insight provided by the theory of Gauss sums [99, 100]. Pushing the analogy further, it may be interesting to understand the behaviour of $\mathcal{Z}_{U(1),N_f}$ when k becomes irrational. This is not allowed in gauge theory for compact gauge group. However, the iterative application of the elementary Fourier transform identity

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \frac{e^{i\pi a_1 x^2 + i\pi a_2 y^2 + i2\pi xy}}{\prod_{j=1}^{N_f} 2\cosh \pi(x+m_j)} = \sqrt{\frac{i}{a_2}} \int_{-\infty}^{+\infty} dx \frac{e^{i\pi\left(a_1 - \frac{1}{a_2}\right)x^2}}{\prod_{j=1}^{N_f} 2\cosh \pi(x+m_j)}$$

[3] $\log \mathcal{Z}_{U(1),N_f}$ does not take a simple form, which prevents us from reading off the precise form of the mixed flavour-R Chern–Simons couplings.

allows to interpret the partition function at rational k, with continued fraction expansion

$$k = \frac{\kappa}{\varrho} = a_1 - \cfrac{1}{a_2 - \cfrac{1}{\cdots - \cfrac{1}{a_n}}}$$

as a chain of $U(1)_{a_p}$ theories at integer Chern–Simons levels a_p, $p = 1, \ldots, n$, with matter insertion only at the first node. Notice that this theory would correspond to a completely disconnected quiver (no bi-fundamentals), and the various nodes are coupled only through the mixed Chern–Simons terms $k_{p,p+1} = 1$. The Chern–Simons level of the original theory attains an irrational value in the limit of infinitely many coupled Chern–Simons theories.

It would therefore be desirable to look further into the behaviour of $\mathcal{Z}_{U(1),N_f}$ when the number of integers a_p in the continued fraction expansion of k is increased, and eventually understand the $n \to \infty$ limit.

2.3.4 Abelian quivers

We consider now Chern–Simons theories classified by Dynkin diagram of type A_r, which correspond to linear quivers. We consider Abelian theories with gauge group $G = U(1)^r$.

Abelian A_2 theory

The first example is a two node quiver with Abelian gauge group, see Figure 2.9. The matter content consists only of the bi-fundamental hypermultiplet joining the nodes, to which we assign a mass m. We set the FI parameters to zero, as they can be reintroduced at the end by the usual shift of the masses and overall coefficient, as in (2.3.4).

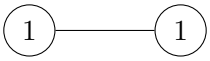


Figure 2.9. Abelian A_2 quiver.

The partition function is:

$$\begin{aligned}\mathcal{Z}_{U(1)^2}(\vec{k}, m) &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \frac{e^{i\pi k_1 x^2 + i\pi k_2 y^2}}{2\cosh \pi(x - y + m)} \\ &= \int_{-\infty}^{+\infty} dy e^{i\pi k_2 y^2 + \pi(y-m)} \mathcal{I}_{k_1}\left(y - m, \frac{1}{2}\right).\end{aligned}$$

We take for concreteness $k_1 = \frac{\kappa_1}{\varrho_1} > 0$ with either κ_1 even or ϱ_1 odd. This restriction is not necessary, but simplifies the expressions as we do not need to carry factors $(-1)^{\kappa_1(\varrho_1 - 1)}$. From (2.2.3) and a rescaling of the integration variable, we get:

$$\begin{aligned}\mathcal{Z}_{U(1)^2}(\vec{k}, m) = \frac{1}{\kappa_1} \Bigg[& e^{i\pi k_1 (m^2 - \frac{1}{4}} \sum_{\beta=0}^{\varrho_1} (-e^{-2\pi k_1 m})^\beta e^{i\pi k_1 \beta(\beta-1)} \mathcal{I}_{k_{\text{eff}}}(0, \check{\xi}_1(\beta)) \\ & + i\sqrt{\frac{i\varrho_1}{\kappa_1}} \sum_{\alpha=1}^{\kappa_1} q_1^{-\frac{1}{2}\left(\alpha - \frac{1}{2}\right)^2} (-e^{-2\pi m})^{\left(\alpha - \frac{1}{2}\right)} \mathcal{I}_{k'_{\text{eff}}}\left(0, \check{\xi}_2(\alpha)\right) \Bigg],\end{aligned}$$

where

$$k_{\text{eff}} = \frac{k_2\varrho_1 + \kappa_1}{\kappa_1^2}, \quad \check{\xi}_1(\beta) = \frac{1}{\varrho_1}\left(-im + \beta - \frac{1}{2}\right),$$
$$k'_{\text{eff}} = \frac{k_2}{\kappa_1^2}, \qquad \check{\xi}_2(\alpha) = \frac{1}{\kappa_1}\left(\alpha - \frac{1}{2}\right).$$

The solution can be made explicit plugging (2.2.6). When $k_1 + k_2 = 0$, it takes a much simpler form. We introduce both the mass and the FI parameter explicitly, assume $k > 0$ without loss of generality, and write

$$\begin{aligned}\mathcal{Z}_{U(1)^2}(\vec{k}, m, \zeta) &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \frac{e^{i\pi k(x^2-y)+i2\pi\zeta(x+y)}}{2\cosh\pi(x-y+m)} \\ &= \int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} dx \frac{e^{-i\pi kv^2+i2\pi\zeta v-i2\pi x(kv-\zeta)}}{2\cosh\pi(v-m)} \\ &= \frac{e^{i\pi\frac{\zeta^2}{k}}}{2k\cosh\pi\left(m-\frac{\zeta}{k}\right)},\end{aligned}$$

where we have used the centre of mass variable $v = y - x$.

In Subsection 2.3.4 we present the computations of the lowest-rank non-Abelian A_2 theory extending the ideas presented here.

Abelian A_3 theory

The Abelian A_3 quiver is depicted in Figure 2.10. We turn on a real FI parameter ζ in the middle node, and give masses m_1 and m_2 to the hypermultiplets.

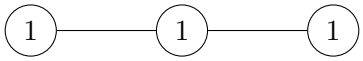


Figure 2.10. Abelian A_3 quiver.

The partition function is:

$$\mathcal{Z}_{U(1)^3}(\vec{k}, \zeta, \vec{m}) = \int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \frac{e^{i\pi\left(k_1v^2+k_2x^2+k_3y^2\right)+i2\pi\zeta x}}{2\cosh\pi(v-x+m_1)2\cosh\pi(x-y+m_2)}.$$

Instead of directly applying (2.2.6), we first use the change of variables

$$v' = v - x, \quad y' = y - x \tag{2.3.13}$$

(henceforth we drop the prime). We work under the assumption [74]

$$\sum_{p=1}^{3} k_p = 0.$$

Integrating over x we get

$$\mathcal{Z}_{U(1)^3}(\vec{k}, \zeta, \vec{m}) = \frac{1}{|k_1|}\int_{-\infty}^{+\infty} dv \frac{e^{i\pi k_{\text{eff}}v^2}}{2\cosh\pi(v-m_2)2\cosh\pi\left(\frac{k_3}{k_1}v - m_2 + \frac{\zeta}{k_1}\right))},$$

where we have defined the effective Chern–Simons level

$$k_{\text{eff}} = \frac{k_3}{k_1}(k_1 - k_3).$$

At this point, from the denominator, we see that the tractable cases correspond to $k_3 = \pm k_1$. The first choice, $k_3 = k_1$, means that we restrict to the one-parameter family of theories with Chern–Simons levels

$$(k_1, k_2, k_3) = (k, -2k, k),$$

in which case we get

$$\mathcal{Z}_{U(1)^3}((k,-2k,k),\zeta,(m_1,m_2)) = \frac{1}{|k|}\mathcal{Z}^{\text{SQED}}_{N_f=2}\left(-m_1 - \frac{\zeta}{k}, -m_2\right),$$

where we have recognized the partition function of a single-node theory without Chern–Simons term and two fundamental flavour of mass $-m_1 - \frac{\zeta}{k}$ and $-m_2$, respectively. Note that the hypermultiplet is off-shell, as it does not respect the $SU(2)$ flavour symmetry, unless we tune $\frac{\zeta}{k} = -m_1 - m_2$. We can safely turn off the FI parameter ζ, as it only shifts m_1, and it is convenient to introduce an FI parameter $\tilde{\zeta}$ in the third node. We get [101]

$$\mathcal{Z}_{U(1)^3}((k,-2k,k),\zeta,(m_1,m_2)) = \frac{(e^{i2\pi m_2\tilde{\zeta}} - e^{i2\pi m_1\tilde{\zeta}})}{4i|k|\sinh\pi(m_2-m_1)\sinh(\pi\tilde{\zeta})}.$$

The other tractable case corresponds to the one-parameter family of Chern–Simons theories with levels

$$(k_1, k_2, k_3) = (k, 0, -k).$$

In this case $k_{\text{eff}} = -2k$, and the $U(1)^3$ partition function is given by

$$\mathcal{Z}_{U(1)^3}((k,0,-k),\zeta,(m_1,m_2)) = \frac{1}{2|k|\sinh\pi\left(m_2+m_1+\frac{\zeta}{k}\right)}\left[\mathcal{I}_{k_{\text{eff}}}(-m_1,1) - \mathcal{I}_{k_{\text{eff}}}(-m_2,1)\right],$$

which, up to the factor $|k|^{-1}$, is the partition function of the A_1 Abelian theory with $N_f = 2$ studied in [62] and in Subsection 2.3.2, at level $k_{\text{eff}} = -2k$.

A third instance in which the Abelian $U(1)^3$ theory is exactly solvable corresponds to the so-called Model III of Jafferis and Yin [65], with Chern–Simons levels $\vec{k} = (1,-1,1)$. This theory is dual to SQED with two fundamental hypermultiplets and no Chern–Simons couplings [65]. The equality of the two partition functions, up to a phase, is easily proved from their integral representation,

$$\begin{aligned}\mathcal{Z}^{\text{SQED}}_{N_f=2}(m',\zeta') &= e^{i2\pi m'\zeta'}\int_{-\infty}^{+\infty} dx \frac{e^{i2\pi\zeta' x}}{[2\cosh\pi(x+2m')][2\cosh(\pi x)]} \\ &= \frac{1}{\sqrt{i}}\left[e^{-i\pi m_1 m_2}\ \mathcal{Z}_{U(1)^3}((1,-1,1),\vec{\zeta}=\vec{0},m_1,m_2)\right]_{m_1=\zeta',m_2=2m'}\end{aligned} \tag{2.3.14}$$

with the last equality following from the change of variables (2.3.13). The proof extends straightforwardly to the vev of a Wilson loop charged under one of the three $U(1)$'s.

An exact evaluation of $\mathcal{Z}^{\text{SQED}}_{N_f=2}$ has been given in [101, 102]. In turn, we are able to evaluate the partition function on the A_3 side using (2.2.6):

$$\mathcal{Z}_{U(1)^3}((1,-1,1),\vec{m}) = e^{\pi(-m_1+m_2)}\int_{-\infty}^{+\infty} dx\ e^{-i\pi x^2+2\pi x}\mathcal{I}_{+1}\left(x-m_1,\frac{1}{2}\right)\mathcal{I}_{+1}\left(x+m_2,\frac{1}{2}\right)$$

which, using (2.3.21), becomes

$$\mathcal{Z}_{U(1)^3}((1,-1,1),\vec{m}) = \frac{1}{2\sinh\pi(m_1+m_2)}\left[Z_1^{\rm JY}(m_1,m_2)+Z_2^{\rm JY}(m_1,m_2)+Z_3^{\rm JY}(m_1,m_2)\right], \tag{2.3.15}$$

where we have defined

$$\begin{aligned}
Z_1^{\rm JY}(m_1,m_2) &\equiv \mathcal{I}_{+1}(m_1,1-im_1+im_2)-\mathcal{I}_{+1}(-m_2,1-im_1+im_2)\\
Z_2^{\rm JY}(m_1,m_2) &\equiv ie^{i\pi\left(m_1^2+m_2^2\right)}\left[\mathcal{I}_{+1}(m_1,1-im_1+im_2)-\mathcal{I}_{+1}(-m_2,1-im_1+im_2)\right]\\
Z_3^{\rm JY}(m_1,m_2) &\equiv -\sqrt{i}\int_{-\infty}^{+\infty}dx e^{2\pi x}\left(e^{i\pi m_1^2-i2\pi x m_1}+e^{i\pi m_2^2+i2\pi x m_2}\right)\left[\frac{1}{e^{2\pi x}+e^{2\pi m_1}}-\frac{1}{e^{2\pi x}+e^{-2\pi m_2}}\right]
\end{aligned}$$

The first piece, which we have named $Z_1^{\rm JY}$, is given in (2.3.22) and contributes

$$Z_1^{\rm JY}(m_1,m_2) = \sqrt{-i}\left[\frac{1}{1-e^{2\pi m_1}}-\frac{1}{1-e^{-2\pi m_2}}+\frac{e^{-i\pi m_1^2}}{2\sinh(\pi m_1)}-\frac{e^{-i\pi m_2^2}}{2\sinh(\pi m_2)}\right].$$

The second piece is

$$Z_2^{\rm JY}(m_1,m_2) = \sqrt{-i}e^{i2\pi m_1 m_2}\left[\frac{1}{e^{2\pi m_2}-1}-\frac{e^{-i\pi m_2^2}}{2\sinh(\pi m_2)}-\frac{1}{e^{-2\pi m_2}-1}-\frac{e^{i\pi m_1^2}}{2\sinh(\pi m_1)}\right].$$

The last contribution is

$$Z_3^{\rm JY}(m_1,m_2) = \sqrt{i}\left[\frac{e^{i\pi m_1^2+i2\pi m_1 m_2}}{2\sinh(\pi m_1)}-\frac{e^{-i\pi m_1^2}}{2\sinh(\pi m_1)}\right]+\sqrt{-i}\left[\frac{e^{i\pi m_2^2+i2\pi m_1 m_2}}{2\sinh(\pi m_2)}-\frac{e^{-i\pi m_2^2}}{2\sinh(\pi m_2)}\right].$$

Plugging these three expressions back in (2.3.15) and simplifying, we get

$$\mathcal{Z}_{U(1)^3}((1,-1,1),\vec{m}) = \sqrt{-i}\frac{\left(e^{-i2\pi m_1 m_2}-1\right)}{[2\sinh(\pi m_1)][2\sinh(\pi m_2)]}. \tag{2.3.16}$$

From (2.3.14), the result we find agrees with [101, 102].

Abelian ABJM

We consider mass-deformed Abelian ABJM theory. This is $U(1)_k \times U(1)_{-k}$ Chern–Simons theory with two massive bi-fundamental hypermultiplets, represented in Figure 2.11.

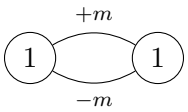


Figure 2.11. Mass deformed Abelian ABJM theory.

The partition function of the theory is:

$$\mathcal{Z}_{\rm ABJ(1|1)}(k,m) = \int_{-\infty}^{+\infty}dy\int_{-\infty}^{+\infty}dx\frac{e^{i\pi k(x^2-y^2)}}{4\cosh\pi(x-y+m)\cosh\pi(x-y-m)} \tag{2.3.17}$$

where the variables x and y parametrize the two $\mathfrak{u}(1)$'s. From (2.2.5) we rewrite it as:

$$\mathcal{Z}_{\rm ABJ(1|1)}(k,m) = \frac{1}{2\sinh(2\pi m)}\int_{-\infty}^{+\infty}dy e^{-i\pi k y^2}\left[\mathcal{I}_k(y-m,1)-\mathcal{I}_k(y+m,1)\right].$$

As one may expect, the contribution from a single node coincides with the partition function of $U(1)_k$ theory with two massive hypermultiplets with masses $y \pm m$. Without loss of generality, we take $k > 0$ and, from (2.2.6) together with (2.2.3) we get

$$\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k,m) = \frac{1}{2\sinh(2\pi m)} \left\{ i e^{i\pi k\left(m^2-\frac{1}{4}\right)} \left[\int_{-\infty}^{+\infty} dy \frac{e^{-i2\pi kmy}}{2\sinh(\pi k(y-m))} - \int_{-\infty}^{+\infty} dy \frac{e^{i2\pi kmy}}{2\sinh(\pi k(y+m))} \right] \right.$$
$$\left. + \sqrt{\frac{i}{k}} \sum_{\alpha=0}^{k-1} (-1)^\alpha q^{-\frac{\alpha^2}{2}} \int_{-\infty}^{+\infty} dy e^{-i\pi k y^2} \left[\frac{e^{2\pi\alpha(y+m)}}{e^{2\pi k(y+m)}-1} - \frac{e^{2\pi\alpha(y-m)}}{e^{2\pi k(y-m)}-1} \right] \right\}.$$

The two integrals in the first line are the Fourier transform of $\sinh(\pi x)$ and are immediately solved. The two integrals in the second line, after a change of variables $y' = k(y \pm m)$ are reduced again to Mordell integrals:

$$\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k,m) = \frac{1}{2k\sinh(2\pi m)} \left\{ e^{-i\pi k\left(m^2+\frac{1}{4}\right)} \tanh(\pi m) \right.$$
$$\left. + e^{-i\pi km^2} \sqrt{\frac{i}{k}} \sum_{\alpha=0}^{k-1} (-1)^\alpha q^{-\frac{\alpha^2}{2}} \left[\Psi_-\left(-\frac{\alpha}{k} - im, 0; 1, k\right) - \Psi_-\left(-\frac{\alpha}{k} + im, 0; 1, k\right) \right] \right\}.$$

Plugging the solution (2.2.4) and after some simplification,

$$\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k,m) = \frac{e^{-i\pi km^2}}{2k\sinh(2\pi m)} \left\{ e^{-i\pi\frac{k}{4}} \tanh(\pi m) \right.$$
$$+ \sum_{\alpha=0}^{k-1} (-1)^\alpha \left[e^{i\pi km^2} \left(\frac{e^{2\pi m\alpha}}{(-1)^k e^{2\pi km} - 1} - \frac{e^{-2\pi m\alpha}}{(-1)^k e^{-2\pi km} - 1} \right) \right.$$
$$\left. \left. + i\sqrt{\frac{i}{k}} \sum_{\beta=1}^{k} q^{-\frac{(\alpha-\beta)^2}{2}} \left(\frac{e^{2\pi m\beta}}{(-1)^k e^{2\pi km} - 1} - \frac{e^{-2\pi m\beta}}{(-1)^k e^{-2\pi km} - 1} \right) \right] \right\}.$$

The second line is a geometric sum, with a prefactor $e^{i\pi km^2}$. Using the Gauss sum identity (2.2.14) to sum over α in the third line, we find another geometric sum, over β this time, which cancels the contribution of the first line. After these simplifications we get:

$$\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k,m,\zeta=0) = \frac{1}{4k\cosh(\pi m)^2}.$$

In general, unitary $\widehat{\mathsf{A}}_r$ quivers have topological symmetry $[\prod_{p=0}^r U(1)_{\mathrm{top},p}]/U(1)$. This allows us to introduce an FI parameter ζ turning on a background twisted vector multiplet for the $U(1)_{\mathrm{top}}$ topological symmetry of ABJM. This can be reabsorbed in a simple change of variables, and the result is directly obtained from above replacing $\pm m \mapsto \frac{2\zeta}{k} \pm m$. We get

$$\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k,m,\zeta) = \frac{1}{4k\cosh\pi\left(m-\frac{2\zeta}{k}\right)\cosh\pi\left(m+\frac{2\zeta}{k}\right)}. \tag{2.3.18}$$

The partition function, as written in (2.3.17), is invariant under $k \leftrightarrow -k$ but the final expression (2.3.18) is not because, without loss of generality, we have assumed $k > 0$ in the intermediate steps. The result agrees with [64], where the answer was obtained in a straightforward way using a change of variables $x' = x - y$ in (2.3.17). Nevertheless, with our approach we can consider the more general case with arbitrary rational k_1 and k_2, which corresponds to deform the gravity dual

by a Romans mass $F_0 = k_1 + k_2$, cf. Subsection 2.2.1. Letting $k_2 \neq -k_1$ and also allowing generic masses m_1, m_2 and a FI parameter ζ, the partition function is

$$\begin{aligned}\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k_1,k_2,\vec{m},\zeta) &= \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \frac{e^{i\pi k_1 x^2 + i\pi k_2 y^2 + i2\pi\zeta(x+y)}}{2\cosh\pi(x-y+m_1)2\cosh\pi(y-x+m_2)} \\ &= e^{-i\pi\frac{\zeta^2}{k_{\mathrm{eff}}}} \int_{-\infty}^{+\infty} dv \int_{-\infty}^{+\infty} dy \frac{e^{i\pi k_1 v^2 + i\pi(k_1+k_2)y^2 - i2\pi k_1 vy}}{2\cosh\pi(v+m_+)2\cosh\pi(v+m_-)} \\ &= e^{-i\pi\frac{\zeta^2}{k_{\mathrm{eff}}}} \sqrt{\frac{i}{k_1+k_2}} \int_{-\infty}^{+\infty} dv \frac{e^{i\pi k_{\mathrm{eff}} v^2}}{2\cosh\pi(v+m_+)2\cosh\pi(v+m_-)}.\end{aligned}$$

To pass from the first to the second line we have used the change of variables [64]

$$v = \left(x + \frac{\zeta}{k_1}\right) - \left(y + \frac{\zeta}{k_2}\right)$$

together with a redefinition of the parameters

$$m_+ := m_1 - \frac{\zeta}{k_1} + \frac{\zeta}{k_2}, \quad m_- := -m_2 - \frac{\zeta}{k_1} + \frac{\zeta}{k_2}, \quad k_{\mathrm{eff}} = \left(\frac{1}{k_1} + \frac{1}{k_2}\right)^{-1}.$$

In the last line, we recognize the partition function of the $U(1)$ Chern–Simons theory with two fundamentals at level k_{eff}, studied in Subsection 2.3.2. For generic k_1 and $k_2 \neq k_1$ the effective Chern–Simons level k_{eff} is rational, and we assume $k_{\mathrm{eff}} = \frac{\kappa}{\varrho} > 0$. The partition function is

$$\begin{aligned}\mathcal{Z}_{\mathrm{ABJ}(1|1)}(k_1,k_2,\vec{m},\zeta) = &\frac{e^{-i\pi\frac{\zeta^2}{k_{\mathrm{eff}}}}}{2\sinh\pi(m_1+m_2)} \left\{ \frac{1}{(-1)^{\kappa(-1\varrho)} e^{-2\pi\kappa m_+} - 1} \left[\frac{1}{\sqrt{\kappa}} \sum_{\alpha=0}^{\kappa-1} (-e^{-2\pi m_+})^\alpha e^{i\pi\frac{\varrho}{\kappa}\alpha^2} \right.\right. \\ &\left.\left. + \frac{1}{\sqrt{i\varrho}} \sum_{\beta=1}^{\varrho} e^{-i\pi\frac{\kappa}{\varrho}\beta(\beta-1) - 2\pi\frac{\kappa}{\varrho}m_+\beta} \right] - (\text{ replace } m_+ \text{ with } m_-) \right\}.\end{aligned}$$

The theory has a well defined $m \to 0$ limit. For $k_2 = -k_1$ equation (2.3.18) gives directly $\frac{1}{4k}$, while the limit $m \to 0$ for generic k_1 and k_2 is given in Subsection (2.3.18) making use of the Gauss sum identity (2.2.14), and follows straightforwardly from [62].

Non-Abelian A_2 theory

This Subsection contains an example of the application of the ideas of this section to a non-Abelian quiver. We consider the simplest such theory, the A_2 quiver with gauge group $U(1)_{k_1} \times U(2)_{k_2}$ and without any additional insertion, as in Figure 2.12. The bi-fundamental hypermultiplet has a real mass m.

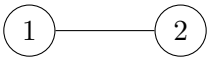


Figure 2.12. The simplest non-Abelian A_2 quiver.

The partition function is

$$\mathcal{Z}_{U(2)\times U(1)}(k_1,k_2,m) = \frac{1}{2!}\int_{\mathbb{R}^2} dx_1 dx_2 \int_{-\infty}^{+\infty} dy \frac{e^{i\pi k_1(x_1^2+x_2^2)+i\pi k_2 y}\ (2\sinh\pi(x_1-x_2))^2}{2\cosh\pi(x_1-y+m)2\cosh\pi(x_2-y+m)}.$$

A change of variables

$$u = x_1 - y, \quad v = x_2 - y, \quad y' = y + \frac{k_1}{2k_1 + k_2}(u + v)$$

allows to directly integrate out y', leaving

$$\mathcal{Z}_{U(2)\times U(1)}(k_1, k_2, m) = \sqrt{\frac{i}{2k_1 + k_2}} \int_{\mathbb{R}^2} \frac{dudv}{2} \frac{e^{i\pi\left(k_1 - \frac{k_1^2}{2k_1+k_2}\right)(u^2+v^2) - i2\pi\frac{k_1^2}{2k_1+k_2}uv} (2\sinh\pi(x_1 - x_2))^2}{2\cosh\pi(u+m)2\cosh\pi(v+m)}.$$

We discuss the two cases $k_1 + k_2 = 0$ and $k_1 + k_2 \neq 0$ separately.

When $k_1 = -k_2 \equiv k$, the Chern–Simons coupling disappears after integrating over y'. Expanding $\sinh\pi(u-v)^2$ and integrating over v we get

$$\begin{aligned}\mathcal{Z}_{U(2)\times U(1)}(k, -k, m) &= -\sqrt{\frac{i}{k}} \int_{-\infty}^{+\infty} du \frac{e^{i2\pi kmu}(e^{2\pi(u+m)} + 1)}{2\cosh\pi(u+m)2\cosh(\pi ku)} \\ &= -\sqrt{\frac{i}{k}} \frac{e^{\pi m}}{2k\left[\cosh(\pi m)\cos\left(\frac{\pi}{2k}\right) + i\sinh(\pi m)\sin\left(\frac{\pi}{2k}\right)\right]}.\end{aligned}$$

When $k = \pm 1$ the partition function takes the specially simple form

$$\mathcal{Z}_{U(2)\times U(1)}(\pm 1, \mp 1, m) = \frac{\sqrt{\mp i}}{|k|(1 - e^{-2\pi m})}.$$

When $k_1 \neq -k_2$ we have to invoke the Mordell integrals. It is convenient to slightly deform the denominator, replacing

$$\prod_{a=1}^{2} 2\cosh\pi(x_a - y + m) \mapsto 2\cosh\pi(x_1 - y + m_1)2\cosh\pi(x_2 - y + m_2),$$

and eventually take the limit $m_1, m_2 \to m$ in the final expression. Besides, it is also more efficient to integrate first over x_1 and x_2 obtaining

$$\begin{aligned}\mathcal{Z}_{U(2)\times U(1)}(\vec{k}, \vec{m}) = 2e^{-\pi(m_1+m_2)} \int_{-\infty}^{+\infty} dy\, e^{i\pi k_2 y^2 + 2\pi y} \Bigg[&\mathcal{I}_{k_1}\left(y - m_1, \frac{3}{2}\right)\mathcal{I}_{k_1}\left(y - m_2, -\frac{1}{2}\right) \\ &- \mathcal{I}_{k_1}\left(y - m_1, \frac{1}{2}\right)\mathcal{I}_{k_1}\left(y - m_2, \frac{1}{2}\right)\Bigg].\end{aligned}$$

The $\mathcal{I}_{k_1}$ integrals give an overall denominator

$$\frac{1}{[e^{2\pi k_1(y-m_1)} + 1][e^{2\pi k_1(y-m_2)} + 1]},$$

whence we see that, thanks to the splitting of the masses, the last integral over y can be solved again using the formula (2.2.6), this time with a rational effective Chern–Simons level $\frac{k_2}{k_1}$. The resulting expression is a long multiple sum, which however admits a well-defined limit $m_1, m_2 \to m$, despite an overall factor $[2\sinh\pi(m_2 - m_1)]^{-1}$, which can be dealt with in exactly the same manner as we have done in Subsection 2.3.2. We conclude mentioning that the argument presented here is easily extended to ABJ theory with ranks 1 and 2 and arbitrary, possibly rational Chern–Simons levels k_1, k_2, although it requires a convenient rewriting of the denominator and produces twice the number of terms than the theory with a single bi-fundamental that we have just discussed.

2.3.5 Abelian quivers at $k = \pm 1$

Beyond selected example that can be analyzed with the methods herein for a whole family of Chern–Simons levels $\vec{k}$, the iterative application of Mordell's formula gives the quiver partition function when the Chern–Simons levels are an alternating string of $+1$ and -1,

$$\vec{k} = (1, -1, 1, \ldots, -1). \tag{2.3.19}$$

In particular, when the rank is even, say $2r$, then $\vec{k}$ consists of alternating $+1$ and -1, with exactly r of each sign. When the rank is odd, say $2r+1$, then we take the middle node without Chern–Simons couplings, in order to ensure

$$\sum_{p=1}^{2r+1} k_p = 0$$

for every rank. With such choice, the quiver is invariant under $\vec{k} \leftrightarrow -\vec{k}$. This symmetry is the diagonal action of the S-duality in the space of couplings. Let us stress that the restrictive choice of $\vec{k}$ is a sufficient condition that ensures the solvability through iterative application of Mordell's formula, but not necessary, as proved explicitly in the previous Subsections.

With this condition, an example of theory solvable with the methods presented in the present work is the linear A_r quiver, with gauge group $U(1)^r$, represented in Figure 2.13.

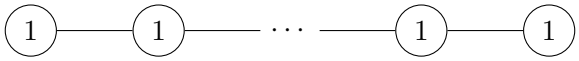


Figure 2.13. Abelian A_r quiver.

On one hand, inspection of formula (2.2.5) has led us to a sufficient condition for the partition function of a linear quiver Chern–Simons theory to be solved by iterative application of Mordell integrals. On the other hand, these theories are simple enough to be studied from a different angle. Let us focus on the even rank case. The partition function of the A_{2r} quiver in Figure 2.13 with Chern–Simons levels (2.3.19) is

$$\mathcal{Z}_{U(1)^{2r}}(\vec{m}) = \int_{-\infty}^{+\infty} dx_1 e^{i\pi x_1^2} \prod_{p=2}^{2r} \int_{-\infty}^{+\infty} dx_p \frac{e^{i\pi(-1)^{p-1}x_p^2}}{2\cosh\pi(x_p - x_{p-1} + m_{p-1})}.$$

We change variables

$$\begin{aligned} v_1 &= x_1 \\ v_2 &= x_2 - v_1 \\ v_3 &= x_3 - (v_2 + v_1) \\ &\vdots \\ v_{2r} &= x_{2r} - \sum_{p=1}^{2r} v_p \end{aligned}$$

and get

$$\mathcal{Z}_{U(1)^{2r}}(\vec{m}) = \int_{-\infty}^{+\infty} dv_1 e^{-i2\pi v_1(v_2+v_4+\cdots+v_{2r})} \prod_{p=2}^{2r} \int_{-\infty}^{+\infty} dv_p \frac{(\text{ CS couplings })}{2\cosh\pi(v_p + m_{p-1})},$$

with the bracket containing the Chern–Simons couplings and mixed Chern–Simons couplings in terms of the new variables $(v_2, \dots, v_{2r})$. The denominator, which carries the matter dependence, is completely factorized. The integral over the first variable yields a constraint on the variables at even nodes,

$$\delta\left(\sum_{p'=1}^{r} v_{2p'}\right). \tag{2.3.20}$$

Besides, one can check that, thanks to the choice (2.3.19), there is no Chern–Simons level at the odd nodes, except for mixed Chern–Simons couplings

$$\exp\left(-i2\pi v_p \sum_{p'=(p-1)/2}^{r} v_{2p'}\right), \qquad p \text{ odd}.$$

Thus the integral over v_p can be solved straightforwardly for all odd p, yielding

$$\int_{-\infty}^{+\infty} dv_p \frac{e^{-i2\pi v_p \sum_{p'=(p-1)/2}^{r} v_{p'}}}{2\cosh\pi(v_p + m_{p-1})} = \frac{1}{2\cosh\pi\left(\sum_{p'=(p-1)/2}^{r} v_{2p'} + m_{p-1}\right)}, \qquad p \text{ odd}.$$

We are left with the integral over the variables $v_{2p'}$, $p' = 1, \dots, r$, but we have the delta function (2.3.20) to get rid of one of the variables, for example v_2. The advantage is that all the Chern–Simons couplings are cancelled by (2.3.20), and we find:

$$\mathcal{Z}_{U(1)^{2r}}(\vec{m}) = \int_{\mathbb{R}^{r-1}} \frac{1}{2\cosh\pi(\sum_{p'=1}^{r} v_{2p'} - m_1)} \prod_{p'=2}^{r} \frac{dv_{2p'}}{2\cosh\pi\left(\sum_{s=p'}^{r} v_{2s} + m_{2p'-2}\right) 2\cosh\pi\left(v_{2p'} + m_{2p'-1}\right)}$$

Therefore the Chern–Simons interactions can be removed from the computations, which are now reduced to $r-1$ integrals. We change again variables

$$\begin{aligned} y_1 &= v_{2r} \\ y_2 &= v_{2r-2} + y_1 \\ y_3 &= v_{2r-4} + y_2 \\ &\vdots \\ y_{r-1} &= v_4 + y_{r-2} \end{aligned}$$

and arrive at

$$\mathcal{Z}_{U(1)^{2r}}(\vec{m}) = \int_{\mathbb{R}^{r-1}} \frac{1}{2\cosh\pi(y_1 + m_1'')} \prod_{p=1}^{r-1} \frac{dy_p}{2\cosh\pi(y_p + m_p')2\cosh\pi(y_p - y_{p+1} + m_p'')}$$

where in the formula $y_r \equiv 0$ and we have renamed the masses

$$m_p' = m_{2r-2p}, \quad m_p'' = m_{2r-2p+1}.$$

In the latter form, we recognize the partition function of a linear quiver gauge theory of type A_{r-1}, without Chern–Simons term and with additional fundamental matter insertions, one at each node except for the first and last node, that yield two fundamentals. This is represented in Figure 2.14. This last theory has manifest $\mathcal{N} = 4$ theory, which was expected from the choice (2.3.19). The partition function in this new form can be evaluated introducing FI terms $\zeta_{p'}$ [101], which can

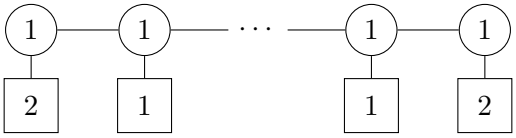


Figure 2.14. Abelian A_{r-1} quiver with one fundamental at each interior node and two fundamentals at the outermost nodes. In this picture the Chern–Simons levels are all set to zero.

either be related to FI couplings in the original theory or we can take the limit $\zeta_{p'} \to 0$ at the end. Note also that both the iterative application of Mordell formula and the method of [101] require the masses to be generic, but the limit of equal masses can be safely taken at the end of the calculations.

For the special case $r = 2$ our formula states the equality of the partition function of the A_4 quiver with alternating Chern–Simons levels $+1$ and -1 with that of SQED with three fundamental flavours ($N_f = 3$ can be seen by direct computations, or starting with $r = 2$ and ungauging the second node in the A_2 quiver), which are known to be dual [65].

The particularly suitable choice of Chern–Simons levels (2.3.19) allows us to study a much wider class of quivers, such as extended $\widehat{A}_r$ quivers with insertion of fundamental matter at any node, as in Figure 2.15. Specializing to $n_p = 0$ for all p, the resulting theories are Abelian sub-cases of [75, 76]. The Abelian $\widehat{A}_r$ quiver with Chern–Simons levels (2.3.19) corresponds to the gauge theoretical realization of the M-crystal model [103] derived in [104, 105]. Although a complete analytical solution seems hard to find, it should be possible to obtain explicit solutions for every r through an algorithmic iteration of formula (2.2.6).

The building blocks in the solution are the integrals $\mathcal{I}_k(y, \check{\xi})$ defined in (2.2.5) at $k = \pm 1$ and $\check{\xi} = 1$, or $\check{\xi} = \frac{1}{2}$ for boundary nodes of a linear quiver without additional matter insertion. They are evaluated as:

$$\mathcal{I}_k\left(y, \frac{1}{2}\right) = \frac{ik}{e^{2\pi y} + 1}\left[1 - e^{ki\pi\left(y^2 + \frac{1}{4}\right)}\right], \qquad k \in \{\pm 1\} \tag{2.3.21}$$

and

$$\mathcal{I}_k\left(y, 1\right) = \frac{e^{\frac{i\pi}{4}k}}{e^{2\pi y} - 1}\left[-1 + e^{ki\pi y^2 + \pi y}\right], \qquad k \in \{\pm 1\}\,. \tag{2.3.22}$$

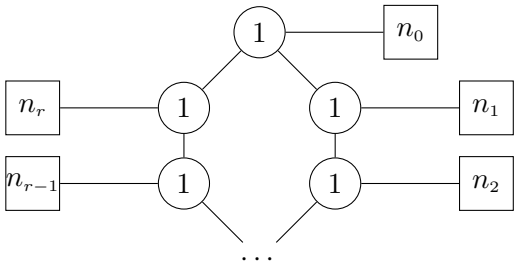


Figure 2.15. Abelian $\widehat{A}_r$ extended quiver.

2.4 Wilson loops in ABJ theory

This section is dedicated to the study of vacuum expectation values of $\frac{1}{2}$-BPS Wilson loops in ABJ(M) theories [73, 82], whenever the Wilson loop is in a type of representation of $U(N_1|N_2)$ called *typical* representation (also known as *long* representation in more physical settings). This

distinction between types of representations emerges when considering Lie supergroups and superalgebras and has not been discussed in the context of Wilson loops of ABJ(M) theories before. Hence, we explain this first.

2.4.1 On Lie superalgebras representations

While every finite-dimensional $\mathfrak{g}$-module of a semi-simple Lie algebra $\mathfrak{g}$ is completely reducible (that is, every representation decomposes into a direct sum of irreducible representations), this no longer holds for Lie superalgebras. A consequence of the classical Djokovic–Hochschild theorem [106] states that all simple Lie superalgebras, with the exception of the family $\{\mathfrak{osp}(1,2n), n \geq 1\}$ of ortho-symplectic Lie superalgebras, have indecomposable (that is, not completely reducible) representations.

This leads to the definition of two types of irreducible representations for a Lie superalgebra $\mathfrak{g}$. Let μ be a highest weight for a finite dimensional irreducible representation $\mathcal{R}(\mu)$ of $\mathfrak{g}$. If the representation cannot be extended to an indecomposable representation of $\mathfrak{g}$, then it is called a typical representation. These are the ones that satisfy the usual properties of the irreducible representations of a Lie algebra. More involved are the atypical representations, which can be extended, with another $\mathfrak{g}$-module, in a manner that the new representation is an indecomposable representation of $\mathfrak{g}$. Atypical representations appear, for example, in the decomposition of the tensor product of two typical representations.

By focusing on Wilson loops with typical representations we will be able to exploit a powerful mathematical factorization property for the characters of such representations [47].

2.4.2 Wilson loops in typical representations

$\frac{1}{2}$-BPS Wilson loops in ABJ(M) theories can be constructed as the trace of the holonomy of a $\mathfrak{u}(N_1|N_2)$-valued superconnection [73]. We therefore consider an irreducible representation $\mathcal{R}(\mu)$ of the supergroup $U(N_1|N_2)$ with highest weight labelled by a partition μ. We henceforth identify $\mathcal{R}(\mu) \simeq \mu$, further identified with the Young diagram representing the partition μ.

The vev of the $\frac{1}{2}$-BPS Wilson loop in the representation μ is [69]:

$$\langle W_\mu \rangle_{N_1,N_2;k} = \frac{1}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)}(k)} \int_{\mathbb{R}^{N_1}} d^{N_1}x \int_{\mathbb{R}^{N_2}} d^{N_2}y \; \mathfrak{s}_\mu \left(e^{2\pi x}|e^{2\pi y}\right) e^{i\pi k\left(\sum_{a=1}^{N_1} x_a^2 - \sum_{\dot{a}=1}^{N_2} y_{\dot{a}}^2\right)} \\ \times \frac{\prod_{1\leq a<b\leq N_1} \left(2\sinh\pi(x_b - x_a)\right)^2 \prod_{1\leq \dot{a}<\dot{b}\leq N_2} \left(2\sinh\pi(y_{\dot{b}} - y_{\dot{a}})\right)^2}{\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2} \left(2\cosh\pi(x_a - y_{\dot{a}})\right)^2}.$$

$\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)}(k)$ is the ABJ partition function, and we are denoting $\langle \cdots \rangle_{N_1,N_2;k}$ the vevs taken in $U(N_1)_k \times U(N_2)_{-k}$ ABJ theory. Indices associated to the first node are labelled $a, b, \ldots$ while indices corresponding to the second node are labelled by $\dot{a}, \dot{b}, \ldots$, hence undotted indices are always meant to run from 1 to N_1 and dotted indices run from 1 to N_2. Moreover, $\mathfrak{s}_\mu(\cdot|\cdot)$ is the supersymmetric Schur polynomial [90, 92] (also known as hook Schur polynomial) associated to the partition μ, and $e^{2\pi x}$ and $e^{2\pi y}$ stand for $(e^{2\pi x_1}, \ldots, e^{2\pi x_{N_1}})$ and $(e^{2\pi y_1}, \ldots, e^{2\pi y_{N_2}})$ respectively. Vevs of correlators of Wilson loops are taken inserting additional supersymmetric Schur polynomials in the matrix model. Notice that if $N_1 = 0$ or $N_2 = 0$ the supersymmetric Schur polynomial degenerates in the usual Schur polynomial, and the vev of a Wilson loop in $U(N_1)_k$ or $U(N_2)_{-k}$ pure Chern–Simons theory is recovered.

We now assume μ to be a typical representation of $U(N_1|N_2)$, which implies that its associated Young diagram fills the upper-left $N_1 \times N_2$ rectangle. These representations have the remarkable

factorization property [47, Thm 6.20]

$$\mathfrak{s}_\mu(X|Y) = \mathfrak{s}_\gamma(X)\mathfrak{s}_{\eta'}(Y)\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2}(X_a + Y_{\dot{a}})\,, \tag{2.4.1}$$

with $\mu = (\kappa+\gamma)\sqcup\eta$, with κ the $N_1\times N_2$ rectangular Young diagram, γ the Young diagram consisting of the boxes of μ on the right of κ and η the Young diagram consisting of the boxes below κ, as in Figure 2.16. The representation η' appearing in the factorization formula (2.4.1) is the conjugate representation of η, corresponding to the conjugate Young diagram.

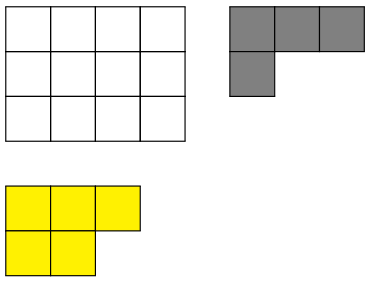


Figure 2.16. Decomposition of a typical (i.e. long) representation μ. In this example, $N_1 = 4, N_2 = 3$, the representation $\mu \simeq (7,5,4,3,2)$ is decomposed into $\kappa \simeq (4,4,4)$ (white), $\gamma \simeq (3,1)$ (gray) and $\eta \simeq (3,2)$ (yellow). Note that in the decomposition of $\mathfrak{s}_\mu$ it appears $\eta' \simeq (2,2,1)$, and not η.

2.4.3 Two Wilson loops

One can foresee from (2.4.1) that part of the contribution from a long representation μ will cancel against the contribution from a bi-fundamental hypermultiplet. When the correlator of two Wilson loops is considered, one gets rid of the denominator in the two-matrix model, simplifying the computations. Taking the vev $\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k}$, with $\vec{\mu} := (\mu,\tilde{\mu})$ a pair of long representations, and using (2.4.1) we obtain:

$$\begin{aligned}\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k} = \frac{1}{\mathcal{Z}_{\mathrm{ABJ}}}\int_{\mathbb{R}^{N_1}}\mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_{\tilde{\gamma}}(e^{2\pi x})\prod_{1\le a<b\le N_1}(2\sinh\pi(x_b-x_a))^2\prod_{a=1}^{N_1}e^{i\pi k x_a^2+2\pi N_2 x_a}\,dx_a\\ \times\int_{\mathbb{R}^{N_2}}\mathfrak{s}_{\eta'}(e^{2\pi y})\mathfrak{s}_{\tilde{\eta}'}(e^{2\pi y})\prod_{1\le\dot{a}<\dot{b}\le N_2}\left(2\sinh\pi(y_{\dot{b}}-y_{\dot{a}})\right)^2\prod_{\dot{a}=1}^{N_2}e^{-i\pi k y_{\dot{a}}^2+2\pi N_1 y_{\dot{a}}}\,dy_{\dot{a}}.\end{aligned} \tag{2.4.2}$$

The correlator of two such Wilson loops in ABJ theory is therefore factorized into two pairs of Wilson loops, one pair for each node. Shifting variables and using basic properties of the Schur polynomials [90] we obtain:

$$\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k} = C^{\vec{\mu}}_{N_1,N_2;k}\,\frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}}\langle W_\gamma W_{\tilde{\gamma}}\rangle_{N_1;k}\langle W_{\eta'}W_{\tilde{\eta}'}\rangle_{N_2;-k}. \tag{2.4.3}$$

Here $\mathcal{Z}_{N_p,k}$ is the partition function of pure $U(N_p)$ bosonic Chern–Simons theory at renormalized level k, and $\langle\cdots\rangle_{N_p,k}$ is the average in the pure Chern–Simons theory at node $p=1,2$. The shift of variables moves the integration cycle away from the real axis, but it can be translated back without changing the answer. The overall coefficient arising from the shift of variables is

$$C^{\vec{\mu}}_{N_1,N_2;k} = \exp\left[\frac{i\pi}{k}\left(N_2\left(N_1N_2+2|\vec{\gamma}|\right)-N_1\left(N_1N_2+2|\vec{\eta}'|\right)\right)\right],$$

where $|\vec{\gamma}|$ is a shorthand for $|\gamma|+|\tilde{\gamma}|$, and the same for $|\vec{\eta}'|$. Recall that $|\gamma|$ is the number of boxes in the Young diagram γ. Closely related results have been obtained in [107], where the operator formalism was used to prove the factorization of the Hopf link invariant.

The factorization property (2.4.3) is stable under deformation of the gravity dual theory by a Romans mass, taking different levels k_1, k_2. The procedure goes identically as above and gives

$$\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k_1,k_2} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_{k_1,k_2}}}\, C^{\vec{\mu}}_{N_1,N_2;k_1,k_2}\, \langle W_\gamma W_{\tilde{\gamma}}\rangle_{N_1;k_1}\langle W_{\eta'} W_{\tilde{\eta}'}\rangle_{N_2;k_2}, \tag{2.4.4}$$

with refined coefficient

$$C^{\vec{\mu}}_{N_1,N_2;k_1,k_2} = \exp\left[\frac{i\pi N_2}{k_1}\left(N_1N_2+2|\vec{\gamma}|\right)+\frac{i\pi N_1}{k_2}\left(N_1N_2+2|\vec{\eta}'|\right)\right].$$

The expression (2.4.4) can be further reduced using a character expansion:

$$\langle W_\gamma W_{\tilde{\gamma}}\rangle_{N_1;k_1} = \sum_\nu \mathsf{N}_{\gamma\tilde{\gamma}}{}^{\nu}\langle W_\nu\rangle_{N_1;k_1},$$

with $\mathsf{N}_{\gamma\tilde{\gamma}}{}^{\nu}$ the Littlewood–Richardson coefficients, and analogously for $\langle W_{\eta'}W_{\tilde{\eta}'}\rangle_{N_2;k_2}$. The vev of a Wilson loop in Chern–Simons theory along an unknot wrapping a great circle is known [83], see (2.2.1), and the final form of (2.4.4) is:

$$\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k_1,k_2} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_{k_1,k_2}}} C^{\vec{\mu}}_{N_1,N_2;k_1,k_2}\sum_{\nu,\tilde{\nu}}\mathsf{N}_{\gamma\tilde{\gamma}}{}^{\nu}\mathsf{N}_{\eta\tilde{\eta}}{}^{\tilde{\nu}}(\dim_{q_1}\nu)(\dim_{q_2}\tilde{\nu})e^{i\pi\left[\frac{\mathsf{C}_{2;N_1}(\nu)}{k_1}+\frac{\mathsf{C}_{2;N_2}(\tilde{\nu})}{k_2}\right]}.$$

There exists an equivalent derivation, which consists in inverting the variables of one of the two Schur polynomials in each integrals in (2.4.2), using the identity

$$\mathfrak{s}_\nu(X_1^{-1},\dots,X_N^{-1}) = \prod_{a=1}^N X_a^{-\nu_1}\mathfrak{s}_{\nu^*}(X_1,\dots,X_N), \tag{2.4.5}$$

with the starred partition defined as

$$\nu^* = (\nu_1-\nu_N,\nu_1-\nu_{N-1},\dots,\nu_1-\nu_2). \tag{2.4.6}$$

We work directly with generic k_1, k_2 as the computations are identical. Exploiting (2.4.5) we recognize in each factorized integral the vev of a Wilson loop wrapping a Hopf link in pure Chern–Simons theory [107]:

$$\langle W_\mu W_{\tilde{\mu}}\rangle_{N_1,N_2;k_1,k_2} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_{k_1,k_2}}}\, e^{-i\pi\left[\frac{N_1}{k_1}\tilde{\gamma}_1^2+\frac{N_2}{k_2}(\tilde{\eta}_1')^2\right]} C^{\vec{\mu}}_{N_1,N_2;k_1,k_2}\, \langle W_{\gamma\tilde{\gamma}^*}\rangle_{N_1;k_1}\langle W_{\eta'(\tilde{\eta}')^*}\rangle_{N_2;k_2}. \tag{2.4.7}$$

Inverting one of the two Wilson loops

A different correlator of two $\frac{1}{2}$-BPS Wilson loops than (2.4.2) was considered in [108], with one loop carrying inverted variables, mimicking the Hopf link invariant of [107]. This correlator has the integral representation

$$\langle W_\mu \overline{W}_{\tilde{\mu}}\rangle = \frac{1}{\mathcal{Z}_{\mathrm{ABJ}}}\int_{\mathbb{R}^{N_1}}\int_{\mathbb{R}^{N_2}}\mathfrak{s}_\mu(e^{2\pi x}|e^{2\pi y})\mathfrak{s}_{\tilde{\mu}}(e^{-2\pi x}|e^{-2\pi y})\prod_{a=1}^{N_1}e^{i\pi k_1 x_a^2}\,dx_a\prod_{\dot{a}=1}^{N_2}e^{i\pi k_2 y_{\dot{a}}^2}\,dy_{\dot{a}}$$
$$\frac{\prod_{1\le a<b\le N_1}\left(2\sinh\pi(x_b-x_a)\right)^2\prod_{1\le \dot{a}<\dot{b}\le N_2}\left(2\sinh\pi(y_{\dot{b}}-y_{\dot{a}})\right)^2}{\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2}\left(2\cosh\pi(x_a-y_{\dot{a}})\right)^2}$$

where in the left-hand side we have omitted the subscript, $\langle W_\mu \overline{W}_{\tilde{\mu}} \rangle \equiv \langle W_\mu \overline{W}_{\tilde{\mu}} \rangle_{N_1,N_2;k_1,k_2}$, to avoid clutter. We have also considered generic Chern–Simons levels k_1, k_2 as we have seen that the argument holds with no difference. Using (2.4.1) on both supersymmetric Schur polynomial, with

$$\mathfrak{s}_{\tilde{\mu}}(e^{-2\pi x}|e^{-2\pi y}) = \mathfrak{s}_{\tilde{\gamma}}(e^{-2\pi x})\mathfrak{s}_{\tilde{\eta}'}(e^{-2\pi y}) \prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2} (2\cosh \pi(x_a - y_{\dot{a}}))\, e^{-\pi x_a - \pi y_{\dot{a}}},$$

we get

$$\begin{aligned}\langle W_\mu \overline{W}_{\tilde{\mu}} \rangle = \frac{1}{\mathcal{Z}_{\text{ABJ}}} \int_{\mathbb{R}^{N_1}} \mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_{\tilde{\gamma}}(e^{-2\pi x}) \prod_{1\le a<b\le N_1} (2\sinh \pi(x_b - x_a))^2 \prod_{a=1}^{N_1} e^{i\pi k_1 x_a^2}\, dx_a \\ \times \int_{\mathbb{R}^{N_2}} \mathfrak{s}_{\eta'}(e^{2\pi y})\mathfrak{s}_{\tilde{\eta}'}(e^{-2\pi y}) \prod_{1\le \dot{a}<\dot{b}\le N_2} \big(2\sinh \pi(y_{\dot{b}} - y_{\dot{a}})\big)^2 \prod_{\dot{a}=1}^{N_2} e^{i\pi k_2 y_{\dot{a}}^2}\, dy_{\dot{a}}.\end{aligned}$$

We find that the factorization persists, but the observables we get now are Hopf link invariants in $U(N_1)_{k_1}$ and $U(N_2)_{k_2}$ pure Chern–Simons theory, instead of the correlator of two unlinked unknots:

$$\langle W_\mu \overline{W}_{\tilde{\mu}} \rangle_{N_1,N_2;k_1,k_2} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\text{ABJ}(N_1|N_2)_{k_1,k_2}}} \langle W_{\gamma\tilde{\gamma}} \rangle_{N_1;k_1} \langle W_{\eta'\tilde{\eta}'} \rangle_{N_2;k_2}.$$

We could as well run the argument that led to (2.4.7) backwards. Inverting the variables in one of the two (ordinary) Schur polynomials in each integral using (2.4.5) disentangles the Hopf link and gives the correlator of two circular Wilson loops,

$$\langle W_\mu \overline{W}_{\tilde{\mu}} \rangle = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\text{ABJ}(N_1|N_2)_{k_1,k_2}}} e^{i\pi\left[\frac{N_1}{k_1}\tilde{\gamma}_1^2 + \frac{N_2}{k_2}(\tilde{\eta}_1')^2\right]} \langle W_\gamma W_{\tilde{\gamma}^*} \rangle_{N_1;k_1} \langle W_{\eta'} W_{(\tilde{\eta}')^*} \rangle_{N_2;k_2}.$$

The upshot is that having the variables of one of the two supersymmetric Schur polynomials inverted has the effect to switch the role of the partitions $\tilde{\gamma}$ and $\tilde{\eta}'$ with that of the starred ones $\tilde{\gamma}^*$ and $(\tilde{\eta}')^*$.

2.4.4 Three or more Wilson loops

Consider three long $U(N_1|N_2)$ representations $\vec{\mu} = \big(\mu^{(1)}, \mu^{(2)}, \mu^{(3)}\big)$, and let

$$\big\langle W_{\vec{\mu}} \big\rangle \equiv \left\langle \prod_{j=1}^{3} W_{\mu^{(3)}} \right\rangle_{N_1,N_2;k_1,k_2} \tag{2.4.8}$$

denote the correlator of three $\frac{1}{2}$-BPS Wilson loops carrying the representations $\vec{\mu}$ in ABJ theory with ranks N_1 and N_2, and we have allowed generic Chern–Simons levels k_1 and k_2. We also denote for shortness $\mathfrak{s}_{\vec{\gamma}}(e^{2\pi x}) = \prod_{j=1}^{3} \mathfrak{s}_{\gamma^{(j)}}(e^{2\pi x})$, and likewise for $\mathfrak{s}_{\vec{\eta}'}(e^{2\pi y})$. The correlator of the three Wilson loops, using (2.4.1), is

$$\begin{aligned}\big\langle W_{\vec{\mu}} \big\rangle = \frac{1}{\mathcal{Z}_{\text{ABJ}}} \int_{\mathbb{R}^{N_1}} \int_{\mathbb{R}^{N_2}} \prod_{1\le a<b\le N_1} (2\sinh \pi(x_b - x_a))^2 \prod_{1\le \dot{a}<\dot{b}\le N_2} \big(2\sinh \pi(y_{\dot{b}} - y_{\dot{a}})\big)^2 \\ \times \mathfrak{s}_{\vec{\gamma}}(e^{2\pi x})\mathfrak{s}_{\vec{\eta}'}(e^{2\pi y}) \left[\prod_{a=1}^{N_1}\prod_{\dot{a}=1}^{N_2} \big(e^{2\pi x_a} + e^{2\pi y_{\dot{a}}}\big)\, e^{2\pi(x_a + y_{\dot{a}})}\right] \prod_{a=1}^{N_1} e^{i\pi k x_a^2}\, dx_a \prod_{\dot{a}=1}^{N_2} e^{-i\pi k y_{\dot{a}}^2}\, dy_{\dot{a}}.\end{aligned}$$

The term in square bracket on the second line is a symmetric polynomials both in the variables $e^{2\pi x_a}$ and $e^{2\pi y_{\dot a}}$, and we can expand it in the Schur basis using the dual Cauchy identity (2.2.12):

$$\left[\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2}\left(e^{2\pi x_a}+e^{2\pi y_{\dot a}}\right)e^{2\pi(x_a+y_{\dot a})}\right]=\left(\prod_{a=1}^{N_1}e^{2\pi N_2 x_a}\right)\left(\prod_{\dot a=1}^{N_2}e^{3\pi N_1 y_{\dot a}}\right)\sum_{\nu}\mathfrak{s}_{\nu'}(e^{2\pi x})\mathfrak{s}_{\nu}(e^{-2\pi y}), \tag{2.4.9}$$

with the sum running over all partition of length at most $\min\{N_1,N_2\}$. At this point, the correlator is given by a finite sum of terms, each one completely factorized between the two nodes. We can exploit (2.4.5) to invert the variables in the second Schur polynomial in (2.4.9), and get the partition ν^* instead of ν. We find

$$\langle W_{\vec\mu}\rangle=\frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\rm ABJ}}\sum_{\nu}\tilde C_0(\nu_1)\tilde C_1\left(\vec\gamma,\nu\right)\tilde C_2\left(\vec\eta',\nu^*\right)\left\langle W_{\nu'}W_{\vec\gamma}\right\rangle_{N_1;k_1}\left\langle W_{\nu^*}W_{\vec\eta'}\right\rangle_{N_2;k_2}, \tag{2.4.10}$$

where the coefficients are defined as

$$\begin{aligned}\tilde C_0(\nu_1)&=\exp\left[i\pi\frac{N_1N_2^2}{k_1}+i\pi\frac{N_2}{k_2}\left(\frac{3}{2}N_1-\nu_1\right)^2\right],\\ \tilde C_1\left(\vec\gamma,\nu\right)&=\exp\left[i2\pi\frac{N_2}{k_1}\left(|\vec\gamma|+|\nu|\right)\right],\\ \tilde C_2\left(\vec\eta',\nu^*\right)&=\exp\left[i2\pi\frac{N_1}{k_2}\left(|\vec\eta|+|\nu^*|\right)\right].\end{aligned}$$

We are using, as in (2.4.8), the shorthand notation $W_{\vec\gamma}\equiv\prod_j W_{\gamma^{(j)}}$, $|\vec\gamma|=\sum_j|\gamma^{(j)}|$ and so on. We have also used $|\eta'|=|\eta|$, but note that $|\nu^*|\neq|\nu|$.

Formula (2.4.10) is factorized into two correlators of four ordinary Wilson loops is two pure Chern–Simons theories, disconnected and without matter. Each correlator can be further simplified expanding pairwise the products of two Schur polynomials in the Schur basis, using the Littlewood–Richardson rule. Repeating this step twice reduces completely the vev $\langle W_{\vec\mu}\rangle$ to a finite sum of products of two ordinary Wilson loop vevs is two pure Chern–Simons theories:

$$\begin{aligned}\langle W_{\vec\mu}\rangle=\frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\rm ABJ}}\sum_{\nu}\tilde C_0(\nu_1)\tilde C_1\left(\vec\gamma,\nu\right)&\left[\sum_{\tilde\nu,\hat\nu,\check\nu}\mathsf{N}_{\nu'\gamma^{(1)}}{}^{\tilde\nu}\mathsf{N}_{\gamma^{(2)}\gamma^{(3)}}{}^{\hat\nu}\mathsf{N}_{\tilde\nu\hat\nu}{}^{\check\nu}\ \langle W_{\check\nu}\rangle_{N_1;k_1}\right]\\ &\times\tilde C_2\left(\vec\eta',\nu^*\right)\left[\sum_{\tilde\sigma,\hat\sigma,\check\sigma}\mathsf{N}_{\nu^*\eta^{(1)\prime}}{}^{\tilde\sigma}\mathsf{N}_{\eta^{(2)\prime}\eta^{(3)\prime}}{}^{\hat\sigma}\mathsf{N}_{\tilde\sigma\hat\sigma}{}^{\check\sigma}\ \langle W_{\check\sigma}\rangle_{N_2;k_2}\right].\end{aligned}$$

The Wilson loop vevs are known, cf. (2.2.1), and the coefficients $\mathsf{N}_{\mu\nu}{}^{\tilde\nu}$ are the Littlewood–Richardson coefficients, and recall that the sum over ν only includes a finite number of terms.

From the derivation, it is clear that the method applies to the correlator of any number of Wilson loops greater than two. Consider ABJ theory with ranks N_1 and N_2 and Chern–Simons levels k_1 and k_2. Let $\vec\mu$ be a set of $n_{\rm W}\geq 2$ irreducible typical $U(N_1|N_2)$ representations, and take the correlator of the $n_{\rm W}$ $\frac{1}{2}$-BPS Wilson loops in the representations $\vec\mu$. The recipe to compute the correlator is:

- apply the factorization (2.4.1) to all the $n_{\rm W}$ supersymmetric Schur polynomials, and
- simplify two of the products arising from (2.4.1) with the denominator coming from the bi-fundamental hypermultiplets.

- Apply $n_{\rm W}-2$ times the dual Cauchy identity (2.2.12) to expand all the remaining products in the numerator in the Schur basis.
- Use (2.4.5) to bring all the Schur polynomials with variables $e^{-2\pi x}$ or $e^{-2\pi y}$ into functions of $e^{2\pi x}$ and $e^{2\pi y}$.
- Expand the product of ordinary Schur polynomials pairwise using the Littlewood–Richardson rule. Repeat this step until the products are completely reduced.
- The final result is a finite sum of Wilson loop vevs in pure Chern–Simons theory, wrapping a great circle in $\mathbb{S}^3$.

Besides, we notice that if some of the supersymmetric Schur polynomials have inverted variables [108], the recipe does not change and they are taken care of in the fourth step.

As a sample application, consider the particular case of four rectangular $N_1\times N_2$ Young diagrams, $\vec{\mu}=(\kappa,\kappa,\kappa,\kappa)$. From (2.4.1) we obtain

$$\begin{aligned}\langle (W_\kappa)^4\rangle = \frac{1}{\mathcal{Z}_{\rm ABJ}}\int_{\mathbb{R}^{N_1}}\int_{\mathbb{R}^{N_2}}\prod_{1\leq a<b\leq N_1}\left(2\sinh\pi(x_b-x_a)\right)^2\prod_{1\leq \dot a<\dot b\leq N_2}\left(2\sinh\pi(y_{\dot b}-y_{\dot a})\right)^2\\ \times\left[\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2}2\cosh\pi(x_a-y_{\dot a})\right]^2\prod_{a=1}^{N_1}e^{i\pi k x_a^2+4\pi N_2 x_a}\,dx_a\prod_{\dot a=1}^{N_2}e^{-i\pi k y_{\dot a}^2+4\pi N_1 y_{\dot a}}\,dy_{\dot a},\end{aligned}$$

which, except for the normalization by $\mathcal{Z}_{\rm ABJ}$, is the partition function of pure $U(N_1+N_2)$ Chern–Simons theory on the lens space $L(2,1)\simeq\mathbb{S}^3/\mathbb{Z}_2$, evaluated in the background of a fixed, generic flat connection that breaks the gauge symmetry

$$U(N_1+N_2)\longrightarrow U(N_1)\times U(N_2).$$

Following the steps listed above, we get

$$\begin{aligned}\langle (W_\kappa)^4\rangle = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\rm ABJ}}\sum_{\nu,\tilde\nu}\hat C(\nu,\tilde\nu)\left[\sum_{\hat\nu}\mathsf{N}_{\nu'\tilde\nu'}{}^{\hat\nu}(\dim_{q_1}\hat\nu)q_1^{-\frac{1}{2}\mathsf{C}_{2;N_1}(\hat\nu)}\right]\\ \times\left[\sum_{\hat\sigma}\mathsf{N}_{\nu^*\tilde\nu^*}{}^{\hat\sigma}(\dim_{q_1}\hat\sigma)q_1^{-\frac{1}{2}\mathsf{C}_{2;N_1}(\hat\sigma)}\right],\end{aligned}$$

with coefficient

$$\hat C(\nu,\tilde\nu)=\exp\left[i\pi\frac{N_1}{k_1}\left(N_2^2+2|\nu|+2|\tilde\nu|\right)+i\pi\frac{N_2}{k_2}\left((2N_1-\nu_1-\tilde\nu_1)^2+|\nu^*|+|\tilde\nu^*|\right)\right].$$

The complete partition function of pure $U(N)$ Chern–Simons theory on $L(2,1)$ is obtained from this expression, dropping the overall normalization and summing over all N_1 and N_2 with $N_1+N_2=N$ fixed.

2.4.5 Necklace quivers

We now discuss the insertion of supersymmetric Schur polynomials in the matrix model describing quiver Chern–Simons theories

$$U(N_0)_{k_0}\times U(N_1)_{k_1}\times\cdots\times U(N_r)_{k_r}.$$

We focus for clarity on an extended $\widehat{A}_r$-type quiver, periodically identifying the nodes $r+1 \equiv 0$, being the discussion for linear quivers completely analogous. Let us fix $p \in \{0, \dots, r\}$ and consider a typical $U(N_p|N_{p+1})$ representation μ. The average of the supersymmetric Schur polynomial $\mathfrak{s}_\mu$ is

$$\langle \mathfrak{s}_\mu \rangle = \int_{\mathbb{R}^{r+1}} \mathfrak{s}_\mu \left(e^{2\pi x_p}|e^{2\pi x_{p+1}}\right) \prod_{p=1}^{r} \prod_{a=1}^{N_p} e^{i\pi k_p x_a^2} dx_a \\ \frac{\prod_{1\le a<b\le N_p} \left(2\sinh \pi(x_{p,a}-x_{p,b})\right)^2 \prod_{1\le \dot{a}<\dot{b}\le N_{p+1}} \left(2\sinh \pi(x_{p+1,a}-x_{p+1,b})\right)^2}{\prod_{a=1}^{N_p}\prod_{\dot{a}=1}^{N_{p+1}} 2\cosh \pi(x_{p,a}-x_{p+1,\dot{a}})}.$$

The identity (2.4.1) has the net effect to cut the edge joining the p^{th} node to the $(p+1)^{\text{th}}$, leaving behind the correlator of two Wilson loops, one in the $U(N_p)$ representation γ and the other in the $U(N_{p+1})$ representation η', computed in a A_{r+1} linear quiver gauge theory.

The correlator of more than one supersymmetric Schur polynomial, taken in typical representations of different supergroups $U(N_p|N_{p+1})$, cuts the edges joining each pair of nodes involved in the definition of the supersymmetric Schur polynomials. The final expression is factorized into the correlators of Wilson loops in disconnected linear quivers, with the loop operator inserted at the first or last node of each sub-quiver.

Consider, for example, a necklace quiver with four nodes, and take a typical $U(N_0|N_1)$ representations μ and a typical $U(N_1|N_2)$ representations $\tilde{\mu}$, as in Figure 2.17. We find

$$\langle \mathfrak{s}_\mu \mathfrak{s}_{\tilde{\mu}} \rangle_{\widehat{A}_3} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{A_3}}{\mathcal{Z}_{\widehat{A}_3}} \langle W_\gamma W_{\tilde{\eta}'} \rangle_{A_3} \langle W_{\eta'} W_{\tilde{\gamma}} \rangle_{N_1;k_1}.$$

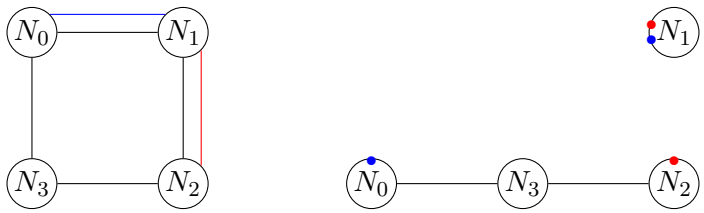


Figure 2.17. Left: $\widehat{A}_3$ quiver with two supersymmetric Schur polynomial insertions, represented as a blue and a red line respectively. Right: the same quantity is factorized into two disjoint sub-quivers, with blue and red dots denoting ordinary Schur polynomial insertions.

The special case

$$\left\langle \prod_{p=0}^{r} \mathfrak{s}_{\mu^{(p)}} \right\rangle \qquad \text{with } \mu^{(p)} \text{ a typical } U(N_p|N_{p+1}) \text{ representation}$$

is completely factorized into correlators of pairs of Wilson loops in ordinary, bosonic $U(N_p)$ Chern–Simons theory with renormalized Chern–Simons level k_p, for all $p = 0, 1, \dots, r$.

2.5 Schur expansion and its perturbative meaning

In this section, we exploit the Cauchy identity (2.2.10) in different classes of Chern–Simons-matter theories and uncover a relation between the partition function of such theories and formal power series encoding topological invariants of simple links and (un)knots. As we will see, the series that appear are coarse-grained versions of generating functions. The invariants we obtain on the right-hand side are associated to either the unknot, a collection of unlinked unknots, or the Hopf link,

coloured by $U(N)$ or $SU(N)$ representations. If we denote by t the variable in the generating-like series of such link invariants, we find that it is related to the physical quantities of the gauge theory we started with through

$$t = -e^{2\pi m} \tag{2.5.1}$$

where m is a real mass parameter. If there are more mass parameters, associated to the Cartan subalgebra of the flavour symmetry, we get a collection $\{t_j\} = \{-e^{2\pi m_j}\}$.

The simple rewriting

$$t = e^{-i2\pi\left(\frac{1}{2}+im\right)} \equiv e^{-i2\pi\lambda}$$

shows that t is a fugacity for the variable $\lambda = \frac{1}{2} + im$, which is ubiquitous in the calculations of Section 2.3. More accurately stated, in Sections 2.2.3 and 2.3 we have found that the results are functions of the fugacity $e^{i2\pi\lambda \, \mathrm{sign}\Re k}$, but it is in fact a matter of conventions whether we choose to expand in positive or negative powers of t, as will be clear from the examples below.

As we already pointed out in Subsection 2.3.3, the partition function is holomorphic in λ, which is precisely the holomorphic variables of Jafferis [96], but further constrained by the $\mathcal{N} \geq 3$ supersymmetry in all the theories considered in the present work. Besides, we have found holomorphy in the vertical strip $\{0 < \Re\lambda < 1, -\infty < \Im\lambda < +\infty\}$ [61], thus the partition functions are holomorphic functions of $t \in \mathbb{C} \setminus \mathbb{R}_{\geq} 0$.

The fact that turning off background values for the flavour symmetry corresponds to "take the Euler characteristic", $t \to -1$, may point toward an interpretation in terms of categorification of link invariants [109, 110], although not in the direction of the Khovanov–Rozansky homology. However, as we will see explicitly in the examples below, the quantities we obtain with our prescription have a too simple structure to capture homological data. In conclusion, there are obstructions in embedding the results presented in this section into some homological theory of knots.

Before diving into the detailed analysis, a remark is in order. It is important to bear in mind that the Cauchy identity (2.2.10) is algebraic, and is meant as an equality of the coefficients of the book-keeping variables $\{t_j\}$ order by order in a (possibly formal) series expansion.[4] We will use the symbol "$\overset{\text{pert.}}{=}$" to signify that the equality between the left- and the right-hand side will be understood as equating the coefficients of each variable t_j order by order. Note that the distinction between perturbative and non-perturbative in all the formulas in this section is meant as functions of the fugacities $\{t_j\}$ of the global symmetries, and not as functions of the gauge or Chern–Simons couplings.

A toy example: Dawson's integral

To set the ground for the Schur expansion of physically sensible theories in the forthcoming Subsections, we firstly present our argument in a toy model. Consider the integral

$$F_{\text{Dawson}}(t^{-1}) = \int_{-\infty}^{+\infty} \frac{dx}{\sqrt{\pi}} \frac{e^{-x^2}}{x + t^{-1}} = \underbrace{\sqrt{\pi} e^{-i\frac{\pi}{2}\mathrm{sign}(t) - \frac{1}{t^2}}}_{\text{non-pert.}} + t + \frac{1}{2}t^3 + \frac{3}{4}t^5 + \frac{15}{8}t^7 + \frac{105}{16}t^9 + \dots \tag{2.5.2}$$

known as Dawson's integral [111]. The Gaussian damping term plays the role of the Chern–Simons coupling in this toy example, and moreover we ave chosen to write t^{-1} instead of t to mimic what we get from massive hypermultiplets in the physical theories. In the right-hand side we have identified a non-perturbative part in t and a formal power series in t. The customary expansion of

[4] The dual Cauchy identity (2.2.12), instead, is a finite sum and this issue does not show up.

a Stieltjes transform such as (2.5.2) consists in considering the denominator as a geometric series, giving:

$$F_{\text{Dawson}}(t^{-1}) \overset{\text{pert.}}{=} t + \frac{1}{2}t^3 + \frac{3}{4}t^5 + \frac{15}{8}t^7 + \frac{105}{16}t^9 + \dots$$

The agreement of this solution with (2.5.2) can be checked to arbitrarily high order in t, once the non-perturbative term is discarded.

2.5.1 Schur expansion of A_1 theories with adjoint matter

Schur expansion: $SU(2)$ Chern–Simons theory with one adjoint hypermultiplet

Let us consider the partition function of $SU(2)_k$ Chern–Simons theory with one adjoint hypermultiplet. We turn on a real mass m associated to the $U(1)$ flavour symmetry rotating the adjoint, and define the fugacity $t = -e^{2\pi m}$, as in (2.5.1). The partition function is

$$\begin{aligned}\mathcal{Z}_{SU(2),1_{\text{adj}}}(m) &= \int_{-\infty}^{+\infty} dx_1 \int_{-\infty}^{+\infty} dx_2\ \delta(x_1+x_2) \frac{(2\sinh\pi(x_1-x_2))^2\ \ e^{i\pi k(x_1^2+x_2^2)}}{(2\cosh\pi(x_1-x_2+m))(2\cosh\pi(x_2-x_1+m))} \\ &= \int_{-\infty}^{+\infty} dx \frac{4\sinh(2\pi x)^2\ e^{i2\pi k x^2}}{(2\cosh\pi(2x+m))(2\cosh\pi(2x-m))}.\end{aligned}$$

Rewriting the denominator and using the Cauchy identity (2.2.10) we arrive at

$$\frac{\mathcal{Z}_{SU(2)_k,1_{\text{adj}}}(m)}{\mathcal{Z}_{SU(2)_k}} \overset{\text{pert.}}{=} \sum_{\nu=0}^{\infty} t^{\nu+1} \langle W_{\nu\nu}\rangle_{SU(2)}, \tag{2.5.3}$$

where we have used the definition (2.5.1). The sum runs over isomorphism classes of irreducible $SU(2)$ representations, in one-to-one correspondence with non-negative integers ν. We recognize the generating function of the vevs of a Wilson loops running along a Hopf link in $\mathbb{S}^3$, computed in $SU(2)$ Chern–Simons theory with renormalized coupling $k = k_{\text{bare}} + 2$. These vevs in turn are given by coloured Jones polynomials [42].

Schur expansion: $SU(N)$ Chern–Simons theory with one adjoint hypermultiplet

We now generalize the discussion above to higher rank, considering $SU(N)$ theory. We may consider $U(N)$ theory as well, and the procedure goes through in precisely the same way.

The partition function of $SU(N)$ Chern–Simons theory coupled to one adjoint is

$$\mathcal{Z}_{SU(N)_k,1_{\text{adj}}}(m) = \int_{\mathbb{R}^N} d^N x\ \delta\left(\sum_{a=1}^N x_a\right) \frac{\prod_{1\le a\neq b\le N} 2\sinh\pi(x_a-x_b)}{\prod_{a,b=1}^N 2\cosh(x_a-x_b+m)}\ e^{i\pi k\sum_{a=1}^N x_a^2}.$$

The usual manipulations on the denominator, taking advantage of the δ function in the integrand to simplify the expression, and the application of the Cauchy identity (2.2.10) lead us to

$$\prod_{a=1}^N \prod_{b=1}^N \left(1 + e^{2\pi x_a} e^{-2\pi x_b + 2\pi m}\right)^{-1} = \sum_{\nu} \mathfrak{s}_\nu(e^{2\pi x}) \mathfrak{s}_\nu(-e^{-2\pi x + 2\pi m}),$$

where the sum is over irreducible representation of $SU(N)$, which are equivalently represented by Young diagrams with at most $N-1$ rows. We also adopted a shorthand notation $\mathfrak{s}_\nu(e^{2\pi x}) := \mathfrak{s}_\nu(e^{2\pi x_1}, e^{2\pi x_2}, \dots, e^{2\pi x_N})$. We obtain the expansion of the partition function

$$\frac{\mathcal{Z}_{SU(N)_k,1_{\text{adj}}}(m)}{\mathcal{Z}_{SU(N)_k}} \overset{\text{pert.}}{=} t^{\frac{N(N-1)}{2}} \sum_{\nu} t^{|\nu|} \langle W_{\nu\nu}\rangle_{SU(N)}. \tag{2.5.4}$$

We have used the definition (2.5.1) of the fugacity t, and $|\nu|$ is the number of boxes in the Young diagram ν.

We find a (formal) polynomial in two variables (q,t), which as a function of the variable t, looks similar to a generating function of HOMFLY-PT polynomials of the Hopf link coloured by $SU(N)$ representation. Note that the Hopf link is non-generic, since it is yields equal representations on the two components. Note also that it is not truly a generating function, because each summand is weighted by $t^{|\nu|}$, which does not distinguish between representation with the same value of $|\nu|$.

Schur expansion: $SU(N)$ Chern–Simons theory with N_{adj} adjoint hypermultiplets

The computations can be extended to an arbitrary number $N_{\text{adj}} = n$ of adjoint hypermultiplets with generic masses. The $SU(N)$ partition function is

$$\mathcal{Z}_{SU(N)_k,1_{\text{adj}}}(m) = \int_{\mathbb{R}^N} d^N x\ \delta\left(\sum_{a=1}^N x_a\right)\ \frac{\prod_{1\le a\neq b\le N} 2\sinh\pi(x_a-x_b)}{\prod_{j=1}^{N_{\text{adj}}}\prod_{a,b=1}^N 2\cosh(x_a-x_b+m_j)}\ e^{i\pi k\sum_{a=1}^N x_a^2}.$$

We mimic the steps above and apply the Cauchy identity n times, arriving at

$$\frac{\mathcal{Z}_{SU(N)_k,N_{\text{adj}}}(m)}{\mathcal{Z}_{SU(N)_k}} \overset{\text{pert.}}{=} \left(\prod_{j=1}^n t_j^{\frac{N(N-1)}{2}}\right)\ \sum_{\nu^{(1)}} t_1^{|\nu^{(1)}|}\cdots\sum_{\nu^{(n)}} t_n^{|\nu^{(n)}|}\left\langle\prod_{j=1}^n W_{\nu^{(j)}\nu^{(j)}}\right\rangle_{SU(N)}.$$

The average computes the correlator of $n = N_{\text{adj}}$ pairwise unlinked Hopf links, each one with equally coloured components.

2.5.2 Schur expansion of A_1 theories with fundamental matter

Schur expansion: $U(1)$ theory with N_f hypermultiplets

We now go back to the Abelian A_1 Chern–Simons theory with N_f massive hypermultiplets, discussed in Subsection 2.3.2. We assume an even number of hypermultiplets $N_f = 2n$ and write the partition function in the form

$$\mathcal{Z}_{U(1),2n}(k,\vec{m}) = \int_{-\infty}^{+\infty} dx \frac{e^{i\pi k x^2+2\pi n x}}{\prod_{j=1}^{2n}(1-t_j e^{2\pi x})},$$

where $t_j = -e^{2\pi m_j}$, as defined in (2.5.1), and we used $\sum_{j=1}^{N_f} m_j = 0$ to drop an overall factor. We now exploit the Cauchy identity (2.2.10). We thus write

$$\prod_{j=1}^{2n}\left(1-t_j e^{2\pi x}\right)^{-1} = \sum_{\nu=0}^{\infty}\mathfrak{s}_\nu(e^{2\pi x})\mathfrak{s}_\nu(t_1,\dots,t_{2n}) \tag{2.5.5}$$

and obtain:

$$\frac{\mathcal{Z}_{U(1),2n}(k,\vec{m})}{\mathcal{Z}_{\text{CS}(1)_k}} \overset{\text{pert.}}{=} \sum_{\nu=0}^{\infty}\mathfrak{s}_\nu(t_1,\dots,t_{2n})\langle W_{\nu+n}\rangle_{U(1)_k}$$

where $\langle W_{\nu+n}\rangle_{U(1)_k}$ stands for the vev of the Wilson loop in the $U(1)$ representation corresponding to $\nu+n\in\mathbb{Z}_{>0}$ computed in pure Chern–Simons theory on $\mathbb{S}^3$ at level k. Recall that the fugacities t_j are defined in (2.5.1) as minus the fugacities for the maximal torus of the flavour symmetry.

This Abelian case is particularly simple: recall from (2.2.11) that equation (2.5.5) gives in fact the generating function of the homogeneous symmetric polynomials $\mathfrak{h}_\nu(t_1,\dots,t_{2n})$ [90], and besides the Wilson loop is captured by a simple Gaussian integral. We get:

$$\begin{aligned}\mathcal{Z}_{U(1),2n}(k,\vec{m}) &\overset{\text{pert.}}{=} \sqrt{\frac{i}{k}} e^{\frac{i\pi}{k}n^2} \sum_{\nu=0}^{\infty} e^{\frac{i\pi}{k}(\nu^2+2\nu n)} \mathfrak{h}_\nu(t_1,\dots,t_{2n}) \\ &= \sqrt{\frac{i}{k}} \left\{ e^{\frac{i\pi}{k}n^2} - e^{\frac{i\pi}{k}(n+1)^2}\mathfrak{h}_1(t_1,\dots,t_{2n}) + e^{\frac{i\pi}{k}(n+2)^2}\mathfrak{h}_2(t_1,\dots,t_{2n}) + \dots \right\},\end{aligned} \tag{2.5.6}$$

where $\mathfrak{h}_1(t_1,\dots,t_{2n}) = \sum_{j=1}^{2n} t_j$, $\mathfrak{h}_2(t_1,\dots,t_{2n}) = \sum_{1\le j\le l\le 2n} t_j t_l$ and so on, and recall that the number of flavours is $N_f = 2n$. The result is a symmetric polynomial in the fugacities t_j.

We can compare the result (2.5.6) with the exact one obtained in Section 2.3.2, but in doing so we have to bear in mind a few caveats:

- While the physical parameters satisfy $\prod_{j=1}^{2n} t_j = 1$, we should treat these as formal indeterminates, thus expanding for each t_j independently.
- The formal expansion is in positive powers of t_j, hence it will be compared with $k < 0$ in Subsection 2.3.2. We could as well have begun with the expansion in negative powers of t_j, to be compared with $k > 0$ in 2.3.2. In each case, the choice must be made at the beginning, through the manipulations of the denominator before plugging the identity (2.2.10). Nevertheless, the summation variable ν plays the role of a real irreducible $U(1)$ representation, and through the isomorphism with its conjugate representation we could extract the expansion for $k > 0$.
- The Schur expansion will miss non-perturbative terms in t_j, namely those $\propto e^{-i\pi k\left(\frac{1}{2}+im_j\right)^2}$.

After elementary manipulations of the result in Subsection 2.3.2 and dropping non-perturbative terms, the expansion relative to a single t_j is

$$\frac{1}{1-(-t_j)^{|k|}} \sum_{\alpha=0}^{|k|-1} t_j^\alpha e^{\frac{i\pi}{k}(\alpha+n)^2} \prod_{s\neq j} \sum_{\beta_s=0}^{\infty} \left(\frac{t_s}{t_j}\right)^{\beta_s}.$$

The prefactor should be expanded as a geometric series to compare with the Schur expansion. In this way, the terms $t_j^{\beta|k|}$ kick in extending the summation range beyond $\alpha = |k|-1$,

$$\sum_{\beta=0}^{\infty} \sum_{\alpha=0}^{|k|-1} (-1)^{\beta|k|} e^{\frac{i\pi}{k}(\alpha+n)^2} t_j^{\alpha+\beta|k|} \prod_{s\neq j} \sum_{\beta_s=0}^{\infty} \left(\frac{t_s}{t_j}\right)^{\beta_s}. \tag{2.5.7}$$

To check the agreement of the two expressions requires care in the power counting. So, for example, tho compare at order t_j^1, one should take into account all the combinations, which in particular include a term from all the homogeneous polynomials in the Schur expansion, which contribute

$$e^{\frac{i\pi}{k}(n+1)^2} t_j + \sum_{s\neq j} e^{\frac{i\pi}{k}(n+2)^2} t_j t_s + \sum_{s_1,s_2\neq j} e^{\frac{i\pi}{k}(n+3)^2} t_j t_{s_1} t_{s_2} + \dots$$

For $\mathfrak{h}_\nu$ with $\nu > |k|-1$ write $\nu = \alpha + \beta|k|$ and use

$$e^{\frac{i\pi}{k}(n+\nu)^2} = e^{\frac{i\pi}{k}(n+\alpha)^2 + i\pi\beta^2|k|} = (-1)^{\beta|k|} e^{\frac{i\pi}{k}(n+\alpha)^2}.$$

On the side of the exact evaluation (2.5.7) in turn we see that all α and β contribute, as they are partially cancelled by the $t_j^{-\beta_s}$. Term by term comparison shows that the Schur expansion

correctly reproduces the exact answer, with the non-perturbative contributions already discarded. Let us stress once again that the agreement is understood in an algebraic sense, reading off the coefficients of the multiple expansion in $\{t_j\}$.

To conclude the analysis of the present theory, we note that the same expressions have been analyzed in [112], in the context of topological strings with non-compact branes. To make contact with that setting we specialize the masses

$$m_j = m\left(n - j + \frac{1}{2}\right)$$

(recall that $N_f = 2n$) and define $t = -e^{2\pi m}$. The homogeneous polynomials become a q-binomial, with q-parameter t:

$$\mathfrak{h}_\nu\left(t^{n-\frac{1}{2}}, \dots, t^{-n+\frac{1}{2}}\right) = \begin{bmatrix} n \\ \nu \end{bmatrix}_t .$$

Then, our expressions differ from [112] only in the Gaussian term in the sum. This mismatch is exactly the factor due to the difference in the framing, as the Wilson loop vev on $\mathbb{S}^3$ in [112] is computed in the natural framing instead of the matrix model framing.

Schur expansion: $U(N)$ and $SU(N)$ theory with N_f hypermultiplets

The manipulations above have been presented in the Abelian theory for clarity, but are straightforwardly generalized to the non-Abelian setting. The partition function of $U(N)_k$ Chern–Simons theory with N_f fundamental hypermultiplets, studied in Section 2.3.2, is more suitably written for our purposes in the form

$$\mathcal{Z}_{U(N),N_f} = \int_{\mathbb{R}^N} \frac{\prod_{1\le a<b\le N} \left(2\sinh \pi(x_b - x_a)\right)^2}{\prod_{a=1}^N \prod_{j=1}^{N_f} \left(1 - t_j e^{2\pi x_a}\right)} \prod_{a=1}^N e^{i\pi k x_a^2 + \pi N_f x_a}\, dx_a.$$

When the gauge group is $SU(N)$ the partition function includes a δ-function $\delta\left(\sum_{a=1}^N x_a\right)$ in the measure.

Using the Cauchy identity (2.2.10), we identify the average of a Schur polynomial in the Chern–Simons random matrix ensemble, which computes the vev of a Wilson loop. Note however that in principle we cannot reabsorb the $\pi N_f x_a$ in the exponential because it would move the integration contour away from the real axis, and the integrand has poles in the complex plane. Equivalently, the problem can be seen reabsorbing the shift into a redefinition of the masses, which would acquire half-integer imaginary part, rendering the integrand singular. To handle this, we pass from $q = e^{-i2\pi/k}$ to $q = e^{-g}$, $g > 0$. Doing so, we can safely complete the square in the matrix model, and the change of variables shifts $2\pi m_j \mapsto 2\pi m_j + \frac{g}{2} N_f$.

This problem does not arise in the $SU(N)$ theory, since the δ-constraint on the eigenvalues would cancel the linear shift, and we are allowed to work directly with q root of unity.

With this distinction in mind, we get for the $U(N)$ case

$$\frac{\mathcal{Z}_{U(N),N_f}}{\mathcal{Z}_{\mathrm{CS}(N)}} \overset{\text{pert.}}{=} q^{-\frac{N_f^2}{8}} \sum_\nu \mathfrak{s}_\nu\left(q^{-\frac{N_f}{2}} t_1, \dots, q^{-\frac{N_f}{2}} t_{N_f}\right) \langle W_\nu \rangle_{\mathrm{CS}(N)}. \tag{2.5.8}$$

The sum runs over Young diagrams associated to irreducible $U(N)$ representations, and the basic properties of the symmetric polynomials imply that all contributions with

$$\mathrm{length}(\nu) > \min\{N, N_f\}$$

vanish. The average $\langle\cdots\rangle_{\mathrm{CS}(N)}$ means the vev in $U(N)$ Chern–Simons theory with real $q = e^{-g}$.

The overall factor in (2.5.8) is reminiscent of the effective Chern–Simons coupling associated to a mixed flavour-R contact term [98]. Besides, we again notice how the result is more naturally written in terms of fugacities for the holomorphic variables $\lambda_j = \frac{1}{2} + im_j$ rather than for the masses m_j alone. The q-shift of the mass parameters seem likewise to originate from an effective coupling for the background fields. This q-shift can be brought out of the Schur polynomials and contributes a factor $q^{-\frac{N_f}{2}|\nu|}$ to each summand.

The Wilson loop vev is known [83] and has been presented in equation (2.2.1), which we report here for clarity:

$$\langle W_\nu\rangle_{\mathrm{CS}(N)} = (\dim_q \nu)\; q^{-\frac{1}{2}\mathsf{C}_{2;N}(\nu)}.$$

In the $SU(N)$ theory instead we obtain

$$\frac{\mathcal{Z}_{SU(N),N_f}}{\mathcal{Z}_{SU(N)_k}} \overset{\text{pert.}}{=} \sum_\nu \mathfrak{s}_\nu\left(t_1,\dots,t_{N_f}\right)\langle W_\nu\rangle_{SU(N)_k}. \tag{2.5.9}$$

The difference, besides the overall factor $q^{-\frac{N_f^2}{8}}$, is the specialization of the variables in the argument of the Schur polynomial, which are not renormalized by a q-shift.

We have therefore written the partition function of Chern–Simons theory with N_f fundamental hypermultiplets as a generating-like function of unknot invariants. From (2.5.8) we can also obtain the Schur expansion of the A_2 quiver theory, simply dropping the constraint $\prod_{j=1}^{N_f} t_j = (-1)^{N_f}$ (this would introduce a factor $\prod_{j=1}^{N_f}(-t_j)^N$ in the matrix model, which we have set to 1) and gauging the $U(N_f)$ symmetry. Adding a Chern–Simons term to the newly gauge node and using (2.2.1) we find for the A_2 quiver $U(N_1)\times U(N_2)$ Chern–Simons theory

$$\frac{\mathcal{Z}_{A_2}}{\mathcal{Z}_{\mathrm{CS}(N_1)}\mathcal{Z}_{\mathrm{CS}(N_2)}} \overset{\text{pert.}}{=} q_1^{-\frac{N_f^2}{8}}\sum_\nu\left(-q_1^{-\frac{N_2}{2}}\right)^{|\nu|}(\dim_{q_1}\nu)(\dim_{q_2}\nu)\; q_1^{-\frac{1}{2}\mathsf{C}_{2;N_1}(\nu)}q_2^{-\frac{1}{2}\mathsf{C}_{2;N_2}(\nu)}$$

where q_1 and q_2 are the q-parameters of the two pure Chern–Simons theories obtained removing the edge joining the two nodes of the A_2 quiver.

Schur expansion: $SU(N)$ theory with fundamental and adjoint hypermultiplets

We can consider a theory with both N_f fundamental and N_{adj} adjoint hypermultiplets. We will limit ourselves to $N_{\mathrm{adj}} = 1$, being the effect of adding more adjoint matter studied in Subsection 2.5.1. We work with gauge group $SU(N)$ for concreteness, being the $U(N)$ theory completely analogous, up to a change of variables which generates a q-shift of the fugacities t_j.

$$\mathcal{Z}_{SU(N),N_f,1_{\mathrm{adj}}}(\vec{m},m_0) = \int_{\mathbb{R}^N}\delta\left(\sum_{a=1}^N x_a\right)\frac{\prod_{1\le a<b\le N}\left(2\sinh\pi(x_b-x_a)\right)^2}{\prod_{a,b=1}^N\left(1-t_0e^{2\pi x_a-2\pi x_b}\right)}\prod_{a=1}^N\frac{e^{i\pi k x_a^2}\,dx_a}{\prod_{j=1}^{N_f}\left(1-t_je^{2\pi x_a}\right)}.$$

Here the variables $\{t_j\}$ are as in (2.5.1), and we have denoted m_0 the mass of the adjoint and $t_0 = -e^{2\pi m_0}$ the corresponding fugacity. Combining the manipulations of Subsection 2.5.1 with those of 2.5.2 we arrive at

$$\frac{\mathcal{Z}_{SU(N),N_f,1_{\mathrm{adj}}}(\vec{m},m_0)}{\mathcal{Z}_{SU(N)_k}} \overset{\text{pert.}}{=} \sum_{\mu,\nu} t_0^{|\mu|}\mathfrak{s}_\nu(t_1,\dots,t_{N_f})\langle W_\mu W_{\nu\mu}\rangle_{SU(N)_k}.$$

From the matrix model description we see that, adding fundamental matter to the theory with one adjoint, we have produced more interesting observables, which are correlators of two Wilson loops, one along an unknot and one along a Hopf link, with the latter not necessarily coloured by two equal representations.

Schur expansion: 4d $\mathcal{N}=4$ super-Yang–Mills with defects

We now apply the ideas presented in this section to a special case of four-dimensional gauge theory, namely $\mathcal{N}=4$ $U(N)$ super-Yang–Mills (SYM) on $\mathbb{S}^4$ with codimension-one matter defects placed at the equatorial $\mathbb{S}^3 \subset \mathbb{S}^4$ [113]. The partition function of such theory, as obtained from localization, is [113, 114]

$$\mathcal{Z}^{\text{4d+defect}}_{U(N),N_f} = \int_{\mathbb{R}^N} \prod_{1\le a<b\le N} (x_a - x_b)^2 \prod_{a=1}^{N} \frac{e^{-\frac{8\pi^2}{g_{\text{4d}}} x_a^2}\, dx_a}{\prod_{j=1}^{N_f} 2\cosh \pi(x_a + m_j)}.$$

Applying identical manipulations as in Subsection 2.5.2, we arrive at a perturbative expansion in the parameters t_j, exactly as in the purely 3d framework, but now the summands are vevs of Wilson loops computed in 4d $\mathcal{N}=4$ SYM (with $q_{\text{4d}} = e^{-g_{\text{4d}}/16\pi^2}$):

$$\frac{\mathcal{Z}^{\text{4d+defect}}_{U(N),N_f}}{\mathcal{Z}^{\text{4d }\mathcal{N}=4}_{U(N),N_f}} \overset{\text{pert.}}{=} q_{\text{4d}}^{-\frac{N_f^2}{8}} \sum_{\nu} \mathfrak{s}_\nu \left(q_{\text{4d}}^{-\frac{N_f}{2}} t_1, \dots, q_{\text{4d}}^{-\frac{N_f}{2}} t_{N_f} \right) \langle W_\nu \rangle^{\text{4d }\mathcal{N}=4}_{U(N)}.$$

2.5.3 Schur expansion of necklace quiver theories

The focus of this Subsection is on quiver gauge theories of type $\widehat{A}_r$.

Schur expansion: ABJ

We now consider the mass-deformed ABJ theory, whose partition function reads:

$$\begin{aligned}\mathcal{Z}_{\text{ABJ}(N_1|N_2)}(k,m) = &\int_{\mathbb{R}^{N_1}} d^{N_1}\vec{x} e^{i\pi k \sum_{a=1}^{N_1} x_a^2} \int_{\mathbb{R}^{N_2}} d^{N_2}\vec{y} e^{-i\pi k \sum_{\dot a=1}^{N_2} y_{\dot a}^2} \\ &\times \frac{\prod_{1\le a<b\le N_1} \left(2\sinh\pi(x_b - x_a)\right)^2 \prod_{1\le \dot a<\dot b\le N_1} \left(2\sinh\pi(y_{\dot b} - y_{\dot a})\right)^2}{\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2} 2\cosh\pi(x_a - y_{\dot a} + m_-)\; 2\cosh\pi(x_a - y_{\dot a} + m_+)}\end{aligned}$$

where the physical values of the masses are $m_\pm = \pm m$, but here we treat them as independent. This can be achieved turning on a FI parameter which, upon changing variables, shifts the real masses. The second line is more conveniently written as

$$e^{\pi(N_1+N_2)(m_+ - m_-)} \left[\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2} \left(1 + e^{2\pi x_a} e^{-2\pi(y_{\dot a} + m_-)}\right) \left(1 + e^{-2\pi x_a} e^{2\pi(y_{\dot a} + m_+)}\right) \right]^{-1}.$$

We now apply the Cauchy identity (2.2.10)

$$\left[\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2} \left(1 + e^{\pm 2\pi x_a} e^{\mp 2\pi(y_{\dot a} + m_\mp)}\right) \right]^{-1} = \sum_{\nu} \mathfrak{s}_\nu(-e^{\pm 2\pi x}) \mathfrak{s}_\nu(e^{\mp 2\pi(y + m_\mp)}),$$

adopting the usual shorthand $e^{2\pi x}$ for $(e^{2\pi x_1}, \dots, e^{2\pi x_{N_1}})$ and likewise for $e^{2\pi y}$, and the sum runs over all partitions ν with

$$\text{length}(\nu) \le \min(N_1, N_2). \tag{2.5.10}$$

Therefore, bringing the common factors $e^{\mp 2\pi m_\mp}$ out of the Schur polynomials in $e^{2\pi y}$ we get

$$\begin{aligned}\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)}(k,m) \overset{\text{pert.}}{=} e^{\pi(N_1+N_2)(m_+-m_-)} \sum_\mu \sum_\nu (-e^{2\pi m_+})^{|\mu|}(-e^{-2\pi m_-})^{|\nu|} \\ \times \int_{\mathbb{R}^{N_1}} \mathfrak{s}_\mu(e^{2\pi x})\mathfrak{s}_\nu(e^{-2\pi x}) \prod_{1\le a<b\le N_1} \left(2\sinh \pi(x_b-x_a)\right)^2 \prod_{a=1}^{N_1} e^{i\pi k x_a^2} dx_a \\ \times \int_{\mathbb{R}^{N_2}} \mathfrak{s}_\mu(e^{-2\pi y})\mathfrak{s}_\nu(e^{2\pi y}) \prod_{1\le \dot a<\dot b\le N_2} \left(2\sinh \pi(y_{\dot b}-y_{\dot a})\right)^2 \prod_{\dot a=1}^{N_2} e^{-i\pi k y_{\dot a}^2} dy_{\dot a}.\end{aligned}$$

We find that the integrals are factorized into vevs of Wilson loops in pure Chern–Simons theory [107] at each node:

$$\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)}(k,m) \overset{\text{pert.}}{=} e^{\pi(N_1+N_2)(m_+-m_-)} \sum_\mu \sum_\nu (-e^{2\pi m_+})^{|\mu|}(-e^{2\pi m_-})^{-|\nu|} \langle W_{\mu\nu}\rangle_{N_1;k} \langle W_{\nu\mu}\rangle_{N_2;-k}$$

where the equality is understood order by order in the Laurent expansion in the parameters $t_\pm = -e^{2\pi m_\pm}$, and the two vevs compute Hopf link invariants respectively in $U(N_1)$ and $U(N_2)$ pure Chern–Simons theory on $\mathbb{S}^3$ with renormalized levels k and $-k$. Note also how the roles of the two representations μ, ν are swapped between the two nodes. The restriction (2.5.10), which arises here from an elementary property of the symmetric polynomials, matches with the analysis of the quiver variety of $\widehat{A}_1$, which only includes $U(N)$ representations for $N = \min\{N_1, N_2\}$.

The result does not rely on the specific choice of Chern–Simons levels and immediately extends to generic (k_1, k_2).

Schur expansion: necklace quivers

ABJ theory belongs to the class of extended $\widehat{A}_r$ quivers. We now show how the Schur expansion holds for the whole $\widehat{A}_r$ family of theories, with mass deformation and without any additional matter content beyond the bi-fundamental hypermultiplets linking the gauge nodes, as depicted in quiver notation in Figure 2.18. These $\mathcal{N} = 3$ Chern–Simons theories have been constructed in [75, 76]. The result we find is a series expansion in the parameters $t_p = -e^{2\pi m_p}$, with the coefficients being vevs of a Wilson loop in pure Chern–Simons theory with gauge group $U(N_p)$ and level k_p.

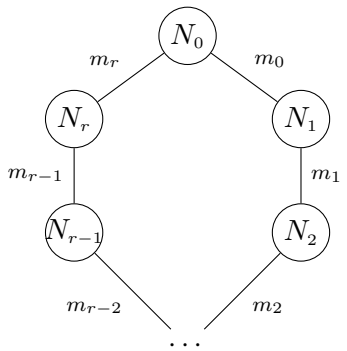


Figure 2.18. Mass-deformed non-Abelian $\widehat{A}_r$ extended quiver.

The partition function of the theory is:

$$\mathcal{Z}_{\widehat{A}_r}(\vec{k},\vec{m}) = \int_{\mathbb{R}^{N_0}} d^{N_0}\vec{x}_0 \int_{\mathbb{R}^{N_1}} d^{N_1}\vec{x}_1 \cdots \int_{\mathbb{R}^{N_r}} d^{N_r}\vec{x}_r \prod_{p=0}^{r} e^{i\pi k_p \sum_{a=1}^{N_p} x_{p,a}^2} \times \prod_{p=0}^{r} \frac{\prod_{1\le a<b\le N_p} \left(2\sinh\pi(x_{p,b}-x_{p,a})\right)^2}{\prod_{a=1}^{N_p}\prod_{\dot{b}=1}^{N_{p+1}} 2\cosh\pi(x_{p+1,\dot{b}}-x_{p,a}+m_p)},$$

with periodic identification of the labels, $r+1 \equiv 0$. At the level of the matrix model, the eigenvalues associated to each gauge node interact among themselves as in pure $U(N_p)_{k_p}$ Chern–Simons theory, and also interact with the nearest neighbours through the denominator.

We apply the Cauchy identity (2.2.10) at each edge of the quiver in Figure 2.18, expanding in the fugacities associated to the masses of the bi-fundamental hypermultiplets. We obtain the expressions

$$\begin{aligned}\frac{1}{\prod_{a=1}^{N_p}\prod_{\dot{b}=1}^{N_{p+1}} 2\cosh\pi(x_{p,a}-x_{p+1,\dot{b}}+m_p)} &= \prod_{a=1}^{N_p}\prod_{\dot{b}=1}^{N_{p+1}} \frac{e^{\pi(x_{p,a}+m_p)-\pi x_{p+1,\dot{b}}}}{1+e^{2\pi(x_{p,a}+m_p)}e^{-2\pi x_{p+1,\dot{b}}}} \\ &= e^{\pi(N_{p+1}+N_p)m_p - \pi N_{p+1}\sum_{a=1}^{N_p} x_{p,a} - \pi N_p \sum_{\dot{b}=1}^{N_{p+1}} x_{p+1,\dot{b}}} \\ &\times \sum_{\nu^{(p)}} (-1)^{|\nu^{(p)}|} \mathfrak{s}_{\nu^{(p)}}(e^{2\pi(x_p+m_p)}) \mathfrak{s}_{\nu^{(p)}}(e^{-2\pi x_{p+1}}),\end{aligned}$$

where the sum runs over partitions $\nu^{(p)}$ satisfying

$$\text{length}(\nu^{(p)}) \le \min(N_p, N_{p+1}). \tag{2.5.11}$$

Each set of variables $e^{2\pi x_p}$ appears with plus sign in the exponent in the Schur $\mathfrak{s}_{\nu^{(p)}}$ and with minus sign in $\mathfrak{s}_{\nu^{(p-1)}}$. Besides, as above, we have written $\mathfrak{s}_{\nu^{(p)}}(e^{2\pi x_p})$ as a shorthand for the Schur polynomial in the N_p variables $(e^{2\pi x_{p,1}},\dots,e^{2\pi x_{p,N_p}})$. Putting all such contributions together we get

$$\begin{aligned}\mathcal{Z}_{\widehat{A}_r}(\vec{k},\vec{m}) &\overset{\text{pert.}}{=} \int_{\mathbb{R}^{N_0}}\int_{\mathbb{R}^{N_1}}\cdots\int_{\mathbb{R}^{N_r}} \prod_{p=0}^{r}\prod_{a=1}^{N_p} e^{\left[i\pi k_p x_{p,a}^2 + \pi(N_{p+1}-N_{p-1})x_{p,a}\right]}\, dx_{p,a} \\ &\times \prod_{p=0}^{r} \prod_{1\le a<b\le N_p} \left(2\sinh\pi(x_{p,b}-x_{p,a})\right)^2 \\ &\times e^{\pi\sum_{p=0}^{r}(-1)^p m_p(N_{p+1}+N_p)} \sum_{\vec{\nu}} (-1)^{|\vec{\nu}|} \prod_{p=0}^{r} e^{(-1)^p 2\pi m_p |\nu^{(p)}|}\, \mathfrak{s}_{\nu^{(p-1)}}(e^{-2\pi x_p}) \mathfrak{s}_{\nu^{(p)}}(e^{2\pi x_p}),\end{aligned}$$

with the sum running over $(r+1)$-tuples of partitions

$$\vec{\nu} = (\nu^{(0)},\dots,\nu^{(r)}),$$

with all partitions $\nu^{(p)}$ constrained according to (2.5.11). The integrals are now suitably factorized in each summand. Completing the squares in the Gaussian term at each node and comparing with [107], we obtain

$$\begin{aligned}\frac{\mathcal{Z}_{\widehat{A}_r}(\vec{k},\vec{m})}{\prod_{p=0}^{r}\mathcal{Z}_{\text{CS}(N_p);k_p}} &\overset{\text{pert.}}{=} e^{\pi\sum_{p=0}^{r}\left[m_p(N_{p+1}+N_p)+\frac{i}{2k_p}N_p(N_{p+1}-N_{p-1})^2\right]} \\ &\times \sum_{\vec{\nu}} \prod_{p=0}^{r} t_p^{|\nu^{(p)}|}\, e^{\frac{i}{2k_p}(N_{p+1}-N_{p-1})\left(|\nu^{(p)}|-|\nu^{(p-1)}|\right)} \left\langle W_{\nu^{(p)}\nu^{(p-1)}}\right\rangle_{N_p;k_p},\end{aligned} \tag{2.5.12}$$

where the average in each summand is the vev of a Wilson loop in $U(N_p)_{k_p}$ Chern–Simons theory, computing the Hopf link invariant in the representations $(\nu^{(p)}, \nu^{(p-1)})$. As always, the two sides of the equality are understood as formal series expansions in the parameters $t_p = -e^{2\pi m_p}$. These global symmetry fugacities serve as book-keeping variables in the expansion, while all other variables are integrated. Furthermore, if we think of each $\langle W_{\bullet\bullet}\rangle_{N_p;k_p}$ as a ring homomorphism from the ring of $U(N_p)$ representations to $\mathbb{C}[q_p, q_p^{-1}]$, we notice the emergence of a trace of the product of $r+1$ such maps as a direct consequence of the quiver being necklace-shaped. This trace is taken on the ring of $U(N_{\max})$ representations, with $N_{\max} = \max_p N_p$ and $\langle W_{\mu\nu}\rangle_{N_p;k_p}$ understood to vanish if either μ or ν is not a $U(N_p)$ representation. The trace structure appears more clearly when $N_p = N$ and $k_p = \pm k$ for all $p = 0, 1, \dots, r$.

From the properties of pure Chern–Simons theory and its relation with the level k WZW model [42], only *integrable* representations contribute to each Hopf link invariant. This introduces an effective "mod k_p" periodicity [89] of the coefficients of each t_p.

Schur expansion: the M-crystal model

A simple yet interesting example of the above setting corresponds to the Abelian model with alternating ± 1 Chern–Simons levels, $\vec{k} = (+1, -1, \dots, -1)$. This quiver gauge theory describes the M-crystal model [104, 105], see for example Figure 2.19. Specializing the computations above and after a few simplifications we get

$$\frac{\mathcal{Z}_{\widehat{A}_r}(\pm 1, \vec{m})}{\prod_{p=0}^{r} \mathcal{Z}_{\mathrm{CS}(1);(-1)^p}} \overset{\text{pert.}}{=} e^{2\pi|\vec{m}|} \sum_{\vec{\nu}\in\mathbb{Z}_{\geq 0}^{r+1}} \prod_{p=0}^{r} \left(-e^{2\pi m_p \nu^{(p)} + i2\pi(-1)^p \nu^{(p-1)}\nu^{(p)}}\right).$$

The series is clearly not convergent, but this was expected as the right-hand side has the meaning of an algebraic expansion in multiple variables. In conclusion, a perturbative expansion of the partition function of the M-crystal model has all the terms $t_0^{\nu^{(0)}} \cdots t_r^{\nu^{(r)}}$ with coefficient 1.

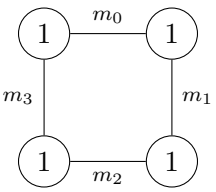


Figure 2.19. Mass-deformed $\widehat{A}_3$ extended quiver. For $\vec{k} = (1, -1, 1, -1)$ the associated Chern–Simons theory is the gauge theoretical realization of the M-crystal model with four vertices.

A simple generalization of the above formula to the Abelian necklace quiver with arbitrary $\vec{k}$ gives

$$\frac{\mathcal{Z}_{\widehat{A}_r}(\vec{k}, \vec{m})}{\prod_{p=0}^{r} \mathcal{Z}_{\mathrm{CS}(1);k_p}} \overset{\text{pert.}}{=} (-1)^r \left(\prod_{p=0}^{r} t_p\right) \sum_{\vec{\nu}\in\mathbb{Z}_{\geq 0}^{r+1}} \prod_{p=0}^{r} t_p^{\nu^{(p)}} e^{-i\pi\nu^{(p)}\left(\frac{1}{k_p}+\frac{1}{k_{p-1}}\right)+i\frac{2\pi}{k_p}\nu^{(p)}\nu^{(p-1)}}. \tag{2.5.13}$$

When $r = 1$ and $k_2 = -k_1$ we get the Abelian ABJM theory, exactly solved in Subsection 2.3.4. As a consistency check, we expand the geometric series

$$\sum_{\alpha=0}^{k-1} \frac{t_j^\alpha}{t_j^k - 1} = -\sum_{\alpha=0}^{k-1}\sum_{\beta=0}^{\infty} t_j^{\alpha+\beta k} = -\sum_{\alpha=0}^{\infty} t_j^\alpha$$

in the answer from Subsection 2.3.4, and confirm that the Schur expansion reproduces the correct coefficients to all orders in t_1, t_2, although, as expected, it misses all the terms proportional to $e^{i\pi m_j^2}$.

2.5.4 Schur expansion for Wilson loops

It is possible to combine the ideas used in this section with those of Section 2.4 to study Wilson loops.

We come back to the setting of Section 2.4 and consider the vev of a single $\frac{1}{2}$-BPS Wilson loop in ABJ theory, with ranks N_1 and N_2. We assume the Wilson loop carries a typical representation μ of the supergroup $U(N_1|N_2)$ [73]. We write

$$\begin{aligned}\frac{\mathfrak{s}_\mu(e^{2\pi x}|e^{2\pi y})}{\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2}(2\cosh\pi(x_a-y_{\dot a}))^2} &= \mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_{\eta'}(e^{2\pi y})\prod_{a=1}^{N_1}\prod_{\dot a=1}^{N_2}\frac{e^{2\pi x_a+2\pi y_{\dot a}}}{e^{2\pi x_a}+e^{2\pi y_{\dot a}}}\\ &= \mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_{\eta'}(e^{2\pi y})\left(\prod_{a=1}^{N_1}e^{2\pi x_a}\right)\sum_\nu(-1)^{|\nu|}\mathfrak{s}_\nu(e^{2\pi x})\mathfrak{s}_\nu(e^{-2\pi y}),\end{aligned}$$

where the first equality follows from the factorization property (2.4.1), while to pass from the first to the second line we have used the Cauchy identity (2.2.10) and brought out the factor (-1) from $\mathfrak{s}_\nu(-e^{-2\pi y})$. The sum over ν runs over all partitions

$$\{\nu\ :\ \text{length}(\nu)\le\min(N_1,N_2)\}\,.$$

It is important to stress the difference between the results we present in this subsection and the ones we have found in Section 2.4. There, the correlator of two or more Wilson loops in ABJ has been taken into account, and the factorization of the final result into vevs of Wilson loops in Chern–Simons theories without matter is exact. Here, instead, we consider a single Wilson loop in ABJ, and we use the Cauchy identity to expand the interaction between the two nodes. In turn, the latter is the Schur expansion of the A_2 quiver of Subsection 2.5.2.

With the Schur expansion, the expectation value of the Wilson loop decomposes into a sum of contributions, indexed by the partition ν, factorized into two multiple integrals, one for each node:

$$\langle W_\mu\rangle_{N_1,N_2;k}\overset{\text{pert.}}{=}\frac{1}{\mathcal{Z}_{\text{ABJ}}}\sum_\nu(-1)^{|\nu|}\mathcal{W}^{(1)}_{\gamma\nu}\mathcal{W}^{(2)}_{\nu\eta'},\tag{2.5.14}$$

$$\begin{aligned}\mathcal{W}^{(1)}_{\gamma\nu}&:=\int_{\mathbb{R}^{N_1}}\prod_{1\le a<b\le N_1}(2\sinh\pi(x_b-x_a))^2\ \ \mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_\nu(e^{2\pi x})\prod_{a=1}^{N_1}e^{i\pi kx_a^2+2\pi x_a}\,dx_a,\\ \mathcal{W}^{(2)}_{\nu\eta'}&:=\int_{\mathbb{R}^{N_2}}\prod_{1\le\dot a<\dot b\le N_1}\left(2\sinh\pi(y_{\dot b}-y_{\dot a})\right)^2\ \ \mathfrak{s}_\nu(e^{2\pi y})\mathfrak{s}_{\eta'}(e^{-2\pi y})\prod_{\dot a=1}^{N_2}e^{-i\pi ky_{\dot a}^2}\,dy_{\dot a}.\end{aligned}$$

The first function corresponds to integration over the Cartan subalgebra of $\mathfrak{u}(N_1)$ and the second to integration over the Cartan subalgebra of $\mathfrak{u}(N_2)$. In the integral over the second node, we have reflected variables $y_{\dot a}\mapsto -y_{\dot a}$. The term $\prod_a e^{2\pi x_a}$ in $\mathcal{W}^{(1)}_{\gamma\nu}$ can be removed with a shift of variables and translating back the integration cycle onto $\mathbb{R}$, obtaining

$$\mathcal{W}^{(1)}_{\gamma\nu}=e^{\frac{i\pi}{k}(N_1+2|\gamma|+2|\nu|)}\int_{\mathbb{R}^{N_1}}\mathfrak{s}_\gamma(e^{2\pi x})\mathfrak{s}_\nu(e^{2\pi x})\prod_{1\le a<b\le N_1}(2\sinh\pi(x_b-x_a))^2\prod_{a=1}^{N_1}e^{i\pi kx_a^2}dx_a.$$

Rectangular partition

The simplest case is the expectation value of a Wilson loop in a rectangular representation $\mu = \kappa$, so $\gamma = \emptyset = \eta$. We get:

$$\begin{aligned}\mathcal{W}^{(1)}_{\emptyset\nu} &= e^{\frac{i\pi}{k}(N_1+2|\nu|)} \int_{\mathbb{R}^{N_1}} \mathfrak{s}_\nu(e^{2\pi x}) \prod_{1\le a<b\le N_1} (2\sinh\pi(x_b-x_a))^2 \prod_{a=1}^{N_1} e^{i\pi k x_a^2} dx_a \\ &= \mathcal{Z}_{N_1;k}\; q^{-\frac{\mathsf{C}_{2;N_1}(\nu)}{2}-\frac{N_1}{2}-|\nu|} \dim_q \nu\end{aligned}$$

and

$$\begin{aligned}\mathcal{W}^{(2)}_{\nu\emptyset} &= \int_{\mathbb{R}^{N_2}} d^{N_2}y\; \mathfrak{s}_\nu(e^{2\pi x})\; e^{-\sum_{\dot a=1}^{N_2} i\pi k y_{\dot a}^2} \prod_{1\le \dot a<\dot b\le N_2} \big(2\sinh\pi(y_{\dot b}-y_{\dot a})\big)^2 \\ &= \mathcal{Z}_{N_2;-k}\; q^{\frac{\mathsf{C}_{2;N_2}(\nu)}{2}} \dim_{q^{-1}} \nu\end{aligned}$$

In both evaluations, the second line follows from the Wilson loop vev (2.2.1) and $\mathcal{Z}_{N_1,k}$ and $\mathcal{Z}_{N_2,-k}$ are the corresponding normalizations. Noting that $\dim_{q^{-1}}\nu = \dim_q \nu$, the vev of the Wilson loop in a rectangular representation κ of the supergroup $U(N_1|N_2)$ is then

$$\langle W_\kappa\rangle_{N_1,N_2;k} \overset{\text{pert.}}{=} \frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}} q^{-\frac{N}{2}} \sum_\nu (-q)^{-|\nu|} (\dim_q\nu)^2\, q^{\frac{\mathsf{C}_{2;N_2}(\nu)-\mathsf{C}_{2;N_1}(\nu)}{2}}.$$

In particular, for ABJM theory, $N_1 = N = N_2$, the quadratic Casimir cancels and we arrive at the simpler formula

$$\langle W_\kappa\rangle_{N,N;k} \overset{\text{pert.}}{=} \frac{\mathcal{Z}_{N;k}\mathcal{Z}_{N;-k}}{\mathcal{Z}_{\mathrm{ABJM}(N)_k}} q^{-\frac{N}{2}} \sum_\nu (-q)^{-|\nu|} (\dim_q\nu)^2.$$

Arbitrary typical representation

We now tackle the general case of a typical (long) but otherwise arbitrary representation μ, and give two equivalent, and in fact related, evaluations of the vev of the Wilson loop in ABJ theory.

Both approaches require to invert the variables in a Schur polynomial, which can be done using the identity (2.4.5).

The first procedure mimics [107], and extends the result to the unknot Wilson loop. Inverting variables in $\mathcal{W}^{(1)}_{\gamma\nu}$ using (2.4.5) we identify $\mathcal{W}^{(1)}_{\gamma\nu}$ and $\mathcal{W}^{(2)}_{\nu\eta'}$ with Hopf link invariants computed in $U(N_1)_k$ and $U(N_2)_{-k}$ Chern–Simons theory on $\mathbb{S}^3$, respectively. Explicitly:

$$\langle W_\mu\rangle_{N_1,N_2;k} = \frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}} \sum_\nu C_{\gamma\nu}(q)\; \langle W_{\gamma\nu^*}\rangle_{N_1;k}\langle W_{\nu\eta'}\rangle_{N_2;-k}, \tag{2.5.15}$$

where the averages in the sum are the Hopf link invariants and the summands are weighted by

$$C_{\gamma\nu}(q) = (-1)^{|\nu|} q^{-(1+\nu_1)\left(|\gamma|+|\nu|+\frac{1+\nu_1}{2}N_1\right)}. \tag{2.5.16}$$

The partition ν^* has been defined in (2.4.6). In the operator formalism the expansion (2.5.15) takes the form

$$\langle 0|TST|\mu\rangle_{N_1,N_2;k} = \frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}} \sum_\nu C_{\gamma\nu}(q)\; \langle\eta|TST|\nu\rangle_{N_1,k}\langle\nu^*|TST|\gamma\rangle_{N_2,-k},$$

where T, S are the $SL(2,\mathbb{Z})$ modular matrices. Note that the two ν's are treated differently: one is considered as a $U(N_1)$ representation and the other as a $U(N_2)$ representation, with the latter twisted by the starred partition. The appearance of the operator TST rather than S is because the matrix model presentation computes the observables in a special instance of the Seifert framing, rather than in the natural $\mathbb{S}^3$ framing.

The alternative path consists in applying the inversion formula (2.4.5) to $\mathfrak{s}_\nu(e^{2\pi y})$. Similar manipulations lead to:

$$\langle W_\mu\rangle_{N_1,N_2;k} = \frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}}\, q^{-\frac{N_1}{2}-|\gamma|}\sum_\nu \tilde{C}_{\nu\eta'}(q)\langle W_\gamma W_\nu\rangle_{N_1;k}\langle W_{\nu^*}W_{\eta'}\rangle_{N_2;-k} \tag{2.5.17}$$

with coefficient

$$\widetilde{C}_{\nu\eta'}(q) = (-q)^{|\nu|}q^{-\nu_1\left(|\eta'|+|\nu|-\frac{\nu_1}{2}N_2\right)}.$$

The expression (2.5.17) is expressed as a sum of correlators of two pairs of (unlinked) unknots, one pair in each Chern–Simons theory. These correlators can be further reduced with a character expansion in the Schur basis:

$$\langle W_\mu\rangle_{N_1,N_2;k} = \frac{\mathcal{Z}_{N_1;k}\mathcal{Z}_{N_2;-k}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_k}}\, q^{-\frac{N_1}{2}-|\gamma|}\sum_\nu \widetilde{C}_{\nu\eta'}(q)\sum_{\tilde\nu,\hat\nu}\mathsf{N}_{\gamma\nu}{}^{\tilde\nu}\langle W_{\tilde\nu}\rangle_{N_1;k}\mathsf{N}_{\eta'\nu^*}{}^{\hat\nu}\langle W_{\hat\nu}\rangle_{N_2;-k},$$

where, as above, $\mathsf{N}_{\gamma\nu}{}^{\tilde\nu}$ are the Littlewood–Richardson coefficients.

The solvability is again preserved if we turn on a Romans mass in the dual theory as prescribed in [71]. The above computation is straightforwardly generalized to $k_2 \neq -k_1$ and gives:

$$\langle W_\mu\rangle_{N_1,N_2;k_1,k_2} = \frac{\mathcal{Z}_{N_1;k_1}\mathcal{Z}_{N_2;k_2}}{\mathcal{Z}_{\mathrm{ABJ}(N_1|N_2)_{k_1,k_2}}}\sum_\nu C_{\gamma\nu}(q_1)\,\langle W_{\gamma\nu^*}\rangle_{N_1;k_1}\langle W_{\nu\eta'}\rangle_{N_2;k_2}$$

with summands weighted by (2.5.16) with $q = q_1 = e^{-\frac{i2\pi}{k_1}}$, hence independent of k_2.

Our derivation complements previous results [115, 116] extending the analysis to a broader class of representations.

2.5.5 Comments on the $U(N)$ theory with N_f fundamental hypermultiplets

As a final observation, and departing from the previous use of Schur expansions, we discuss further the partition function $\mathcal{Z}_{U(N),N_f}$. The expression (2.5.8) appears in topological string theory [112] in the study of non-compact branes on the resolved conifold. There, the fugacities t_j correspond to diagonal holonomies of the gauge fields along a circle $\mathbb{S}^1$, determined as the locus where a non-compact brane intersects $\mathbb{S}^3$. Replacing a brane with an anti-brane in the framework of [112] corresponds here to exchange the Cauchy identity (2.2.10) with the dual Cauchy identity (2.2.12), which describes the Schur expansion of the matrix model

$$\mathcal{Z}^{\mathrm{ferm.}}_{U(N),N_f} = \int_{\mathbb{R}^N}\prod_{1\le a<b\le N}(2\sinh\pi(x_b-x_a))^2\prod_{a=1}^N\left[\prod_{j=1}^{N_f}2\cosh\pi(x_a+m_j)\right]e^{-\frac{1}{2g}(2\pi x_a)^2}\,dx_a.$$

The choice of notation "ferm." for this matrix model will be justified momentarily.

In the Abelian theory, in particular, replacing a brane with an anti-brane [112] switches from the generating function of the complete homogeneous symmetric polynomials to that of the elementary symmetric polynomials $\mathfrak{e}_\alpha$ [90], cf. (2.2.13).

We compare now $\mathcal{Z}_{U(N),N_f}$ with $\mathcal{Z}^{\mathrm{ferm.}}_{U(N),N_f}$. The former is the partition function of $U(N)$ Chern–Simons theory at level k on $\mathbb{S}^3$ coupled to N_f fundamental hypermultiplets. We have

introduced the latter to mimic the pair of identities (2.2.10)-(2.2.12) at the level of matrix integrals. Nevertheless, there are several physical motivations to study both $\mathcal{Z}_{U(N),N_f}$ and $\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}$.

As we have already mentioned, in topological string theory on the conifold it is important to have both functions [112, 117]. Moreover, the matrix model $\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}$ with all the masses vanishing, has been studied in [118] in the context of fermionic quantum mechanics, and solved in [119] for any $\{m_j\} \subset \mathbb{R}^{N_f}$. A third motivation for the introduction of the "fermionic" partition function comes from looking at each summand in the Schur expansions. Consider a fixed ν in the sums over representations which corresponds to a symmetric $SU(N)$ representation. The associated reduced coloured knot invariants have been categorified in [120, 121]. The corresponding homologies posses a mirror symmetry which exchanges the symmetric representation ν with the totally antisymmetric representation ν'. Generalizing this operation to the present context, replacing one representation by its conjugate, switches from the Cauchy identity (2.2.10), to the dual Cauchy identity (2.2.12).

A fourth, heuristic argument to consider the pair $\mathcal{Z}_{U(N),N_f}$ $\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}$ is presented below.

Averages of characteristic polynomials

It has been shown in [95] that $\mathcal{Z}_{U(N),N_f}$ computes the average of the inverse of the product of characteristic polynomials in the Stieltjes–Wigert random matrix ensemble, which describes the Chern–Simons matrix model [84]. Explicitly:

$$\frac{\mathcal{Z}_{U(N),N_f}}{\mathcal{Z}_{\text{CS}(N)}} \propto \left\langle \left[\prod_{j=1}^{N_f} \det\left(\widetilde{t}_j^{\vee} - \mathbf{X}\right) \right]^{-1} \right\rangle_{\text{SW}(N)},$$

with $\mathbf{X}$ a random Hermitian matrix whose eigenvalues are $(x_1, \ldots, x_N)$. The spectral parameters $\widetilde{t}_j^{\vee}$ are related to the physical quantities through

$$\widetilde{t}_j^{\vee} = -q^{-N-\frac{N_f}{2}} e^{-2\pi m_j} = q^{-N-\frac{N_f}{2}} t_j^{-1},$$

with $q = e^{-g}$.

The average of the inverse product of characteristic polynomials in the Hermitian random matrix ensemble with Stieltjes–Wigert weight is calculated exactly, and is a $N_f \times N_f$ determinant:

$$\frac{\mathcal{Z}_{U(N),N_f}}{\mathcal{Z}_{\text{CS}(N)}} = \frac{c_{N,N_f}}{\prod_{1\le j<l\le N_f}\left(\widetilde{t}_j^{\vee} - \widetilde{t}_l^{\vee}\right)} \det_{1\le j,l\le N_f} \left[\mathfrak{p}^{\vee}_{N+l-1}\left(\widetilde{t}_j^{\vee}\right)\right], \tag{2.5.18}$$

where $\mathfrak{p}^{\vee}_n(\widetilde{t}^{\vee})$ are the Cauchy transform of the Stieltjes–Wigert orthogonal polynomials, and the constant c_{N,N_f} in (2.5.18) does not depend on the spectral parameters $\widetilde{t}_j^{\vee}$. We refer to [95] for more details, proofs and references.

We obtain the analogous expression for the other matrix integral considered, in terms of averages of products of characteristic polynomials in the Stieltjes–Wigert ensemble [119]:

$$\frac{\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}}{\mathcal{Z}_{\text{CS}(N)}} \propto \left\langle \prod_{j=1}^{N_f} \det\left(\widetilde{t}_j - \mathbf{X}\right) \right\rangle_{\text{SW}(N)},$$

Here, the spectral parameters $\widetilde{t}_j$ are related to the parameters of the gauge theory as

$$\widetilde{t}_j = -q^{-N+\frac{N_f}{2}} e^{-2\pi m_j} = q^{-N}\left(q^{-\frac{N_f}{2}} t_j\right)^{-1}.$$

The average of the product of characteristic polynomials is explicitly given by a $N_f \times N_f$ determinant

$$\frac{\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}}{\mathcal{Z}_{\text{CS}(N)}} = \frac{c^{\text{ferm.}}_{N,N_f}}{\prod_{1\leq j<l\leq N_f}(\widetilde{t}_j - \widetilde{t}_l)} \det_{1\leq j,l\leq N_f} \left[\mathfrak{p}_{N+l-1}\left(\widetilde{t}_j\right)\right],$$

where $\mathfrak{p}_n(\widetilde{t})$ are the Stieltjes–Wigert polynomials, and $c^{\text{ferm.}}_{N,N_f}$ is a numerical constant. We refer to [119] for details and a detailed list of references. A closely related result was obtained in [117], in the context of topological string theory on the conifold.

Bosonic versus fermionic matrix models

We can recast the two expressions in a unified formalism, integrating over auxiliary variables:

$$\mathcal{Z}^{\epsilon}_{U(N),N_f} = \int d\mathbf{X}\; e^{-\frac{1}{2g}\text{Tr}(\log \mathbf{X})^2} \prod_{j=1}^{N_f} \int e^{-\bar{\psi}^j\left(\mathbf{X}-\widetilde{t}^{\epsilon}_j\right)\psi^j} \prod_{a=1}^{N} \frac{d\bar{\psi}^j_a d\psi^j_a}{2\pi}. \tag{2.5.19}$$

In this expression, $\epsilon \in \{\pm 1\}$, with $\epsilon = -1$ giving $\mathcal{Z}_{U(N),N_f}$ and $\epsilon = +1$ giving $\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}$. The spectral parameters are respectively $\widetilde{t}^{\vee}_j$ and $\widetilde{t}_j$ for $\epsilon = -1, +1$. The integration is over N_f N-component vectors $\psi^j = (\psi^j_a)_{a=1,\ldots,N}$, for $j = 1, \ldots, N_f$, and their conjugates $\bar{\psi}^j = (\bar{\psi}^j_a)_{a=1,\ldots,N}$. These vectors have Grassmann-even entries when $\epsilon = -1$ and Grassmann-odd entries when $\epsilon = +1$. We recall that $\widetilde{t}^{\vee}_j, \widetilde{t}_j < 0$ from their definition in terms of the physical variables, which guarantees that the inner integral in (2.5.19) is well posed. Besides, we have dropped an overall constant.

Written in the form (2.5.19), we see that switching from the Cauchy identity (2.2.10) to the dual Cauchy identity (2.2.12) passes from the Schur expansion of the matrix model (2.5.19) with bosonic fields to the Schur expansion of (2.5.19) with fermionic fields.[5]

The ideas of the present subsection can be applied to 4d $\mathcal{N} = 4$ SYM with codimension-one matter defects sitting on a great $\mathbb{S}^3$ inside $\mathbb{S}^4$. The analogue of (2.5.19) is

$$\mathcal{Z}^{\text{4d+defect},\epsilon}_{U(N),N_f} = \int d\mathbf{X} \int \frac{d\bar{\psi}d\psi}{(2\pi)^N}\; e^{-\text{Tr}\left[\frac{8\pi^2}{g_{\text{4d}}}\mathbf{X}^2 + \sum_{j=1}\bar{\psi}^j\left(e^{\mathbf{X}}-\widetilde{t}^{\epsilon}_j\right)\psi^j\right]}. \tag{2.5.20}$$

Here $\epsilon = -1$ corresponds to the physical 4d $\mathcal{N} = 4$ theory and $\epsilon = +1$ is its counterpart using a fermionic matrix representation. Compared to the purely three-dimensional theory, we have removed the $\log^2$-interaction, at the cost of an exponential term in the action. Rescaling $\mathbf{X}_{ab} \to \sqrt{g_{\text{4d}}/16\pi^2}\mathbf{X}_{ab}$ we can expand the interaction term in (2.5.20) in a power series in

$$g_{\text{eff}} = \frac{\sqrt{g_{\text{4d}}}}{4\pi}.$$

The resulting effective action includes infinitely many vertices:

$$\bar{\psi}^j\left(e^{\mathbf{X}} - \widetilde{t}^{\epsilon}_j\right)\psi^j = \bar{\psi}^j_a\left(1 - \widetilde{t}^{\epsilon}_j\right)\delta_{ab}\psi^j_b \;+\; g_{\text{eff}}\bar{\psi}^j_a X_{ab}\psi^j_b \;+\; \frac{g^2_{\text{eff}}}{2}\bar{\psi}^j_a X_{ac}X_{cb}\psi^j_b + \ldots$$

and can be analyzed by standard perturbative techniques in random matrix theory [122].

The upshot of this digression is that we may as well describe four-dimensional $\mathcal{N} = 4$ SYM with defects using random matrix theory, and equivalently represent it as a theory of massive scalars $\bar{\psi}^j, \psi^j$ in the vector representation of $U(N)$ which interact with a (zero-dimensional) gluon $\mathbf{X}$ in the adjoint representation of $U(N)$. We also naturally get an associated theory in which the bosons are replaced by zero-dimensional fermions.

[5] The suggestive form (2.5.19) does not seem to allow a unified treatment of fermionic and bosonic versions of the quantum mechanical model of [118], because the matrix model representation $\mathcal{Z}^{\text{ferm.}}_{U(N),N_f}$ has been originally derived using a Wick rotation that is forbidden in the bosonic counterpart.

2.A Mordell integrals at $\lambda \neq 0$

In this appendix we explain some subtlety related to the integrals (2.2.3)-(2.2.4). We mainly review results from [61] and comment on how to properly combine them.

In his seminal paper, Mordell gave the formulas [61, Eq.s (8.1)-(8.2)]

$$\Psi_+(\xi) := \int_{-\infty}^{+\infty} dx \frac{e^{i\pi\frac{\kappa}{\varrho}x^2-2\pi x\xi}}{e^{2\pi x}-1} = \frac{1}{e^{i\pi\varrho(2\xi-\kappa)}-1}\left[-\sqrt{\frac{i\varrho}{\kappa}}\sum_{\alpha=1}^{\kappa} e^{i\pi\frac{\varrho}{\kappa}(\xi+\alpha)^2} + i\sum_{\beta=1}^{\varrho} e^{i\pi\beta\left(2\xi-\frac{\kappa}{\varrho}\beta\right)}\right], \tag{2.A.1}$$

$$\Psi_-(\xi) := \int_{-\infty}^{+\infty} dx \frac{e^{-i\pi\frac{\kappa}{\varrho}x^2-2\pi x\xi}}{e^{2\pi x}-1} = \frac{1}{e^{i\pi\varrho(2\xi-\kappa)}-1}\left[\sqrt{-\frac{i\varrho}{\kappa}}\sum_{\alpha=0}^{\kappa-1} e^{-i\pi\frac{\varrho}{\kappa}(\xi-\alpha)^2} + i\sum_{\beta=1}^{\varrho} e^{i\pi\beta\left(2\xi+\frac{\kappa}{\varrho}\beta\right)}\right], \tag{2.A.2}$$

valid for $\kappa, \varrho \in \mathbb{Z}_{>0}$. The integration contour can be taken either along the real axis avoiding $x = 0$ by a small semicircle, or on a straight line inclined with respect to the real axis and intersecting the imaginary axis between 0 and $-i$. The inclination should be a negative angle for Ψ_+ and a positive angle for Ψ_-. We follow this latter choice, and represent the inclined straight line in Figure 2.20. The final result is independent of the angle θ between the integration axis and the real axis.

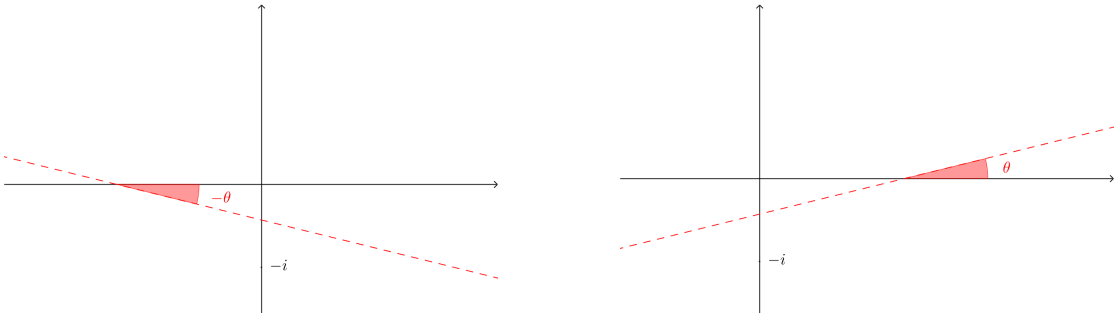


Figure 2.20. Choice of integration contour for the Mordell integrals, shifted and rotated by a small angle with respect to the real axis. Left: contour for Ψ_+, rotated by a negative angle $-\theta < 0$. Right: contour for Ψ_-, rotated by a positive angle $\theta > 0$.

On the other hand, the integral

$$\tilde{\Psi}(\lambda,\xi) := \int_{\mathbb{R}-i\lambda} dx \frac{e^{i\pi\tilde{\kappa}x^2-2\pi x(\xi+\tilde{k}\lambda)}}{e^{2\pi x}-1}$$

with

$$\Im(\tilde{\kappa}) > 0, \qquad 0 < \Re\lambda < 1$$

is equivalent to [61, Eq. (3.8)]

$$\tilde{\Psi}(\lambda,\xi) = e^{i\pi\lambda(2+2\xi+\tilde{\kappa}\lambda)} \int_{-\infty}^{+\infty} dx \frac{e^{i\pi\tilde{\kappa}x^2-2\pi x\xi}}{e^{2\pi x}-e^{i2\pi\lambda}}$$

now with the integration cycle along the real axis. The proof of this formula [61] makes explicit that one can move the original integration contour in the region $0 < \Re\lambda < 1$ without changing the result. The same is true for the contour of $\Psi_\pm$, where we are free to chose where to intersect the imaginary axis. We can therefore introduce the parameter λ also in $\Psi_\pm$, obtaining the integrals

defined in (2.2.3)-(2.2.4). However, in order not to get out of the proper region, we must impose θ-dependent restrictions on λ, as can be seen from Figure 2.21. So the formula (2.2.3) for Ψ_+ hold for

$$r\sin\theta < \Re\lambda < 1 + r\sin\theta, \qquad -r\cos\theta < \Im\lambda < r\cos\theta$$

where $r \geq 0$ is arbitrary, and similarly for Ψ_-. Rotating $\theta \to 0^+$ we recover the constraints $0 < \Re\lambda < 1$ and $\Im\lambda$ arbitrary for $\tilde{\Psi}$ defined above. In particular we can fix $0 < \Re\lambda < 1$ and $0 < \theta \ll \frac{\pi}{2}$, so that we are free to chose $\Im\lambda$ arbitrarily large. With this choice, we change variables in (2.A.1) and recover (2.2.3), and likewise in (2.A.2) to recover (2.2.4), with the integration contour now arbitrarily close to the real axis.

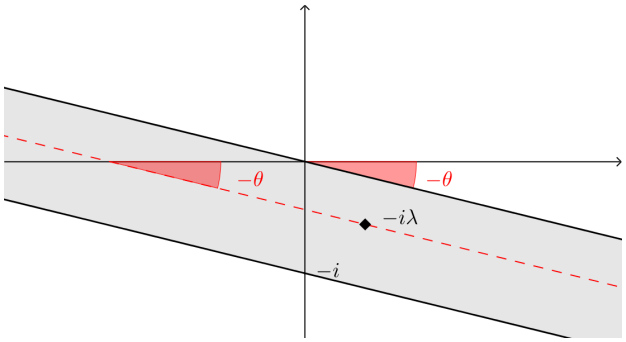


Figure 2.21. Choice of integration contour for the Mordell integral Ψ_+. The angle $0 < \theta < \frac{\pi}{2}$ is arbitrary, and $-i\lambda$ must lie in the shaded region.

Another subtle aspect is that the denominators in the right-hand side of (2.2.3)-(2.2.4) seem to have a sign ambiguity when κ is a multiple of ϱ and $|\Re\xi| = \frac{1}{2}$ and $\Re\lambda = \frac{1}{2}$. This happens because the result for $\Re\left(\xi - \frac{\kappa}{\varrho}\lambda\right) \in \mathbb{Z}$ is obtained by analytic continuation, and this should be performed at the end of the computations. The result is unique and unambiguous if we move slightly away from such points, for example by a shift $\xi \mapsto \xi + \varepsilon$ for a small ε, and take the limit at the end. Stated differently, the apparent sign ambiguity would only be an artefact of the intermediate steps and will disappear after simplifications in the final answer. We have also checked it for the solutions in Section 2.3.

Chapter 3

Complex (super)-matrix models with external sources and q-ensembles of Chern–Simons and ABJ(M) type

3.1 Introduction to the chapter

The idea of introducing non-zero commutators among position or momentum coordinates in different directions goes back to the 1940s [123]. In the late 1990s there was a boost of interest in noncommutative field theories, in great part due to the fact that low energy string theory can be related to noncommutative field theory [124, 125]. However, it was soon established that the expectation that washing out the space-time points could weaken UV divergences in quantum field theory, and consequently simplify renormalization, did not work as expected. Rather the opposite turned out to hold: renormalization gets harder due to noncommutativity, because in planar diagrams of a perturbative expansion the UV divergences simply persist. Second, in the non-planar diagrams, they tend to "mix" with IR divergences [126]. We refer to [127, 128] for classical reviews of the topic, and to [129] for insights into the relation between fuzzy spaces and matrix models.

Langmann, Szabo and Zarembo (LSZ) introduced and studied a scalar field theory on the Moyal plane, and showed that its partition function admits a matrix model representation [48, 49]. The LSZ matrix model has the explicit form

$$\mathcal{Z}_{\mathrm{LSZ}}\left(E,\tilde{E}\right)=\int \mathscr{D}M\mathscr{D}M^{\dagger}\ \exp\left(-N\,\mathrm{Tr}\left\{MEM^{\dagger}+M^{\dagger}\tilde{E}M+\widehat{V}\left(M^{\dagger}M\right)\right\}\right), \tag{3.1.1}$$

where M is a $N\times N$ complex matrix, $E,\tilde{E}$ are external matrices with real eigenvalues determined by the model and $\widehat{V}$ is a polynomial potential with quadratic and quartic interaction terms. Recently, a Hermitian matrix model with external field E and quartic interaction has been studied in the context of noncommutative scalar field theories [130–132]. It is worthwhile to stress that, despite the similarities, the fact that the LSZ model is a complex random matrix ensemble leads to a different analysis and very different set of results. In particular, we will establish a relationship between this model and a family of matrix models that appear in Chern–Simons theory and are related to q-deformed random matrix ensembles.

We will also consider a supersymmetric extension of the LSZ model (sLSZ), which we define as the supermatrix model

$$\mathcal{Z}_{\mathrm{sLSZ}}\left(E,\tilde{E}\right)=\int \mathscr{D}M\mathscr{D}M^{\dagger}\ \exp\left(-(N_1+N_2)\,\mathrm{STr}\left\{MEM^{\dagger}+M^{\dagger}\tilde{E}M+\widehat{V}\left(M^{\dagger}M\right)\right\}\right), \tag{3.1.2}$$

where now M is a $(N_1+N_2)\times(N_1+N_2)$ complex supermatrix and STr is the supertrace. The details are given in Section 3.4.

On the other hand, a number of different matrix models have been studied in gauge theory, more precisely in the study of certain topological and supersymmetric gauge theories in compact three-manifolds such as Seifert manifolds. The simplest case is that of $\mathbb{S}^3$, where the partition function of $U(N)$ Chern–Simons theory, admits the expression [133]

$$\mathcal{Z}_{\mathrm{CS}}=\int_{\mathbb{R}^N}\prod_{1\leq a<b\leq N}\left(2\sinh\left(\frac{x_a-x_b}{2}\right)\right)^2\prod_{a=1}^{N}e^{-\frac{x_a^2}{2g_s}}\mathrm{d}x_a, \tag{3.1.3}$$

with g_s a coupling constant, which can be related to the level $k\in\mathbb{Z}$ of Chern–Simons theory by $g_s=\frac{2\pi\mathrm{i}}{N+k}$. A review of early results is [134]. The matrix model description can also be obtained and further understood by using different types of localization of the path integral [135–137].

Interestingly, expressions such as (3.1.3), or generalizations thereof, with Schur polynomial insertions, describing then Wilson loop observables in the Chern–Simons theory, are completely solvable using random matrix theory methods [84, 83, 138]. We will give an interpretation of the LSZ matrix model in terms of probabilities in the random matrix description of the observables of $U(N)$ Chern–Simons theory on $\mathbb{S}^3$. This type of probability is a classical object in random matrix theory [39]. However, we will be naturally lead to the consideration of atypically large fluctuations rather than any bulk spectral quantity (like for the GUE model [139], the model (3.1.3) being a q-deformation of the GUE). In addition, we shall see that different insertions of external matrices $E,\tilde{E}$ will be related to different Chern–Simons observables.

Further examples beyond the topological $U(N)_k$ Chern–Simons theory are the ABJ(M) theories [43, 44], supersymmetric $U(N_1)_k\times U(N_2)_{-k}$ Chern–Simons-matter theories preserving twelve supercharges. The partition function of ABJ theory on $\mathbb{S}^3$ is [58] (cf. Subsection 2.2.1)

$$\begin{aligned}\mathcal{Z}_{\mathrm{ABJ}}=\int_{\mathbb{R}^{N_1}}\int_{\mathbb{R}^{N_2}}&\frac{\prod_{1\leq a<b\leq N_1}\left(2\sinh\left(\frac{x_a-x_b}{2}\right)\right)^2\prod_{1\leq\dot{a}<\dot{b}\leq N_2}\left(2\sinh\left(\frac{w_{\dot{a}}-w_{\dot{b}}}{2}\right)\right)^2}{\prod_{a=1}^{N_1}\prod_{\dot{b}=1}^{N_2}\left(2\cosh\left(\frac{x_a-w_{\dot{b}}}{2}\right)\right)^2}\\&\times\prod_{a=1}^{N_1}e^{-\frac{x_a^2}{2g_s}}\mathrm{d}x_a\prod_{\dot{a}=1}^{N_2}e^{-\frac{w_{\dot{a}}^2}{2g_s}}\mathrm{d}w_{\dot{a}},\end{aligned} \tag{3.1.4}$$

where, again, we have continued the Chern–Simons level into the string coupling $g_s>0$. It was shown in [69] that equation (3.1.4) defines a supermatrix model related to the ordinary matrix model for $U(N)$ Chern–Simons theory on the lens space $\mathbb{S}^3/\mathbb{Z}_2$ [140, 141]. The approach of [69] has been applied in [68] to evaluate exactly the ABJ partition function. We will show how the relation uncovered between LSZ and Chern–Simons matrix models extends to the supermatrix models (3.1.2)-(3.1.4).

In this chapter, we establish the relationship between the two sets of matrix models described above. This is a random matrix result, linking two families of models: complex matrix models with external sources on one hand, q-ensembles that appear in Chern–Simons theory on the other. Importantly, the relationship is not between the same observables. On one hand, we have partition functions (albeit generalized via different choices of external matrices) and on the other hand probabilities in the q-ensembles, corresponding to different observables in the Chern–Simons theory. This result can also be appreciated independently of the gauge theoretic origin of both sets of models. The Chern–Simons matrix model for example, plays a prominent role in the subject of non-intersecting Dyson Brownian motion [142, 143], see [144, 145]. It also appears in the study of Riemannian Gaussian distributions and their geometry, with applications to data analysis, as we have shown in [12].

The chapter is organized as follows. A field-theoretical background for the matrix model of interest is provided in Section 3.2, where we review the derivation of the LSZ matrix model from a noncommutative scalar theory. After that, we analyze the LSZ matrix model and explain its close relation with the Chern–Simons matrix model in Section 3.3. We emphasize that the connection holds in a generalized sense, as the external matrices are not restricted to come from a kinetic operator of a scalar field theory, and different spectra of the external matrices will be interpreted in terms of different observables in Chern–Simons theory. In Subsection 3.3.4 we show that the generalized LSZ partition function encodes certain probabilities in the random matrix description of topological invariants computed from Chern–Simons theory. Then, in Section 3.4, we extend the analysis to the sLSZ supermatrix model, and establish an analogous relation between the sLSZ generalized partition function and observables in ABJ theory. These are the central results of the present chapter. Finally, the Appendix 3.A contains technical details about Schur polynomials.

3.2 Noncommutative scalar theory with background field

In this section we review the LSZ model: in the first subsection, the geometric construction of the Moyal plane in the presence of a background magnetic field is sketched, while the second subsection is dedicated to the construction of a scalar field theory with quartic interaction. This provides a motivation and a physical background for the study of the matrix model (3.1.1), that will be thoroughly analyzed in Section 3.3.

The content of this section follows [49], although for the derivation of the matrix model in Subsection 3.2.2 we use a slightly different formalism than the original work, that will allow us a more direct comparison with the results in the next section.

3.2.1 Moyal plane with magnetic field

The Moyal plane is defined through the commutation relations

$$\left[q^j, q^k\right] = \mathrm{i}\theta\epsilon^{jk},$$

where θ is the essential parameter of the theory, with dimension of length squared. A Moyal plane can always be seen as a harmonic system, in the sense that passing to dimensionless complex coordinates one has $[z, \bar{z}] = 1$. The noncommutative plane needs not to arise from a modification of space-time. In fact, an example is given by a particle moving on a plane with a magnetic field of intensity B in the transverse direction; the momentum space then becomes noncommutative $\mathbb{R}^2$, as the momentum operators modify according to:

$$p_j \mapsto P_j := p_j - \frac{1}{2}B\epsilon_{jk}q^k.$$

In this case the covariant momenta P_j satisfy the commutation relation $[P_j, P_k] = -\mathrm{i}B\epsilon_{jk}$.

If the two frameworks are put together, that is, a transverse magnetic field is plugged in on a noncommutative plane, three possible harmonic oscillator pictures arise:

(i) on the two-dimensional position space, with annihilation and creation operators given by the complex coordinates as above;

(ii) on the two-dimensional momentum space, with annihilation and creation operator defined analogously;

(iii) a pair of canonical harmonic oscillators, one on each phase space plane.

However, the most suitable choice for us is none of them, and we will take a mixture of all these ingredients to form two commuting copies of annihilation and creation operators, in such a way that the problem decouples into two one-dimensional harmonic systems. To do so, define:

$$z := \frac{q^1 + \mathrm{i}q^2}{\sqrt{2\theta}}, \quad \bar{z} := \frac{q^1 - \mathrm{i}q^2}{\sqrt{2\theta}},$$
$$v := \frac{p_1 + \mathrm{i}p_2}{\sqrt{2\theta^{-1}}}, \quad \bar{v} = \frac{p_1 - \mathrm{i}p_2}{\sqrt{2\theta^{-1}}},$$

and use them to introduce the operators:

$$a_1 = \frac{z + \mathrm{i}v}{\sqrt{2}}, \quad a_1^\dagger = \frac{\bar{z} - \mathrm{i}\bar{v}}{\sqrt{2}},$$
$$a_2 = \frac{\bar{z} + \mathrm{i}\bar{v}}{\sqrt{2}}, \quad a_2^\dagger = \frac{z - \mathrm{i}v}{\sqrt{2}}.$$

Straightforward calculations provide:

$$\left[a_\alpha, a_\beta^\dagger\right] = \delta_{\alpha\beta},$$
$$[a_\alpha, a_\beta] = 0 = \left[a_\alpha^\dagger, a_\beta^\dagger\right],$$

for $\alpha, \beta = 1, 2$, hence we got a pair of decoupled harmonic oscillators.

Remark. Lifting the obstruction θ shifts the canonical symplectic structure on the cotangent bundle. It turns out that such shifted 2-form is still symplectic. One can then rotate to Darboux coordinates so that the new symplectic structure on the phase space $T^*\mathbb{R}^2$ is block-diagonal. The calculations above are precisely the explicit change of coordinates.

Consider now the differential operator D_j associated to the covariant momenta P_j, and $\tilde{D}_j$ analogous but carrying a reflected magnetic field $-B$. If we take the arbitrary combination $-\sigma D^2 - \tilde{\sigma}\tilde{D}^2$ and evaluate it at the symmetric point $\sigma = \tilde{\sigma} = \frac{1}{2}$, we obtain:

$$\left(-\sigma D^2 - \tilde{\sigma}\tilde{D}^2\right)_{\sigma=\tilde{\sigma}=\frac{1}{2}} = |\vec{p}|^2 + \frac{B^2}{4}|\vec{q}|^2 = \theta^{-1}\left(\frac{B^2\theta^2}{4}\{z, \bar{z}\} + \{v, \bar{v}\}\right),$$

where the curly bracket in the right-hand side stands for anticommutation. On the other hand, in terms of the harmonic oscillators description, we have:

$$\sum_{\alpha=1}^{2} a_\alpha^\dagger a_\alpha = \frac{1}{2}\left(\{z, \bar{z}\} + \{v, \bar{v}\} - \mathrm{i}[v, \bar{z}] + \mathrm{i}[z, \bar{v}]\right),$$

which means

$$\left(-\sigma D^2 - \tilde{\sigma}\tilde{D}^2\right)_{\sigma=\tilde{\sigma}=\frac{1}{2}} = \frac{2}{\theta}\sum_{\alpha=1}^{2}\left(a_\alpha^\dagger a_\alpha + \frac{1}{2}\right) \tag{3.2.1}$$

at points $B^2\theta^2 = 4$. The preferred curves $\frac{B^2\theta^2}{4} = 1$ correspond to the self-dual points of the Langmann–Szabo symmetry [146]. The theory is independent of the actual choice of curve in parameter space we restrict to, namely $B = \pm 2\theta^{-1}$. In fact, the two theories we obtain are equivalent, in the sense that they are dual descriptions of the same theory [146]. The invariance reflects the fact that the operators $D_j, \tilde{D}_j$ only differ by a reflection $B \mapsto -B$, thus the symmetric choice $\sigma = \tilde{\sigma}$ drops the dependence on the sign of the magnetic field.

3.2.2 LSZ model

Given a scalar field Φ on the Moyal plane, we can expand it in terms of the Landau basis, consisting of eigenstates of both harmonic oscillators, as:

$$\Phi = \sum_{\ell_1,\ell_2=1}^{\infty} M_{\ell_1\ell_2} |\ell_1, \ell_2\rangle.$$

This expression naturally defines an infinite matrix M associated to the field Φ. Now recall the kinetic operator in (3.2.1); using the property

$$a_\alpha^\dagger a_\alpha |\ell_1, \ell_2\rangle = (\ell_\alpha - 1)\, |\ell_1, \ell_2\rangle, \qquad \ell_\alpha = 1, 2, \ldots, \ \alpha = 1, 2,$$

it is possible to write:

$$\begin{aligned}\left(\sum_{\alpha=1}^{2}\left(a_\alpha^\dagger a_\alpha + \frac{1}{2}\right)\right)\Phi &= \sum_{\ell_1,\ell_2} M_{\ell_1\ell_2}\left\{\left(\ell_1 - \frac{1}{2}\right) + \left(\ell_2 - \frac{1}{2}\right)\right\}|\ell_1, \ell_2\rangle \\ &= \sum_{\ell_1,\ell_2}\left(\ell_1 - \frac{1}{2}\right)\left\{M_{\ell_1\ell_2}|\ell_1, \ell_2\rangle + M_{\ell_2\ell_1}|\ell_2, \ell_1\rangle\right\}.\end{aligned}$$

Therefore one obtains:

$$\begin{aligned}\Phi^\dagger\left(\sum_{\alpha=1}^{2}\left(a_\alpha^\dagger a_\alpha + \frac{1}{2}\right)\right)\Phi &= \sum_{\ell_1,\ell_2}\left(\ell_1 - \frac{1}{2}\right)\left\{M^\dagger_{\ell_2\ell_1} M_{\ell_1\ell_2} + M_{\ell_2\ell_1} M^\dagger_{\ell_1\ell_2}\right\} \\ &= \mathrm{Tr}\left\{M^\dagger E M + M E M^\dagger\right\},\end{aligned} \tag{3.2.2}$$

where in the last line we have introduced the diagonal matrix

$$E_{\ell_1\ell_2} = \left(\ell_1 - \frac{1}{2}\right)\frac{4\pi}{N}\delta_{\ell_1\ell_2}. \tag{3.2.3}$$

Consider the action [48, 49]

$$\begin{aligned}S_{\mathrm{LSZ}}\left(\Phi, \Phi^\dagger\right) = \int_{\mathbb{R}^2_\theta}\Big\{&\frac{1}{2}\Phi^\dagger\left(-\sigma D^2 - \tilde{\sigma}\tilde{D}^2\right)\Phi + \frac{1}{2}\Phi\left(-\sigma D^2 - \tilde{\sigma}\tilde{D}^2\right)\Phi^\dagger \\ &+ m_0^2\Phi^\dagger\Phi + \frac{g_0^2}{2}\left(\Phi^\dagger\Phi\right)^2\Big\}.\end{aligned}$$

Evaluated at the symmetric point $\sigma = \tilde{\sigma} = \frac{1}{2}$ it becomes:

$$S_{\mathrm{LSZ}}\left(\Phi, \Phi^\dagger\right) = N\,\mathrm{Tr}\left\{M^\dagger E M + M E M^\dagger + \widehat{m}^2 M^\dagger M + \frac{\widehat{g}^2}{2}\left(M^\dagger M\right)^2\right\} \tag{3.2.4}$$

where we have introduced the dimensionless couplings

$$\widehat{m}^2 = \left(\frac{2\pi\theta}{N}\right)m_0^2, \qquad \widehat{g}^2 = \left(\frac{2\pi\theta}{N}\right)g_0^2.$$

We have also used equations (3.2.1) and (3.2.2), and the matrix E defined in (3.2.3). In Section 3.3 we will solve the matrix model with action (3.2.4).

Notice that, to regularize the integral, we truncate the matrix M to its top-left $N \times N$ block, which introduces a finite cutoff at short distance $\sqrt{\frac{2\pi\theta}{N}}$. The full theory is recovered in the large N limit. As expected from general features of noncommutative field theory [128], the original noncommutativity of the phase space is eventually encoded in the noncommutativity of matrix multiplication. Consistently, integrals over the spacetime become traces.

3.3 Solving the matrix model

As we have seen in Section 3.2, different noncommutative field theories reduce to a matrix model of the form (3.1.1)

$$\mathcal{Z}_{\mathrm{LSZ}}\left(E,\tilde{E}\right)=\int \mathscr{D}M\mathscr{D}M^{\dagger}\ \exp\left(-N\,\mathrm{Tr}\left\{ME M^{\dagger}+M^{\dagger}\tilde{E}M+\widehat{V}\left(M^{\dagger}M\right)\right\}\right)$$

in terms of $N\times N$ complex matrices, depending on the insertion of two external matrices. The potential $\widehat{V}\left(M^{\dagger}M\right)$ is a quadratic polynomial with dimensionless coefficients

$$\widehat{V}\left(M^{\dagger}M\right)=\widehat{m}^{2}\left(M^{\dagger}M\right)+\frac{\widehat{g}^{2}}{2}\left(M^{\dagger}M\right)^{2}. \tag{3.3.1}$$

We also let the external fields $E,\tilde{E}$ have arbitrary eigenvalues which, for convenience and consistently with the noncommutative field theory setting (recall (3.2.3)), we write as $\frac{4\pi}{N}\eta_{\ell},\frac{4\pi}{N}\tilde{\eta}_{\ell}$ respectively, for $\ell=1,\dots,N$. In particular, the original LSZ model, discussed in Section 3.2, corresponds to external matrices with $\eta_{\ell}=\tilde{\eta}_{\ell}$ given by consecutive integers plus a constant shift, see (3.2.3). The unconventional presence of $\tilde{E}$ explicitly breaks $U(N)$ symmetry, but the system is still tractable.

We now reduce this matrix model to an ordinary multiple integral in terms of the spectra of the external fields, and show that the results can be written in terms of the observables of $U(N)$ Chern–Simons theory in $\mathbb{S}^3$, with $q=e^{-g_s}$ real.

3.3.1 General solution

As shown in [49], we can approach the solution using the singular value decomposition of a generic $N\times N$ complex matrix M:

$$M=U_{1}^{\dagger}\mathrm{diag}\left(\lambda_{1},\dots,\lambda_{N}\right)U_{2},$$

with $U_{\alpha=1,2}$ unitary matrices and $\lambda_{\ell}\geq 0$.[6] The measure becomes

$$\mathscr{D}M\ \mathscr{D}M^{\dagger}=\left[\mathrm{d}U_{1}\right]\left[\mathrm{d}U_{2}\right]\prod_{\ell=1}^{N}\mathrm{d}y_{\ell}\ \Delta_{N}\left[y\right]^{2},$$

where $[\mathrm{d}U_{\alpha}]$ is the invariant Haar measure over $U(N)$, and $y_{\ell}:=\lambda_{\ell}^{2}$ and

$$\Delta_{N}\left[y\right]=\prod_{1\leq\ell<\ell'\leq N}\left(y_{\ell}-y_{\ell'}\right) \tag{3.3.2}$$

is the Vandermonde determinant.

As shown in [49], the use of this transformation implies that integrals over $U_{\alpha=1,2}\in U(N)$ decouple over the two types of external field terms. Denoting by $Y=\mathrm{diag}\left(y_{1},\dots,y_{N}\right)$, the angular degrees of freedom U_{α} can be integrated out using the Harish-Chandra–Itzykson–Zuber (HCIZ) formula [148, 149]:

$$\int_{U(N)}\left[\mathrm{d}U_{2}\right]\exp\left\{-N\,\mathrm{Tr}\left(EU_{2}^{\dagger}YU_{2}\right)\right\}=\mathcal{C}_{N}\frac{\det_{1\leq\ell,\ell'\leq N}\left(e^{-4\pi\eta_{\ell}y_{\ell'}}\right)}{\Delta_{N}\left[\eta\right]\Delta_{N}\left[y\right]}, \tag{3.3.3}$$

[6] See [147] for comments on this parametrization with regards to the more usual one in terms of eigenvalues [39]. In any case, this transformation is much used and very useful when studying complex matrix models, such as the LSZ.

where we have used the explicit form of the eigenvalues of E and denoted

$$\mathcal{C}_N := (4\pi)^{-\frac{N(N-1)}{2}} \prod_{j=1}^{N-1} j! = (4\pi)^{-\frac{N(N-1)}{2}} G(N+1),$$

where $G(\cdot)$ is the Barnes G-function.

Analogous expression is obtained for integration over U_1 replacing E by $\tilde{E}$. For a potential $\widehat{V}\left(M^\dagger M\right)$ as in (3.3.1), one gets:

$$\begin{aligned}\mathcal{Z}_{\mathrm{LSZ}}\left(E,\tilde{E}\right) = &\int_{[0,\infty)^N} \Delta_N\left[y\right]^2 \prod_{\ell=1}^{N} e^{-\left(\widehat{m}^2 y_\ell + \frac{\widehat{g}^2}{2} y_\ell^2\right)} \mathrm{d}y_\ell \\ &\times \int \left[\mathrm{d}U_2\right] \exp\left(-NEU_2^\dagger Y U_2\right) \int \left[\mathrm{d}U_1\right] \exp\left(-N\tilde{E}U_1^\dagger Y U_1\right)\end{aligned} \tag{3.3.4}$$

Plugging HCIZ (3.3.3) into (3.3.4), simplifying a Vandermonde squared and suitably rescaling the integration variables, we note that the partition function for the matrix model is:

$$\mathcal{Z}_{\mathrm{LSZ}}\left(E,\tilde{E}\right) = \frac{\mathcal{C}'_N}{\Delta_N\left[\eta\right]\Delta_N\left[\tilde{\eta}\right]} \int_{[0,\infty)^N} \det_{1\le\ell,\ell'\le N}\left(e^{-\eta_\ell y_{\ell'}}\right) \det_{1\le\ell,\ell'\le N}\left(e^{-\tilde{\eta}_\ell y_{\ell'}}\right) \prod_{\ell=1}^{N} e^{-m^2 y_\ell - \frac{g^2}{2} y_\ell^2} \,\mathrm{d}y_\ell, \tag{3.3.5}$$

where we recall that η_ℓ (respectively $\tilde{\eta}_\ell$) stand for the eigenvalues of E (respectively $\tilde{E}$), up to a factor $\frac{4\pi}{N}$, and with coefficients redefined as:

$$m^2 = \frac{\widehat{m}^2}{4\pi}, \qquad g^2 = \frac{\widehat{g}^2}{(4\pi)^2}, \tag{3.3.6}$$

and

$$\mathcal{C}'_N = (4\pi)^{-N}\mathcal{C}_N^2 = \frac{G(N+1)^2}{(4\pi)^{N^2}} = \frac{2^{-N(N-1)}\pi^N}{\mathrm{vol}\left(U(N)\right)^2},$$

where $\mathrm{vol}(\cdot)$ is the volume of the gauge group. Note that this normalization is essentially the square of the partition function of the Gaussian unitary ensemble (GUE) [39].

In the theory of non-intersecting Brownian motion, the determinants in (3.3.5) are very familiar. This whole theory of determinantal processes is known to be directly related to $U(N)$ Chern–Simons theory on $\mathbb{S}^3$ and with the Wess–Zumino–Witten (WZW) model [144], where such connection was shown to follow from specializations of the determinants

$$\det_{1\le\ell,\ell'\le N}\left(e^{-\eta_\ell y_{\ell'}}\right), \qquad \det_{1\le\ell,\ell'\le N}\left(e^{-\tilde{\eta}_\ell y_{\ell'}}\right)$$

in (3.3.5). However, it is more direct to show the relation through the corresponding matrix model formulation. Recall for this the definition of a Schur polynomial [90]

$$s_\mu\left(x_1,\ldots,x_N\right) = \frac{\det_{1\le\ell,\ell'\le N}\left(x_\ell^{\mu_{\ell'}+N-\ell'}\right)}{\det_{1\le\ell,\ell'\le N}\left(x_\ell^{N-\ell'}\right)},$$

and hence we rewrite the scaled eigenvalues $\eta_\ell, \tilde{\eta}_\ell$ of the external matrices $E, \tilde{E}$, assuming they are integers, as:

$$\begin{aligned}\eta_\ell &= \mu_\ell + N - \ell, \\ \tilde{\eta}_\ell &= \nu_\ell + N - \ell,\end{aligned} \tag{3.3.7}$$

for $\ell = 1, \dots, N$. Without loss of generality, we can relabel the eigenvalues so that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_N \geq 0$, and likewise for $\nu_1 \geq \cdots \geq \nu_N$. Then, using

$$\prod_{1\leq \ell<\ell'\leq N} (e^{y_\ell} - e^{y_{\ell'}})^2 = \prod_{1\leq \ell<\ell'\leq N} \left(2\sinh\left(\frac{y_\ell - y_{\ell'}}{2}\right)\right)^2 \prod_{m=1}^{N} e^{(N-1)y_m}$$

and reflecting the variables $y_\ell \mapsto -y_\ell$, we immediately have:

$$\begin{aligned}\mathcal{Z}_{\mathrm{LSZ}}(\mu,\nu) = \frac{(-1)^N \mathcal{C}'_N}{\Delta_N[\eta]\Delta_N[\tilde{\eta}]} \int_{(-\infty,0]^N} \prod_{1\leq \ell<\ell'\leq N} \left(2\sinh\left(\frac{y_\ell - y_{\ell'}}{2}\right)\right)^2 \prod_{\ell=1}^{N} \mathrm{d}y_\ell \; e^{\beta y_\ell - \frac{g^2}{2} y_\ell^2} \\ \times s_\mu\left(e^{y_1}, \dots, e^{y_N}\right) s_\nu\left(e^{y_1}, \dots, e^{y_N}\right),\end{aligned} \tag{3.3.8}$$

with $\beta = m^2 - N + 1$.[7] We henceforth write $\mathcal{Z}_{\mathrm{LSZ}}(\mu,\nu)$ for $\mathcal{Z}_{\mathrm{LSZ}}\left(E, \tilde{E}\right)$, stressing the dependence on the partitions $\mu = (\mu_1, \dots, \mu_N)$ and $\nu = (\nu_1, \dots, \nu_N)$. Except for the integration domain, the expression (3.3.8) is close to the general version of the $U(N)$ Chern–Simons on $\mathbb{S}^3$ matrix model, with two different insertions of Schur polynomials, whose evaluation gives the topological invariant of a pair of unknots [134, 107] carrying the $U(N)$ representations μ and ν. Furthermore, if the external matrix E has positive integer eigenvalues while $\tilde{E}$ has negative integers eigenvalues, or vice versa, the relation then is with the Hopf link invariant carrying representations μ and ν.

Instead, after a shift of variables in (3.3.8) and using the identity (3.A.1), we obtain the matrix model representation

$$\begin{aligned}\mathcal{Z}_{\mathrm{LSZ}}(\mu,\nu) = \mathcal{A}_N(\mu,\nu) \int_{(-\infty,-\gamma]^N} \prod_{j=1}^{N} \mathrm{d}x_j \; e^{-\frac{g^2}{2}\sum_{j=1}^N x_j^2} \prod_{1\leq j<k\leq N} \left(2\sinh\left(\frac{x_j - x_k}{2}\right)\right)^2 \\ \times s_\mu\left(e^{x_1}, \dots, e^{x_N}\right) s_\nu\left(e^{x_1}, \dots, e^{x_N}\right),\end{aligned}$$

where $\gamma = \beta/g^2$ and

$$\mathcal{A}_N(\mu,\nu) := \frac{(-1)^N \mathcal{C}'_N}{\Delta_N[\eta]\Delta_N[\tilde{\eta}]} \; \exp\left(\frac{\gamma^2 N}{2} + \gamma\left(|\mu| + |\nu|\right)\right). \tag{3.3.9}$$

Notice also that the Vandermonde factors in the denominator of (3.3.9), which depend exclusively on the eigenvalues of the external matrices, can be written, using Weyl's denominator formula (see Appendix 3.A), as

$$\begin{aligned}\Delta_N[\eta] &= \prod_{1\leq \ell<\ell'\leq N} \left(\mu_\ell - \mu_{\ell'} - \ell + \ell'\right) = G(N+1)\dim\mu, \\ \Delta_N[\tilde{\eta}] &= \prod_{1\leq \ell<\ell'\leq N} \left(\nu_\ell - \nu_{\ell'} - \ell + \ell'\right) = G(N+1)\dim\nu.\end{aligned}$$

These Barnes G-functions in (3.3.9) cancel against the ones coming from the double application of HCIZ formula. We finally obtain

$$\mathcal{Z}_{\mathrm{LSZ}}(\mu,\nu) = C_N(\gamma, |\mu|, |\nu|) \; \frac{\mathcal{Z}_{\mathrm{CS}} \left\langle W_\mu W_\nu \; \mathbb{1}_{(-\infty,-\gamma]^N} \right\rangle_{\mathrm{CS}}}{\dim\mu \; \dim\nu}, \tag{3.3.10}$$

[7] With our definition, η_ℓ and $\tilde{\eta}_\ell$ are non-negative integers, which is the physical choice motivated by the LSZ model. If we want them to be non-positive integers, we simply do not reflect y_ℓ and define $\eta_\ell, \tilde{\eta}_\ell$ with opposite sign. The discussion would be exactly the same, except for the integration domain being $[0, \infty)$.

where the relation between the partitions μ, ν and the external matrices of the generalized LSZ model is given in equation (3.3.7), and

$$C_N(\gamma, |\mu|, |\nu|) = \frac{(-1)^N e^{\frac{\gamma^2 N}{2} + \gamma(|\mu|+|\nu|)}}{(4\pi)^{N^2}}.$$

In formula (3.3.10), $\mathcal{Z}_{\text{CS}}$ is the $U(N)$ Chern–Simons partition function on $\mathbb{S}^3$, which is a quantum topological invariant also known as Witten–Reshetikhin–Turaev invariant [42, 150, 151], defined as the matrix model (3.1.3) and whose explicit evaluation we give below. Besides, W_μ is the trace of the holonomy of the gauge connection along an unknot inside $\mathbb{S}^3$, taken in the $U(N)$ representation corresponding to the partition μ, and likewise for W_ν. In (3.3.10), the Chern–Simons coupling is $g_s = \frac{1}{g^2}$. In Chern–Simons theory, g_s is related to the Chern–Simons level k by $g_s = \frac{2\pi \mathrm{i}}{N+k}$, while the real string coupling constant g_s is used when describing topological strings. That is the same type of description here, since g^2 is real. Finally, $\langle \cdots \rangle_{\text{CS}}$ in (3.3.10) means the average in the Chern–Simons matrix model (3.1.3), and $\mathbb{1}_{(-\infty,-\gamma]^N}$ is the N-dimensional indicator function

$$\mathbb{1}_{(-\infty,-\gamma]^N}(x_1, \ldots, x_N) = \prod_{j=1}^{N} \mathbb{1}_{(-\infty,-\gamma]}(x_j) = \begin{cases} 1, & x_j \leq -\gamma \quad \forall\, j = 1, \ldots, N, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, $\left\langle W_\mu W_\nu \mathbb{1}_{(-\infty,-\gamma]^N} \right\rangle_{\text{CS}}$ would correspond to the two-unknot invariant, but averaged only using variables $x_j \leq -\gamma$ (instead of $x_j \in \mathbb{R}$). Its relation with the actual invariant of a pair of unknots is further discussed in Subsection 3.3.4 through the lenses of random matrix theory.

3.3.2 Quantum dimensions

An important particular case of the above general setting is when one of the partitions in (3.3.7) is void. That is, in terms of LSZ theory, one has an external matrix with the equispaced spectra and the other one generalized with a partition, see the definitions (3.2.3) and (3.3.7). This case corresponds, as we shall see, to quantum dimensions in the Chern–Simons interpretation [83]. Quantum dimension of a representation associated to the partition μ is given by the following hook-content formula [90, 83]

$$\dim_q \mu := \prod_{\square \in \mu} \frac{\lfloor N + c(\square) \rfloor_q}{\lfloor h(\square) \rfloor_q},$$

where for each box $\square \equiv (j,k)$ of the Young diagram determined by μ, the quantity $h(\square) := \mu_j + \mu'_k - j - k + 1$ is the hook length, with the prime meaning conjugate diagram, and $c(x) := j - k$ is known as the content of the box x. The operation $\lfloor \cdot \rfloor_q$ denotes the symmetric q-number, that is

$$\lfloor n \rfloor_q = \frac{q^{n/2} - q^{-n/2}}{q^{1/2} - q^{-1/2}}.$$

In Chern–Simons theory on $\mathbb{S}^3$, the unknot invariant is given by quantum dimensions [134, 83]. Since one of the two external matrices has harmonic oscillator spectrum the matrix model above reduces to

$$\mathcal{Z}_{\text{LSZ}}(\mu, \emptyset) = \mathcal{A}_N(\mu, \emptyset) \int_{(-\infty,-\gamma]^N} \prod_{1 \leq j < k \leq N} \left(2 \sinh\left(\frac{x_j - x_k}{2} \right) \right)^2 \prod_{j=1}^{N} e^{-\frac{g^2}{2} x_j^2} \mathrm{d}x_j \\ \times s_\mu\left(e^{x_1}, \ldots, e^{x_N}\right),$$

whose evaluation leads to

$$\mathcal{Z}_{\rm LSZ}\left(\mu,\emptyset\right)=C_N(\gamma,|\mu|,0)\,\frac{\mathcal{Z}_{\rm CS}\,\left\langle W_\mu\,\mathbb{1}_{(-\infty,-\gamma]^N}\right\rangle_{\rm CS}}{\dim\mu}. \tag{3.3.11}$$

This should be compared with the unknot invariant computed in Chern–Simons theory, which differs from the present setting in the fact that the integral is taken over $\mathbb{R}^N$ instead of $(-\infty,-\gamma]^N$. We will come back to this point in Subsection 3.3.4. The exact evaluation of the unknot invariant gives [83]

$$\langle W_\mu\rangle=q^{-\frac{1}{2}C_2(\mu)}\dim_q\mu,$$

where $q=e^{-1/g^2}$. Besides, in the expression above the term

$$C_2\left(\mu\right)=(N+1)\,|\mu|+\sum_{\ell=1}^{N}\left(\mu_\ell^2-2\ell\mu_\ell\right)$$

is the $U(N)$ quadratic Casimir of the representation μ, labelled by the Young diagram associated to the partition μ, with μ_ℓ boxes in the ℓ-th row, with rows understood to be aligned on the left.

The appearance of quantum dimensions is interesting in that they appear as well in the study of noncommutative gauge theories through the analysis of D-branes [152, 153]. However, as we have seen in Section 3.2, only the simpler setting, where the two external matrices are equal and have harmonic oscillator spectra, is directly linked to a noncommutative scalar theory.

3.3.3 Chern–Simons and LSZ partition function

We have shown in Section 3.2 that our study of noncommutative scalar field theory naturally leads to a LSZ matrix model with $E=\tilde{E}$ and spectra $\frac{4\pi}{N}\left(\ell-\frac{1}{2}\right)$ for $\ell=1,...,N$, cf. (3.2.3). Thus, we consider now the case in which both partitions are void: this corresponds to the two external matrices having harmonic oscillator spectra. In particular, from (3.3.7), we have that $\eta_\ell=N-\ell$ for $\ell=1,...,N$. The fact that the two physical spectra have an overall energy shift only has an impact at the level of renormalization of the mass parameter. This follows immediately from a simple property of Schur polynomials, given in Appendix 3.A.

Then one obtains the matrix model without Schur polynomial insertions, and the corresponding matrix integral is related to the one for the Chern–Simons partition function, given in (3.1.3) (recall that $g_s=1/g^2$). The Chern–Simons matrix model has the exact solution [84]:

$$\mathcal{Z}_{\rm CS}=\left(\frac{2\pi}{g^2}\right)^{N/2}N!\;e^{\frac{N(N+1)(N-1)}{6g^2}}\prod_{j=1}^{N-1}\left(1-q^j\right)^{N-j}, \tag{3.3.12}$$

with $q=e^{-1/g^2}$ as above. The product can also be written as a q-deformed Barnes function, which, in the limit $g\to\infty$ (which is $q\to 1$), reduces to the Barnes G-function. Then, the LSZ matrix model partition function is

$$\mathcal{Z}_{\rm LSZ}\left(\emptyset,\emptyset\right)=\frac{(-1)^N e^{\frac{\gamma^2 N}{2}}}{(4\pi)^{N^2}}\;\mathcal{Z}_{\rm CS}\;\langle\mathbb{1}_{(-\infty,-\gamma]^N}\rangle_{\rm CS}. \tag{3.3.13}$$

Notice that, apart from the simple prefactor, the ratio between the LSZ and the Chern–Simons partition function is the average $\langle\mathbb{1}_{(-\infty,-\gamma]^N}\rangle_{\rm CS}$ of the N-dimensional indicator function. This aspect is further analyzed the next subsection.

3.3.4 Probabilistic interpretation

The main result (3.3.10) and its specializations (3.3.11) and (3.3.13) admit an interpretation in terms of probabilities of large deviations of the smallest or largest eigenvalue of Hermitian random matrices. Consider a generic weight function $w(x)$, and let $\mathcal{Z}_w$ be the associated Hermitian random matrix ensemble,

$$\mathcal{Z}_w = \int_{\mathbb{R}^N} \Delta_N[x]^2 \prod_{j=1}^{N} w(x_j)\, \mathrm{d}x_j. \tag{3.3.14}$$

The probability that the largest eigenvalue of a random matrix in the ensemble (3.3.14) is smaller than a given threshold $s \in \mathbb{R}$ is

$$\begin{aligned} \mathrm{Prob}_w(x_{\max} \leq s) &= \mathrm{Prob}_w(x_1 \leq s, \dots, x_N \leq s) \\ &= \frac{1}{\mathcal{Z}_w} \int_{(-\infty,s]^N} \Delta_N[x]^2 \prod_{j=1}^{N} w(x_j) \mathrm{d}x_j = \left\langle \mathbb{1}_{(-\infty,s]^N} \right\rangle_w, \end{aligned} \tag{3.3.15}$$

where in the last expression $\langle\cdot\rangle_w$ means the average in the ensemble (3.3.14) and $\mathbb{1}_{(-\infty,s]^N}$ is the N-dimensional indicator function. In the Coulomb gas picture, replacing the weight $w(x)$ by $w(x)\mathbb{1}_{(-\infty,s]}(x)$ introduces a hard wall placed at $x = s$ which leaves the charges on its left. See [154] for the large N limit of matrix models in presence of hard walls. For the GUE, the probabilities of large fluctuations at large N have been found in [155, 139].

We immediately see from equation (3.3.13) that

$$\frac{\mathcal{Z}_{\mathrm{LSZ}}(\emptyset,\emptyset)}{\mathcal{Z}_{\mathrm{CS}}} = \frac{(-1)^N e^{\frac{\gamma^2 N}{2}}}{(4\pi)^{N^2}}\, \mathrm{Prob}_{\mathrm{CS}}(x_{\max} \leq -\gamma),$$

hence the ratio between the partition function of the LSZ model and that of Chern–Simons theory is effectively computing the probability of large deviations of eigenvalues in the Chern–Simons ensemble (3.1.3) (or, strictly speaking, in the Stieltjes–Wigert ensemble [84], see below), up to a completely determined, parameter-dependent overall factor. More in general, formula (3.3.10) states that

$$\frac{\mathcal{Z}_{\mathrm{LSZ}}(\mu,\nu)}{\mathcal{Z}_{\mathrm{CS}}} = C_N(\gamma, |\mu|, |\nu|)\, \frac{\langle W_\mu W_\nu \rangle_{\mathrm{CS}}}{\dim \mu \dim \nu}\, \mathrm{Prob}_{\mathrm{CS};W_\mu W_\nu}(x_{\max} \leq -\gamma),$$

where by $\mathrm{Prob}_{\mathrm{CS};W_\mu W_\nu}(\cdot)$ we mean the probability in the matrix ensemble computing the invariant of two unknots, that is, in the Chern–Simons ensemble (3.1.3) with the insertion of two Schur polynomials. This probability must be normalized by $\mathcal{Z}_{\mathrm{CS}} \cdot \langle W_\mu W_\nu \rangle_{\mathrm{CS}}$ and not only $\mathcal{Z}_{\mathrm{CS}}$. Therefore, the LSZ matrix model with general assignment of the external matrices, divided by the Chern–Simons partition function, is proportional to the two-unknot invariant weighted by the probability of large deviations in the random matrix description of such topological invariant. The proportionality constant yields an elementary dependence on the size N and on the free parameters of the generalized LSZ theory.

We emphasize that the LSZ partition function does not compute the (typically small) fluctuations of the largest eigenvalue around the edge of the eigenvalue distribution. Equivalently, in the Coulomb gas picture, the LSZ partition function does not describe the fluctuations of the rightmost charge around the equilibrium. What it gives is the probability of an atypically large fluctuation, with the greatest eigenvalue moving deep into the bulk of the eigenvalue distribution. This is a q-analogue of the large fluctuations in the GUE discussed in [155, 139] (see also [156, 157]). Besides, we underline that, differently from [139] where a large deviation from the

equilibrium configuration is of order $\sim \sqrt{N}$, in the q-analogue the support of the eigenvalue density grows as Ng_s, thus large deviations from the equilibrium are of order $\sim N$.

Now, if we consider equation (3.3.5) without Schur insertions, $\mu = \emptyset = \nu$, then, changing variables $u_j = e^{y_j - (m^2+2N-1)/g^2}$ [84], it is well-known that we have a standard-random matrix ensemble with a log-normal weight, named Stieltjes–Wigert (SW) ensemble. That is:

$$\mathcal{Z}_{\rm LSZ}(\emptyset,\emptyset) = \mathcal{B}_N \int_{[s,\infty)^N} \prod_{1\le j<k\le N} (u_j-u_k)^2 \prod_{j=1}^N e^{-\frac{1}{2g_s}(\ln u_j)^2} \mathrm{d}u_j, \tag{3.3.16}$$

for $g_s = 1/g^2$ and $s = e^{\frac{m^2+2N-1}{g^2}}$, and with

$$\mathcal{B}_N := (4\pi)^{-N^2} \exp\left[\frac{N}{2g^2}\left(m^2-1\right)\left(m^2+2N-1\right)\right].$$

Note that a SW ensemble is not centered around 0, as the weight is supported on $u \ge 0$. Therefore, up to a proportionality factor, the LSZ partition function $\mathcal{Z}_{\rm LSZ}(\emptyset,\emptyset)$ normalized by the SW partition function gives the probability of atypically large fluctuations of the eigenvalues away from the left edge at $u = 0$, with the smallest eigenvalue deep into the bulk, $u_{\min} \ge s$, in a Stieltjes–Wigert ensemble:

$$\frac{\mathcal{Z}_{\rm LSZ}(\emptyset,\emptyset)}{\mathcal{Z}_{\rm SW}} = \mathcal{B}_N \, \mathrm{Prob}_{\rm SW}\left(u_{\min} \ge s\right).$$

To evaluate this quantity numerically at finite N, it is convenient to rewrite the integrals in the numerator and denominator of $\mathrm{Prob}_{\rm SW}\left(u_{\min} \ge s\right)$ as determinants [39], obtaining:

$$\begin{aligned}\mathrm{Prob}_{\rm SW}\left(u_{\min} \ge s\right) &= \frac{\det_{1\le j,k\le N}\left[\int_s^\infty u^{j+k-2} e^{-\frac{1}{2g_s}(\ln u)^2}\mathrm{d}u\right]}{\det_{1\le j,k\le N}\left[\int_0^\infty u^{j+k-2} e^{-\frac{1}{2g_s}(\ln u)^2}\mathrm{d}u\right]} \\ &= 2^{-N}\frac{\det_{1\le j,k\le N}\left[e^{\frac{g_s}{2}(j+k-1)^2}\mathrm{erfc}\left(\sqrt{\frac{g_s}{2}}(m^2+2N-j-k)\right)\right]}{\det_{1\le j,k\le N}\left[e^{\frac{g_s}{2}(j+k-1)^2}\right]}.\end{aligned}$$

In the latter expression, $\mathrm{erfc}(z) = 1-\mathrm{erf}(z)$ is the complementary error function. The denominator is known, and is readily extracted from (3.3.12). We obtain:

$$\begin{aligned}\mathrm{Prob}_{\rm SW}\left(u_{\min} \ge s\right) = &\frac{q^{\frac{N}{3}(2N^2-1)}}{2^N(1-q)^{\frac{N}{2}(N-1)}\prod_{j=1}^{N-1}(1-q^j)^{N-j}} \\ &\times \det_{1\le j,k\le N}\left[e^{\frac{g_s}{2}(j+k-1)^2}\mathrm{erfc}\left(\sqrt{\frac{g_s}{2}}(m^2+2N-j-k)\right)\right],\end{aligned}$$

and we recall that $q = e^{-g_s}$ and the argument of erfc is related to s through $\ln s = g_s(m^2+2N-1)$. We plot the logarithm of this probability at finite N in Figure 3.1. It is also clear from the definition of s that the role of any fixed m^2 becomes less relevant as N is increased, unless m^2 itself is increased linearly with N. This aspect is shown in Figure 3.2.

We stress once more that the present setting is very different from the study of (typical) small fluctuation of the largest or smallest eigenvalue around the edge, which are suppressed by inverse powers of N. See [139, 156] for thorough discussion on this point, and [158] for further insights in the theory of large deviations. In the more general case, $\mathcal{Z}_{\rm LSZ}(\mu,\nu)$ is proportional to the probability of a large fluctuation of the eigenvalues away from the left edge in a SW ensemble deformed by the insertion of two Schur polynomials $s_\mu(u_1,\ldots,u_N)$ and $s_\nu(u_1,\ldots,u_N)$.

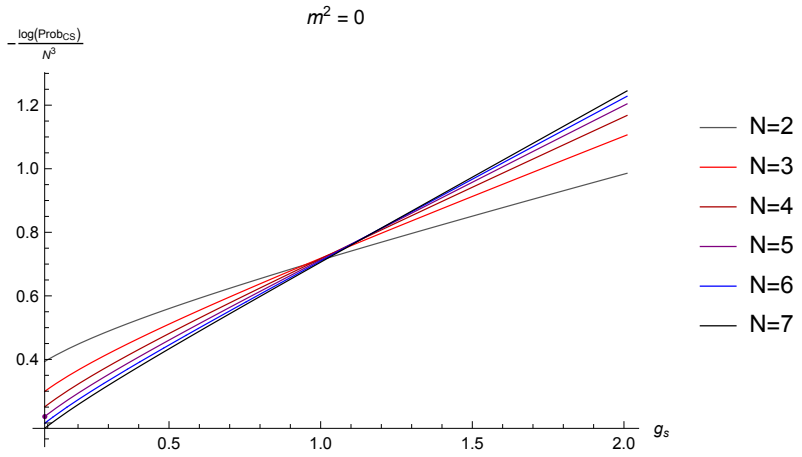


Figure 3.1. $-\frac{1}{N^3}\ln\frac{\mathcal{Z}_{\text{LSZ}}(\emptyset,\emptyset)}{\mathcal{B}_N\mathcal{Z}_{\text{SW}}}$ as a function of the coupling g_s at $m^2 = 0$. The curves correspond to N from 2 to 7.

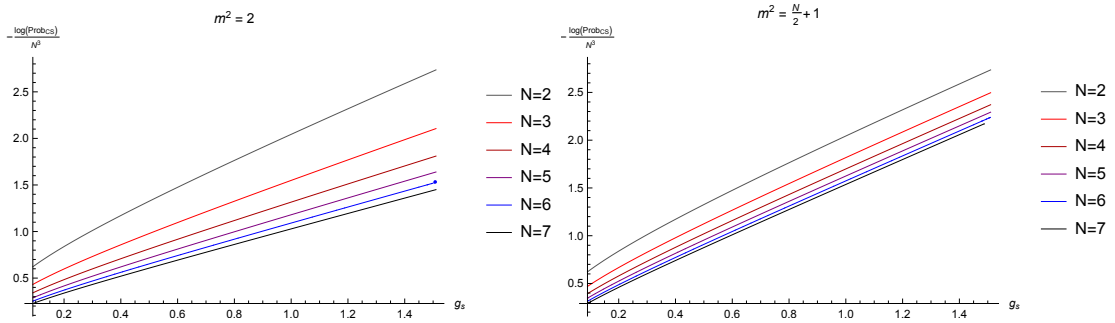


Figure 3.2. $-\frac{1}{N^3}\ln\frac{\mathcal{Z}_{\text{LSZ}}(\emptyset,\emptyset)}{\mathcal{B}_N\mathcal{Z}_{\text{SW}}}$ as a function of the coupling g_s at $m^2 = 2$ (left) and $m^2 = \frac{N}{2} + 1$ (right). The curves correspond to N from 2 to 7.

3.3.5 Large N limit

In the LSZ model, the parameter N regularizes the path integral of the noncommutative field theory, as discussed in Section 3.2. Therefore, it is natural to consider the large N limit of $\mathcal{Z}_{\text{LSZ}}(\emptyset,\emptyset)$, and more generally of $\mathcal{Z}_{\text{LSZ}}(\mu,\nu)$. In turn, the probabilistic interpretation of the Chern–Simons theory observables also calls for a large N analysis, from the perspective of large deviations theory [158].

The knowledge of the large N solution of the matrix model $\mathcal{Z}_{\text{LSZ}}(\mu,\nu)$ on either the complex matrix model side or on the side of the probability in Chern–Simons observables, would directly provide the solution on the other side. However, there are difficulties in solving the large N limit from both perspectives. As mentioned in Subsection 3.3.1, and already observed in the original work [49], a difficulty in the study of the LSZ partition function, generalized to arbitrary external matrices $E, \tilde{E}$, comes from the presence of $\tilde{E}$, which prevents a reformulation of the model in terms of a single Hermitian matrix $M^\dagger M$. Therefore, it becomes hard to solve the large N limit of the matrix model explicitly, using for example the loop equations (that are zero-dimensional Schwinger–Dyson equations) [159] when $\tilde{E} \neq 0$. For $\tilde{E} = 0$, the large N solution has been found in [49], but this choice does not yield meaningful observables in Chern–Simons theory.

Taking the converse route, we could try to analyze the large N asymptotics of (3.3.16), adapting the argument of [139] to the present case. This would amount to solve a constrained extrem-

ization problem defined as follows. Write

$$t \equiv g_s N = N/g^2, \qquad z \equiv \log s = g_s(m^2 + 2N - 1)$$

and take the large N 't Hooft limit of (3.3.16) keeping t and z fixed. Let $\rho_{\mathrm{SW}}(u)$ be the constrained SW eigenvalue density at large N, supported on the interval $[z, L(z)]$. We have stressed the dependence of the upper bound $L(z)$ on z, and let the more standard dependence on the 't Hooft coupling t implicit. Then $\rho_{\mathrm{SW}}(u)$ solves the saddle point equation

$$\mathrm{P}\int_z^{L(z)} \mathrm{d}u' \, \frac{\rho_{\mathrm{SW}}(u')}{u - u'} = \frac{1}{2t}\frac{\ln u}{u}, \tag{3.3.17}$$

with the symbol $\mathrm{P}\int$ meaning the Cauchy principal value integral. The upper boundary of the support $L(z)$ is fixed as a function of z and t by the normalization condition

$$\int_z^{L(z)} \mathrm{d}u \; \rho_{\mathrm{SW}}(u) = 1.$$

This problem cannot be solved by the method of [139], because of the non-polynomial form of the right-hand side of (3.3.17). More precisely, the extremization problem described by the saddle point equation (3.3.17) does not satisfy the hypothesis of Tricomi's theorem. A complete solution at large N would entail an extension of the method applied in [139, 155] to q-ensembles.

There are, nevertheless, two simplifying limits in which the matrix model becomes tractable at large N:

- The $g^2 \to \infty$ limit. The interaction term in the LSZ action dominates, and the fields become non-dynamical. This corresponds to $g_s \to 0$, that is, $q \to 1$ from below. In this limit, we recover the Gaussian ensemble from the SW ensemble (up to an overall factor, readable from (3.3.12)), and the results of [139] directly apply to the present setting.
- The $g^2 \to 0$ limit. The quartic interaction should become tractable in standard perturbation theory in the LSZ field theory. From the Chern–Simons perspective, this corresponds to $g_s \to \infty$, equivalently $q \to 0$ from above, and simplifications take place.

In the first case, the eigenvalue density is [139]

$$\lim_{g_s \to 0^+} \rho_{\mathrm{SW}}(u) = \frac{1}{2\pi}\sqrt{\frac{L(z) - u}{u - z}}\,[L(z) - z + 2u]\,,$$

with $L(z) = \frac{2}{3}\sqrt{z^2 + 6} + \frac{1}{3}z$, and now $z \equiv \gamma/\sqrt{N}$. From [139, Eq. (58)] and a change of variables $x \mapsto \sqrt{2g_s}x$ we have

$$\begin{aligned}
&\ln \mathrm{Prob}_{\mathrm{CS}}(x_{\max} \le -\gamma) = \ln \mathrm{Prob}_{\mathrm{CS}}(x_{\min} \ge \gamma) \\
&\approx \frac{N^2}{2}\ln\frac{g_s}{2} - \frac{N^2}{54}\left[36z^2 - z^4 + z(z^2 + 15)\sqrt{z^2 + 6} + 27\left[\ln 18 - 2\ln\left(\sqrt{z^2 + 6} - z\right)\right]\right]
\end{aligned}$$

in the $N \to \infty$ and $g_s \to 0$ limit. This formula together with (3.3.13) describes the large N limit of the LSZ model in the strong interaction regime $g^2 \to \infty$.

The converse limit $g_s \to \infty$, that is, $g^2 \to 0$, can be analyzed as well. Changing variables $x \mapsto \sqrt{2g_s}x$ in the Chern–Simons ensemble and using

$$2\sinh\left(\sqrt{\frac{g_s}{2}}(x_j - x_k)\right) \approx \exp\left(\sqrt{\frac{g_s}{2}}|x_j - x_k|\right)$$

when $g_s \to \infty$, we get

$$\mathcal{Z}_{\rm LSZ}(\emptyset,\emptyset) = \frac{(-1)^N e^{\frac{\gamma^2 N}{2}}}{(4\pi)^{N^2}} \left(\frac{g_s}{2}\right)^{\frac{N^2}{2}} \int_{(-\infty,-\gamma]^N} \prod_{j=1}^N \mathrm{d}x_j \ \exp\left[-\sum_{j=1}^N x_j^2 + \sqrt{\frac{g_s}{2}} \sum_{j\neq k} |x_j - x_k|\right].$$

The latter expression describes the partition function of a constrained one-dimensional Coulomb gas [160, 161], corresponding to placing a hard wall at $x = -\gamma$ in a model due to Baxter [162]. Here we derive the large N asymptotics at leading order for this constrained model.

Assuming x grows as $x = \xi N^\alpha$ in the large N limit, for some $\alpha > 0$ and fixed ξ, the first term in the exponential grows as $N^{1+2\alpha}$, while the second term grows as $N^{2+\alpha}$. A non-trivial saddle point exists for $\alpha = 1$. Therefore, the large N limit in this large g_s regime is governed by the eigenvalue density $\rho_0(\xi)$ that solves the saddle point equation

$$\sqrt{\frac{g_s}{2}} \int_z^{L(z)} \mathrm{d}\xi' \ \rho_0(\xi') \mathrm{sign}\left(\xi - \xi'\right) = \xi,$$

with, this time, $z = \gamma/N$, for the scaling of γ with N to be consistent with the growth of the eigenvalues. Splitting the integral in two pieces, with $\xi' < \xi$ and $\xi' > \xi$ respectively, and taking the derivative of the saddle point equation in the interior of the domain, $z < \xi < L(z)$, we find

$$\rho_0(\xi) = \frac{1}{\sqrt{2g_s}}.$$

The normalization condition then imposes

$$\int_z^{L(z)} \mathrm{d}\xi \ \rho_0(\xi) = 1 \implies L(z) - z = \sqrt{2g_s}.$$

With this eigenvalue density we obtain

$$\ln\left[(-1)^N \mathcal{Z}_{\rm LSZ}(\emptyset,\emptyset)\right] \approx -N^3 \left[z^2 + z\sqrt{2g_s} + \frac{g_s}{3}\right]$$

at leading order in N and in the large g_s regime.

The complete solution will then interpolate between this two limiting situations. The N^2-behaviour will correspond to small 't Hooft coupling, $t \to 0$, while letting t grow linearly with N will give back the N^3-behaviour.

3.4 Solving the supermatrix model

The goal of the present section is to extend the analysis of the LSZ matrix model developed in Section 3.3 to the supermatrix model

$$\mathcal{Z}_{\rm sLSZ}\left(E,\tilde{E}\right) = \int \mathscr{D}M\mathscr{D}M^\dagger \ \exp\left(-(N_1+N_2)\,\mathrm{STr}\left\{MEM^\dagger + M^\dagger \tilde{E} M + \widehat{V}\left(M^\dagger M\right)\right\}\right)$$

introduced in (3.1.2). The integral is over complex supermatrices of size $(N_1+N_2)\times(N_1+N_2)$, and the external supermatrices $E, \tilde{E}$ have N_1+N_2 real eigenvalues each, which we write as $\frac{4\pi}{N_1+N_2}(\eta_\ell, \xi_r)$ and $\frac{4\pi}{N_1+N_2}\left(\tilde{\eta}_\ell, \tilde{\xi}_r\right)$ for E and $\tilde{E}$ respectively. We use indices $\ell = 1,\dots,N_1$ and $r = 1,\dots,N_2$. The potential $\widehat{V}$ has a quadratic and a quartic interaction, as in (3.3.1).

In the next subsection we present a few generalities about supermatrices, then we will extend the derivation of Section 3.3 to the sLSZ supermatrix model, this time establishing a connection with ABJ(M) theory.

3.4.1 Supermatrix models

We now introduce supermatrices and supermatrix models [163, 164]. We consider the general case of $(N_1+N_2)\times(N_1+N_2)$ supermatrices, which can be defined in the block form

$$M=\begin{pmatrix} A & \psi \\ \chi & B \end{pmatrix},$$

where A and B are respectively $N_1\times N_1$ and $N_2\times N_2$ complex matrices with bosonic entries, and ψ and χ are respectively $N_2\times N_1$ and $N_1\times N_2$ matrices with fermionic entries. The supermatrix M is acted on by the unitary supergroup $U(N_1|N_2)_{\text{left}}\times U(N_1|N_2)_{\text{right}}$. The supertrace operation is

$$\operatorname{STr} M=\operatorname{Tr} A-\operatorname{Tr} B,$$

and the integration measure $\mathscr{D}M\mathscr{D}M^\dagger$ is the product of Haar measures

$$\mathscr{D}M\mathscr{D}M^\dagger=\mathscr{D}A\mathscr{D}A^\dagger\mathscr{D}B\mathscr{D}B^\dagger\mathscr{D}\psi\mathscr{D}\chi.$$

We will need the supersymmetric version of the Harish-Chandra–Itzykson–Zuber formula (3.3.3), which reads [165, 166]

$$\begin{aligned}&\int_{U(N_1|N_2)}[\mathrm{d}U]\exp\left\{-(N_1+N_2)\operatorname{STr}(EUYU)\right\}\\&=\mathbf{C}_{N_1N_2}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(y_\ell-z_r)}{\Delta_{N_1}[y]\Delta_{N_2}[z]}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(\eta_\ell-\xi_r)}{\Delta_{N_1}[\eta]\Delta_{N_2}[\xi]}\det_{1\le\ell,\ell'\le N_1}\left(e^{-4\pi\eta_\ell y_{\ell'}}\right)\det_{1\le r,r'\le N_2}\left(e^{4\pi\xi_r z_{r'}}\right),\end{aligned}\tag{3.4.1}$$

where E and Y are supermatrices with eigenvalues $\frac{4\pi}{N}(\eta_\ell,\xi_r)$ and (y_ℓ,z_r) respectively, for $\ell=1,\dots,N_1$ and $r=1,\dots,N_2$, and the coefficient is

$$\mathbf{C}_{N_1N_2}=\frac{(4\pi)^{N_1N_2}}{(4\pi)^{\frac{N_1(N_1-1)}{2}}(-4\pi)^{\frac{N_2(N_2-1)}{2}}}G(N_1+1)G(N_2+1),$$

where $G(\cdot)$ is as usual the Barnes G-function. The factors Δ_N are Vandermonde determinants, introduced in (3.3.2). We note the appearance of the terms

$$\frac{\Delta_{N_1}[y]\Delta_{N_2}[z]}{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(y_\ell-z_r)},$$

and likewise for (η,ξ), which is the superdeterminant (called Berezinian) version of the Vandermonde. When $N_2=0$, formula (3.4.1) reduces to the well known HCIZ formula (3.3.3).

3.4.2 General solution

We now focus on the supermatrix model $\mathcal{Z}_{\text{sLSZ}}$ defined in (3.1.2), and solve it as we have done in Section 3.3 for the LSZ model.

We first rewrite the partition function as[8]

$$\begin{aligned}\mathcal{Z}_{\text{sLSZ}}\left(E,\tilde{E}\right)=&\int_{[0,\infty)^{N_1}}\int_{[0,\infty)^{N_2}}\frac{\Delta_{N_1}[y]^2\,\Delta_{N_2}[z]^2}{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(y_\ell-z_r)^2}\prod_{\ell=1}^{N_1}e^{-\widehat{m}^2y_\ell-\frac{\widehat{g}^2}{2}y_\ell^2}\mathrm{d}y_\ell\prod_{r=1}^{N_2}e^{\widehat{m}^2z_r+\frac{\widehat{g}^2}{2}z_r^2}\mathrm{d}z_r\\&\times\int_{U(N_1|N_2)}[\mathrm{d}U_2]\exp\left(-NEU_2^\dagger YU_2\right)\int_{U(N_1|N_2)}[\mathrm{d}U_1]\exp\left(-N\tilde{E}U_1^\dagger YU_1\right),\end{aligned}$$

[8] The Jacobian is the squared Vandermonde Berezinian, see [163, Sec. 4] and references therein.

and then apply the supersymmetric HCIZ formula (3.4.1) twice. As in Section 3.3, we simplify the Jacobian with the denominator coming from (3.4.1), and obtain

$$\begin{aligned}\mathcal{Z}_{\text{sLSZ}}\left(E,\tilde{E}\right) &= \mathbf{C}^2_{N_1N_2}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(\eta_\ell-\xi_r)(\tilde{\eta}_\ell-\tilde{\xi}_r)}{\Delta_{N_1}[\eta]\Delta_{N_1}[\xi]\Delta_{N_2}[\tilde{\eta}]\Delta_{N_2}[\tilde{\xi}]}\\ &\times\int_{[0,\infty)^{N_1}}\prod_{\ell=1}^{N_1}e^{-\widehat{m}^2y_\ell-\frac{\widehat{g}^2}{2}y_\ell^2}\mathrm{d}y_\ell\ \det_{1\le\ell,\ell'\le N_1}\left(e^{-4\pi\eta_\ell y_{\ell'}}\right)\det_{1\le\ell,\ell'\le N_1}\left(e^{-4\pi\tilde{\eta}_\ell y_{\ell'}}\right)\\ &\times\int_{[0,\infty)^{N_2}}\prod_{r=1}^{N_2}e^{\widehat{m}^2z_r+\frac{\widehat{g}^2}{2}z_r^2}\mathrm{d}z_r\ \det_{1\le r,r'\le N_2}\left(e^{4\pi\xi_r z_{r'}}\right)\det_{1\le r,r'\le N_2}\left(e^{4\pi\tilde{\xi}_r z_{r'}}\right).\end{aligned}$$

We now assume the eigenvalues of the external supermatrices $E,\tilde{E}$ are $\frac{4\pi}{N_1+N_2}$ times integers, and we rewrite them in the form

$$\begin{aligned}\eta_\ell &= \mu_{1;\ell}+N_1-\ell, &\qquad \xi_r &= \mu_{2;r}+N_2-r\\ \tilde{\eta}_\ell &= \nu_{1;\ell}+N_1-\ell, &\qquad \tilde{\xi}_r &= \nu_{2;r}+N_2-r,\end{aligned}\tag{3.4.2}$$

which is the obvious extension of (3.3.7). In our conventions, the partitions with index 1 have rows labelled by $\ell=1,\dots,N_1$ and partitions with index 2 have rows labelled by $r=1,\dots,N_2$. We again recognize the Schur polynomials,

$$\begin{aligned}\mathcal{Z}_{\text{sLSZ}}\left(E,\tilde{E}\right) &= \mathbf{C}'_{N_1N_2}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(\eta_\ell-\xi_r)(\tilde{\eta}_\ell-\tilde{\xi}_r)}{\Delta_{N_1}[\eta]\Delta_{N_1}[\xi]\Delta_{N_2}[\tilde{\eta}]\Delta_{N_2}[\tilde{\xi}]}\\ &\times\int_{[0,\infty)^{N_1}}\Delta_{N_1}[e^y]^2\ s_{\mu_1}(e^{-y_1},\dots,e^{-y_{N_1}})s_{\nu_1}(e^{-y_1},\dots,e^{-y_{N_1}})\prod_{\ell=1}^{N_1}e^{-m^2y_\ell-\frac{g^2}{2}y_\ell^2}\mathrm{d}y_\ell\\ &\times\int_{[0,\infty)^{N_2}}\Delta_{N_2}[e^z]^2\ s_{\mu_2}(e^{z_1},\dots,e^{z_{N_2}})s_{\nu_2}(e^{z_1},\dots,e^{z_{N_2}})\prod_{r=1}^{N_2}e^{m^2z_r+\frac{g^2}{2}z_r^2}\mathrm{d}z_r\end{aligned}$$

where we defined the parameters m and g as in (3.3.6) to reabsorb the factor 4π, and

$$\mathbf{C}'_{N_1N_2}=(4\pi)^{-(N_1+N_2)}\mathbf{C}^2_{N_1N_2}=\frac{G(N_1+1)^2G(N_2+1)^2}{(4\pi)^{(N_1-N_2)^2}}.$$

We see that the present setting closely resembles what we obtained in Section 3.3, now with two sets of integration variables and two pairs of Schur polynomials. In fact, the integrals over the two sets of variables are factorized, and we may give the result as a product of two copies of the result in Section 3.3.

However, we follow a different path and assemble the two pairs of Schur polynomials into two supersymmetric Schur polynomials labelled by representations of the supergroup $U(N_1|N_2)$, see [47, 167] for definitions and properties. Irreducible $U(N_1|N_2)$ representations are classified in typical and atypical, and here we need the typical ones. The two typical representations that appear in our computations correspond to the Young diagrams

$$\begin{aligned}\mu &= (\kappa+\mu_1)\sqcup\lambda,\\ \nu &= (\tilde{\kappa}+\nu_1)\sqcup\tilde{\lambda},\end{aligned}$$

where κ and $\tilde{\kappa}$ are rectangular $N_1\times N_2$ diagrams. This means that the Young diagram μ is composed by a rectangle κ, a Young diagram μ_1 on the right of it and another diagram λ below it,

and analogously for the Young diagram ν (see Subsection 2.4.2). For these typical representations, the supersymmetric Schur polynomials decompose as [47, 167]

$$S_\mu(e^{y_1},\dots,e^{y_{N_1}}|e^{z_1},\dots,e^{z_{N_2}}) = s_{\mu_1}(e^{y_1},\dots,e^{y_{N_1}})s_{\lambda'}(e^{z_1},\dots,e^{z_{N_2}})\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(e^{y_\ell}+e^{z_r}),$$
$$S_\nu(e^{y_1},\dots,e^{y_{N_1}}|e^{z_1},\dots,e^{z_{N_2}}) = s_{\nu_1}(e^{y_1},\dots,e^{y_{N_1}})s_{\tilde\lambda'}(e^{z_1},\dots,e^{z_{N_2}})\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(e^{y_\ell}+e^{z_r}),$$

where S_μ, S_ν are the supersymmetric Schur polynomials and $\lambda',\tilde\lambda'$ are the conjugate partitions to $\lambda,\tilde\lambda$. Therefore, identifying the generic $\lambda,\tilde\lambda$ to be μ_2' and ν_2' in our case, and following the same manipulations as in Section 3.3, we rewrite $\mathcal{Z}_{\rm sLSZ}$ as

$$\begin{aligned}\mathcal{Z}_{\rm sLSZ}(\mu,\nu) = \mathbf{A}_{N_1,N_2}(\mu,\nu)\times\int_{(-\infty,-\gamma_1]^{N_1}}\int_{(\gamma_1,\infty]^{N_2}}\prod_{a=1}^{N_1}e^{-\frac{g^2}{2}x_a^2}\mathrm{d}x_a\prod_{\dot a=1}^{N_2}e^{\frac{g^2}{2}w_{\dot a}^2}\mathrm{d}w_{\dot a}\\ \times\frac{\prod_{1\le a<b\le N_1}\left(2\sinh\left(\frac{x_a-x_b}{2}\right)\right)^2\ \prod_{1\le\dot a<\dot b\le N_2}\left(2\sinh\left(\frac{w_{\dot a}-w_{\dot b}}{2}\right)\right)^2}{\prod_{a=1}^{N_1}\prod_{\dot b=1}^{N_2}\left(2\cosh\left(\frac{x_a-w_{\dot b}}{2}\right)\right)^2}\\ \times S_\mu(e^{x_1},\dots,e^{x_{N_1}}|e^{w_1},\dots,e^{w_{N_2}})S_\nu(e^{x_1},\dots,e^{x_{N_1}}|e^{w_1},\dots,e^{w_{N_2}}),\end{aligned}\tag{3.4.3}$$

with $\gamma_\alpha = m^2 - N_\alpha + 1$, $\alpha = 1,2$, and the overall coefficient being

$$\begin{aligned}\mathbf{A}_{N_1,N_2}(\mu,\nu) &= (-1)^{N_1}\mathbf{C}'_{N_1N_2}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(\eta_\ell-\xi_r)(\tilde\eta_\ell-\tilde\xi_r)}{\Delta_{N_1}[\eta]\Delta_{N_1}[\xi]\Delta_{N_2}[\tilde\eta]\Delta_{N_2}[\tilde\xi]}\\ &\quad\times\exp\left(\frac{\gamma_1^2N_1-\gamma_2^2N_2}{2}+\gamma_1\left(|\mu_1|+|\nu_1|\right)+\gamma_2\left(|\mu_2|+|\nu_2|\right)\right)\\ &= (-1)^{N_1}\frac{\prod_{\ell=1}^{N_1}\prod_{r=1}^{N_2}(\eta_\ell-\xi_r)^2(\tilde\eta_\ell-\tilde\xi_r)^2}{(4\pi)^{(N_1-N_2)^2}\dim\mu\ \dim\nu}\\ &\quad\times\exp\left(\frac{\gamma_1^2N_1-\gamma_2^2N_2}{2}+\gamma_1\left(|\mu_1|+|\nu_1|\right)+\gamma_2\left(|\mu_2|+|\nu_2|\right)\right).\end{aligned}$$

For the second equality, we have written the contribution from the eigenvalues of the external supermatrices $E,\tilde E$ in terms of the partitions μ,ν, and noted again that the Barnes G-functions coming from the dimensions of the representations cancel with those arisen from the supersymmetric HCIZ formula, precisely as in the ordinary matrix model of Section 3.3.

To arrive at (3.4.3) we have first shifted both sets of variables and then reflected the first set. If we go back to equation (3.4.2) and define $\mu_{2;r},\nu_{2;r}$ from $\xi_r,\tilde\xi_r$ with opposite sign, we get to an expression analogous to (3.4.3) but with both $(x_1,\dots,x_{N_1})$ and $(w_1,\dots,w_{N_2})$ integrated over the same domain. In each case, we have to first change variables and then insert the supersymmetric Schur factorization identity.

Comparing with [107], we rewrite the expression (3.4.3) in the form

$$\mathcal{Z}_{\rm sLSZ}(\mu,\nu) = \mathbf{A}_{N_1,N_2}(\mu,\nu)\ \mathcal{Z}_{\rm ABJ}\ \left\langle W_\mu W_\nu\ \mathbb{1}_{(-\infty,-\gamma_1]^{N_1}}(x)\mathbb{1}_{[\gamma_2,\infty)^{N_2}}(w)\right\rangle_{\rm ABJ},\tag{3.4.4}$$

where $\langle\cdots\rangle_{\rm ABJ}$ is the average in the ABJ matrix model (3.1.4), and $\mathbb{1}_{(-\infty,-\gamma_1]^{N_1}}(x)$ (respectively $\mathbb{1}_{[\gamma_2,\infty)^{N_2}}(w)$) is the N_1-dimensional (N_2-dimensional) indicator function. Besides, W_μ is the trace of the holonomy of a superconnection along an equatorial circle inside $\mathbb{S}^3$, in the representation μ of the supergroup $U(N_1|N_2)$, which describes a Wilson loop in ABJ(M) theory [73, 82]. Previous

works on averages of supersymmetric Schur polynomials over the ABJ(M) ensemble include [168–170].

The probabilistic interpretation of the result (3.4.4) in terms of large deviations away from the equilibrium in the ABJ(M) matrix model with supersymmetric Schur insertions, follows directly from the discussion in Subsection 3.3.4. In particular

$$\frac{\mathcal{Z}_{\text{sLSZ}}(\mu,\nu)}{\mathcal{Z}_{\text{ABJ}}} = \mathbf{A}_{N_1,N_2}(\mu,\nu)\ \langle W_\mu W_\nu\rangle_{\text{ABJ}}\ \text{Prob}_{\text{ABJ};W_\mu W_\nu}(x_{\max}\leq -\gamma_1, w_{\min}\geq\gamma_2),$$

hence the supermatrix model (3.1.2) measures the probability of atypically large deviations from the equilibrium in the random matrix description of two supersymmetric Wilson loops carrying two $U(N_1|N_2)$ typical representations. The specialization to two void partitions implies

$$\frac{\mathcal{Z}_{\text{sLSZ}}(\emptyset,\emptyset)}{\mathcal{Z}_{\text{ABJ}}} = \mathbf{A}_{N_1,N_2}(\emptyset,\emptyset)\ \text{Prob}_{\text{ABJ}}(x_{\max}\leq -\gamma_1, w_{\min}\geq\gamma_2).$$

Remark. Throughout this section we defined and analyzed a supermatrix version of the LSZ model. It would be interesting to derive the supermatrix model (3.1.2) from an extension of the LSZ scalar theory to a noncommutative superspace.

3.A Schur polynomials

Here we include explicit formulas involving Schur polynomials [90].

Spectral shift and rectangular Schur

A simple identity of Schur polynomials quickly shows what occurs if, in the case of a equispaced, harmonic oscillator spectra, we have a global overall shift in the spectrum (that is, a different zero point energy). If we have a rectangular partition of length N, $(l,l,...,l)$ which we denote by l^N then, assuming that λ is a partition of length equal to or lower than N, it holds

$$s_{\mu+l^N}(e^{x_1},...,e^{x_N}) = \left(\prod_{a=1}^{N} e^{l x_a}\right) s_\mu(e^{x_1},...,e^{x_N}),$$

Therefore, an overall spectral shift by an integer l in one external matrix, corresponds to a renormalization of the mass parameter $\widehat{m}^2 \mapsto \widehat{m}^2 - l$.

Another useful identity is [90]

$$s_\mu(ce^{x_1},\dots,ce^{x_N}) = c^{|\mu|}s_\mu(e^{x_1},\dots,e^{x_N}). \tag{3.A.1}$$

Dimensions

The value of $s_\mu(1,\dots,1)$ gives the dimension of the irreducible representation of $U(N)$ with highest weight μ. Weyl's denominator formula states that

$$s_\mu(\underbrace{1,...,1}_{N}) = \prod_{1\leq a<b\leq N}\left(\frac{\mu_a - a - \mu_b + b}{a-b}\right) = \frac{1}{G(N+1)}\prod_{1\leq a<b\leq N}(\mu_a-\mu_b+b-a),$$

where $G(\cdot)$ is the Barnes G-function.

Part II

Large N phase transitions in matrix models and gauge theories

Chapter 4

Prologue to Part II

This chapter serves as an introduction to Part II, which is devoted to the study of the large N limit of a wealth of matrix models, with focus on the implications from the point of view of phase transitions and critical phenomena [33, 34]. We begin by discussing the large N limit in some generality. Then, we give an extremely brief review of the theory of phase transitions. We conclude this overview recalling the analogy between matrix ensembles and log gases in Section 4.4.

4.1 On large N limits

Let us begin with a lightning introduction to the large N limit, with focus on gauge theories with simple compact gauge group for concreteness.

A standard approach in QFT is to compute correlation functions in a perturbative expansion in the coupling constants of the model. Higher orders in perturbation theory correspond to a higher number of loops in the Feynman diagrammatic expansion. Therefore, by construction, any computation in perturbation theory gives as output a Maclaurin expansion around a free field theory. Assume now that one is interested in a gauge theory, and denote by N the rank of the gauge group. Instead of expanding in the coupling constants, consider the Maclaurin expansion of the physical observables in the parameter $1/N$. Each order in $1/N$ includes contributions from Feynman diagrams with any number of loops, thus remaining valid away from the weak coupling regime. An additional advantage of large N is that the expansion parameter is not renormalized.

Depending on the specific model under consideration, there may exist different ways to take a large N limit. However, two of them stand out, which always exist in theories with matrix degrees of freedom. Throughout the dissertation we adopt a QFT-oriented nomenclature to distinguish these two limiting procedures:

- the *strict* large N limit, in which the number N of degrees of freedom is sent to infinity keeping all other parameters fixed;
- the *'t Hooft* large N limit [17–19], in which the number N of degrees of freedom is sent to infinity and, at the same time, the couplings are tuned in such a way that all contributions are kept of the same order in N.

In the context of unitary matrix models, the strict limit is also sometimes referred to as Szegő limit, because the result is given by the celebrated Szegő theorem [171]. Besides, in the mathematics literature the 't Hooft limit is oftentimes simply referred to as large N limit. As a further remark, note that here we include in the 't Hooft limit various limiting procedures that may have different

physical meaning. For example, the Veneziano limit is nothing but a special case of 't Hooft limit from our standpoint.

The 't Hooft limit is in fact more general, with the strict limit recovered as a first order approximation. In absence of phase transitions, we can nevertheless extract the results in the 't Hooft limit from those in the strict limit by analytic continuation. This aspect is more extensively elaborated upon in [11].

Other limits, such as the M-theory limit [172], are recovered at strong or weak coupling from the 't Hooft limit, although more precise asymptotics can be obtained by directly working in the appropriate regime [172, 173].[9] We refer to the lecture notes [175] for a thorough discussion on the large N limits in various models, to [27] for an overview of the techniques to solve matrix models at large N, to [176] for the specialization of these techniques to q-ensembles, that play a central role in the dissertation, and finally to [177, 178] for the recovery of the M-theory limit as a 't Hooft limit at strong coupling.

A further limit of crucial importance and with far-reaching physical applications is the double-scaling limit. Originally pioneered to study two-dimensional gravity, it has been introduced for Hermitian [179–181] and unitary matrix models [182], and later refined and applied to myriad situations related to random matrix theory. A notable achievement is the solution of the long-standing problem of the longest increasing subsequence of a random permutation [183] (see [184] for an overview of this subject).

Succinctly, the double-scaling consists in taking the 't Hooft limit while concurrently tuning a control parameter towards one of its critical values. It is customary to take the control parameter to be a physically meaningful parameter, such as the 't Hooft gauge coupling, although it can in principle be chosen to describe an arbitrary curve in parameter space. As a side remark, we note that in the physics literature it is sometimes misleadingly termed double-scaling limit a 't Hooft limit with multiple 't Hooft parameters. As clarified above, and repeatedly mentioned in the following, one should think of the latter procedure as the proper way of taking the 't Hooft limit, whereas the double-scaling is a more subtle limit, not guaranteed to exist in general, which explores a distinct physical regime and hence yields different physical consequences [179–181].

It is worthwhile to emphasize that, with the exception of Chapter 7, our large N will always refer to a large number of colours in the physical parlance. That is, we deal with integrals over rank-N matrix degrees of freedom, dictated by the gauge group, and work in the limit of large matrix size [18, 19]. The phrasing "large N limit" is, however, equally used to indicate the vector large N limit, following the early literature on the $O(N)$ model [33, 185]. In such case, one has fields that transform in the vector representation of $O(N)$ or in the fundamental representation of $U(N)$, the groups being global symmetries of the theory and not gauge groups. Again we refer to [175] for a comparison of the various large N limits.

The large N limit considered in Chapter 7 is a vector limit, as opposed to the rest of the manuscript. We keep the gauge group of small rank but work in the limit of a large flavour symmetry, taking a 't Hooft-type scaling.[10]

[9] It is fair to say that one not seldom has to take additional perturbative expansions in a 't Hooft coupling, to solve the large N limit in closed form. Therefore, the comparison between the 't Hooft and M-theory limits is usually an uneasy task, because of the thin overlap of the regions of validity for the assumptions made on the two sides. A successful comparison of the 't Hooft and M-theory limit in the large N Chern–Simons-matter quivers of Chapter 2 has been worked out in [174].

[10] We stress the conceptual similarity with recent works [186–189] discussing the vector large N limit of the $O(N)$ model (or of its $U(N)$ analogue) in a 't Hooft-type limit with the central charge scaled linearly with N. In the case of interest to us, extended supersymmetry allows us to obtain exact results in the whole parameter space, as well as to compute $1/N$ corrections.

4.2 $SU(N)$ versus $U(N)$

Standard lore claims that $SU(N)$ and $U(N)$ are indistinguishable at leading order in the $1/N$ expansion. One of the main results of this thesis is to defy this belief.

The statement is supported by a solid argument, that we now formulate in three ways. Let us work at the level of Lie algebras, $\mathfrak{u}(N) \cong \mathfrak{su}(N) \oplus \mathfrak{u}(1)$, in order to neglect the global structure of the gauge group.

(i) The difference between $\mathfrak{u}(N)$ and $\mathfrak{su}(N)$ is a diagonal $\mathfrak{u}(1) \subset \mathfrak{u}(N)$, whose contribution is negligible compared to the rest of the gauge group.

(ii) In a diagrammatic $1/N$ expansion, the only difference between $\mathfrak{u}(N)$ and $\mathfrak{su}(N)$ comes from the propagator of a gauge-singlet meson. The meson propagator can be attached to matter legs in the diagrammatic expansion, but not to gauge boson legs, thus it will only appear in diagrams suppressed by a factor $1/N$.

(iii) The difference between $\mathfrak{u}(N)$ and $\mathfrak{su}(N)$ is the tracelessness condition on the latter. This can be enforced at the level of Lagrangian density by introducing a Lagrange multiplier (shown here for simplicity in the case of a scalar field ϕ in the adjoint representation of $\mathfrak{u}(N)$)

$$\xi \mathrm{Tr}(\phi), \tag{4.2.1}$$

which only grows linearly in N, thus suppressed by a factor $1/N$ compared to the contribution from the adjoint representation of $\mathfrak{su}(N)$.

Point (ii) is the particle theorist's rephrasing of (i), while point (iii) presents a pragmatic approach useful for computations, especially in matrix models.

We now look deeper into the statements above. To take the large N limit of $\mathfrak{u}(N)$ corresponds, at leading order, to work in the direct limit Lie algebra $\mathfrak{u}(\infty)$ defined as the inductive limit of

$$\mathfrak{u}(1) \subset \cdots \subset \mathfrak{u}(N) \subset \mathfrak{u}(N+1) \subset \cdots$$

In this regime, the fields attain a saddle point value that minimizes the effective action, which is a configuration in $\mathfrak{u}(\infty)$ also known as master field.

In general, though, the saddle point equation (or master field equation) for $\mathfrak{u}(\infty)$ needs not admit a traceless solution. Therefore, to simply neglect the difference between $\mathfrak{u}(N)$ and $\mathfrak{su}(N)$ may, and sometimes will, as shown in Chapters 5-6, result in a non-traceless solution for a $\mathfrak{su}(N)$ problem.

The puzzle is rooted in the definition of what is meant by $\mathfrak{su}(\infty)$. If we define it to be the same as $\mathfrak{u}(\infty)$, then the statement that $U(N)$ and $SU(N)$ agree at leading order is tautologically correct, but we have to accept that physical observables in a $SU(N)$ gauge theory are dominated by non-traceless field configurations.

A reasonable alternative is to define $\mathfrak{su}(\infty)$ to be the traceless subspace of $\mathfrak{u}(\infty)$. This is the definition we take on in this work.

In order to keep track of the traceless condition on $\mathfrak{su}(N)$ at large N, we redefine the Lagrange multiplier in (4.2.1) as $\xi = N\tilde{\xi}$, with $\tilde{\xi}$ fixed. While this may seem ad hoc, let us emphasize that ξ is a background field whose normalization is completely arbitrary. Moreover, the crucial argument in favour of the 't Hooft scaling of the Lagrange multiplier ξ comes from the seminal work [18]. Let us focus again on our scalar example for simplicity, and study the $\mathfrak{u}(N)$-valued field ϕ. It was shown in [18] that the proper normalization of the Lagrangian density in the 't Hooft large N limit is $N\mathrm{Tr}(V(\phi))$, where $V(\cdot)$ is an arbitrary polynomial with coefficients independent of N. From that point of view, there is no reason not to add a linear term $\tilde{\xi}\phi$ to $V(\phi)$. The saddle point

configuration will then depend on $\tilde{\xi}$, and it must be fixed to a value that extremizes the effective action.

To sum up, we propose that the Lagrange multiplier enforcing the tracelessness condition on $\mathfrak{su}(N)$-valued fields should be scaled in a 't Hooft way, to guarantee that the large N saddle point configuration in a $\mathfrak{su}(N)$ gauge theory is traceless. Not only the procedure is self-consistent, but it is natural from a mathematical point of view when dealing with random matrix ensembles.

4.3 On phase transitions

The goal of this section is to give a very short account of the theory of phase transitions in QFT. For an extensive discussion we refer to the monograph [34].

Phase transitions are a hallmark of theories with infinitely many degrees of freedom. Whenever a computation in a QFT reduces to a matrix model, i.e. to zero spacetime dimensions, phase transitions can only appear at large N.

A phase transition is the property that, depending on the value of a control parameter λ, a system shows different responses if the large N limit is taken in different ways. Order parameters discern among the phases. The presence of a phase transition corresponds to a new saddle point configuration for the fields appearing as the control parameter λ crosses a critical value $\lambda_{\rm cr}$.

1) The typical situation that leads to a first order phase transition is an effective action $S_{\rm eff}$ that has several local minima. One of them, the absolute minimum, will yield the leading contribution at large N. Varying λ, the value of $S_{\rm eff}$ at one of the local minima can decrease, so that the corresponding point in field space becomes the new absolute minimum, as sketched in the left panel of Figure 4.1. The system experiences a sharp transition in this case, with the leading contribution to physical observables jumping discontinuously in field space. Order parameters are then discontinuous at this transition.

2) Second order phase transitions are associated with spontaneous symmetry breaking. The prototypical setup is a model in which $S_{\rm eff}$ has only one minimum for $\lambda < \lambda_{\rm cr}$. However, when $\lambda > \lambda_{\rm cr}$, new absolute minima emanate from the previous saddle point configuration. This is represented in the right panel of Figure 4.1. In this case, the order parameters are continuous at $\lambda = \lambda_{\rm cr}$, because the two saddle point configurations in the two phases are not at finite distance. However, their derivatives will not be continuous.

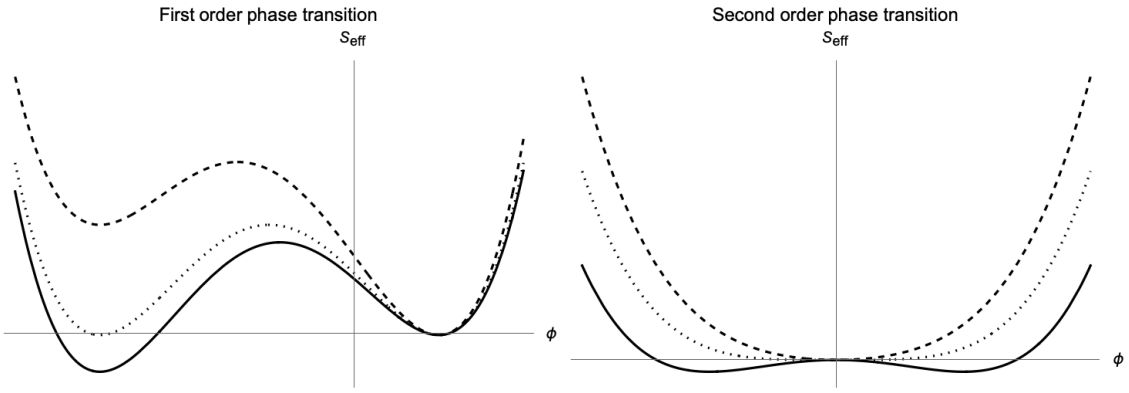


Figure 4.1. Cartoon representation of a first (left) and a second (right) order phase transition. The three curves correspond to $\lambda < \lambda_{\rm cr}$ (dashed), $\lambda = \lambda_{\rm cr}$ (dotted), and $\lambda > \lambda_{\rm cr}$ (solid).

3) Third order phase transitions are ubiquitous in matrix models, and they can be triggered by various and diverse mechanisms.

- The effective action S_{eff} does not change its shape as λ crosses λ_{cr}. Nevertheless, the minimization problem is constrained by additional conditions, which cannot be fulfilled by the absolute minimum of S_{eff} at $\lambda > \lambda_{\text{cr}}$. The system is then forced into a new phase, dominated by the configuration that minimizes S_{eff} in the region allowed by the constraints. Examples of phase transitions induced by this mechanism include:
  - ▷ the Gross–Witten–Wadia transition [190–192], in which the constraint originates from the compactness of the integration domain, that upper bounds the distance between two degrees of freedom;
  - ▷ the Douglas–Kazakov transition [193], in which the constraint originates from the discreteness of the integration domain, that bounds below the distance between two degrees of freedom;
  - ▷ the soft-edge to hard-edge transition common in Coulomb gases with a hard wall [160].

  Transitions of these types are prominent in Chapters 8-9.

  Clearly, a model may be subject to more than one constraint, which generically leads to a number of phase transitions and to a richer phase diagram. It could also happen, though, that one of the constraints is always satisfied, an explicit example will be briefly discussed in Chapter 8.
- Fields that are heavy for $\lambda < \lambda_{\text{cr}}$, and thus are integrated out and contribute a one-loop determinant to S_{eff}, become light at $\lambda > \lambda_{\text{cr}}$. The saddle point configuration is reorganized accordingly. These transitions feature in Chapters 5-6.

4.4 Log gases

Random matrix theory is, concisely, the study of correlation functions

$$\langle f \rangle_N = \frac{1}{N! \mathcal{Z}_N} \int_{\mathcal{C}^N} f(u_1, \dots, u_N) e^{-N^2 S_{\text{eff}}(u)} \prod_{a=1}^{N} \mathrm{d}u_a,$$
$$\mathcal{Z}_N = \frac{1}{N!} \int_{\mathcal{C}^N} \mathrm{d}^N u e^{-N^2 S_{\text{eff}}(u)},$$

where the characteristics of the underlying physical system are encoded in the integration cycle $\mathcal{C}$ and in the effective action S_{eff}.

It is instructive to think of such matrix ensembles as one-dimensional Coulomb systems of N particles, whose positions are represented by the eigenvalues u_a, $a = 1, \dots, N$ [40]. In this analogy, the vast majority of matrix models discussed throughout correspond to log gases of particles that interact only through their mutual Coulomb forces, and are immersed in a background confining potential $V(u)$. Explicitly,

$$S_{\text{eff}}(u) = \frac{\beta}{N^2} \sum_{a=1}^{N} V(u_a) - \frac{1}{N^2} \sum_{1 \le a \ne b \le N} \Phi(|u_a - u_b|),$$

with β the inverse temperature. Φ encodes the one-dimensional Coulomb interaction [40], and thus is essentially determined by the topology of $\mathcal{C}$. We are mainly interested in three cases:

(i) $\mathcal{C} = \mathbb{R}$ and $\Phi(|u_a - u_b|) = \log|u_a - u_b|$, describing a Coulomb gas on the real line. In fact, in Part III we will meet a variation of this setup, in which $\mathcal{C} = \mathbb{Z}$, so the eigenvalues, hence the particles in the log gas analogy, are constrained to live on a lattice and cannot get closer than the lattice spacing.

(ii) $\mathcal{C} = \mathbb{S}^1$ and $\Phi(|u_a - u_b|) = \log|u_a - u_b|$. In this case, it is convenient to change the measure to $\frac{\mathrm{d}u_a}{2\pi \mathrm{i} u_a}$. The change of variables $u_a = e^{\mathrm{i}\theta_a}$ yields $\Phi(|u_a - u_b|) = \log 2|\sin\frac{\theta_a - \theta_b}{2}|$.

(iii) The particles live on a two-dimensional cylinder, but are forced to stay on a line. In this case $\mathcal{C} = \mathbb{R}$, but the cylindrical Coulomb interaction reads $\Phi(|u_a - u_b|) = \log 2|\sinh\frac{u_a - u_b}{2}|$ [40]. This is the log gas model that corresponds to Chern–Simons theory on $\mathbb{S}^3$ [133, 84].

Let us now review the large N limit from the log gas point of view. Taking the number of particles to be large at finite temperature, $0 < \beta < \infty$ fixed, the Boltzmann factor $e^{-\beta V(u)}$ suppresses the excited configurations, and the thermodynamic equilibrium is dominated by the repulsive Coulomb interaction. The particles are spread to infinity with a large separation among each other, if $\mathcal{C}$ is unbounded, or are distributed with equal spacing if $\mathcal{C}$ is bounded. For the special case $\mathcal{C} = \mathbb{S}^1$, the average density of particles is $\frac{1}{2\pi}$.

A more interesting limit corresponds to drive the system to zero temperature at the same rate as $N \to \infty$, that is, we keep $t = \frac{N}{\beta}$ fixed. This is precisely the 't Hooft limit introduced in [18] and discussed in Section 4.1. In this procedure, the equilibrium configuration is determined by the competition between the confining potential $V(u)$, which tends to gather the particles at its bottom, and the Coulomb repulsion, which tends to spread the particles far away from each other. The resulting equilibrium is described by particles at separated points that fill the wells around the minima of the potential. Increasing the couplings in the potential shrinks the width of the region onto which the particles are spread.

It is clear that the topology of $\mathcal{C}$ may introduce additional constraints on the equilibrium configuration. Consider for instance the setup of point (ii), so $\mathcal{C} = \mathbb{S}^1$, at very small t. The particles are gathered in a neighbourhood of the absolute minimum of the potential. As we increase the control parameter t, the width of the interval on which the particles are placed increases, due to the dominance of the repulsive force. Eventually, at $t = t_{\mathrm{cr}}$, the particles cover the whole $\mathbb{S}^1$. They cannot be spread further apart for $t > t_{\mathrm{cr}}$, thus the resulting density of particles is a deformation of the uniform density that takes into account the attraction toward the minimum of the potential. The transition between the two types of equilibrium is the Gross–Witten–Wadia transition [190, 192], or a generalization thereof.

An alternative example is provided by the setup of point (i), with $\mathcal{C} = \mathbb{Z}$. Taking $t \to 0$ to push the particles towards the minimum of $V(u)$, the equilibrium configuration breaks down and the particles must reorganize themselves when the average distance becomes smaller than the lattice spacing. This mechanism is the log gas view of the Douglas–Kazakov transition [193, 194] and its generalizations.

To summarize, the large N 't Hooft limit in matrix models corresponds to the thermodynamic limit of a one-dimensional Coulomb gas, which is simultaneously tuned to zero temperature. The topology of the space on which the gas lives may induce phase transitions.

Chapter 5

Phases of three-dimensional supersymmetric Chern–Simons-matter theories

5.1 Introduction to the chapter

The study of supersymmetric gauge theories in compact manifolds has experienced great progress on the last decade, due to the development of the localization method [56], which leads to a much simplified description of the original functional integral describing the observables of the gauge theories. The resulting object to analyze is a finite-dimensional integral of random matrix type.

An important stream of research, that emerged by analyzing such matrix model description, brought upon by localization, is the discovery and characterization of large N quantum phase transitions of supersymmetric gauge theories on spheres, typically in the large radius limit. Aspects of this research line have been especially investigated in four-dimensional $\mathcal{N} = 2$ theories, where it was found to be quite a generic feature of theories with massive matter [195–201]. Exceptions to this pattern are the superconformal theories, such as four-dimensional $\mathcal{N} = 2$ $SU(N)$ theory with $N_f = 2N$ hypermultiplets in the fundamental representation. This peculiarity is indeed expected on physical grounds, as we will explain below.

It was observed in these works that the quantum phase transitions originate from resonances emerging when the 't Hooft coupling is such that the saddle points of the action hit a singularity of the Coulomb branch, where massless modes appear. In other words, there are critical values for the coupling such that, when crossed, field configurations with extra massless multiplets contribute to the saddle-point, leading to non-analytic behaviour of supersymmetric observables, such as the partition function or the vacuum expectation value (vev) of Wilson loops.

In this chapter we shall study three-dimensional $\mathcal{N} > 2$ theories on $\mathbb{S}^3$ [58]. In three dimensions, $U(N)$ Chern–Simons-matter theories have been studied at large N and large volume in [202, 62, 203–207]. Results in five dimensions have been obtained in [208, 4].

We will follow and extend [202], studying the phase structure of $U(N)$ Chern–Simons-matter with massive fundamental hypermultiplets, but for arbitrary masses of the hypermultiplets and with a Fayet–Iliopoulos (FI) parameter turned on. Then, we analyze vevs of Wilson loops in the fundamental or antisymmetric representation at large N, adopting the now standard and widespread technique described in [209] for the latter. This approach is based on studying the matrix model average of the generating function of the Wilson loops in antisymmetric representations. In mathematical terms, these are the generating functions of elementary symmetric polynomials. The study of Wilson loops in large representations is comparatively less developed

than the case of the Wilson loop in the fundamental representation, although in four dimensions, there are a number of works in the higher-rank case in the last years [210, 211]. The case of a three dimensional Chern–Simons theory is understudied in comparison and we tackle it here. We then have an additional scaling parameter to play with, namely $\kappa = K/N$, where K denotes the rank of the Wilson loop representation. As explained in [210, 211], the existence of this parameter makes these Wilson loops good probes of the critical behaviour of the theory.

We push our analysis beyond $\mathcal{N} = 3$ Chern–Simons matter theories and discuss three-dimensional $\mathcal{N} = 4$ SQCD. We study this theory in a twofold approach: either directly at large radius, taking a suitable limit for the Chern–Simons coupling, or by first solving the saddle point equation at large N and finite radius. The two derivations are shown to agree in the large radius limit. To our knowledge, the eigenvalue density for 3d $\mathcal{N} = 4$ SQCD has not appeared in the literature, although its qualitative form may have been guessed from comparison with related earlier works [202, 212].

With these results at hand, we will comment on similarities and differences among the solved theories in three, four and five dimensions. One major outcome of our analysis is to test and explicitly confirm expectations based on physical intuition in the large volume limit. In particular, we find a dramatic change in the phase diagram of a theory when the Chern–Simons term is set to zero.

The rest of the chapter is organized as follows. In the next section we introduce the partition function of 3d $\mathcal{N} = 3$ $U(N)$ Chern–Simons-matter theory on $\mathbb{S}^3$ and take its large N and large radius limits. In Section 5.3 we derive the phase diagram of $U(N)$ Chern–Simons-matter theories in the 't Hooft limit with analytically continued Chern–Simons coupling. We begin by rederiving the results of [202], as a warm-up example to present our machinery. Then, we extend the results to the generic situation. The main physical lessons inferred from our analysis are detailed in Section 5.3.5. The analysis of fundamental and antisymmetric Wilson loop vevs is given in Section 5.4. Finally, we discuss three–dimensional SQCD with $N_f = 2N$ flavours in Section 5.5.

5.2 Gauge theories with large rank on the three-sphere

As already reviewed in Section 2.2, the partition function of any $\mathcal{N} \geq 2$ gauge theory on $\mathbb{S}^3$ localizes to an ordinary integral [58]. The integration variable is the real scalar field in the $\mathcal{N} = 2$ vector multiplet, that we denote by ϕ throughout this Chapter.[11] In this chapter we only consider $U(N)$ gauge theories with N_f hypermultiplets in the fundamental representation. The resulting partition function takes the schematic form

$$\mathcal{Z}_{\mathbb{S}^3} = \frac{1}{N!}\int_{-\infty}^{\infty} \mathrm{d}\phi_1 \cdots \int_{-\infty}^{\infty} \mathrm{d}\phi_N \; Z_{\text{class}}(\phi) Z^{\text{vec}}_{\text{1-loop}}(\phi) Z^{\text{hyp}}_{\text{1-loop}}(\phi). \tag{5.2.1}$$

Here

$$Z_{\text{class}}(\phi) = \prod_{a=1}^{N} e^{-V(\phi_a)},$$
$$V(\phi) = i\pi k (r\phi)^2 + i2\pi r^2 \xi \phi,$$

where r is the radius of $\mathbb{S}^3$, the first piece is the Chern–Simons term and the second piece is the FI term, with FI parameter ξ of mass dimension one. Besides, assuming $\mathcal{N} \geq 3$ supersymmetry,

[11] The real scalar in the $\mathcal{N} = 2$ vector multiplet is more often denoted by σ, with ϕ referring to a $SU(2)$ triplet of fields. However, to uniform our conventions with the five-dimensional setting discussed in Chapter 6, we adopt the symbol ϕ for the real scalar.

the one-loop fluctuation determinants are given by [58]

$$Z^{\text{vec}}_{\text{1-loop}}(\phi) = \prod_{1\leq a<b\leq N} \left(2\sinh \pi r(\phi_a - \phi_b)\right)^2$$
$$Z^{\text{hyp}}_{\text{1-loop}}(\phi) = \prod_{\alpha=1}^{F}\prod_{a=1}^{N} \left(2\cosh \pi r(\phi_a + m_\alpha)\right)^{-n_\alpha} .$$

Here we are assuming without loss of generality that there are F distinct real mass scales $\{m_\alpha\}_{\alpha=1,\dots,F}$, and that n_α hypermultiplets have the same mass m_α. Clearly, the total number of hypermultiplets is

$$N_f = \sum_{\alpha=1}^{F} n_\alpha .$$

Let us comment on our sign convention for the masses. Consistently with Chapters 2 and 6, we are assuming that the hypermultiplets transform in the fundamental representation of both the gauge group and the flavour symmetry, hence we get a relative plus sign in the contribution $\phi_a + m_\alpha$. Sometimes, in the literature one finds the expression $\phi_a - m_\alpha$, which corresponds to assume the matter field in the fundamental representation of the $U(N)$ gauge group but in the anti-fundamental representation of the flavour symmetry group. The latter convention is consistent with producing the flavour symmetry from ungauging a $U(N)$ gauge group. Of course, $m_\alpha \in \mathbb{R}$, so there is no difference in the two conventions. Moreover, this issue would not appear in $Sp(N)$ and $SO(N)$ gauge theories, for which the fundamental, respectively vector, representation is (pseudo-)real.

For vanishing Chern–Simons level, $k = 0$, the theory is SQCD with $\mathcal{N} = 4$ supersymmetry. We focus now on the case $k \neq 0$ and only come back to SQCD below.

A first convenient step is to reabsorb the FI parameter into a change of variables $\phi'_a = \phi_a + \frac{\xi}{k}$, which yields:

$$\mathcal{Z}_{\mathbb{S}^3} = \frac{e^{-\mathrm{i}\pi\frac{N}{k}(r\xi)^2}}{N!}\int_{\mathbb{R}^N} \frac{\prod_{1\leq a<b\leq N}\left(2\sinh \pi r(\phi'_a - \phi'_b)\right)^2}{\prod_{\alpha=1}^{F}\prod_{a=1}^{N}\left(2\cosh \pi r(\phi'_a + m'_\alpha)\right)^{n_\alpha}} \prod_{a=1}^{N} e^{i\pi k(r\phi'_a)^2}\mathrm{d}\phi'_a \tag{5.2.2}$$

where we have defined

$$m'_\alpha = m_\alpha - \frac{\xi}{k}.$$

Therefore, the net effect of turning on an FI parameter is to shift all the masses, up to the overall exponential factor which is to be interpreted as a background Chern–Simons term for the Abelian twisted vector multiplet to which the real scalar ξ belongs. This simple effect is not surprising but is very different from what one gets in a 3d $\mathcal{N} = 4$ $U(N)$ theory without Chern–Simons terms. Essentially, in the Chern–Simons theory under consideration, if we start with a $U(N)$ gauge theory with $SU(N_f)$ flavour symmetry, turning on a FI parameters is equivalent to turn on the central $U(1) \subset U(N_f)$. Therefore, we will henceforth simply consider a $U(N_f)$ flavour symmetry and neglect the FI term, with the understanding that the latter one has been turned on and reabsorbed as shown above (with a few caveats to be discussed shortly).

At this point, simply rewrite (5.2.2) as (we drop henceforth all the $'$)

$$\mathcal{Z}_{\mathbb{S}^3} = \frac{e^{-\mathrm{i}\pi\frac{N}{k}(r\xi)^2}}{N!}\int_{\mathbb{R}^N} \mathrm{d}^N\phi\; e^{-N^2 S_{\text{eff}}(\phi)},$$

with

$$S_{\text{eff}}(\phi) = \frac{1}{N}\sum_{a=1}^{N}\left[-\mathrm{i}\pi\frac{k}{N}(r\phi_a)^2 + \sum_{\alpha=1}^{F}\frac{n_\alpha}{N}\log 2\cosh\pi r(\phi_a + m_\alpha)\right] - \frac{1}{N^2}\sum_{1\leq a\neq b\leq N}\log 2\sinh\pi r\,|\phi_a - \phi_b|\,.$$

In the large N limit the integrand is suppressed as $\sim e^{-N^2(\cdots)}$, so that the leading order contribution comes from the saddle points of the effective action S_{eff}. It is convenient and customary, though, to first consider the analytic continuation in the Chern–Simons coupling k. For later convenience, we introduce a massive 't Hooft parameter λ through the replacement

$$\frac{\mathrm{i}N}{kr} = \lambda,$$

and we will keep λ fixed in our large N limit. Our choice differs from the standard conventions for the Chern–Simons 't Hooft coupling, which is typically taken to be adimensional. Here the appearance of a new massive parameter is simply due to counting powers of r, and will play a crucial role in the ensuing analysis. Notice also that the FI parameter only enters through the ratio $\frac{\xi}{k}$. This means that, upon analytically continuing k, we are also led to analytically continue the FI parameter. Besides, as we take $k\to\infty$ in the 't Hooft limit, we must also scale ξ linearly with N, otherwise its contribution would be sub-leading in the large N limit. This is foreseen already before reabsorbing ξ with the change of variables, and is consistent with the general analysis of Chapter 4, which suggests that, in the 't Hooft limit, all terms in the effective action should be scaled so to contribute at the same order in N.

Additionally, we take a Veneziano scaling, with

$$\frac{n_\alpha}{N} \equiv \zeta_\alpha \text{ fixed} \quad \forall\alpha = 1,\dots,F.$$

The Veneziano parameters ζ_α are unconstrained in the present case, although in absence of a Chern–Simons term they must satisfy $\sum_{\alpha=1}^{F}\zeta_\alpha \geq 2$, which follows directly from the convergence of the initial integral (5.2.2).

The last ingredient we need is the eigenvalue density

$$\rho(\phi^\star) = \frac{1}{N}\sum_{a=1}^{N}\delta(\phi^\star - \phi_a),$$

which, by definition, is normalized to 1, and is compactly supported in the large N limit. Here we have introduced the variable $\phi^\star$, which runs continuously on $\mathrm{supp}\rho$, to distinguish it from the N-component scalar field ϕ.

At this point we take the large N limit of the matrix model (5.2.2). The eigenvalue density allows us to gather the N equations $\frac{\partial S_{\text{eff}}}{\partial\phi_a} = 0$ into a single singular integral equation for ρ, called the saddle point equation (SPE):

$$\mathrm{P}\int \mathrm{d}\psi\rho(\psi)\coth\pi r(\phi^\star - \psi) = \frac{\phi^\star}{\lambda} + \sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{2}\tanh\pi r(\phi^\star + m_\alpha), \tag{5.2.3}$$

to be satisfied at every $\phi^\star\in\mathrm{supp}\rho$. From now on, we will simply replace the notation $\phi^\star$ with ϕ to avoid clutter.

Gaining inspiration from earlier works, we do not try to directly solve (5.2.3). Instead, we approximate the flat space dynamics by taking the large radius limit, referred to as decompactification limit. Notice that, in our version of the 't Hooft limit, we keep λ finite as $r\to\infty$, which corresponds to take the usual, adimensional 't Hooft coupling $\frac{N}{k}\to\infty$ at the same rate [202].

For large argument, the hyperbolic functions in (5.2.3) tend to ± 1 and are replaced by sign functions, significantly simplifying the SPE. We thus arrive at:

$$\int \mathrm{d}\psi \rho(\psi) \mathrm{sgn}(\phi - \psi) = \frac{\phi}{\lambda} + \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \mathrm{sgn}(\phi + m_\alpha). \tag{5.2.4}$$

This equation is the central object of study for the rest of the chapter. The appearance of sign functions on the right-hand side of (5.2.4) is responsible for the appearance of phase transitions. Indeed, assume a given m_α lies outside $\mathrm{supp}\rho$. Then, the corresponding sign function is effectively independent of ϕ. However, if we move m_α, we expect a phase transition when it enters in $\mathrm{supp}\rho$, because it introduces a new dependence on ϕ and the eigenvalues rearrange themselves in a new saddle point configuration.

5.3 Phase diagram of Chern–Simons-matter theories

5.3.1 Two mass scales: Symmetric case

Let us begin our detailed analysis considering a model with two opposite mass scales, that is $F = 2$, $m_1 = -m_2 \equiv m$. In addition, consider a $\mathbb{Z}_2$-symmetric setup in which the two Veneziano parameters are taken equal, $\zeta_1 = \zeta_2 \equiv \zeta$. This theory was first considered in [202, 62] and we revisit it here.

The SPE reduces to

$$\int \mathrm{d}\psi \rho(\psi) \mathrm{sgn}(\phi - \psi) = \frac{\phi}{\lambda} + \frac{\zeta}{2} \left[\mathrm{sgn}(\phi + m) + \mathrm{sgn}(\phi - m) \right]. \tag{5.3.1}$$

Thanks to the discrete reflection symmetry, we can restrict to $m > 0$ without loss of generality. We begin considering the region of very large mass, $m \to +\infty$. In this limit, we expect that the hypermultiplets decouple and we are left with a pure Chern–Simons theory. The eigenvalue density is thus supported on a single interval $[A, B]$. Decreasing m from infinity but keeping $m > B$ and $-m < A$, the sign functions cancel each other. Differentiating (5.3.1) we find

$$\rho(\phi) = \frac{1}{2\lambda}, \qquad -A = B = \lambda,$$

with the values of A and B fixed by normalization and by symmetry. Alternatively, without relying on the $\mathbb{Z}_2$ symmetry, that will not be available in general, we can plug the solution for ρ back into (5.3.1). Since the solution has been obtained differentiating (5.3.1) once, this procedure endows us with a further consistency condition. Complementing this condition with the normalization condition, we find a system of two linear equations for the variables A and B [3, 4]. This feature, in particular, suggests that all solutions in this regime will be one-cut, i.e. $\mathrm{supp}\rho = [A, B]$ consists of a single interval.

The solution was derived under the assumption $m > B$, thus holds for

$$m > m_{\mathrm{cr},1} = \lambda.$$

If, instead, we assume $A < -m < m < B$, (5.3.1) is solved analogously, but this time the sign functions will contribute. Differentiating (5.3.1) once we get

$$\rho(\phi) = \frac{1}{2\lambda} + \frac{\zeta}{2} \left[\delta(\phi + m) + \delta(\phi - m) \right].$$

Plugging this expression back into (5.3.1) imposes $A = -B$, as predicted by symmetry arguments, while the normalization condition reads

$$\int_A^B \mathrm{d}\phi\rho(\phi) = \frac{B-A}{2\lambda} + \frac{\zeta}{2} + \frac{\zeta}{2} = 1,$$

that is $-A = B = \lambda(1-\zeta)$. Therefore, this solution holds in the region

$$0 \leq m < m_{\mathrm{cr},2} = \lambda(1-\zeta).$$

This cannot be the end of the story, because a solution in the region $\lambda(1-\zeta) < m < \lambda$ is still lacking.

The remaining possibility, overlooked above, is $-m = A$. To address this phase, we have to slightly change our strategy, for two reasons. Firstly, we can only take derivatives of at (5.3.1) inside the support but not at its boundary. Secondly, A and B are fixed, this time, so that we cannot adjust them to normalize the eigenvalue density. We are led to the ansatz

$$\rho(\phi) = \frac{1}{2\lambda} + c_A\left[\delta(\phi+m) + \delta(\phi-m)\right],$$

where we have used the standard symmetry argument to predict equal coefficients for the two δ-functions. The normalization condition in this case is

$$\int_A^B \mathrm{d}\phi\rho(\phi) = \frac{m}{\lambda} + 2c_A = 1,$$

leading to $c_A = \frac{1}{2}\left(1 - \frac{m}{\lambda}\right)$.

To sum up, we have obtained three phases, with critical curves $m = \lambda$ and $m = \lambda(1-\zeta)$, and eigenvalue density

$$\rho(\phi) = \frac{1}{2\lambda} + c_A\left[\delta(\phi+m) + \delta(\phi-m)\right], \quad c_A = \begin{cases} 0 & m > \lambda \\ \frac{1}{2}\left(1 - \frac{m}{\lambda}\right) & \lambda(1-\zeta) < m < \lambda \\ \frac{\zeta}{2} & 0 \leq m < \lambda(1-\zeta). \end{cases} \tag{5.3.2}$$

Moving $m < 0$, the theory simply passes through identical phases. We notice that the third phase only exists if $\zeta < 1$. We have thus recovered the phase diagram of [202], which we represent in Figure 5.1.

5.3.2 Free energy and phase transitions in the symmetric case

The eigenvalue density derived in the previous subsection can be used to evaluate the free energy and vacuum expectation values (vev) of operators in the theory. The sphere free energy is defined as

$$\mathcal{F}_{\mathbb{S}^3} = -\frac{1}{\pi r N^2}\log\mathcal{Z}_{\mathbb{S}^3} - \lambda\left(\frac{r\xi}{N}\right)^2,$$

where the shift by the last term is included to get rid of the overall factor in (5.2.2), irrelevant for the ensuing analysis.

Before computing $\mathcal{F}_{\mathbb{S}^3}$, we focus on the vev of the two-point function $\frac{1}{N}\mathrm{Tr}\phi^2$. It is obtained from the expression

$$\langle\phi^2\rangle = \frac{\partial\mathcal{F}_{\mathbb{S}^3}}{\partial(\lambda^{-1})}$$

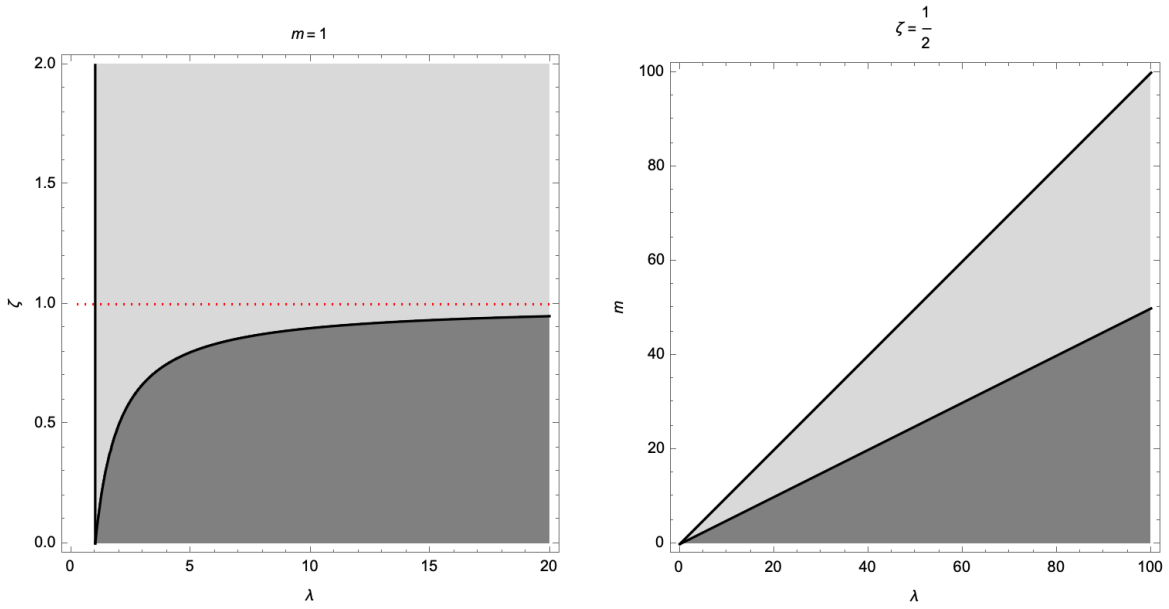


Figure 5.1. Phase structure for two equal sets of hypermultiplets. Left: (λ, ζ)-plane at $m = 1$. The dotted red line is $\{\zeta = 1\}$. Right: (λ, m)-plane at $\zeta = \frac{1}{2}$. As $\zeta \uparrow 1$, the lower critical line is rotated towards the $\{m = 0\}$ axis and eventually disappears.

for any N, and we evaluate it at large N. Exploiting (5.3.2) we get

$$\frac{\partial \mathcal{F}_{\mathbb{S}^3}}{\partial(\lambda^{-1})} = \int_A^B \mathrm{d}\phi \rho(\phi)\phi^2 = \begin{cases} \frac{\lambda^2}{3} & m > \lambda \\ m^2 - \frac{2m^3}{3\lambda} & \lambda(1-\zeta) < m < \lambda \\ (1-\zeta)^3\frac{\lambda^2}{3} + \zeta m^2 & 0 \leq m < \lambda(1-\zeta). \end{cases} \tag{5.3.3}$$

This order parameter is continuous and differentiable both in λ^{-1} and m at the two critical hypersurfaces in parameter space. The second derivative is discontinuous with a finite jump, thus yielding a third order phase transition [202]. We also notice that the discontinuity in the second and higher derivative at the second critical locus $m = \lambda(1-\zeta)$ vanishes at $\zeta \uparrow 1$, as expected.

Let us now consider the free energy. At large N and in the large r approximation it is given by

$$\mathcal{F}_{\mathbb{S}^3} = \int_A^B \mathrm{d}\phi \rho(\phi) \left\{ \frac{\phi^2}{\lambda} + \zeta \left[|\phi + m| + |\phi - m| \right] - \int_A^B \mathrm{d}\psi \rho(\psi) |\phi - \psi| \right\}.$$

Plugging the eigenvalue density (5.3.2) into this expression gives:

$$\mathcal{F}_{\mathbb{S}^3} = \begin{cases} 2\zeta m - \frac{\lambda}{3} & m > \lambda \\ \frac{m}{3}\left(6\zeta - \frac{m(m-3\lambda)}{\lambda^2} - 3\right) & \lambda(1-\zeta) < m < \lambda \\ -\frac{\lambda}{3}(1-\zeta)^3 + \zeta\frac{m^2}{\lambda} + \zeta^2 m & 0 \leq m < \lambda(1-\zeta). \end{cases} \tag{5.3.4}$$

This expression is continuous and differentiable up to the second derivatives, both with respect to λ and m, again showing that the phase transition is third order. Its first derivative in λ^{-1} agrees with (5.3.3), as expected.

We now comment on (5.3.4) in detail, emphasizing lessons that will turn out to hold in the generic case. First of all, let us notice that, in the phase $m > \lambda$, the hypermultiplets are heavier than the saddle point configuration and decouple, whence the simple m-dependence in the free energy. The interesting quantity in that phase is thus $\mathcal{F}_{\mathbb{S}^3} - 2\zeta m$, in which the contribution from

background fields is discarded. We observe that, taking $m \to 0$ first, the free energy trivializes in the second phase and reduces to $-\frac{\lambda}{3}(1-\zeta)^3$. Sending $\zeta \to 0$ this agrees with the renormalized expression at $m > \lambda$. This is a non-trivial consistency check. Indeed, for large m, we expect the hypermultiplets to decouple and to get a split contribution from the matter fields and from a pure Chern–Simons theory. In turn, going first to the conformal point $m \to 0$ and then sending $\zeta \to 0$, we are removing the matter fields from the system, thus we should recover a pure Chern–Simons theory. The expression for $\mathcal{F}_{\mathbb{S}^3}$ is consistent with this expectation, and (after taking into account the $1/r$ factor in our normalization) also agrees with the sphere free energy of pure Chern–Simons theory at leading order in the strong coupling limit [176].

In addition, once the correct renormalization of $\mathcal{F}_{\mathbb{S}^3}$ in the first phase is taken into account, (5.3.4) satisfies the three-dimensional $\mathcal{F}$-theorem [173, 213]

$$\mathcal{F}_{\mathbb{S}^3}^{(\mathrm{IR})} < \mathcal{F}_{\mathbb{S}^3}^{(\mathrm{UV})}. \tag{5.3.5}$$

It follows directly from the positivity of (5.3.3) that $\mathcal{F}_{\mathbb{S}^3}$ is monotonically decreasing in the massive parameter λ, thus fulfilling (5.3.5) not only at the fixed points but at all energy scales.

5.3.3 Two mass scales: Generic case

Let us now go back to the SPE (5.2.4) and solve it without assuming a $\mathbb{Z}_2$ parity symmetry. We take $F = 2$ as before, but now we let the masses m_1, m_2 as well as ζ_1 and ζ_2 arbitrary. We assume for simplicity $m_1 > m_2$, the converse case being recovered by simply swapping $\zeta_1 \leftrightarrow \zeta_2$ in the expressions below. As already argued, this more generic setup includes the deformation of the previous case in which the $\mathbb{Z}_2$ symmetry is explicitly broken by the FI term.

We begin in a region with m_1 very large and negative and m_2 large and positive. The sign functions in the SPE are constant in ϕ and cancel each other. We thus again find a solution

$$\rho_{\mathrm{I}}(\phi) = \frac{1}{2\lambda}, \quad \mathrm{supp}\rho = [-\lambda, \lambda],$$

valid for $m_1 < -\lambda$ and $m_2 > \lambda$. We call this region Phase I.

At this point we may increase m_1 until it reaches $-\lambda$, or decrease m_2 until it reaches λ. Let us assume the former. Gaining insight from the results in Subsection 5.3.1, we expect a new phase with $B = -m_1$. However, differently from the previous case, A is not hit by m_2 and is thus not affected by the phase transition (A is independent of m_1 in this phase). Differentiating the SPE once we arrive at

$$\rho_{\mathrm{II}_1}(\phi) = \frac{1}{2\lambda} + c_1\delta(\phi + m_1), \quad \mathrm{supp}\rho = [-\lambda, -m_1].$$

Imposing the normalization condition fixes c_1, through

$$\frac{\lambda - m_1}{2\lambda} + c_1 = 1 \quad \Longrightarrow \quad c_1 = \frac{1}{2}\left(1 + \frac{m_1}{\lambda}\right).$$

If, instead, we decrease m_2 keeping m_1 large, negative and fixed, by an analogous argument we obtain

$$\rho_{\mathrm{II}_2}(\phi) = \frac{1}{2\lambda} + \frac{1}{2}\left(1 - \frac{m_2}{\lambda}\right)\delta(\phi + m_1), \quad \mathrm{supp}\rho = [-m_2, \lambda].$$

We call the two phases so obtained Phase II_1 and II_2, making reference to which mass triggers the transition.

From either Phase II_α, two scenarios disclose. Either we move the same mass further, until it enters in the bulk of suppρ, or we move the other mass. Let us begin from the latter scenario.

Starting from either Phases II_α we arrive at a new phase, that we denote Phase III, in which the eigenvalue density is

$$\rho_{\mathrm{III}}(\phi) = \frac{1}{2\lambda} + \sum_{\alpha=1}^{2} c_\alpha \delta(\phi + m_\alpha), \quad \mathrm{supp}\rho = [-m_2, -m_1],$$

where the coefficients are

$$c_1 = \frac{1}{2}\left(1 + \frac{m_1}{\lambda}\right), \quad c_2 = \frac{1}{2}\left(1 - \frac{m_2}{\lambda}\right).$$

Alternatively, from Phase II_α we may enter Phase IV_α, $\forall \alpha = 1, 2$, in which

$$\rho_{\mathrm{IV}_\alpha}(\phi) = \frac{1}{2\lambda} + \frac{\zeta_\alpha}{2}\delta(\phi + m_\alpha),$$

where the support is determined by normalization to be

$$[A, B] = \begin{cases} [-\lambda, \lambda(1-\zeta_1)] & \text{Phase } \mathrm{IV}_1 \ , \\ [-\lambda(1-\zeta_2), \lambda] & \text{Phase } \mathrm{IV}_2 \ . \end{cases}$$

At this point, two more phases disclose. Namely, Phase V_α can be reached from Phase IV_α moving $m_{\alpha'}$ for $\alpha' \neq \alpha$, as well as from Phase III moving m_α, $\forall \alpha = 1, 2$. Phase V_α is characterized by the mass $-m_\alpha$ being in the interior of $\mathrm{supp}\rho$, that is $A < -m_\alpha < B$, and the other mass stuck at one boundary. Proceeding as above, we get

$$\rho_{\mathrm{V}_\alpha}(\phi) = \frac{1}{2\lambda} + c_1\delta(\phi + m_1) + c_2\delta(\phi + m_2), \qquad (c_1, c_2) = \begin{cases} \left(\frac{\zeta_1}{2}, \frac{1}{2}\left(1 - \frac{m_2}{\lambda}\right)\right) & \text{Phase } \mathrm{V}_1 \ , \\ \left(\frac{1}{2}\left(1 + \frac{m_1}{\lambda}\right), \frac{\zeta_2}{2}\right) & \text{Phase } \mathrm{V}_2 \ . \end{cases}$$

There is one further phase that can be accessed, that we denote Phase VI, in which both masses are in the interior of $\mathrm{supp}\rho$. We find

$$\rho_{\mathrm{VI}}(\phi) = \frac{1}{2\lambda} + \sum_{\alpha=1}^{2} \frac{\zeta_\alpha}{2}\delta(\phi + m_\alpha), \quad \mathrm{supp}\rho = [-\lambda(1-\zeta_2), \lambda(1-\zeta_1)].$$

The phase diagram becomes intricate very fast: every mass scale m_α introduces four critical hyperplanes, at $m_\alpha \in \{\pm\lambda, \pm\lambda(1-\zeta_\alpha)\}$, and the collection of all such critical hyperplanes partitions the parameter space into distinct phases. A summary of the phase diagram in the sample case $\zeta_1 = 2\zeta_2$ and $m_2 = -2m_1$ is reported in Figure 5.2.

5.3.4 Free energy and phase transitions in the generic case

As for the symmetric phase, to probe the critical behaviour of the theory with two mass scales we compute the order parameter $\langle\phi^2\rangle = \frac{\partial \mathcal{F}_{\mathbb{S}^3}}{\partial(\lambda^{-1})}$. We have:

$$\langle\phi^2\rangle = \int_A^B \mathrm{d}\phi \left[\frac{1}{2\lambda} + \sum_{\alpha=1}^{2} c_\alpha \delta(\phi + m_\alpha)\right]\phi^2 = \frac{B^3 - A^3}{2\lambda} + \sum_{\alpha=1}^{2} c_\alpha m_\alpha^2,$$

with A, B, c_1, c_2 computed in the appropriate phase. Explicitly:

$$\langle\phi^2\rangle = \begin{cases} \frac{\lambda^2}{3} & \text{Phase I} \\ \frac{\lambda^3 + 3\lambda m_1^2 + 2m_1^3}{6\lambda} & \text{Phase } \mathrm{II}_1 \\ \frac{3\lambda m_1^2 + 3\lambda m_2^2 + 2m_1^3 - 2m_2^3}{6\lambda} & \text{Phase III} \\ \frac{1}{6}\left[\left((1-\zeta_1)^3 + 1\right)\lambda^2 + 3\zeta_1 m_1^2\right] & \text{Phase } \mathrm{IV}_1 \\ \frac{-(\zeta_1-1)^3\lambda^3 + 3\zeta_1\lambda m_1^2 + 3m_2^2(\lambda - m_2) + m_2^3}{6\lambda} & \text{Phase } \mathrm{V}_1 \\ \frac{1}{6}\left[\left((1-\zeta_1)^3 + (1-\zeta_2)^3\right)\lambda^2 + 3\zeta_1 m_1^2 + 3\zeta_2 m_2^2\right] & \text{Phase VI} \end{cases}$$

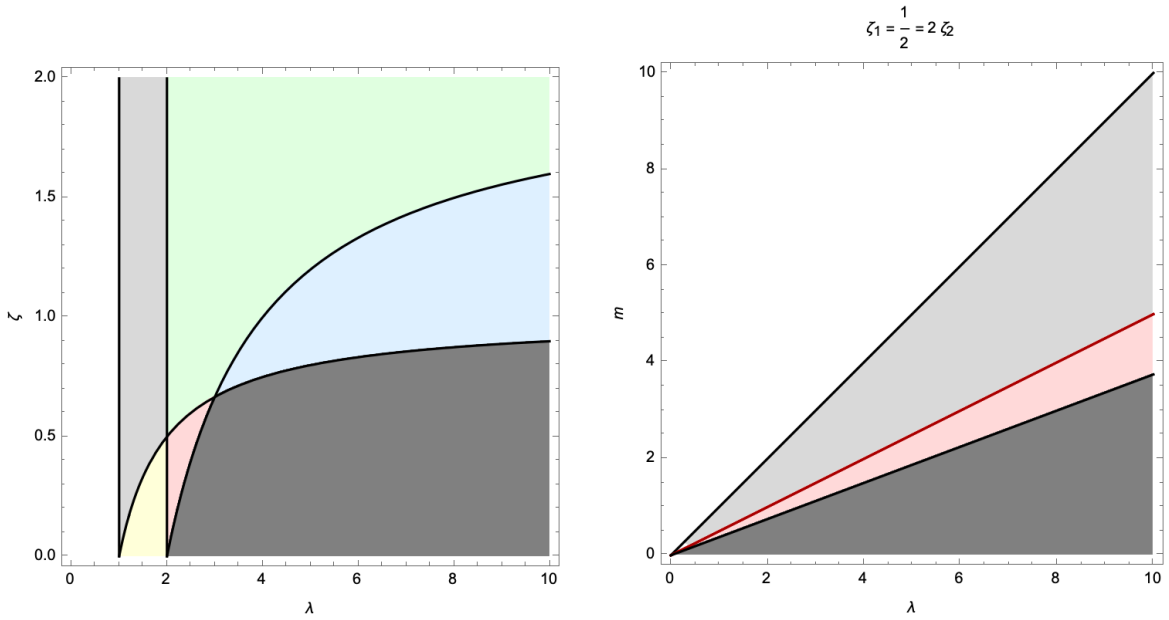


Figure 5.2. Phase structure for two sets of hypermultiplets. Left: (λ, ζ)-plane at $m_1 = -1, m_2 = 2, \zeta_1 = \zeta = 2\zeta_2$. The system can access Phase I (white), Phase II_1 (light gray), Phase III (green), Phase IV_1 (yellow), Phases V_1 and V_2 (red and blue), Phase VI (dark gray). Right: (λ, m)-plane at $\zeta_1 = \frac{1}{2}, \zeta_2 = \frac{1}{4}, m_2 = 2m, m_1 = -m$. For this choice of (ζ_1, ζ_2), the yellow and green phases on the left panel collapse onto the critical line $m = \lambda/2$, highlighted in red in the right panel.

and analogous expression exchanging the labels $1 \leftrightarrow 2$ when $\{\mathrm{II}_2, \mathrm{IV}_2, \mathrm{V}_2\}$ replace $\{\mathrm{II}_1, \mathrm{IV}_1, \mathrm{V}_1\}$. The continuity of the vev follows directly from the continuity of $\rho(\phi)$. Applying $\frac{\partial}{\partial m_\alpha}$ or $\frac{\partial}{\partial(\lambda^{-1})}$ to the expressions in each phase, it is easy to check that $\langle\phi^2\rangle$ is differentiable, and we get a third order phase transition.

5.3.5 Analysis of the phase diagram

The phase diagram obtained for $F = 2$ generalizes to a higher number of mass scales. A phase transition is triggered each time a mass m_α hits a boundary of $[A, B]$ from outside, generating a δ-peak on top of the boundary. Then, a new transition takes place when the peak is moved in the interior of $[A, B]$. All transitions are third order, because of the locality property of the solution: the values of A, B do not depend on the masses, except for the "intermediate" phase in which $A = -m_\alpha$ or $B = -m_\alpha$, independent on the rest of mass scales.

On a technical level, the differentiability of the order parameter stems from the independence of A or B on the masses not involved in the transition, and thus the order of the transition is determined by the locality property of the solution to be third.

We have seen that there is no substantial difference in the symmetric case compared to the more general setting. In particular, this implies that there is no qualitative difference between $SU(N)$ and $U(N)$ Chern–Simons-matter theories, when it comes to critical behaviour. Introducing a Lagrange multiplier to impose the tracelessness condition for $SU(N)$ will affect the explicit form of A and B, but the details are washed out as one moves close to the critical loci. For the technical reason outline above, this claim would not be expected to hold if higher powers of the masses entered in the SPE.

The phase diagrams derived in this section have a neat physical interpretation. In the large radius limit, we have kept a massive parameter λ, which comes from the Chern–Simons coupling but introduces a characteristic mass scale into the problem. Then, each hypermultiplet of mass

$-m_\alpha > \lambda$ is heavy, compared to the characteristic mass scale of the problem, and is integrated out. These hypermultiplets trigger a phase transition when become lighter than the characteristic mass scale λ, and are integrated back in. Their one-loop effect ceases to contribute to the SPE in that case.

5.4 Wilson loops

In Section 5.3 we have probed the phase structure of various $U(N)$ Chern–Simons-matter theories by computing the free energy and the vev of $\frac{1}{N}\mathrm{Tr}\phi^2$. In the present section we extend our analysis by studying a different class of order operators, the Wilson loops. We will consider half-BPS Wilson loops in the fundamental and rank-K antisymmetric representation and compute their vev at large N. In order to preserve half of the supersymmetry, the loops are placed along an equatorial $\mathbb{S}^1 \subset \mathbb{S}^3$.

5.4.1 Wilson loops in the fundamental representation

Fundamental Wilson loop with two symmetric mass scales

To start our analysis of loop operators, we compute the vev of a Wilson loop W_{F} in the fundamental representation of $U(N)$. We first do so in the theory with two opposite masses and equal Veneziano parameters, studied in Subsection 5.3.1. Plugging an operator in the fundamental representation gives a sub-leading contribution to the large N limit, thus the eigenvalue density is not affected by the insertion of the Wilson loop. The vev is thus simply given by

$$\langle W_{\mathsf{F}} \rangle = \int_A^B \mathrm{d}\phi \rho(\phi) e^{2\pi r \phi}.$$

Using the eigenvalue density (5.3.2) we get

$$\langle W_{\mathsf{F}} \rangle = \begin{cases} \frac{1}{2\pi\lambda r}\sinh(2\pi r\lambda) & m > \lambda \\ \frac{1}{2\pi\lambda r}\sinh(2\pi r m) + \left(1 - \frac{m}{\lambda}\right)\cosh(2\pi r m) & \lambda(1-\zeta) < m < \lambda \\ \frac{1}{2\pi\lambda r}\sinh(2\pi r\lambda(1-\zeta)) + \zeta\cosh(2\pi r m) & 0 \leq m < \lambda(1-\zeta). \end{cases} \tag{5.4.1}$$

This is a continuous and differentiable order parameter, with a finite jump in its second derivative. The phase transition in therefore third order, a result which is consistent with and confirms the previous analysis. We note that it would be more appropriate to first replace the hyperbolic functions by exponential ones, because the eigenvalue density has been derived in the large r limit. The final answer is nevertheless robust under exchanging differentiation and large radius limit.[12] The Wilson loop vev (5.4.1) follows a perimeter law, $\log\langle W_{\mathsf{F}}\rangle \propto r$. It is also interesting to rewrite (5.4.1) recalling that $\lambda r = it$, with $t \equiv N/k$ the standard, adimensional 't Hooft coupling, fixed at large N. We obtain

$$\langle W_{\mathsf{F}} \rangle = \frac{\sin(2\pi t)}{2\pi t} \text{ if } mr > t$$

in the large mass phase. For $m \to 0$, instead, one finds $\langle W_{\mathsf{F}}\rangle = 1$ if $\zeta \geq 1$ and $\langle W_{\mathsf{F}}\rangle = \frac{\sin(2\pi t(1-\zeta))}{2\pi t} + \zeta$ if $\zeta < 1$. In all cases, the dependence on r is dropped, in agreement with physical expectations, when the defect theories are conformal.

[12] This corrects a subtlety in [202].

5.4.2 Fundamental Wilson loop with two generic mass scales

Let us now compute the vev $\langle W_{\mathsf{F}}\rangle$ in the case of two families of hypermultiplets with arbitrary masses. Reasoning as above, we get the generic expression:

$$\langle W_{\mathsf{F}}\rangle = \frac{1}{2\pi r\lambda}\left(e^{2\pi rB}-e^{2\pi rA}\right)+\sum_{\alpha=1}^{2} c_\alpha e^{-2\pi r m_\alpha},$$

with endpoints A, B and coefficients c_α depending on the phase. Explicitly:

$$\langle W_{\mathsf{F}}\rangle = \begin{cases} \frac{\sinh(2\pi\lambda r)}{2\pi\lambda r} & \text{Phase I}\\ \frac{2\pi r(\lambda+m_1)e^{-2\pi m_1 r}+e^{-2\pi m_1 r}-e^{-2\pi\lambda r}}{4\pi\lambda r} & \text{Phase II}_1\\ \frac{2\pi r(\lambda+m_2)e^{-2\pi m_1 r}+2\pi r(\lambda-m_2)e^{-2\pi m_2 r}+e^{-2\pi m_1 r}-e^{-2\pi m_2 r}}{4\pi\lambda r} & \text{Phase III}\\ \frac{1}{4}\left[2\zeta_1 e^{-2\pi m_1 r}+\frac{e^{-2\pi\lambda r}\left(e^{-2\pi(\zeta_1-2)\lambda r}-1\right)}{\pi\lambda r}\right] & \text{Phase IV}_1\\ \frac{1}{4}\left[\frac{e^{-2\pi(\zeta_1-1)\lambda r}-e^{-2\pi m_2 r}}{\pi\lambda r}+2\zeta_1 e^{-2\pi r m_1}+2\left(1-\frac{m_2}{\lambda}\right)e^{-2\pi m_2 r}\right] & \text{Phase V}_1\\ \frac{1}{4}\left[2\zeta_1 e^{-2\pi r m_1}+2\zeta_2 e^{-2\pi m_2 r}+\frac{e^{2\pi(\zeta_2-1)\lambda r}\left(e^{-2\pi(\zeta_1+\zeta_2-2)\lambda r}-1\right)}{\pi\lambda r}\right] & \text{Phase VI} \end{cases}$$

and analogous expression when $\{\mathrm{II}_2, \mathrm{IV}_2, \mathrm{V}_2\}$ replace $\{\mathrm{II}_1, \mathrm{IV}_1, \mathrm{V}_1\}$.

A direct computation shows that the order parameter $\langle W_{\mathsf{F}}\rangle$ is continuous and differentiable (both in λ and m_α), proving that the phase transition is third order. We again find perfect agreement with the analysis of the free energy.

5.4.3 Wilson loops in the antisymmetric presentation

We now study the vev of half-BPS Wilson loop in the rank-K antisymmetric representation A_K. Following [209], we introduce the generating function

$$\Phi_{\mathsf{A}}(w) = \left\langle \prod_{a=1}^{N}\left(1+we^{2\pi r\phi_a}\right)\right\rangle, \tag{5.4.2}$$

and we see that the Wilson loop vev is extracted as

$$\langle W_{\mathsf{A}_K}\rangle = \oint_\gamma \frac{\mathrm{d}w}{2\pi \mathrm{i} w^{1+K}}\Phi_{\mathsf{A}}(w).$$

The integration cycle γ is a small loop around the origin in $\mathbb{C}$. We are interested in the large N limit with the ratio

$$\kappa = \frac{K}{N} \text{ fixed}.$$

At large N, the generating function (5.4.2) is simply written as

$$\Phi_{\mathsf{A}}(w) = \exp\left\{N\int_A^B \mathrm{d}\phi\rho(\phi)\log\left(1+we^{2\pi r\phi}\right)\right\}.$$

Using a change of variables

$$\log w = -2\pi r\left[A+z(B-A)\right],$$

with z the holomorphic variable on a multiply-sheeted cover of the cylinder, $\langle W_{\mathsf{A}_K}\rangle$ is recast in the form

$$\langle W_{\mathsf{A}_K}\rangle = \mathrm{i}e^{2\pi rAK+\log(r(B-A))}\oint_\Gamma \mathrm{d}z\exp\left\{N\left[\int_A^B \mathrm{d}\phi\rho(\phi)\log\left(1+e^{2\pi r(\phi-A-z(B-A))}\right)+2\pi z\kappa r(B-A)\right]\right\}. \tag{5.4.3}$$

The original integration cycle γ has been mapped, under the exponential change of variables, onto a circle Γ wrapping the cylinder at fixed $\Re z$, see Figure 5.3. Shrinking γ around $w = 0$ pushes Γ towards $\Re z \to +\infty$. The integrand in (5.4.3) has branch cuts at

$$0 \leq \Re(z) \leq 1 \quad \text{and} \quad \Im(z) = \frac{(2n+1)\pi}{r(B-A)},\ n \in \mathbb{Z}. \tag{5.4.4}$$

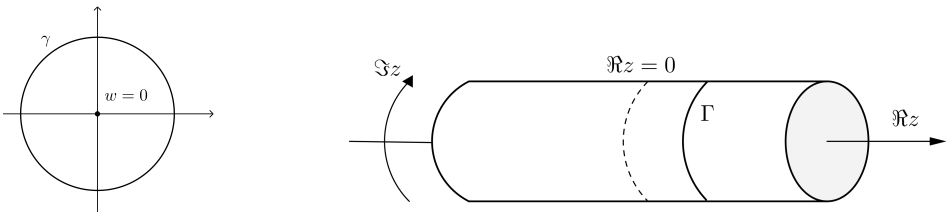


Figure 5.3. The exponential map sends $\mathbb{C}$ to a cover of the cylinder, and the circle γ around $w = 0$ (left) to a circle Γ wrapping the cylinder once (right).

At this point we exploit the fact that we are working at large N, and evaluate the integral in z by a standard steepest descent argument. The saddle point of (5.4.3) is determined by

$$-\int_A^B \mathrm{d}\phi \rho(\phi) \frac{1}{1+e^{2\pi r(\phi - A - z(B-A))}} + \kappa = 0. \tag{5.4.5}$$

We moreover work in the large r approximation. In this regime, direct inspection of (5.4.5) shows that the solution must be looked for in $0 \leq \Re z \leq 1$. Indeed, since $\phi \in [A, B]$,

$$\lim_{r\to\infty} \frac{1}{1+e^{2\pi r(\phi - A - z(B-A))}} = \begin{cases} 0 & \text{if } \Re z < 0 \\ 1 & \text{if } \Re z > 1. \end{cases}$$

Plugging the generic solution for $\rho(\phi)$ with $F = 2$ in (5.4.5), the integral splits into three contributions:

$$\frac{1}{4\pi r\lambda}\mathcal{I}(z) + \sum_{\alpha=1}^{2} c_\alpha \mathcal{J}(z, m_\alpha) = \kappa,$$

where

$$\mathcal{I}(z) = \int_A^B \mathrm{d}\phi \frac{1}{1+e^{2\pi r(\phi - A - z(B-A))}} = 2\pi r(B-A) + \log\left(\frac{1+e^{-2\pi r z(B-A)}}{1+e^{-2\pi r(z-1)(B-A)}}\right),$$

$$\mathcal{J}(z, m_\alpha) = \int_A^B \mathrm{d}\phi \frac{\delta(\phi + m_\alpha)}{1+e^{2\pi r(\phi - A - z(B-A))}} = \frac{1}{1+e^{-2\pi r(m_\alpha + A + z(B-A))}}.$$

An immediate consequence of the analysis so far is that the saddle point lies at $\Im z = 0$ and $0 < \Re z < 1$. In this region, as $r \to \infty$

$$\log\left(\frac{1+e^{-2\pi r z(B-A)}}{1+e^{-2\pi r(z-1)(B-A)}}\right) \approx -\log e^{2\pi r(B-A)(1-z)} \quad \Longrightarrow \quad \mathcal{I}(z) \approx 2\pi r z(B-A).$$

and

$$\mathcal{J}(z, m_\alpha) \approx \theta(z - z_\alpha^\star),$$

where we have introduced the shorthand notation $z_\alpha^\star \equiv -\frac{m_\alpha + A}{B-A}$ and $\theta(\cdot)$ is the step-function defined as

$$\theta(z - z^\star) = \begin{cases} 0 & z < z^\star \\ \frac{1}{2} & z = z^\star \\ 1 & z > z^\star. \end{cases}$$

Thus, in the large r limit, the saddle point equation (5.4.5) simplifies into

$$\frac{B-A}{2\lambda}z+\sum_{\alpha=1}^{2}c_\alpha\theta(z-z_\alpha^\star)=\kappa. \tag{5.4.6}$$

Antisymmetric Wilson loop with two symmetric mass scales: Saddle points

We are now ready to solve (5.4.6) and compute the Wilson loop vev. We start with the theory with $\mathbb{Z}_2$-symmetric assignment of masses.

First, we point out a $\mathbb{Z}_2$ reflection symmetry inherited by the Wilson loop at large N, which remains invariant under the combined transformation $z\mapsto 1-z$, $\kappa\mapsto 1-\kappa$.

In Phase I, $B=-A=\lambda$ and $c_\alpha=0$, leading to a saddle point

$$z=\kappa.$$

Notice that $0\leq\kappa\leq 1$ by definition, hence the solution is consistent with the saddle point being placed in the interval $0\leq z\leq 1$. This solution is also manifestly consistent with the reflection symmetry.

In Phase II, one θ-function contributes identically 1 and the other vanishes identically, because $0<z<1$. The solution to (5.4.6) is thus given by

$$z=\frac{1}{2}+\frac{\lambda}{m}\left(\kappa-\frac{1}{2}\right).$$

The solution exists in this case only if

$$\frac{1}{2}\left(1-\frac{m}{\lambda}\right)\leq\kappa\leq\frac{1}{2}\left(1+\frac{m}{\lambda}\right).$$

If we send $\lambda\to\infty$, a solution exists only if $\kappa=\frac{1}{2}$, in which case $z=\frac{1}{2}$. Remarkably, the unique solution that survives at $\lambda\to\infty$ is the fixed point of the reflection symmetry.

In Phase III, the situation is more involved. Depending on whether

$$(i)\ z<\frac{1}{2}\left(1-\frac{m}{\lambda(1-\zeta)}\right),\quad (ii)\ \frac{1}{2}\left(1-\frac{m}{\lambda(1-\zeta)}\right)<z<\frac{1}{2}\left(1+\frac{m}{\lambda(1-\zeta)}\right),\quad (iii)\ z>\frac{1}{2}\left(1+\frac{m}{\lambda(1-\zeta)}\right),$$

we have that (i) none, (ii) one or (iii) both θ-functions contribute. The three corresponding saddle points are:

$$z=\begin{cases}\frac{\kappa}{1-\zeta} & \text{Region III}_i\\ \frac{2\kappa-\zeta}{2(1-\zeta)} & \text{Region III}_{ii}\\ \frac{\kappa-\zeta}{(1-\zeta)} & \text{Region III}_{iii}\end{cases}$$

The self-consistency condition for the existence of the solution in each region is:

$$\begin{aligned}\text{Region III}_i:&\ \ \kappa\leq\frac{1}{2}\left(1-\zeta-\frac{m}{\lambda}\right)\\ \text{Region III}_{ii}:&\ \ \frac{1}{2}\left(1-\frac{m}{\lambda}\right)\leq\kappa\leq\frac{1}{2}\left(1+\frac{m}{\lambda}\right)\\ \text{Region III}_{iii}:&\ \ \kappa\geq\frac{1}{2}\left(1+\zeta+\frac{m}{\lambda}\right).\end{aligned}$$

We observe that there are regions of width $\frac{\zeta}{2}$, in the domain of the Wilson loop parameter $\kappa\in[0,1]$, for which no consistent solution exists.

Antisymmetric Wilson loop with two symmetric mass scales: Evaluation

Having found the saddle points of the integral representation (5.4.3), we are now in the position to evaluate the Wilson loop vev. We find:

$$\frac{1}{2\pi rN}\log\langle W_{\mathsf{A}_K}\rangle = \kappa\left[A + z(B-A)\right] + \int_A^B \frac{\mathrm{d}\phi}{2\pi r}\rho(\phi)\log\left(1+e^{2\pi r(\phi - A - z(B-A))}\right),$$

at leading order, with z to be replaced by its saddle point value in each phase. Inserting the generic form of $\rho(\phi)$ and simplifying the expression for large r (the asymptotic behaviour of the polylogarithm is used at an intermediate step), we arrive at

$$\frac{1}{2\pi rN}\log\langle W_{\mathsf{A}_K}\rangle = \kappa\left[A + z(B-A)\right] + \frac{(B-A)^2}{4\lambda}(1-z)^2 + \sum_{\alpha=1}^{2} c_\alpha \Theta(m_\alpha; z), \tag{5.4.7}$$

where, for shortness, we have introduced the function

$$\Theta(m;z) = \begin{cases} 0 & \text{if } z > -\frac{m+A}{B-A} \\ -\left[m + A + z(B-A)\right] & \text{if } z < -\frac{m+A}{B-A} \end{cases}$$

whose derivative is closely related to the θ-function above.

In Phase I we plug $z = \kappa$ in (5.4.7) and get

$$\frac{1}{2\pi rN}\log\langle W_{\mathsf{A}_K}\rangle = \lambda\left[3\kappa(\kappa-1)+1\right],$$

which is invariant under the $\mathbb{Z}_2$ action $\kappa \mapsto 1-\kappa$.

In Phase II instead we find

$$\frac{1}{2\pi rN}\log\langle W_{\mathsf{A}_K}\rangle = \frac{1}{4}\left[3(1-2\kappa)^2\lambda - \frac{m^2}{\lambda} + 2m\right],$$

again invariant under the $\mathbb{Z}_2$ action $\kappa \mapsto 1-\kappa$. Moreover, it is continuous and differentiable at the critical point $m = \lambda$.

Finally, in Phase III we find

$$\frac{1}{2\pi rN}\log\langle W_{\mathsf{A}_K}\rangle = \begin{cases} \lambda\left(-(\zeta+3)\kappa + 3\kappa^2 + 1\right) & \mathrm{III}_i \\ \frac{1}{4}\left[\lambda((\zeta-2)\zeta + 12(\kappa-1)\kappa + 4) + 2\zeta m\right] & \mathrm{III}_{ii} \\ \lambda(\zeta(\kappa-1) + 3\kappa(\kappa-1) + 1) & \mathrm{III}_{iii}. \end{cases}$$

Approaching the critical locus $m = \lambda(1-\zeta)$, regions III_i and III_{iii} are ruled out, whilst the region of validity of Phase III_{ii} agrees with that of Phase II. The Wilson loop vev is continuous and differentiable.

Antisymmetric Wilson loop with two generic mass scales

The saddle point analysis above is easily generalized to the case of $F = 2$ families of hypermultiplets of generic masses, although the resulting expressions are more cumbersome.

Phase I is identical to the symmetric case. In Phase II_1 none of the θ-functions contribute, but the value of B changes and (5.4.6) becomes

$$\left(\frac{-m_1+\lambda}{2\lambda}\right) z = \kappa \quad \Longrightarrow \quad z|_{\mathrm{II}_1} = \frac{2\kappa\lambda}{\lambda - m_1}.$$

In Phase II_2 instead, a θ-function contributes and, from (5.4.6), we get

$$\left(\frac{m_2+\lambda}{2\lambda}\right) z + \frac{1}{2}\left(1-\frac{m_2}{\lambda}\right) = \kappa \quad \Longrightarrow \quad z|_{\mathrm{II}_2} = \frac{(2\kappa-1)\lambda+m_2}{\lambda+m_2}.$$

In Phase III, again only $\theta(z-z_2^\star)$ contributes and the solution to (5.4.6) is given by

$$z|_{\mathrm{III}} = \frac{(2\kappa-1)\lambda+m_2}{m_2-m_1}.$$

In Phase IV_1 we have to split in two sub-cases:

$$z|_{\mathrm{IV}_1} = \begin{cases} \frac{\kappa}{1-\frac{\zeta_1}{2}} & \kappa < \frac{1}{2}\left(1-\frac{m_1}{\lambda}\right) \\ \frac{\kappa-\frac{\zeta_1}{2}}{1-\frac{\zeta_1}{2}} & \kappa > \frac{1}{2}\left(1-\frac{m_1}{\lambda}\right)+\frac{\zeta_1}{2} \end{cases}$$

with an interval of width $\frac{\zeta_1}{2}$ of values of κ for which no consistent solution is found. A similar computation works in Phase IV_2.

In Phase V_1 there are again two sub-cases associated to the values of $\theta(z-z_1^\star)$, while $\theta(z-z_2^\star)=1$ in that phase. We obtain:

$$z|_{\mathrm{V}_1} = \begin{cases} \frac{(2\kappa-1)\lambda+m_2}{m_2+\lambda(1-\zeta_1)} & \kappa < \frac{1}{2}\left(1-\frac{m_1}{\lambda}\right) \\ \frac{(2\kappa-1-\zeta_1)\lambda+m_2}{m_2+\lambda(1-\zeta_1)} & \kappa > \frac{1}{2}\left(1-\frac{m_1}{\lambda}\right)+\frac{\zeta_1}{2} \end{cases}$$

with the two sub-cases characterized by the same conditions on κ as in Phase IV_1. In Phase V_2, instead, the two sub-cases corresponds to the two possible values of $\theta(z-z_2^\star)$, while the other step-function vanishes identically in $0<z<1$. Then,

$$z|_{\mathrm{V}_2} = \begin{cases} \frac{2\kappa\lambda}{\lambda(1-\zeta_2)-m_1} & \kappa < \frac{1}{2}\left(1-\zeta_2-\frac{m_2}{\lambda}\right) \\ \frac{(2\kappa-\zeta_2)\lambda}{\lambda(1-\zeta_2)-m_1} & \kappa > \frac{1}{2}\left(1-\frac{m_2}{\lambda}\right). \end{cases}$$

Finally, Phase VI is conceptually analogous to Phase III of the symmetric case, except that now we have to distinguish with step-function is non-vanishing. This produces four sub-cases:

$$z|_{\mathrm{VI}} = \begin{cases} \frac{2\kappa}{2-\zeta_1-\zeta_2} & \kappa < \frac{1}{2}\left(1-\zeta_2-\frac{m_\alpha}{\lambda}\right) \ \forall\alpha=1,2 \\ \frac{2\kappa-\zeta_2}{2-\zeta_1-\zeta_2} & \frac{1}{2}\left(1-\frac{m_2}{\lambda}\right) < \kappa < \frac{1}{2}\left(1-\frac{m_1}{\lambda}\right) \\ \frac{2\kappa-\zeta_1}{2-\zeta_1-\zeta_1} & \frac{1}{2}\left(1-\zeta_2+\zeta_1-\frac{m_1}{\lambda}\right) < \kappa < \frac{1}{2}\left(1-\zeta_2+\zeta_1-\frac{m_2}{\lambda}\right) \\ \frac{2\kappa-\zeta_1-\zeta_2}{2-\zeta_1-\zeta_2} & \kappa > \frac{1}{2}\left(1+\zeta_1-\frac{m_\alpha}{\lambda}\right) \ \forall\alpha=1,2. \end{cases}$$

To compute the vev of the Wilson loop in the antisymmetric A_K representation, it simply remains to plug the correct values of A, B and of the saddle point z into (5.4.7) in each phase. The resulting expressions are lengthy and require a case by case study within each phase, thus we omit them. It is nevertheless straightforward to evaluate $\log\langle W_{\mathsf{A}_K}\rangle$ at large N with the aid of a computer algebra system. What we find is that the Wilson loop vev is everywhere continuous and differentiable. Therefore, the analysis of this additional order operator confirms the existence of a third order phase transition.

5.5 Three-dimensional SQCD

We now leave behind the analysis of $U(N)$ Chern–Simons-matter theories and consider $\mathcal{N}=4$ SQCD, that is, a supersymmetric $U(N)$ gauge theory with N_f hypermultiplets in the fundamental representation. The large N limit of this otherwise extensively studied theory appears to be overlooked in the literature. In particular, the large N eigenvalue density seems not to have been written down explicitly before. In the massless case, a large N solution for the free energy was found in [214] by first solving exactly for the partition function, using a Selberg integral, and then taking $N\to\infty$.

We present here the large N limit of three-dimensional SQCD, for arbitrary r. We will show that, in a large r approximation, the results agree with the $k\to 0$ limit of the results derived previously in the Chern–Simons setting. Our presentation and result are valid for any $N_f\geq 2N$. However, we have been able to obtain closed form expressions only in the balanced case $N_f=2N$, or as long as N_f-2N remains fixed when $N\to\infty$. The solution holds also in the more general case, but then the endpoints of the eigenvalue density are only determined implicitly.

The starting point in the integral representation (5.2.1) for the partition function, now setting the Chern–Simons level and FI parameter to zero. We have:

$$\mathcal{Z}^{\text{SQCD}}_{\mathbb{S}^3}=\frac{1}{N!}\int_{-\infty}^{\infty}\mathrm{d}\phi_1\cdots\int_{-\infty}^{\infty}\mathrm{d}\phi_N\,\frac{\prod_{1\leq a<b\leq N}\left(2\sinh\pi r(\phi_a-\phi_b)\right)^2}{\prod_{a=1}^{N}\prod_{\alpha=1}^{F}\left(2\cosh\pi r(\phi_a+m_\alpha)\right)^{n_\alpha}}, \tag{5.5.1}$$

where, as above, we have assumed F families of hypermultiplets of masses $\{m_\alpha\}_{\alpha=1,\ldots,F}$.

It is convenient to use an exponential change of variables [84] $x_a=e^{2\pi r\phi_a}$. By analogy, to lighten the notation we also set $\mu_\alpha=e^{2\pi r m_\alpha}$. Then, from

$$\begin{aligned}\prod_{1\leq a<b\leq N}\left(2\sinh\pi r(\phi_a-\phi_b)\right)^2&=\prod_{a=1}^{N}x_a^{-(N-1)}\prod_{1\leq a<b\leq N}(x_a-x_b)^2,\\ \prod_{\alpha=1}^{F}\left(2\cosh\pi r(\phi_a+m_\alpha)\right)^{n_\alpha}&=x_a^{-\frac{N_f}{2}}\prod_{\alpha=1}^{F}\left(1+x_a\mu_\alpha\right)^{n_\alpha},\end{aligned}$$

where we have used the constraint $\sum_{\alpha=1}^{F}n_\alpha m_\alpha=0$ from the $SU(N_f)$ flavour symmetry and also $N_f=\sum_{\alpha=1}^{F}n_\alpha$, (5.5.1) is rewritten as

$$\mathcal{Z}^{\text{SQCD}}_{\mathbb{S}^3}=\frac{1}{(2\pi r)^N N!}\int_{(0,\infty)^N}\prod_{1\leq a<b\leq N}(x_a-x_b)^2\prod_{a=1}^{N}\frac{x_a^{\frac{N_f}{2}-N}}{\prod_{\alpha=1}^{F}\left(1+x_a\mu_\alpha\right)^{n_\alpha}}\,\mathrm{d}x_a.$$

From here, we already foresee how the balanced theory $N_f=2N$ will result in more tractable expressions. We henceforth restrict to this case, briefly commenting on the arbitrary $N_f\geq 2N$ case below.

The details of the calculations are very similar to the ones appearing in full detail in the next chapters. For this reason, we only sketch the derivation here.

The argument in Section 5.2 implies that the problem of computing $\mathcal{Z}^{\text{SQCD}}_{\mathbb{S}^3}$ at large N is reduced to solving the SPE

$$\mathrm{P}\int\mathrm{d}y\frac{\hat{\rho}(y)}{x-y}=\sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{2}\frac{1}{x+\mu_\alpha^{-1}}, \tag{5.5.2}$$

where $\zeta_\alpha=\frac{n_\alpha}{N}$, $\forall\alpha=1,\ldots,F$ are the Veneziano parameters, as before. In (5.5.2) we are denoting the eigenvalue density by $\hat{\rho}$ to indicate that it is for the exponential variable x, as opposed to $\rho(\phi)$.

Our goal is to solve (5.5.2) and obtain $\hat{\rho}(x)$. We follow standard methods [27, 176] (see also [202]).

From the insight gained in the previous sections, we expect a one-cut solution for $\hat{\rho}(x)$, with support $[\ell_-, \ell_+] \subseteq (0, \infty)$. Consider the planar resolvent $\omega(z)$, defined as

$$\omega(z) = \int_{\ell_-}^{\ell_+} \mathrm{d}y \frac{\hat{\rho}(y)}{z-y}, \qquad z \in \mathbb{C} \setminus [\ell_-, \ell_+]. \tag{5.5.3}$$

Let us introduce the notation

$$V'(x) = \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \frac{1}{x + \mu_\alpha^{-1}}$$

for the right-hand side of (5.5.2). Combining the SPE (5.5.2) with the definition (5.5.3), we arrive at [27, 176]

$$\omega(z) = \sqrt{(z-\ell_-)(z-\ell_+)} \oint \frac{\mathrm{d}w}{2\pi \mathrm{i}} \frac{V'(w)}{(z-w)\sqrt{(w-\ell_-)(w-\ell_+)}} \tag{5.5.4a}$$

$$= V'(z) + \sqrt{(z-\ell_-)(z-\ell_+)} \oint_{\mathbb{S}^1_\infty} \frac{\mathrm{d}w}{2\pi \mathrm{i}} \frac{V'(w)}{(z-w)\sqrt{(w-\ell_-)(w-\ell_+)}}, \tag{5.5.4b}$$

where the integration contour in the first line encircles the cut $[\ell_-, \ell_+]$ but leaves outside the point $z \in \mathbb{C} \setminus [\ell_-, \ell_+]$, while to pass to (5.5.4b) we have deformed the integration contour to infinity, picking the pole at $w = z$ in the process. We also use that the integrand in (5.5.4a) has no pole at $w = \infty$.

We are left with the task of evaluating the contour integral (5.5.4b) by residues. Only the poles of $V'(w)$ contribute, and we obtain:

$$\omega(z) = V'(z) + \sqrt{(z-\ell_-)(z-\ell_+)} \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \frac{1}{(z+\mu_\alpha^{-1})\sqrt{(\mu_\alpha^{-1}+\ell_-)(\mu_\alpha^{-1}+\ell_+)}}. \tag{5.5.5}$$

For this computation to work, we need to give a small imaginary part to the flavour fugacities μ_α, to ensure $\mu_\alpha \in \mathbb{C} \setminus [\ell_-, \ell_+]$. We then safely send μ_α to real values at the end of the computation, in a zero-dimensional version of the Feynman regularization prescription.

The endpoints $\ell_\pm$ are fixed by comparing the result (5.5.5) with the definition (5.5.3) and imposing the first two orders in the large z asymptotic to agree. Under the assumption $\zeta_{\mathrm{tot}} = 2$, it can be analytically checked that $\ell_- \to 0$ and $\ell_+ \to \infty$ is a solution.[13]

Taking the limits $\ell_- \downarrow 0$, $\ell_+ \uparrow \infty$ in (5.5.5), we deduce from (5.5.3) the eigenvalue density $\hat{\rho}(x)$, through the relation

$$-2\pi \mathrm{i} \hat{\rho}(x) = \lim_{\varepsilon \downarrow 0} \left[\omega(x + \mathrm{i}\varepsilon) - \omega(x - \mathrm{i}\varepsilon) \right], \qquad 0 < x < \infty.$$

After some rewriting and mapping back to the scalar field variable ϕ, the final answer is

$$\rho(\phi) = r \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \frac{1}{\cosh \pi r(\phi + m_\alpha)}, \qquad \phi \in \mathbb{R}\,. \tag{5.5.6}$$

The eigenvalue density $\rho(\phi)$ is properly normalized,

$$\int_{-\infty}^{+\infty} \mathrm{d}\phi \rho(\phi) = \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} = 1,$$

[13] It is convenient to assume that the masses are given in opposite pairs, so that the partition function (5.5.1) has a $\mathbb{Z}_2$ parity symmetry that implies $\ell_- = \ell_+^{-1}$.

the second equality holding by the balancing hypothesis.

As a consistency check, we compare the large r limit of the solution (5.5.6) with the $k \to 0$ limit of the solutions in Section 5.3. To remove the Chern–Simons term, we set $\zeta_{\text{tot}} = 2$ and then send $\lambda \to \infty$. The system is thus always in the intermediate phase, e.g. Phase II in the case of two symmetric mass scales.

At large r, (5.5.6) is exponentially suppressed away from the points $\phi = -m_\alpha$. There, however, it grows linearly in r. At leading order we find

$$\lim_{r\to\infty} \rho(\phi) = \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \delta\left(\phi + m_\alpha\right),$$

finding perfect agreement with the $\lambda \to \infty$ limit of the expressions in Section 5.3.

Free energy

Three-dimensional $\mathcal{N} = 4$ SQCD presents a single phase, both at finite and infinite r, consistent with the expectations. There are no massive couplings, because the gauge coupling goes to infinity at the IR superconformal point. In the large r limit we observe resonances precisely at the value of ϕ at which a hypermultiplet mode becomes massless. However, at finite r, these singularities are smoothened into a one-cut eigenvalue density, peaked around these special points.

We can compute the free energy using (5.5.6). Starting with the massless case $m_\alpha = 0\ \forall \alpha$, we get

$$\mathcal{F}_{\mathbb{S}^3}^{\text{Balanced SQCD}} = 2\int_{-\infty}^{\infty} \mathrm{d}\phi \rho(\phi) \log 2\cosh(\pi r\phi) = 2\log 4,$$

finding agreement with [214]. In fact, in the massless case but arbitrary $N_f \geq 2N$, the leading large N formula $\mathcal{F}_{\mathbb{S}^3}^{\text{SQCD}} = F(\zeta)$ was proven in [214, Eq. (3.10)], where (in our normalization)

$$F(\zeta) = \frac{\zeta^2}{2}\log(2\zeta) + \frac{(\zeta-2)^2}{2}\log\left(\frac{\zeta-2}{2}\right) - (\zeta-1)^2\log(\zeta-1).$$

The positivity and monotonicity of $F(\zeta)$ for $2 \leq \zeta < \infty$ provide a further check of the $\mathcal{F}$-theorem.

For the balanced theory with arbitrary masses, it is simpler to first evaluate the derivative of $\mathcal{F}_{\mathbb{S}^3}^{\text{SQCD}}$ with respect to a mass. It yields:

$$\begin{aligned}\frac{\partial \mathcal{F}_{\mathbb{S}^3}^{\text{SQCD}}}{\partial m_\beta} &= \zeta_\beta \int_{-\infty}^{\infty} \pi r \tanh \pi r\left(\phi + m_\beta\right)\rho(\phi)\mathrm{d}\phi \\ &= -\pi r \zeta_\beta \sum_\alpha \frac{\zeta_\alpha}{2}\frac{e^{\pi r m_\alpha} - e^{\pi r m_\beta}}{e^{\pi r m_\alpha} + e^{\pi r m_\beta}},\end{aligned}$$

where, in the second line, we have used $\cot^{-1}(e^{\pi r m}) + \tan^{-1}(e^{\pi r m}) = \frac{\pi}{2}$ followed by elementary manipulations. Integrating and imposing the condition $\sum_\beta m_\beta = 0$ we arrive at

$$\mathcal{F}_{\mathbb{S}^3}^{\text{SQCD}} = \sum_{\alpha,\beta} \zeta_\alpha \zeta_\beta \log\left(e^{\pi r m_\alpha} + e^{\pi r m_\beta}\right)$$

up to an integration constant, which can be set to zero by comparing with the large mass limit. In the massless limit, instead, we recover the value $4\log 2$, while, for example in the case of two opposite mass scales, $F = 2$, $m_1 = -m_2 \equiv m$, we get $\mathcal{F}_{\mathbb{S}^3}^{\text{SQCD}} = 2\log 2\cosh(\pi r m) + 2\log 2$.

Comments on SQCD with an arbitrary number of flavours

If we allow an arbitrary number $N_f \geq 2N$ of hypermultiplets in the fundamental representation, the SPE (5.5.2) is simply modified into

$$\mathrm{P}\int \mathrm{d}y \frac{\hat{\rho}(y)}{x-y} = -\frac{\zeta_{\mathrm{tot}}-2}{4x} + \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \frac{1}{x+\mu_\alpha^{-1}},$$

with $\zeta_{\mathrm{tot}} = \sum_{\alpha=1}^{F} \zeta_\alpha$ the sum of all the Veneziano parameters. Note that, as should be clear by power counting, it is not necessary to have $N_f = 2N$ for the SPE to simplify, but it suffices that $\frac{N_f}{N} - 2$ is of order $1/N$, hence sub-leading.

The computations go through in exactly the same way, except that $V'(z)$ acquires the extra term $-\frac{\zeta_{\mathrm{tot}}-2}{4z}$. Comparing (5.5.3) with the solution for $\omega(z)$, one derives a pair of equations to be solved for $\ell_\pm$. We have numerically solved the equations for $\ell_\pm$ at fixed masses and at specific values of ζ_α (we used $F = 2$, $\zeta_1 = \zeta_2$ and $m_1 = -m_2 = 1$), which shows how $\ell_- \to 0$ and $\ell_+ \to \infty$ as $\zeta_{\mathrm{tot}} \downarrow 2$, providing further support for the solution found above in the balanced case.

Chapter 6

Phases of five-dimensional supersymmetric Chern–Simons-matter theories

6.1 Introduction to the chapter

Supersymmetric quantum field theories in five and six dimensions are valuable windows onto the dynamics of interacting systems: they are constrained enough to be treated analytically, yet they are rich enough to uncover new phenomena. Five-dimensional $\mathcal{N} = 1$ field theories admit UV completion at superconformal fixed points [215], which are necessarily isolated [216] and strongly coupled [217]. 5d $\mathcal{N} = 1$ gauge theories, which are the main characters of the present work, sit in the IR of such superconformal field theories (SCFTs) and are connected to them by a renormalization group (RG) flow. The Coulomb branches of these theories have a geometric meaning inherited from M-theory compactified on a singular Calabi–Yau threefold X [218]. The extended Kähler cone of X, that we denote $\mathscr{C}(X)$, is the union of chambers that parametrize different crepant resolutions of X, as sketched in Figure 6.1. In the gauge theory description, the extended Kähler cone $\mathscr{C}(X)$ is identified with the extended Coulomb branch and the walls separating two chambers correspond to codimension-one loci on the Coulomb branch at which a state becomes massless [218–220].

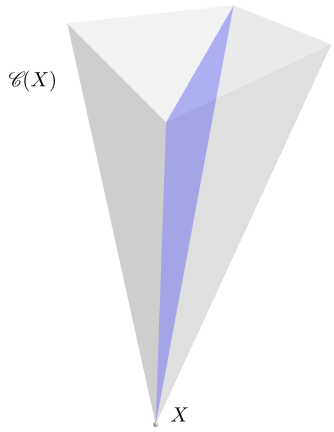


Figure 6.1. Schematic illustration of the extended Kähler cone $\mathscr{C}(X)$ of a singular Calabi–Yau threefold X. Across the blue wall separating distinct chambers, a state becomes massless.

Five-dimensional $\mathcal{N} = 1$ Yang–Mills theories with classical gauge group descend from $6d$

$\mathcal{N} = (1,0)$ SCFTs compactified on a circle, with subsequent RG flows triggered by massive deformations. The combination of geometric and field theoretic perspectives, integrated with new combinatorial tools [221, 222], yields a firm grasp of the gauge theories consistently realized within this framework [219–229].

The importance of weakly coupled, Lagrangian gauge theories resides in the fact that supersymmetry protected quantities carry information on the strongly interacting UV fixed point.

The supersymmetric localization program [38] aims at reducing the path integral description of supersymmetric observables, as the sphere partition function or the vacuum expectation value of Wilson loops, to finite-dimensional integrals. In recent years, a wealth of exact results has been obtained from localization on a broad variety of compact manifolds. Localization on the five-sphere has been carried out in [230–235], see also the review article [236].

The goal of the present work is to analyze the phase structure of the sphere partition function and of half-BPS Wilson loops at large rank of the gauge group. There exists a vast literature discussing the large N behaviour of 5d $\mathcal{N} = 1$ theories on the sphere [237–246], mostly pivoting around the match with the holographic dual. A thorough analysis of the large N phases of certain theories on the five-sphere appears in the work of Minahan and Nedelin [247, 248].

A fruitful approach to study the phase diagram consists in putting the theory on a very large sphere. This procedure, named decompactification limit, has been successfully applied to supersymmetric theories in 4d [195–197, 201, 200, 211, 249, 250] and 3d [202, 62, 203–205, 3]. The first realization of a decompactification limit in 5d is in [208]. Curvature effects are negligible from this vantage point, thus providing a reliable approximation of flat space dynamics without spoiling the computability guaranteed by localization.

In this work, we undertake a systematic study of the phases of five-dimensional $\mathcal{N} = 1$ gauge theories in the decompactification limit. We discuss both the sphere partition function and the vacuum expectation value of Wilson loops, for various choices of gauge group and matter content. It is worthwhile to emphasize that phase transitions are a signature of systems with infinitely many degrees of freedom, whilst localization on $\mathbb{S}^5$ reduces the observables to matrix integrals over zero-modes. For this reason, the large N limit is instrumental for the ensuing analysis and crucial for the appearance of critical loci in parameter space: it will be this limit, rather than the large sphere limit, to give rise to a non-analytic behaviour.

The present chapter is organized as follows. In the rest of this introductory section, we list concisely our main results and mention potential avenues for future research.

The next section is the core of the subsequent analysis. In Subsection 6.2.3, the most general solution to the sphere free energy is obtained, for theories with fundamental hypermultiplets. After that, we discuss half-BPS Wilson loops in the fundamental and in the antisymmetric representation and find the most general solution for these observables in Subsection 6.2.4. Subsection 6.2.5 extends the results to theories with hypermultiplets in representations of higher dimension.

The sections that follow are devoted to a detailed analysis of various gauge theories with gauge group $U(N)$ in Section 6.3, $SU(N)$ in Section 6.4 and the other classical groups in Section 6.5. The text is complemented with two appendices.

6.1.1 Summary of results and outlook

In the study of gauge theories with simple gauge group we find a rich phase diagram, with the models undergoing a phase transition each time a mass parameter is decreased below or above a characteristic scale.

- For gauge group $U(N)$ and hypermultiplets in the fundamental representation, the phase transitions are generically second order. There are, however, exceptions of two types:

(i) In the theory with symmetric assignment of masses the phase transitions are third order;

(ii) In absence of a Yang–Mills term all the phase transitions are third order.

- For gauge groups $SU(N)$, $USp(2N)$, $SO(2N)$ or $SO(2N+1)$ with fundamental hypermultiplets the phase transitions are always third order.
- Expectation values of Wilson loops in the fundamental representation follow a perimeter law. Moreover,
  - ▷ In $U(N)$ theories, their derivative is discontinuous;
  - ▷ They have second order discontinuities when the gauge group is $SU(N)$, $USp(2N)$, $SO(2N)$ or $SO(2N+1)$, and for unitary group in case (i) above.
- Wilson loops in the antisymmetric representation follow a perimeter law and have discontinuities in the first derivative in $U(N)$ theories, and in the second derivative in all other theories.
- For any gauge group and an adjoint hypermultiplet, the theory has a second order phase transition.

We conclude that all gauge theories with a known UV SCFT completion belong to the same universality class. On the contrary, the critical behaviour of $U(N)$ gauge theories depends on the deformation pattern. This raises the question of what kind of UV completion they admit, if at all. It might be that only balanced theories possess a UV fixed point. Another possible explanation is that every $U(N)$ gauge theory descends from a SCFT, but the Abelian factor introduces some subtlety in the order of limits, namely strong coupling and large N limit do not commute. Point (ii) above would fit in this scenario, but other puzzles would remain. In either case, a deeper understanding of these models by diverse methods is highly desirable.

Finally, a systematic understanding of the relation between phase transitions and one-form symmetries, elaborating on the observations in Subsection 6.4.5, is left for future work.

6.2 Gauge theories with large rank on the five-sphere

6.2.1 Coulomb branch localization and large N limit

The moduli spaces of supersymmetric vacua of five-dimensional $\mathcal{N}=1$ gauge theories consist of various branches. Among them, the Coulomb branch is parametrized by the zero-mode of the real scalar ϕ in the $\mathcal{N}=1$ vector multiplet, conjugated in a Cartan subalgebra. For the sake of clarity, the ensuing exposition is based on gauge group $U(N)$, but the aspects we review hold for any compact semi-simple Lie group G.

The Coulomb branch is a wedge inside $\mathbb{R}^{\mathrm{rank}(G)}$ fixed by the choice of Weyl chamber:

$$\mathscr{C}_{\text{gauge}} = \mathbb{R}^{\mathrm{rank}(G)}/\mathrm{Weyl}(G). \tag{6.2.1}$$

It is convenient to consider the extended Coulomb branch of the theory,

$$\mathscr{C}(X) = \mathscr{C}_{\text{gauge}} \times \mathscr{C}_{\text{flavour}}. \tag{6.2.2}$$

In the left-hand side we have adopted the notation $\mathscr{C}(X)$ from M-theory on the singular Calabi–Yau threefold X, and on the right-hand side we have split the extended Coulomb branch into the gauge part, defined in (6.2.1) and parametrized by the dynamical scalar ϕ, and a flavour

part, parametrized by the real scalar fields $\{m_\alpha\}$ in a background vector multiplet for the flavour symmetry group. More generally, one may think of $\mathscr{C}(X)$ as a $\mathscr{C}_{\text{gauge}}$-fibration over the parameter space $\mathscr{C}_{\text{flavour}}$ [223].

Hypermultiplet modes are massive at generic points of the Coulomb branch and become massless at codimension-one loci inside the extended Coulomb branch.

In this work, we analyze the phases of the 5d $\mathcal{N}=1$ gauge theories looking at the matrix model obtained from localization on $\mathbb{S}^5$ [230, 231, 233] (for a review, see [236]). The partition function of the theory in its Coulomb branch localized on $\mathbb{S}^5$ is

$$\mathcal{Z}_{\mathbb{S}^5} = \frac{1}{N!}\int_{-\infty}^{+\infty} \mathrm{d}\phi_1 \cdots \int_{-\infty}^{+\infty} \mathrm{d}\phi_N \; Z_{\text{class}}(\phi) Z^{\text{vec}}_{\text{1-loop}}(\phi) Z^{\text{hyp}}_{\text{1-loop}}(\phi) Z_{\text{inst}}(\phi) \tag{6.2.3}$$

where Z_{class} is the classical contribution by the BPS field configuration, $Z_{\text{1-loop}}$ are the one-loop determinants and Z_{inst} contains the non-perturbative contributions from instantons on $\mathbb{P}^2 \subset \mathbb{S}^5$ [230, 231, 233]. The integration domain has been extended from $\mathscr{C}_{\text{gauge}} \cong \mathbb{R}^N/S_N$ to the whole $\mathbb{R}^N$ using the Weyl invariance of the integral, at the cost of a factor $\frac{1}{N!}$.

The classical piece is

$$Z_{\text{class}}(\phi) = \prod_{a=1}^{N} e^{-V(\phi_a)}, \tag{6.2.4a}$$

$$V(\phi) = \frac{\pi r^3 k}{3}\phi^3 + \frac{8\pi^3 r^3}{g^2_{\text{YM}}}\phi^2. \tag{6.2.4b}$$

k is the Chern–Simons level and g_{YM} is the Yang–Mills coupling, and we will henceforth use the notation

$$h = \frac{8\pi^2}{g^2_{\text{YM}}} \tag{6.2.5}$$

for the inverse gauge coupling, with mass dimension one. The 5d $\mathcal{N}=1$ gauge theories we study, with the exception of those with gauge group $U(N)$, are massive deformations of a UV SCFT, with h determining the scale of such deformation. If the UV completion is a $6d$ $\mathcal{N}=(1,0)$ theory compactified on a circle of radius β, then $h \propto \beta^{-1}$.

The one-loop determinants for gauge group $U(N)$ or $SU(N)$ are [231]:

$$Z^{\text{vec}}_{\text{1-loop}}(\phi) = \prod_{1\le a<b\le N} \left[\sinh \pi r\,(\phi_a-\phi_b)\;\; e^{\frac{1}{2}f(ir(\phi_a-\phi_b))}\right]^2, \tag{6.2.6a}$$

$$Z^{\text{hyp}}_{\text{1-loop}}(\phi) = \prod_{\alpha=1}^{F}\prod_{a=1}^{N} \left[\cosh \pi r\,(\phi_a+m_\alpha)\;\; e^{-f\left(\frac{1}{2}-ir(\phi_a+m_\alpha)\right)-f\left(\frac{1}{2}+ir(\phi_a+m_\alpha)\right)}\right]^{\frac{n_\alpha}{4}}. \tag{6.2.6b}$$

Here we have assumed that the matter content consists of N_f fundamental hypermultiplets with degenerate masses, so that n_α of them have equal mass m_α, and

$$N_f = \sum_{\alpha=1}^{F} n_\alpha.$$

This choice of masses is non-generic, and the singular loci of the Coulomb branch degenerate into walls having n_α layers, as in Figure 6.2.

The function $f(x)$ appearing in the one-loop determinants (6.2.6) has been defined in [230, 231] and comes from the zeta function regularization of the infinite product

$$\prod_{n=1}^{\infty}\left(1-\frac{x^2}{n^2}\right)^{n^2}.$$

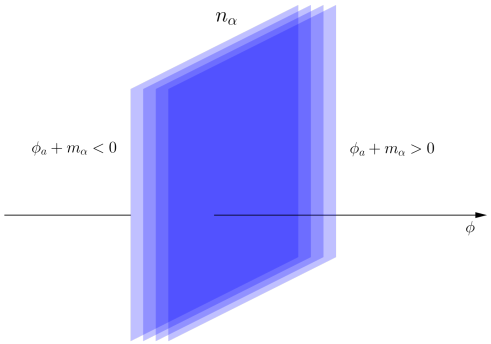


Figure 6.2. When the masses are degenerate, n_α walls inside the Coulomb branch collide.

For our purposes, it suffices to say that it is manifestly even, $f(-x) = f(x)$, and its derivative satisfies

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \pi x^2 \cot(\pi x)\,. \tag{6.2.7}$$

We do not discuss the non-perturbative contributions, since $Z_{\text{inst}} \to 1$ exponentially fast in the setup of this work.[14]

When the gauge group is $SU(N)$, the scalar ϕ must satisfy the constraint

$$\sum_{a=1}^{N} \phi_a = 0,$$

that can be enforced adding a linear term in the potential (6.2.4b) and imposing the independence of the partition function from the Lagrange multiplier. Writing this linear shift as

$$V(\phi) \mapsto V(\phi) + 4\pi r^2 \xi \phi$$

with $V(\phi)$ as in (6.2.4b), we recognize in the Lagrange multiplier ξ a Fayet–Iliopoulos parameter. A geometric reduction from a $U(N)$ to a $SU(N)$ factor in the gauge group has been described in [228], and we revisit their argument in the matrix model language in Appendix 6.A.2.

Before proceeding we notice that, for the integral representation of the partition function to be convergent, one has to impose

$$N - |k| - \frac{1}{2}\sum_{\alpha=1}^{F} n_\alpha \geq 0, \tag{6.2.8}$$

which is a necessary (and believed sufficient) condition for the theory to descend from a non-trivial SCFT in the UV [251].

6.2.2 Large N and decompactification limit

We will take the large N limit of the matrix model (6.2.3) and then compute its decompactification limit $r \to \infty$. Writing the partition function in the form

$$\mathcal{Z}_{\mathbb{S}^5} = \frac{1}{N!}\int_{\mathbb{R}^N} \mathrm{d}^N\phi\; e^{-S_{\text{eff}}(\phi)},$$

[14] Equivalently, in the 't Hooft limit taken below, instantons become infinitely massive and decouple.

we see that the leading contributions in the large N and large r limit come from the stationary points of S_{eff}, while away from these points the integrand is damped as $e^{-r^3 N^2(\cdots)}$. Therefore, the problem is reduced to finding the solutions ϕ^* to the saddle point equations (SPEs)

$$\left.\frac{\partial S_{\text{eff}}(\phi)}{\partial \phi_a}\right|_{\phi=\phi^*} = 0, \qquad a = 1, \dots, N.$$

Let us now look into the simplifications brought in by the large radius limit. Using the parity of f, property (6.2.7) and retaining only the leading contribution at large r, we see that the hypermultiplet and vector multiplet one-loop determinants contribute to the SPE respectively

$$-\frac{\pi r^3}{2} \sum_{\alpha=1}^{N_f} (\phi_a^* + m_\alpha)^2 \operatorname{sgn}(\phi_a^* + m_\alpha),$$
$$\pi r^3 \sum_{b \neq a} (\phi_a^* - \phi_b^*)^2 \operatorname{sgn}(\phi_a^* - \phi_b^*).$$

Putting these terms together with the derivative of the classical piece we arrive at the system of N SPEs

$$k(\phi_a^*)^2 + 2h\phi_a^* + \check{\xi} - \frac{1}{2}\sum_{\alpha=1}^{F} n_\alpha (\phi_a^* + m_\alpha)^2 \operatorname{sgn}(\phi_a^* + m_\alpha) = -\sum_{b \neq a} (\phi_a^* - \phi_b^*)^2 \operatorname{sgn}(\phi_a^* - \phi_b^*) \quad (6.2.9)$$

for $a = 1, \dots, N$. In the latter expression we have introduced the scaled quantity $\check{\xi} = \frac{4\xi}{r}$, to keep track of the Lagrange multiplier at large radius.[15]

Integrating the SPE (6.2.9) and summing over a, we arrive at twice the prepotential of [220]. The factor of two is predicted from the equivariant localization on the round sphere: the partition function only receives contributions from small neighbourhoods of the two fixed points of an isometry rotating a $\mathbb{P}^2$ inside $\mathbb{S}^5$.

6.2.3 Solution

Our goal is to solve the SPE (6.2.9) in a large N 't Hooft limit, with

$$\frac{N}{h} = \lambda \text{ fixed}, \qquad \frac{N}{k} = t \text{ fixed}. \quad (6.2.10)$$

We moreover consider a Veneziano limit, in which the number N_f of fundamental hypermultiplets grows linearly with N, hence we introduce the Veneziano parameters

$$\frac{n_\alpha}{N} = \zeta_\alpha \text{ fixed}, \qquad \forall\ \alpha = 1, \dots, F. \quad (6.2.11)$$

We also keep $\tilde{\xi} = \frac{\check{\xi}}{N}$ fixed. The convergence condition (6.2.8) in this limit becomes

$$\frac{1}{|t|} + \frac{1}{2}\sum_{\alpha=1}^{F} \zeta_\alpha \leq 1. \quad (6.2.12)$$

The large N limit of a 5d $\mathcal{N} = 1$ $U(N)$ Yang–Mills theory in the decompactification regime has been addressed in [208], but only for a very special choice of masses and no Chern–Simons

[15] The linear coupling between ξ and ϕ actually comes from a mixed Chern–Simons term, so it should scale with r^3, as the pure Chern–Simons term. The introduction of a new variable $\check{\xi}$ is an artefact of the normalization, not an additional scaling that we impose.

term. We now derive the phase structure of the most general consistent gauge theory with simple gauge group in the decompactification limit.

Let us introduce the eigenvalue density $\rho(\phi)$, which is normalized:

$$\int \mathrm{d}\phi \rho(\phi) = 1 \tag{6.2.13}$$

and has compact support. The effective action $S_{\text{eff}}(\phi)$ is not an even function, therefore we do not expect the support of the eigenvalue density to be symmetric. Moreover, $\rho(\phi)$ is not required to be a function and, in general, it is sufficient that $\rho(\phi)\mathrm{d}\phi$ is a measure on the union of intervals along a selected integration cycle. Throughout this chapter (as in the previous one), the integration cycle is the real axis and the measure is supported on a compact interval,

$$\mathrm{supp}\rho = [A, B] \subset \mathbb{R}\,.$$

In the scaling limit (6.2.10)-(6.2.11) the system of N saddle point equations is recast into a single integral equation

$$-\int_A^B \mathrm{d}\psi \rho(\psi)\ \left(\phi^\star - \psi\right)^2 \mathrm{sgn}\left(\phi^\star - \psi\right) = \frac{1}{t}\left(\phi^\star\right)^2 + \frac{2}{\lambda}\phi^\star + \tilde{\xi} - \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2}\left(\phi^\star + m_\alpha\right)^2 \mathrm{sgn}\left(\phi^\star + m_\alpha\right) \tag{6.2.14}$$

to be satisfied by every $\phi^\star \in [A, B]$. Here we have denoted $\phi^\star$ the variable running over the continuous spectrum of eigenvalues of ϕ^*, being ϕ^* the solution to the SPE (6.2.9). We have done so in the hope of avoiding confusion between the original N-dimensional integration variable $\phi = (\phi_1, \ldots, \phi_N)$, the fixed N-dimensional saddle point $\phi^* = (\phi_1^*, \ldots, \phi_N^*)$, and the one-dimensional real variable $\phi^\star \in [A, B]$. Henceforth, we will simply use ϕ instead of $\phi^\star$ to reduce clutter.

The mechanism triggering the phase transitions is read off from (6.2.14): the right-hand side changes when a mass parameter crosses A or B, leading to a new eigenvalue density.

Taking three derivatives, we find that the generic solution to the SPE (6.2.14) is

$$\rho(\phi) = c_A \delta\left(\phi - A\right) + c_B \delta\left(\phi - B\right) + \sum_{\alpha=1}^{F} c_\alpha \delta\left(\phi + m_\alpha\right), \tag{6.2.15}$$

with coefficients

$$c_\alpha = \begin{cases} \frac{\zeta_\alpha}{2} & -m_\alpha \in [A, B] \\ 0 & \text{otherwise} \end{cases} \qquad \alpha = 1, \ldots, F.$$

Throughout this work, we define the δ-functions centered at the endpoints of $\mathrm{supp}\rho$ taking the limit from inside the support [208],

$$\delta\left(\phi - A\right) = \lim_{\varepsilon \to 0^+} \delta\left(\phi - (A + \varepsilon)\right); \qquad \delta\left(\phi - B\right) = \lim_{\varepsilon \to 0^+} \delta\left(\phi - (B - \varepsilon)\right).$$

The solution (6.2.15) is given in terms of two coefficients c_A, c_B and two endpoints A, B to be determined. Plugging (6.2.15) back into the cubic equation (6.2.14) yields a system of three equations:

$$-c_A + c_B - \sum_{\alpha=1}^{F} c_\alpha \tilde{s}_\alpha = \frac{1}{t} - \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \tilde{s}_\alpha \tag{6.2.16a}$$

$$c_A A - c_B B - \sum_{\alpha=1}^{F} c_\alpha \tilde{s}_\alpha m_\alpha = \frac{1}{\lambda} - \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \tilde{s}_\alpha m_\alpha \tag{6.2.16b}$$

$$-c_A A^2 + c_B B^2 - \sum_{\alpha=1}^{F} c_\alpha \tilde{s}_\alpha m_\alpha^2 = \tilde{\xi} - \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \tilde{s}_\alpha m_\alpha^2 \tag{6.2.16c}$$

in which we have introduced the shorthand notation

$$\tilde{s}_\alpha = \frac{1}{2}\left[\operatorname{sgn}\left(A+m_\alpha\right)+\operatorname{sgn}\left(B+m_\alpha\right)\right] = \begin{cases} -1 & -m_\alpha > B \\ 0 & A \leq -m_\alpha \leq B \\ +1 & -m_\alpha < A. \end{cases}$$

The normalization condition (6.2.13) applied to (6.2.15) imposes

$$c_A + c_B + \sum_{\alpha=1}^{F} c_\alpha = 1. \tag{6.2.17}$$

Therefore the two coefficients c_A and c_B and the two endpoints A and B of $\mathrm{supp}\rho$ are determined as functions of the gauge theoretical parameters t, λ, $\{\zeta_\alpha, m_\alpha\}$, from the system (6.2.16) completed by the normalization (6.2.17).

Solving (6.2.16a) together with (6.2.17) yields

$$\begin{aligned} c_A &= \frac{1}{2}\left[1 - \frac{1}{t} - \sum_{\alpha=1}^{F}\left(c_\alpha - \tilde{s}_\alpha\left(\frac{\zeta_\alpha}{2} - c_\alpha\right)\right)\right] \\ c_B &= \frac{1}{2}\left[1 + \frac{1}{t} - \sum_{\alpha=1}^{F}\left(c_\alpha + \tilde{s}_\alpha\left(\frac{\zeta_\alpha}{2} - c_\alpha\right)\right)\right]. \end{aligned}$$

To find the endpoints A and B we plug these values in (6.2.16b)-(6.2.16c). The system is quadratic in the variables A and B, thus we find a pair of solutions: at each point in the parameter space, we should retain the one consistent with $A < B$, which must hold by construction. We stress that (6.2.15) has been derived taking derivatives with respect to ϕ, thus working under the assumption that the interior of $\mathrm{supp}\rho$ is not empty. Whenever a consistent pair of endpoints A, B cannot be found, we should drop this assumption and take into account solutions supported at a single point, $\rho(\phi) = \delta(\phi)$.

Phase transitions in the theory are signalled by a non-analytic behaviour of the free energy

$$\mathcal{F}_{\mathbb{S}^5} = -\frac{1}{\pi r^3 N^2}\log\left|\mathcal{Z}_{\mathbb{S}^5}\right|. \tag{6.2.18}$$

In the decompactification and large N 't Hooft limit it becomes

$$\mathcal{F}_{\mathbb{S}^5} = \frac{1}{6}\int \mathrm{d}\phi\rho(\phi)\int \mathrm{d}\psi\rho(\psi)\ |\phi-\psi|^3 + \int \mathrm{d}\phi\rho(\phi)\left[\frac{1}{3t}\phi^3 + \frac{1}{\lambda}\phi^2 - \sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{6}|\phi+m_\alpha|^3\right].$$

The linear term proportional to $\tilde{\xi}$ does not contribute by construction. Using the solution (6.2.15) for $\rho(\phi)$, $\mathcal{F}_{\mathbb{S}^5}$ is found to be

$$\begin{aligned} \mathcal{F}_{\mathbb{S}^5} = {} & \frac{1}{3}\left[c_A c_B|B-A|^3 + \sum_{\alpha=1}^{F} c_\alpha\left(c_A|A+m_\alpha|^3 + c_B|B+m_\alpha|^3 + \sum_{\alpha'=1}^{F}\frac{c_{\alpha'}}{2}|m_\alpha - m_{\alpha'}|^3\right)\right] \\ & + \frac{1}{3t}\left[c_A A^3 + c_B B^3 - \sum_{\alpha=1}^{F} c_\alpha m_\alpha^3\right] + \frac{1}{\lambda}\left[c_A A^2 + c_B B^2 + \sum_{\alpha=1}^{F} c_\alpha m_\alpha^2\right] \\ & - \sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{6}\left[c_A|A+m_\alpha|^3 + c_B|B+m_\alpha|^3 + \sum_{\alpha'=1}^{F} c_{\alpha'}|m_\alpha - m_{\alpha'}|^3\right]. \end{aligned} \tag{6.2.19}$$

Recall that the coefficients c_α vanish unless $A < -m_\alpha < B$, in which case $c_\alpha = \frac{\zeta_\alpha}{2}$. This implies that whenever $A < -m_\alpha < B$ the one-loop contribution of the hypermultiplets of mass m_α is almost entirely cancelled between the first and the last line in (6.2.19). This is consistent with the mass m_α being below the characteristic energy scale of the problem: the hypermultiplet cannot be integrated out, whence no one-loop effect is generated. The cancellation of the one-loop effects between the first and third line of (6.2.19) when $A < -m_\alpha < B$ leaves behind a contribution

$$-\sum_{\alpha'=1}^{F} \frac{\zeta_\alpha \zeta_{\alpha'}}{12} |m_\alpha - m_{\alpha'}|^3 . \tag{6.2.20}$$

It reproduces the one-loop contribution of the massive W-bosons in the background vector multiplet for the flavour symmetry broken by the solution in the phase considered.

The continuity of A and B at each critical surface and the jump by $\frac{\zeta_\alpha}{2}$ of c_A when $-m_\alpha$ crosses A, or of c_B when $-m_\alpha$ crosses B, guarantee the continuity of the free energy at each transition point. Furthermore, the continuity of $\rho(\phi)$ can be used to prove that the transition must be at least second order. This is confirmed by the explicit computations in each case.

In sections 6.3 and 6.4 we consider gauge theories with gauge group $U(N)$ and $SU(N)$ respectively, and with the other classical groups in Section 6.5, and present their large N phase structure explicitly.

$\mathcal{F}$-theorem

The sphere partition function measures the degrees of freedom of a field theory in odd dimensions [213]. In 5d and with normalization (6.2.18), the $\mathcal{F}$-theorem states [213]

$$\mathcal{F}_{\mathbb{S}^5}^{(\mathrm{IR})} > \mathcal{F}_{\mathbb{S}^5}^{(\mathrm{UV})} . \tag{6.2.21}$$

Compelling evidence for this claim has been presented, for instance, in [252, 253]. Inequality (6.2.21) holds when both sides are evaluated at fixed points but, under favourable circumstances, the free energy can be shown to be monotonic all along the RG flow connecting the UV and the IR fixed points. Expression (6.2.19) can be used to provide new support for the $\mathcal{F}$-theorem.

For fixed values of the masses, $\lambda \to 0$ drives the theory to the IR. From (6.2.19) and using the dependence of A and B on λ through (6.2.16), it follows that (6.2.21) is satisfied between any two points on the RG flow. A direct proof of (6.2.21) is less obvious from (6.2.19) at fixed λ and increasing masses, but it can nevertheless be confirmed using the explicit results in Section 6.4.

Remarks on the decompactification limit

We follow the standard nomenclature denoting the large sphere limit as decompactification limit, but it ought to be remarked that the localization procedure requires a (equivariantly) compact topology, and the limit $r \to \infty$ should be really meant as the zero curvature limit $\frac{1}{r} \to 0$. As already emphasized in the Section 6.1, this allows to neglect curvature effects in a controlled way, but only after putting the localization machinery at work on a compact manifold.

A further remark concerns the sign of the Yang–Mills 't Hooft coupling λ. We will consider $\lambda^{-1} \in \mathbb{R}$. The interpretation of this may be puzzling from a field theoretic viewpoint, because then instanton corrections would contribute exponentially (instead of being exponentially suppressed) for $\lambda^{-1} < 0$. Moreover, as reviewed in Appendix 6.A.1, the parameter h in (6.2.5) has the meaning of a volume, thus it should not become negative. Nevertheless, the perturbative partition function can be analytically continued letting $h \in \mathbb{R}$ in (6.2.3) but keeping the non-perturbative quantities, such as instanton masses, as functions of $|h|$. For a thorough discussion on negative Yang–Mills coupling, see [208, 248]. Besides, the SPE (6.2.14) may be likewise analytically continued to negative values of the Veneziano parameters ζ_α.

One last comment is about flavour symmetry. For $U(N)$, the mass parameters belong to a background $SU(N_f)$ vector multiplet. In order not to violate the flavour symmetry, we will assume $N_f = 1 + \sum_{\alpha=1}^{F} n_\alpha$ and give the extra hypermultiplet a mass

$$m_{N_f} = -\sum_{\alpha=1}^{F} n_\alpha m_\alpha .$$

Its contribution is suppressed in the Veneziano limit (6.2.11) and drops out of the SPE.

Remarks on phase transitions and matrix models

As already mentioned, integrals over matrix degrees of freedom do not admit a notion of phase diagram, unless the number N of eigenvalues is sent to infinity. From a field theoretical perspective, a phase structure may originate from the infinite volume limit as well. Phase transitions among distinct chambers of $\mathscr{C}(X)$ in flat space belong to this latter class, while the phase transitions we are concerned with are instead of the first type.

The presence of a phase transition in the decompactification limit does not automatically imply that the transition exists at large N but finite radius. In fact, this implication fails in 3d [202]. Nevertheless, we will now argue that the situation is different in 5d and the transitions discovered with the aid of the decompactification limit persist at finite radius.

A generic effect of finite $\frac{1}{r^2}$ corrections is to smoothen the δ-singularities into peaked curves of finite height and width. In 3d Chern–Simons theories with fundamental hypermultiplets, the solution at large N but finite r is given by a deformation of the pure Chern–Simons eigenvalue density, on top of which a peak forms each time a mass parameter is decreased [202, 3]. The shape of the eigenvalue density is changed without breaking its support [202], therefore there is no phase transition at finite radius.

On the contrary, in 5d we do not have a distribution on top of which the peaks are formed, and we expect that each new peak will produce a new cut in the support of $\rho(\phi)$. Let us elaborate further on this statement. Starting with a pure gauge theory and assuming a very small size of the support, the $\sinh(\phi_a - \phi_b) \approx (\phi_a - \phi_b)$ in (6.2.6a) will dominate against the e^f term, leading to the equilibrium equation of a cubic matrix model. The generic solution is supported on two intervals and the two pieces degenerate into $\delta(\phi - A)$ and $\delta(\phi - B)$ as $\frac{1}{r} \to 0$. A phase transition when the two intervals merge was observed in [247]. In turn, we can work with finite but large enough r to guarantee that the model remains in the two-cut phase. Most importantly, in the one-cut phase $\mathrm{supp}\rho$ is moved away from the real axis [247], thus such solution is discarded by the procedure adopted in the present work.

Decreasing the masses of the hypermultiplets from infinity, new peaks will form on top of the finite radius solution. However, as these new peaks are moved away from one endpoint, they will break the support and produce additional intervals, until they reach the other endpoint and the two intervals merge. In conclusion, the phase transitions uncovered throughout this work are expected to be genuine large N phase transitions, associated with splitting of $\mathrm{supp}\rho$, and not a consequence of the large r approximation.

Phase transitions as the ones observed are ubiquitous in gauge theories with an underlying cohomological structure, that allows to reduce the observables to a matrix model. Prototypical in this respect is the Gross–Witten–Wadia third order phase transition [190–192] in 2d. We ought to emphasize that the integrals from localization of 5d $\mathcal{N} = 1$ gauge theories are not of standard random matrix type, meaning that there seem to be no change of variables to recast the vector multiplet one-loop determinant in the form of a Vandermonde determinant. As a consequence,

the mechanism underlying the phase transitions is inherently technically different from the Gross–Witten–Wadia transition. Nonetheless, a recurrent theme is that phase transitions are triggered by states becoming massless. In the present setup the light states come from the matter sector. Conversely, in pure 2d Yang–Mills theory there are no propagating perturbative particles, thus the transition is induced by instantons [254].

6.2.4 Wilson loops

The eigenvalue density (6.2.15) can be exploited to compute the vacuum expectation value (vev) of Wilson loops on $\mathbb{S}^5$ that preserve half of the supercharges (that is, are half-BPS) in the large r and large N limit. The contribution to the effective action from a Wilson loop in a representation of fixed size is sub-leading and does not alter the eigenvalue density in the large N limit. We conclude that the vev of a Wilson loop in the fundamental representation F is

$$\langle W_{\mathsf{F}} \rangle = \int_A^B \mathrm{d}\phi \rho(\phi) e^{2\pi r \phi} = c_A e^{2\pi r A} + c_B e^{2\pi r B} + \sum_{\alpha=1}^{F} c_\alpha e^{2\pi r m_\alpha}.$$

The continuity of this expression follows from the continuity of A and B at the critical values, together with the jump by $\frac{\zeta_\alpha}{2}$ of c_A or c_B when $-m_\alpha$ crosses A or B respectively. The Wilson loop vev follows a perimeter law, $\log \langle W_{\mathsf{F}} \rangle \approx (2\pi B) r$, as expected and in agreement with [240].

For classical gauge group, it is proven in Subsection 6.4.4 that the Wilson loop is differentiable, as a consequence of the scalar ϕ being traceless.

Wilson loops in large antisymmetric representations

Expectation values of Wilson loops in a given representation whose size grows with N deserve further consideration. Let A_K be the rank-K antisymmetric representation of the gauge group. This implies $0 \leq K \leq N$ for $U(N)$ and $0 \leq K \leq N-1$ for $SU(N)$. We consider a Wilson loop in the representation A_K along a great circle inside $\mathbb{S}^5$.

The formalism to study the vev of such loop operators in the large N limit, with K growing with N, was developed in [209] for 4d $\mathcal{N}=4$ Yang–Mills, and applied to 4d $\mathcal{N}=2$ in [210] and to 3d Chern–Simons theories in [3]. The derivation of [209] does not depend on the specific theory, as long as the Wilson loop vev is localized to a finite-dimensional integral, and directly extends to the present five-dimensional setting, with a few improvements to accommodate a non-even eigenvalue density. The central idea is to introduce the generating function

$$\Phi_{\mathsf{A}}(w) = \left\langle \prod_{a=1}^{N} \left(1 + w e^{2\pi r \phi_a}\right) \right\rangle. \tag{6.2.22}$$

We are interested in the large N with the ratio

$$\kappa = \frac{K}{N} \text{ fixed}, \qquad 0 \leq \kappa \leq 1. \tag{6.2.23}$$

The details are essentially identical to Section 5.4.3, but using the eigenvalue densities of the five-dimensional models.

Plugging the general solution (6.2.15) for the eigenvalue density in the argument of Section 5.4.3, we get

$$\begin{aligned} \langle W_{\mathsf{A}_K} \rangle = e^{rAK} r(B-A) \oint_\Gamma \frac{dz}{2\pi i}\, e^{zN\kappa r(B-A)} \left[1 + e^{-zr(B-A)}\right]^{Nc_A} \\ \times \left[1 + e^{(1-z)r(B-A)}\right]^{Nc_B} \prod_{\alpha=1}^{F} \left[1 + e^{-r(m_\alpha + A + z(B-A))}\right]^{Nc_\alpha}. \end{aligned} \tag{6.2.24}$$

Regardless of the details of each specific phase of any theory, the upshot is that $\log\langle W_{\mathsf{A}_K}\rangle$ grows linearly in N and r, meaning that it follows a perimeter law, and is of the general form

$$\log\langle W_{\mathsf{A}_K}\rangle \approx c_1 rK + c_2 rN$$

at leading order in both N and r, with c_1 and c_2 simple functions of A, B and the masses $\{m_\alpha\}$.

6.2.5 Hypermultiplets in other representations

So far the spotlight has been on theories with fundamental hypermultiplets. We now turn our attention to other types of matter content and analyze the large N limit of $SU(N)$ theories with hypermultiplets in the adjoint, symmetric or rank-two antisymmetric representation.

Adjoint hypermultiplet

We consider Yang–Mills theory with a massive adjoint hypermultiplet [239]. This model has enhanced $\mathcal{N}=2$ supersymmetry at $m=0$. The SPE in the large N decompactification limit is

$$\begin{aligned}\frac{2}{\lambda}\phi = &-\int_A^B \mathrm{d}\psi\rho(\psi)\,(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi)\\ &+\int_A^B \mathrm{d}\psi\rho(\psi)\left[\frac{1}{2}\,(\phi-\psi+m)^2\,\mathrm{sgn}\,(\phi-\psi+m)+\frac{1}{2}\,(\phi-\psi-m)^2\,\mathrm{sgn}\,(\phi-\psi-m)\right].\end{aligned}$$

The Lagrange multiplier $\tilde{\xi}$ has been omitted because the solution turns out to be automatically balanced, with $c_A = c_B$ and $A=-B$.

Without loss of generality we assume $m>0$, and also take $\lambda>0$ for concreteness, being the case $\lambda<0$ completely analogous. It is not hard to check that the eigenvalue density is given by

$$\rho(\phi)=\begin{cases}\frac{1}{2}\delta\left(\phi+\frac{1}{\lambda}\right)+\frac{1}{2}\delta\left(\phi-\frac{1}{\lambda}\right) & m>\frac{2}{\lambda}\\ \frac{1}{2}\delta\left(\phi-m+\frac{1}{\lambda}\right)+\frac{1}{2}\delta\left(\phi+m-\frac{1}{\lambda}\right) & \frac{1}{\lambda}\leq m\leq\frac{2}{\lambda}.\end{cases}$$

At $m=\lambda^{-1}$ nothing special happens, but $-B$ and B cross and we should rename the endpoints of the interval. The free energy in this limit is

$$\mathcal{F}_{\mathbb{S}^5}=\begin{cases}\frac{5}{3}\frac{1}{\lambda^3}-\frac{m}{\lambda^2}-\frac{m^3}{6} & m>\frac{2}{\lambda}\\ -\frac{1}{3}\left(m-\frac{1}{\lambda}\right)^3-\frac{m^3}{6} & 0\leq m\leq\frac{2}{\lambda},\end{cases}\tag{6.2.25}$$

which implies that $\frac{\partial^2\mathcal{F}_{\mathbb{S}^5}}{\partial m^2}$ is discontinuous. The model shows a second order phase transition. We ought to stress that the large N limit we take differs form that in [247], and hence the transition we find is different in nature. Besides, taking $m\to 0$ first in (6.2.25), we are left with a third order phase transition at $\frac{1}{\lambda}\to 0$, which corresponds to pass through a $6d$ $\mathcal{N}=(2,0)$ superconformal point. This transition reflects a flop transition in the dual Calabi–Yau geometry (see Appendix 6.A.1).

The free energy in (6.2.25) is a monotonically increasing function of $\frac{1}{\lambda}$, thus satisfying the $\mathcal{F}$-theorem (6.2.21), discussed in Subsection 6.2.3, all along the RG flow from the SCFT to the deep IR.

The vev of a Wilson loop in the fundamental representation in this model is

$$\langle W_{\mathsf{F}}\rangle=\begin{cases}\cosh\left(\frac{2\pi r}{\lambda}\right) & m>\frac{2}{\lambda}\\ \cosh\left(2\pi r\left(m-\frac{1}{\lambda}\right)\right) & 0\leq m\leq\frac{2}{\lambda}\end{cases}$$

with discontinuous derivative, in agreement with the result for $\mathcal{F}_{\mathbb{S}^5}$.

Antisymmetric or symmetric hypermultiplets

5d $SU(N)$ gauge theories with $n_{\mathsf{A}} \in \{0,1,2\}$ hypermultiplets in the rank-two antisymmetric representation or $n_{\mathsf{S}} \in \{0,1\}$ hypermultiplets in the symmetric representation descend from SCFTs [226]. The free energies of the theories with $n_{\mathsf{S}} = 1$ or $n_{\mathsf{A}} = 1$ differ by terms that are sub-leading at large N and therefore have identical phase diagram. The case $n_{\mathsf{A}} = 2$ does not admit a large Chern–Simons level nor a large number of additional fundamental hypermultiplets. The phase structure of $SU(N)$ theories with (anti-)symmetric matter is derived in Subsection 6.4.5.

6.3 Phases of $U(N)$ theories

In this section, the large N limit (6.2.10)-(6.2.11) of $U(N)$ gauge theories with N_f fundamental flavours is studied. For unitary group, we set $\tilde{\xi} = 0$.

Before delving into the detailed analysis, it is instructive to analyze the solution. When $-m_\alpha \in [A,B]$ for a subset $\mathscr{F} \subset \{1,2,\dots,F\}$ of the F mass scales, we see from (6.2.15) that the eigenvalue density is a sum of δ-functions supported at $-m_\alpha$, for $\alpha \in \mathscr{F}$, as well as at the endpoints of suppρ. The situation is schematically represented in Figure 6.3. The eigenvalues are clustered at $|\mathscr{F}|+2$ points, breaking the $U(N)$ group

$$U(N) \to U\left(c_A N\right) \times U\left(c_B N\right) \times \prod_{\alpha\in\mathscr{F}} U\left(\frac{n_\alpha}{2}\right)$$

with each factor rotating the eigenvalues placed at the support of the corresponding δ-function.

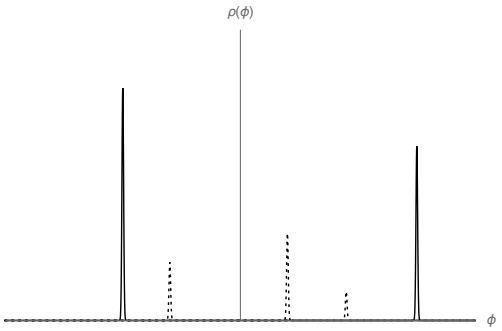


Figure 6.3. Eigenvalue density at large N, in a phase in which three out of the F mass parameters fall inside suppρ. The solid lines are the eigenvalues at the endpoints, the dashed lines are the eigenvalues at $-m_\alpha$. The range of the vertical axis is $\left[0,\frac{1}{2}\right]$ for a better visualization.

Moving a mass, the corresponding δ-function will eventually cross the boundary of suppρ and drop out. When $-m_\alpha$ hits A or B, the corresponding coefficient c_A or c_B jumps by $\frac{\zeta_\alpha}{2}$ in order to preserve the total number N of eigenvalues.

6.3.1 Pure gauge theory

We start our analysis with the pure Yang–Mills-Chern–Simons theory without matter, thus setting $n_\alpha = 0$. This theory lives in the IR of all the other theories with charged hypermultiplets, and is reached giving large masses to the matter fields and integrating them out. The presence of a Chern–Simons level k is therefore necessary, because it is generated dynamically along the RG flow as the effect of integrating out hypermultiplets.

The SPE in pure Yang–Mills-Chern–Simons theory is

$$-\int_A^B \mathrm{d}\psi\rho(\psi)\ \left(\phi-\psi\right)^2 \mathrm{sgn}\left(\phi-\psi\right) = \frac{1}{t}\phi^2 + \frac{2}{\lambda}\phi \tag{6.3.1}$$

and is solved by the ansatz $\rho(\phi) = c_A\delta(\phi - A) + c_B\delta(\phi - B)$. Following the steps described in Subsection 6.2.3, we find

$$c_A = \frac{1}{2} - \frac{1}{2t}, \qquad A = -\frac{t}{\lambda}\left(1 \pm \sqrt{\frac{t+1}{t-1}}\right),$$
$$c_B = \frac{1}{2} + \frac{1}{2t}, \qquad B = -\frac{t}{\lambda}\left(1 \pm \sqrt{\frac{t-1}{t+1}}\right),$$

with the same choice of sign of the square root in A and B. Notice that we have derived the equations assuming $A < B$, and we must retain the solution which respects this hypothesis, depending on $\mathrm{sgn}(\lambda)$. We plot them in Figure 6.4.

Removing the Yang–Mills term sending $|\lambda| \to \infty$, the eigenvalues are attracted to the origin and, for a pure Chern–Simons theory without any mass deformation, the saddle point configuration reduces to the trivial one.

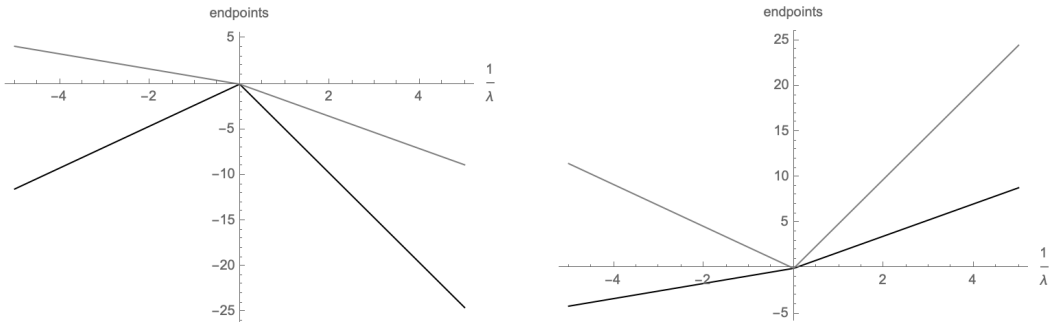


Figure 6.4. Plot of A (black) and B (gray) in the pure gauge theory. Left: $t = 1.3$. Right: $t = -1.3$.

6.3.2 One mass scale

We now consider a single mass scale, $F = 1$. In other words, the theory has N_f hypermultiplets all of equal mass m, and thus a single Veneziano parameter ζ as defined in (6.2.11).

At very large values of the mass, the hypermultiplets can be integrated out to obtain an effective theory with Chern–Simons level $k - \frac{N_f}{2}\mathrm{sgn}(m)$. Therefore, as m is increased from $-\infty$ up to $+\infty$, the effective description interpolates between two different pure Chern–Simons theories. At large gauge coupling, $\lambda \to \pm\infty$, there is no mass scale other than m, thus we expect a phase transition at $m = 0$. Nevertheless, a finite λ^{-1} sets a scale under which the hypermultiplet cannot be integrated out. We now show how this picture is realized.

The SPE reads

$$-\int_A^B \mathrm{d}\psi\rho(\psi)\,(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi) = \frac{1}{t}\phi^2 + \frac{2}{\lambda}\phi - \frac{\zeta}{2}\,(\phi+m)^2\,\mathrm{sgn}\,(\phi+m)\,. \tag{6.3.2}$$

For clarity, we focus first on the limit $|\lambda| \to \infty$, in which the Yang–Mills contribution drops out of the computations, and come back to the more general setting below.

Infinite Yang–Mills 't Hooft coupling

We start increasing m from $-\infty$, which gives $\mathrm{sgn}\,(\phi + m) < 0$. This inequality characterizes the first phase of the theory, which extends as long as $B < -m$. The explicit expressions of the solutions are reported in Appendix 6.B.1, equation (6.B.1).

The solution we have found holds as long as $B < -m$. Increasing m from large negative values, $-m$ will descend and eventually hit suppρ at B. From the explicit form of B in (6.B.1) we find that the inequality $B < -m$ breaks down at $m = 0$.

We may assume the existence of an intermediate phase in which $m \in$ suppρ, but then the solution to (6.3.2) would only be consistent with $A = 0 = B$. Therefore we pass to a new phase, for which $A > -m$ and hence $\text{sgn}(\phi + m) > 0$. The solution is found exactly as before, and is also recovered from the ones at $m < 0$ flipping the sign of the Veneziano parameter, $\zeta \mapsto -\zeta$.

We notice an important aspect: $\rho(\phi)$ is supported on the real line only for $t < -\left(1 - \frac{\zeta}{2}\right)^{-1}$ when $m < 0$, and only for $t > \left(1 - \frac{\zeta}{2}\right)^{-1}$ when $m > 0$. We also find real solutions in the region $0 < t < \left(1 - \frac{\zeta}{2}\right)^{-1}$ when $m < 0$, and with opposite sign when $m > 0$, which however fall out of the window (6.2.12). These solutions should not be discarded in principle, because the matrix model could still be defined at large N if suppρ lies entirely on the positive real axis when $0 < t < \left(1 - \frac{\zeta}{2}\right)^{-1}$, or on the negative real axis for negative t. However, evaluating A and B in that range, we find that the solutions do not satisfy the convergence condition, and therefore are inconsistent with the matrix model we have started with.

We use $\rho(\phi)$ to evaluate the free energy $\mathcal{F}_{\mathbb{S}^5}$ (6.2.18). In the large N and large r limit and at infinite $|\lambda|$, $\mathcal{F}_{\mathbb{S}^5}$ is given by

$$\mathcal{F}_{\mathbb{S}^5}(t, |\lambda| \to \infty, \zeta, m) = \frac{c_A c_B}{3}|B - A|^3 + \frac{1}{3t}\left(c_A A^3 + c_B B^3\right) - \frac{\zeta}{6}\left(c_A|A + m|^3 + c_B|B + m|^3\right),$$

with (c_A, c_B, A, B) functions of the gauge theoretical parameters as given in (6.B.1). The phase transition is third order, as proven by direct calculations, but it can also be predicted looking at the formula for $\mathcal{F}_{\mathbb{S}^5}$. It is a cubic function of $|m|^3$, because A and B are linear functions of m: the expressions up to the second derivative will automatically vanish at $m = 0$, determining the order of the phase transition.

To summarize, at infinite Yang–Mills 't Hooft coupling there are two phases separated by the critical surface $m_{\text{cr}} = 0$. The result is schematically presented in Figure 6.5.

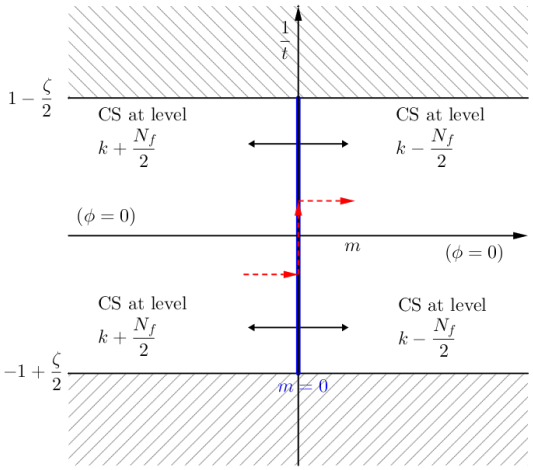


Figure 6.5. Phases diagram of the theory with N_f hypermultiplets all of mass m at $\lambda \to \pm\infty$, in the $(m, \frac{1}{t})$-plane. In the shaded region the matrix model is ill-defined. Crossing the blue wall $m_{\text{cr}} = 0$, the theory undergoes a third order phase transition, indicated by the solid, black arrows in the picture. The dashed, red arrows indicate a phase transition between the two non-trivial regions.

Finite Yang–Mills 't Hooft coupling

We now come back to the more general setting with $|\lambda| < \infty$, hence turning on an additional massive deformation. We start again increasing m from $-\infty$. The first phase is as in the large λ limit studied above, but now for $\lambda^{-1} \neq 0$ the inequality $B < -m$ breaks down at a critical mass $m_{\rm cr,1} < 0$. On the other hand, we could equivalently start decreasing m from $+\infty$, and see that the theory is in a phase equivalent to the second phase above. However, also in this case, the inequality $A > -m$ only holds for $m > m_{\rm cr,2} > 0$.

We see that the theory develops an intermediate phase

$$m_{\rm cr,1} < m < m_{\rm cr,2}, \tag{6.3.3}$$

in which the mass of the hypermultiplets is comparable to scale of the problem, determined by λ^{-1}. The matter fields cannot be integrated out and enter the IR dynamics. The deformation by λ^{-1} has moved the two critical parameters away from $m_{\rm cr} = 0$.

The explicit form of the eigenvalue density $\rho(\phi)$ is given in (6.B.2) in Appendix 6.B.1. As for large λ, we find that the first and third phases are non-trivial only for negative t and for positive t, respectively. Imposing the condition

$$B^{(\rm I)}(t, \lambda, \zeta, m) = -m,$$

with B computed under the assumption $B < -m$, we find the first critical surface $m = m_{\rm cr,1}(t, \lambda, \zeta)$. A direct computation using (6.B.2) gives:

$$m_{\rm cr,1} = \frac{t}{\lambda}\left(1 + \sqrt{\frac{(\zeta-2)t+2}{\lambda^2((\zeta-2)t-2)}}\right) \tag{6.3.4}$$

as plotted in the left panel of Figure 6.6. This solution vanishes for $\lambda \to \pm\infty$, in agreement with the discussion at infinite Yang–Mills 't Hooft coupling.

The second transition point is likewise determined decreasing m from large positive values until the inequality $A > -m$ breaks down, see the right panel of Figure 6.6. Explicitly, this second critical surface is

$$m_{\rm cr,2}(t, \lambda, \zeta) = \frac{t}{\lambda}\left(1 - \sqrt{\frac{(\zeta-2)t-2}{\lambda^2((\zeta-2)t+2)}}\right). \tag{6.3.5}$$

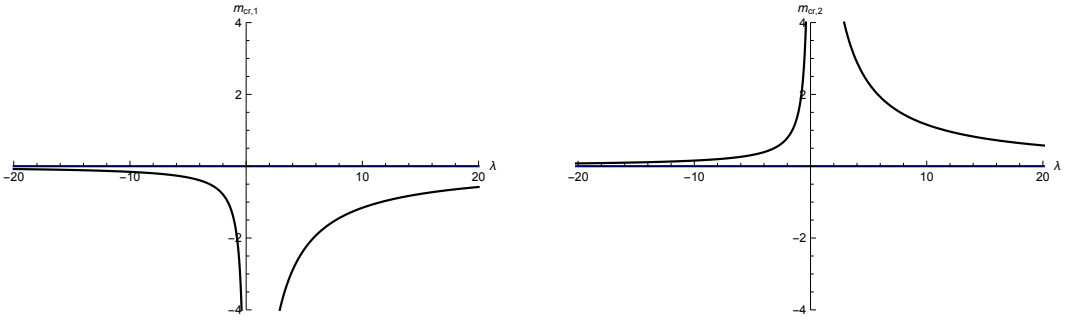


Figure 6.6. Critical surfaces plotted as functions of λ. Left: $m_{\rm cr,1}(t, \lambda, \zeta)$ at $\zeta = \frac{1}{2}$ and $t = -5$. Right: $m_{\rm cr,2}(t, \lambda, \zeta)$ at $\zeta = \frac{1}{2}$ and $t = 5$. The blue horizontal line is the asymptote $m = 0$.

We now pass to the intermediate phase (6.3.3). The solution is well behaved and non-trivial in the whole allowed (ζ, t)-region (6.2.12), and A and B do not depend explicitly on m, as we

already know from the general solution (6.2.15). This region is characterized by $-m \in [A, B]$, and therefore we can as well extract the critical values $m_{\mathrm{cr},1}$ and $m_{\mathrm{cr},2}$ from

$$-m_{\mathrm{cr},1}(t,\lambda,\zeta) = B^{(\mathrm{II})}(t,\lambda,\zeta), \qquad -m_{\mathrm{cr},1}(t,\lambda,\zeta) = A^{(\mathrm{II})}(t,\lambda,\zeta),$$

where $^{(\mathrm{II})}$ means the quantity evaluated in the intermediate phase. The solutions (6.3.4)-(6.3.5) are correctly reproduced.

We compute the free energy $\mathcal{F}_{\mathbb{S}^5}$ at finite λ. In the first and last phase, it has the form

$$\begin{aligned}\mathcal{F}_{\mathbb{S}^5}(t,\lambda,\zeta,m) = {} & \frac{c_A c_B}{3}|B-A|^3 + \frac{1}{3t}\left(c_A A^3 + c_B B^3\right) + \frac{1}{\lambda}\left(c_A A^2 + c_B B^2\right) \\ & - \frac{\zeta}{6}\left(c_A|A+m|^3 + c_B|B+m|^3\right)\end{aligned} \tag{6.3.6}$$

while in the middle phase we obtain

$$\mathcal{F}_{\mathbb{S}^5}(t,\lambda,\zeta,m) = \frac{c_A c_B}{3}|B-A|^3 + \frac{1}{3t}\left(c_A A^3 + c_B B^3 - \frac{\zeta}{2}m^3\right) + \frac{1}{\lambda}\left(c_A A^2 + c_B B^2 + \frac{\zeta}{2}m^2\right). \tag{6.3.7}$$

In these expressions, (c_A, c_B, A, B) are explicitly known functions of the gauge theoretical parameters (t,λ,ζ,m), given in (6.B.2).

Taking the derivative of $\mathcal{F}_{\mathbb{S}^5}$ with respect to m, we find at one critical point

$$\left.\frac{\partial \mathcal{F}_{\mathbb{S}^5}}{\partial m}\right|_{m\uparrow m_{\mathrm{cr},1}} = \frac{2t\zeta}{\lambda^2\left(2+t(2-\zeta)\right)} = \left.\frac{\partial \mathcal{F}_{\mathbb{S}^5}}{\partial m}\right|_{m\downarrow m_{\mathrm{cr},1}}$$

and a closely related expression at the other critical point. The second derivative however is discontinuous, thus we find a pair of second order phase transitions. We summarize the result in Figure 6.7.

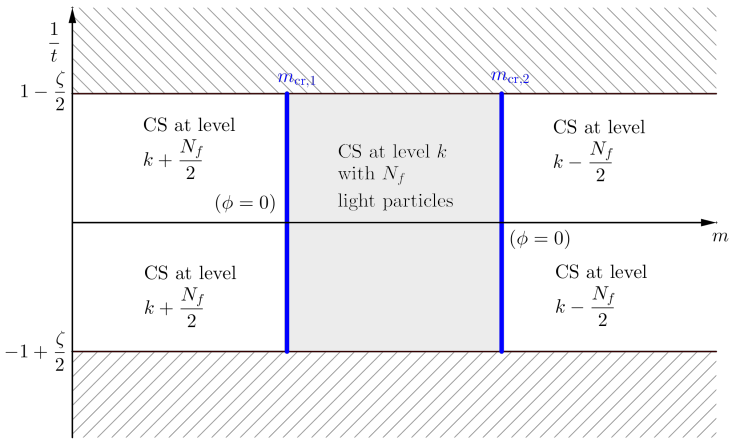


Figure 6.7. Phase diagram of the theory with a single mass scale, plotted in the $\left(m, \frac{1}{t}\right)$-plane.

Large Chern–Simons 't Hooft coupling

We now consider the limit of large Chern–Simons 't Hooft coupling, $t \to \pm\infty$, that realizes the large N limit at fixed Chern–Simons level k.

We start noting that, if $\lambda > 0$ and $\zeta \leq 2$ (possibly analytically continued to negative values), the effective action $S_{\mathrm{eff}}(\phi)$ is non-negative definite, and the large N limit describes trivial dynamics. On the contrary, for $\lambda < 0$, $S_{\mathrm{eff}}(\phi)$ admits a non-trivial saddle point configuration.

The solution is given in (6.B.3). The critical values are

$$m_{\mathrm{cr},1}\left(|t|\to\infty,\lambda,\zeta\right)=\left[\lambda\left(1-\frac{\zeta}{2}\right)\right]^{-1},$$
$$m_{\mathrm{cr},2}\left(|t|\to\infty,\lambda,\zeta\right)=-\left[\lambda\left(1-\frac{\zeta}{2}\right)\right]^{-1}.$$

Recall that $\lambda<0$, so $m_{\mathrm{cr},1}<0$ and $m_{\mathrm{cr},2}=-m_{\mathrm{cr},1}>0$.

In the intermediate phase we find an even $\rho(\phi)$ with symmetric support $A=-B$:

$$\rho(\phi)=\frac{2-\zeta}{4}\left[\delta\left(\phi+B\right)+\delta\left(\phi-B\right)\right]+\frac{\zeta}{2}\delta(\phi+m),\qquad B=-\left[\lambda\left(1-\frac{\zeta}{2}\right)\right]^{-1}.$$

The phase transitions are still second order.

In the intermediate phase, the saddle point configuration clusters the eigenvalues in three peaks, around A, B and $-m$. Approaching a critical value, the peak at $-m$ moves towards A or B, and eventually the two sets of eigenvalues coalesce. The phase transition is thus a signal of the partial restoration of symmetry

$$U\left(\frac{N}{2}-\frac{N_f}{4}\right)^2\times U\left(\frac{N_f}{2}\right)\longrightarrow U\left(\frac{N}{2}-\frac{N_f}{4}\right)\times U\left(\frac{N}{2}+\frac{N_f}{4}\right)$$

in going from the second to the first or third phase. Note that this is a symmetry enhancement because $\frac{N_f}{2}\leq N$.

6.3.3 Two opposite mass scales

We proceed in our analysis breaking the degeneracy in the masses of the hypermultiplets, setting $F=2$ distinct mass scales. We start with a symmetric setting, in which n_1 out of the $N_f<2N$ fundamental hypermultiplets have mass m and the others have mass $-m$. We work in the Veneziano limit (6.2.11) and assume $\zeta_1=\zeta_2\equiv\zeta$ in this symmetric setting. The case with vanishing Chern–Simons level has been addressed in [208], finding two phases separated by a third order transition.

The SPE is

$$-\int_A^B \mathrm{d}\psi\rho(\psi)\left(\phi-\psi\right)^2\mathrm{sgn}\left(\phi-\psi\right)=\frac{1}{t}\phi^2+\frac{2}{\lambda}\phi \\ -\frac{\zeta}{2}\left[\left(\phi+m\right)^2\mathrm{sgn}\left(\phi+m\right)+\left(\phi-m\right)^2\mathrm{sgn}\left(\phi-m\right)\right].\tag{6.3.8}$$

Infinite Yang–Mills 't Hooft coupling

We first consider the limit $|\lambda|\to\infty$. The solution to the SPE (6.3.8) is given in (6.B.4)-(6.B.5). We again find two phases, with a phase transition at $m=0$, as anticipated from general arguments. As discussed in Section 6.2.3, we find a pair of solutions for A and B in each phase. One solution, reported in (6.B.4), is consistent with $t>(1-\zeta)^{-1}$ and the other, reported in (6.B.5), is consistent with $t<-(1-\zeta)^{-1}$. Crossing from one phase to the other, the solutions are mapped consistently.

The free energy and its first and second derivatives are continuous at $m=0$ in this limit, but the third derivative is not. The situation is summarized in Figure 6.8.

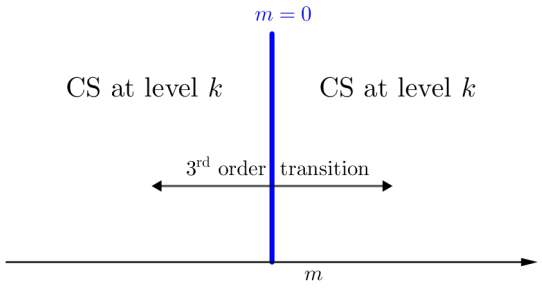


Figure 6.8. Phase diagram of the theory with two opposite mass scales at $|\lambda| \to \infty$: two effective Chern–Simons theories are separated by a third order phase transition at $m = 0$.

Finite Yang–Mills 't Hooft coupling

Reintroducing the mass deformation leading to a Yang–Mills term brings in a new mass scale, and consequently an intermediate phase when m is small compared to λ^{-1}. The phases corresponding to large positive or negative mass are found as for infinite Yang–Mills coupling. The solution is given in (6.B.6).

The asymmetry of suppρ, that is $A \neq -B$, implies that we find different solutions for the critical value $m_{\mathrm{cr},1}$, and the physically realized is the first one for which any of the two inequalities $A > m$ and $B < -m$ breaks down. We find that one scenario is realized for $t\lambda > 0$ and the other for $t\lambda < 0$:

$$m_{\mathrm{cr},1}\,(t,\lambda,\zeta) = \begin{cases} \frac{2t}{\lambda\left(\sqrt{t^2-1}+(2\zeta-1)t-1\right)} & t\lambda > 0 \\ \frac{t\left(\sqrt{t^2-1}+(2\zeta-1)t+1\right)}{\lambda+\lambda t(2(\zeta+(\zeta-1)\zeta t)-1)} & t\lambda < 0. \end{cases}$$

Beyond this first critical point, the system is in a new phase, in which the singularity at $\phi = m$ or at $\phi = -m$ enters in the interval $[A, B]$, while the other singularity falls out of the interval, see (6.B.6) for the explicit solution. The effective theory in this new phase is equivalent to the $F = 1$ theory with renormalized Chern–Simons coupling. We find consistent solutions for $\lambda < 0$. This second phase holds until the second singularity at $\pm m$ (depending on the sign of t) reaches $[A, B]$. For negative t, this means that a new phase transition takes place at $B = -m$, with B computed in the second phase. This equation yields two solutions, but only one is consistent with $m_{\mathrm{cr},1}$. The analogous reasoning applies to the other situation with positive t. We find a second phase transition at

$$m_{\mathrm{cr},2}\,(t,\lambda,\zeta) = \begin{cases} \frac{(\zeta-1)t^2+\sqrt{t^2((\zeta-1)^2t^2-1)}+t}{\lambda+(\zeta-1)\lambda t} & t\lambda > 0 \\ -\frac{\sqrt{t^2((\zeta-1)^2t^2-1)}}{(\zeta-1)\lambda t-\lambda} - \frac{t}{\lambda} & t\lambda < 0, \end{cases}$$

beyond which both m and $-m$ belong to $[A, B]$. The solution is given in (6.B.6). In this third phase $\rho(\phi)$ has the same form for both positive and negative t, and holds for $\lambda^{-1} \in \mathbb{R}$. Increasing m further, the system goes through the same phases in the converse direction, with the role of m and $-m$ swapped. Such behaviour is expected in this especially symmetric case, due to the $\mathbb{Z}_2$ invariance under exchange of the masses, $m_1 \leftrightarrow m_2$.

We summarize the phase structure in Figure 6.9.

The free energy $\mathcal{F}_{\mathbb{S}^5}$ is evaluated using the eigenvalue density $\rho(\phi)$ in each phase, as in (6.2.19). Taking derivatives of the resulting expression we find a third order phase transition for both signs of $t\lambda$. This generalizes the result of [208].

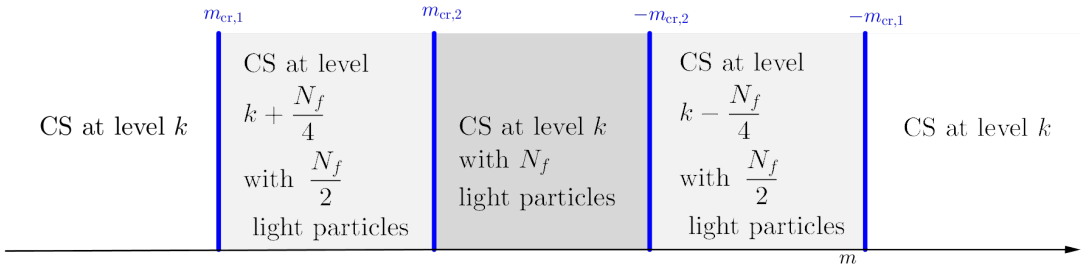


Figure 6.9. Phase diagram of the theory with two opposite mass scales.

Large Chern–Simons 't Hooft coupling

Consider two symmetric masses and $|t| \to \infty$. As in the one-mass setting, a non-trivial saddle point configuration requires $\lambda < 0$. We find a symmetric $\rho(\phi)$ supported on $[-B, B]$, with

$$B = \begin{cases} -\frac{1}{\lambda} - m\zeta & m < m_{\rm cr} \\ -\frac{1}{\lambda(1-\zeta)} & m > m_{\rm cr}. \end{cases}$$

The intermediate phases disappear in this limit, because

$$\lim_{t\to\pm\infty} m_{\rm cr,2}\,(t,\lambda,\zeta) = \lim_{t\to\pm\infty} m_{\rm cr,1}\,(t,\lambda,\zeta) = \frac{1}{\lambda\,(1-\zeta)}.$$

The results of [208] are then recovered.

This model has a $\mathbb{Z}_2$ symmetry. In the intermediate phase, $\rho(\phi)$ presents four clusters of eigenvalues, placed at $\pm B$ and $\pm m$. Approaching the critical locus, the eigenvalues at m and the ones at $-B$ coalesce, and simultaneously the eigenvalues at $-m$ and the ones at B coalesce, see Figure 6.10. The phase transition is thus a signal of the symmetry enhancement

$$U\left(\frac{N}{2} - \frac{N_f}{4}\right)^2 \times U\left(\frac{N_f}{4}\right)^2 \longrightarrow U\left(\frac{N}{2}\right)^2. \tag{6.3.9}$$

The $\mathbb{Z}_2$ symmetry is manifest on both sides of the arrow in (6.3.9).

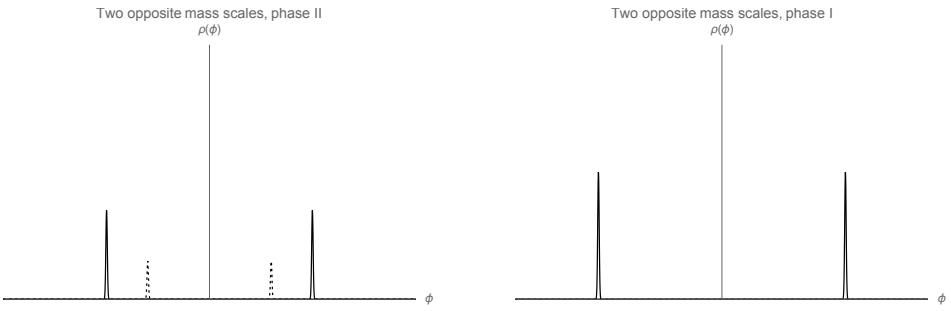


Figure 6.10. Schematic representation of the clustering of eigenvalues at large N in the theory with two opposite mass scales. Left: Eigenvalue density in the intermediate phase, with $-m, m \in [-B, B]$ (dashed lines). Right: Eigenvalue density when $-m, m \notin [-B, B]$. The range of the vertical axis is $[0, 1]$.

6.3.4 Two mass scales

Consider now a generic assignment of masses m_1 and m_2, and two Veneziano parameters ζ_1 and ζ_2. The SPE is

$$-\int_A^B \mathrm{d}\psi\rho(\psi)\,(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi) = \frac{1}{t}\phi^2 + \frac{2}{\lambda}\phi - \frac{\zeta_1}{2}\,(\phi+m_1)^2\,\mathrm{sgn}\,(\phi+m_1) - \frac{\zeta_2}{2}\,(\phi+m_2)^2\,\mathrm{sgn}\,(\phi+m_2)\,. \tag{6.3.10}$$

When $|m_1|$ and $|m_2|$ are both large, we find the usual solution $\rho(\phi)$ with two δ-function singularities at the endpoints. The first phase transition takes place when one of the masses hits $\mathrm{supp}\rho$. We focus first on infinite Yang–Mills 't Hooft coupling limit and reintroduce the corresponding deformation later.

Infinite Yang–Mills 't Hooft coupling

The complete solution to the SPE (6.3.10) in the limit $\frac{1}{\lambda}\to 0$ is given in (6.B.7).

In order to effectively have a single free mass modulus, throughout the present subsection we impose this constraint

$$\zeta_1 m_1 + \zeta_2 m_2 = 0. \tag{6.3.11}$$

The first phase, as usual, arises when the masses fall out of $\mathrm{supp}\rho$, and the solution $\rho(\phi)$ is a sum of δ-functions at the endpoints A and B of the support, reported in equation (6.B.7). At this point, two possible scenarios disclose: moving the values of the masses with the constraint (6.3.11), either the singularity at $\phi=-m_2$ hits $\mathrm{supp}\rho$ from below, or the singularity at $\phi=-m_1$ hits $\mathrm{supp}\rho$ from above. A simple computation imposing (6.3.11) shows that these two scenarios are realized simultaneously: the mass parameter hit the endpoints at $m_2=0=m_1$ for all values of t consistent with (6.2.12). Crossing the critical point, the eigenvalue density in the new phase is obtained via the formal substitution $(\zeta_1,\zeta_2)\leftrightarrow(-\zeta_1,-\zeta_2)$, see (6.B.7).

The free energy $\mathcal{F}_{\mathbb{S}^5}$ is, as usual, a cubic function of A, B and m_1, m_2, and all of them vanish at the critical point. Taking derivatives, we find a third order phase transition. We conclude that the picture is equivalent to the symmetric case studied in Subsection 6.3.3 and summarized in Figure 6.8.

Infinite Yang–Mills 't Hooft coupling revisited

Let us consider the situation in which the massive deformation leading to a Yang–Mills term is removed, $|\lambda|\to\infty$, but dropping the constraint (6.3.11). In this way, we have two real mass parameters to play with.

The explicit solution for $\rho(\phi)$ is found by the standard calculation, and is reported in (6.B.8). Having two real mass moduli, we can either increase m_1 keeping m_2 fixed, or decrease m_2 keeping m_1 fixed, or any linear combination of the two. In the former case, the system undergoes a phase transition when the singularity at $-m_1$ hits B from above, whilst in the latter case a phase transition takes place when the singularity at $-m_2$ hits A from below.

We move m_1 and study the new phase, characterized by $-m_1\in[A,B]$. The explicit solution for $\rho(\phi)$ is reported in (6.B.9). The critical surface is determined imposing $-m_1=B^{(\mathrm{II})}$, with B evaluated in the second phase. If we compute the critical point from the first phase, we get two solutions, and the physical one is the lowest value, that is, the first value for which $-m_1>B$ does not hold as m_1 is increased from $-\infty$. The critical value obtained in this way matches the value of B in the second phase, as required by consistency.

From the second phase, we can either keep $-m_1 \in [A, B]$ and decrease m_2 until $-m_2$ reaches A, or increase m_1 further until $-m_1 < A$. Let us first focus on the former choice. We notice that, decreasing m_2 the support $[A, B]$ of $\rho(\phi)$ shrinks. Therefore, to keep $-m_1 \in [A, B]$ we should in fact decrease m_1 at the same time as we decrease m_2. This procedure leads to a third order phase transition at $m_2 = 0 = m_1$.

Increasing m_1 further with m_2 fixed at a large positive value, the singularity at $-m_1$ eventually reaches the lower boundary of suppρ. This triggers a new phase transition. The critical values for both transitions encountered at fixed m_2 and moving m_1 are linear functions of m_2, see (6.B.9). The rest of the phase diagram is described analogously.

Evaluating the free energy in each phase, we find third order phase transitions.

Finite Yang–Mills 't Hooft coupling

We come back to the general setting reintroducing a Yang–Mills term.

In the middle phase, when both $-m_1$ and $-m_2$ fall inside the support $[A, B]$ of the eigenvalue density, we find

$$c_A = \frac{2-\zeta_1-\zeta_2}{4} - \frac{1}{2t}, \qquad A = -\frac{t}{\lambda}\left(1 \pm \sqrt{\frac{c_B}{c_A}}\right),$$
$$c_B = \frac{2-\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \qquad B = -\frac{t}{\lambda}\left(1 \pm \sqrt{\frac{c_A}{c_B}}\right).$$

In the latter expression the sign must be chosen in consistency with our starting assumption $A < B$, and is the same in both formulas.

We henceforth focus on $\lambda < 0$ for concreteness. In this case, $A < 0$ and $B > 0$ and are explicitly given in (6.B.10). Let us assume we increase m_2 to positive values: the δ-function singularity at $\phi = -m_2$ moves toward the endpoint A and eventually hits the boundary of suppρ at

$$m_{2\mathrm{cr},1} = \frac{t}{\lambda}\left(1 - \sqrt{1 - \frac{4}{(\zeta_1+\zeta_2)\,t - 2t + 2}}\right). \tag{6.3.12}$$

Increasing m_2 beyond this point, the gauge theory enters in a new phase in which the solution, that holds for positive t, is reported in (6.B.11). Notice that B soon becomes negative in this phase as we increase m_2, meaning that we should increase m_1 at the same time so that $-m_1 \in [A, B]$. Then, we can either increase m_1 further or decrease it, driving the system toward a new phase. The explicit results are reported in equations (6.B.13)-(6.B.14).

We can equivalently begin keeping $-m_2 \in [A, B]$ and moving m_1. If we increase $m_1 > 0$, we recover the setting just analyzed, upon relabelling $\zeta_1 \leftrightarrow \zeta_2$. Decreasing $m_1 < 0$, instead, the singularity at $\phi = -m_1$ is moved toward B, and eventually the theory undergoes a phase transition at

$$m_{1\mathrm{cr},1} = -\frac{t}{\lambda}\left(\sqrt{\frac{4}{(\zeta_1+\zeta_2)\,t - 2(t+1)} + 1} - 1\right). \tag{6.3.13}$$

The solution in the new phase, characterized by $-m_2 \in [A, B]$ and $B < -m_1$, is given in (6.B.12), and holds for negative t. The description of the other phases is obtained in a completely analogous fashion.

We plot the phase structure of the $F = 2$ theory in the (m_1, m_2)-plane in Figure 6.11. A more qualitative description of the phases is in Figure 6.12.

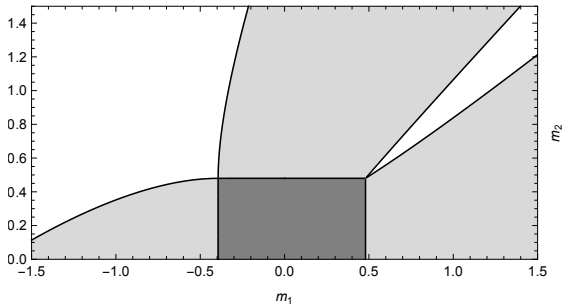


Figure 6.11. Phase diagram in the half-plane $\{m_1 \in \mathbb{R}, m_2 \geq 0\}$. The plot is at $(t, \lambda, \zeta_1, \zeta_2) = (22, -10, \frac{1}{3}, \frac{6}{5})$. In the darker shaded region both singularities lie in the support of the eigenvalue density. In the lighter shaded regions one singularity lies inside and the other lies outside the support. In the white region none of the singularities lies inside the support.

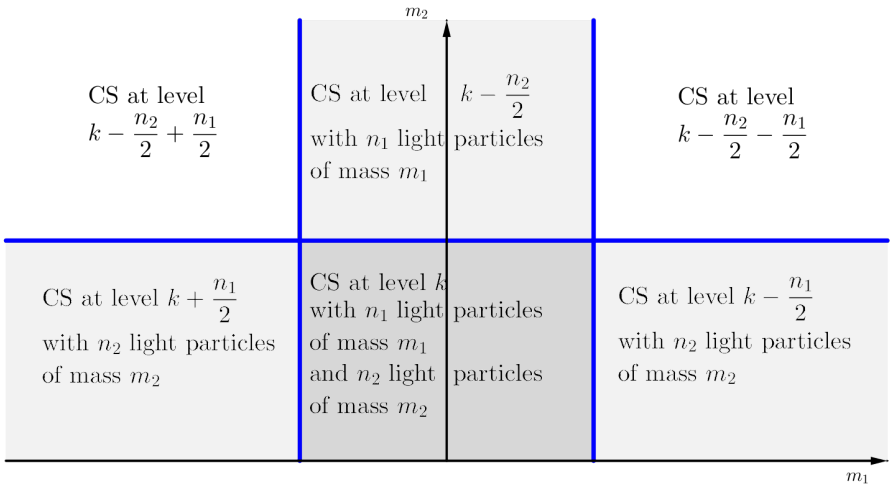


Figure 6.12. Phase diagram of the theory with two mass scales. Note that the solid blue walls, representing the critical surfaces, are not straight lines in the (m_1, m_2)-plane, cf. Figure 6.11.

Limiting cases

The present framework with $F = 2$ encompasses the previously studied theories as special cases. Setting $\zeta_2 = 0$ we expect to recover the $F = 1$ theory of Subsection 6.3.2, whilst setting $\zeta_2 = \zeta_1$ we should recover the symmetric framework of Subsection 6.3.3.

In the first mentioned limiting case we check that, sending $\zeta_2 \downarrow 0$, all the expressions in Appendix 6.B.1 for the $F = 2$ theory reduce to $F = 1$. An alternative approach is to take $|m_2| \to \infty$ and integrate out the massive hypermultiplets. Then we obtain the $F = 1$ theory with a renormalized Chern–Simons 't Hooft coupling $t|_{F=1} = t|_{F=2} - \frac{n_2}{2}$. This result may be visualized comparing the upper strip of Figure 6.12 with the phase diagram of the $F = 1$ theory in Figure 6.7.

To recover the $F = 2$ symmetric case with two opposite masses, set $m_1 = -m_2$. Moving along the diagonal in the $(m_1 \leq 0, m_2 \geq 0)$-quadrant of Figure 6.11 reproduces the phases on the right half $(m_2 \geq 0)$ of Figure 6.9.

6.3.5 Three or more mass scales

The generic solution for F real mass scales is (6.2.15), and the procedure is a direct extension of what we have presented so far. The explicit determination of the phase diagram requires a

detailed case by case study, with each mass m_α moved independently. The upshot is that, for generic λ and $\{m_\alpha\}$, a second order phase transition takes place whenever one of the singularities drops in or out of $[A, B]$.

We have followed a bottom-up approach in our presentation, starting with a pure gauge theory in Subsection 6.3.1 and increasing F. We might have adopted a top-down approach as well, following the RG flow. Indeed, starting with a given F, the other theories with lower $F' < F$ are phases of the original theory, reached giving large mass to $(F - F')$ families of hypermultiplets.

Infinite Yang–Mills 't Hooft coupling and constrained masses

While, as we have shown, the phase transitions are generically second order for $U(N)$, there is a selected sub-class of theories for which we find third order transitions. These are Chern–Simons theories at infinite Yang–Mills coupling and with a single modulus controlling the theory,

$$m_\alpha = x_\alpha m, \qquad x_\alpha \in \mathbb{R},\ \alpha = 1, \dots, F. \tag{6.3.14}$$

Without loss of generality, we impose

$$\sum_{\alpha=1}^{F} x_\alpha^2 = 1,$$

as any scaling of all the x_α together can be absorbed in a redefinition of m. The definition (6.3.14) is a change to polar coordinates in $\mathbb{R}^F$ for each sign of m, with $|m|$ parametrizing the radial direction. Note, however, that the theories we consider allow $m \in \mathbb{R}$.

For very large m, which we take positive for concreteness, the singularities fall out of $\mathrm{supp}\rho$, either above or below depending on the sign of x_α. We get

$$c_A = -\frac{1}{2t} + \frac{1}{2} + \sum_\alpha \frac{\zeta_\alpha}{4}\mathrm{sgn}(x_\alpha), \quad c_B = \frac{1}{2t} + \frac{1}{2} - \sum_\alpha \frac{\zeta_\alpha}{4}\mathrm{sgn}(x_\alpha),$$

while A and B are linear functions of m,

$$A = mA_0,\ B = mB_0, \qquad A_0 \text{ and } B_0 \text{ independent of } m.$$

This implies that, moving m, all the singularities reach the boundary of $\mathrm{supp}\rho$ simultaneously at $m = 0$. Recall that the free energy $\mathcal{F}_{\mathbb{S}^5}$ is a cubic function of A, B and $\{m_\alpha\}$. It follows that the free energy is continuous and vanishing at the critical locus, up to its second derivative. We establish that the phase transition is third order.

In the geometric picture sketched in Appendix 6.A.1, the rewriting (6.3.14) corresponds to take all the Kähler parameters that are dual to non-compact divisors in a resolution of the Calabi–Yau threefold X to be proportional to a single parameter m. Then, sending $\lambda^{-1} \to 0$ first, corresponds to keep the volume $\mathrm{vol}\left(\mathbb{P}^1_0\right)$ of a certain curve $\mathbb{P}^1_0$ finite while by the number of exceptional divisors fibered over it grows to infinity (see Appendix 6.A.1 for notation and definitions). After that, we decrease the Kähler parameter m controlling the volumes of the non-compact exceptional divisors, until it vanishes. Then, the gauge theory undergoes a third order phase transition, which agrees with the expected geometric flop transition.

From the explicit results in Subsection 6.3.4, the present picture with the associated third order transition is expected to hold even dropping the constraint (6.3.14).

6.3.6 Wilson loops

We have argued in Subsection 6.2.4 that Wilson loops are always continuous but generically not differentiable. It is worthwhile to focus on the special instances in which the partition function undergoes a third order transition, and analyze the behaviour of the Wilson loops.

Fundamental Wilson loop: Two opposite mass scales

We consider the theory without Chern–Simons term and with two opposite mass scales, with equal Veneziano parameters, discussed in [208] and revisited in Subsection 6.3.3. The vev of a Wilson loop in the fundamental representation is

$$\langle W_{\mathsf{F}}\rangle = \begin{cases} \cosh\left(2\pi r\left(\zeta m - \frac{1}{\lambda}\right)\right) & |m| > \frac{1}{\lambda(1-\zeta)} \\ (1-\zeta)\cosh\left(\frac{2\pi r}{\lambda(1-\zeta)}\right) + \zeta\cosh\left(2\pi r m\right) & |m| < \frac{1}{\lambda(1-\zeta)}. \end{cases}$$

Taking the logarithm and differentiating, we find

$$\begin{aligned} \frac{\partial}{\partial m}\log\langle W_{\mathsf{F}}\rangle\Big|_{m\downarrow\frac{1}{\lambda(1-\zeta)}} - \frac{\partial}{\partial m}\log\langle W_{\mathsf{F}}\rangle\Big|_{m\uparrow\frac{1}{\lambda(1-\zeta)}} &= 0 \\ \frac{\partial^2}{\partial m^2}\log\langle W_{\mathsf{F}}\rangle\Big|_{m\downarrow\frac{1}{\lambda(1-\zeta)}} - \frac{\partial^2}{\partial m^2}\log\langle W_{\mathsf{F}}\rangle\Big|_{m\uparrow\frac{1}{\lambda(1-\zeta)}} &= \zeta(1-\zeta) \end{aligned}$$

meaning that the Wilson loop vev experiences a second order non-analyticity, one order less than the free energy. Note that the second and higher derivatives vanish as $\zeta \to 1$, because in that case there exists a single phase valid for all m. To conclude, we mention that, as we work in the decompactification limit, the functions $\cosh(2\pi r x)$ should be replaced by $e^{|2\pi x|}$ in all the expression above. Using this substitution before taking the derivatives does not alter the conclusion.

Fundamental Wilson loop: Infinite Yang–Mills coupling

The other situation in which the phase transition is third order is for theories without Yang–Mills 't Hooft coupling, $|\lambda| \to \infty$. We consider the $F = 1$ theory of Section 6.3.2 as an explicit example. The endpoints $A = mA_0$ and $B = mB_0$ are linear functions of m, and

$$\frac{\partial}{\partial m}\log\langle W_{\mathsf{F}}\rangle\Big|_{m\to 0} = 2\pi r\frac{c_A A_0 + c_B B_0}{c_A + c_B} = 2\pi r\left(c_A A_0 + c_B B_0\right).$$

This does not vanish unless $c_A A_0 + c_B B_0 = 0$, and therefore the Wilson loop vev has a first order discontinuity. This may indicate an inconsistency in the strong coupling limit of non-balanced $U(N)$ theories, or at least an ambiguity in the order of strong coupling and large N limits.

Antisymmetric Wilson loop: Pure gauge theory

We now apply the framework presented in Section 6.2.4 to compute the expectation value of half-BPS Wilson loops in antisymmetric representations of large rank.

We begin with the pure gauge theory analyzed in Section 6.3.1. The theory has no mass scales other than the inverse Yang–Mills 't Hooft coupling λ^{-1}, and presents a single phase. Specializing the argument of Section 6.2.4 to such theory without hypermultiplets, we have to evaluate

$$\langle W_{\mathsf{A}_K}\rangle = e^{rAK}\oint\frac{d\tilde{w}}{2\pi i}\,\tilde{w}^{K-1}\left[1+\frac{1}{\tilde{w}}\right]^{\lfloor Nc_A\rfloor}\left[1+\frac{e^{r(B-A)}}{\tilde{w}}\right]^{\lfloor Nc_B\rfloor}$$

keeping the leading contribution at large radius. We observe that such contribution will differ depending on $c_B > \kappa$ or $c_B < \kappa$, where the scaling parameter $\kappa = \frac{K}{N}$ has been introduced in (6.2.23). We find

$$\log\langle W_{\mathsf{A}_K}\rangle = \begin{cases} rBK & K \le \frac{N+k}{2} \\ r\left[AK + (B-A)\left(\frac{N+k}{2}\right)\right] & K > \frac{N+k}{2} \end{cases}$$

with k the Chern–Simons level. The inequalities are understood at large N. Besides, strictly speaking this solution only holds as long as $\kappa \leq c_A+c_B$, but recalling the normalization $c_A+c_B=1$ and that $0 \leq \kappa \leq 1$ by definition, this latter requirement is always satisfied. We stress that the Wilson loop is a continuous but not differentiable function of κ.

This result holds for all gauge theories with massive matter, in the phases in which all the masses fall outside of suppρ, up to a renormalization of the Chern–Simons coupling.

Antisymmetric Wilson loop: One mass scale

We discuss the antisymmetric Wilson loop in the $U(N)$ theory with a single mass scale m. In the first and last phase, with $-m > B$ and $-m < A$ respectively, the solution is analogous to the one for the pure gauge theory, upon replacement $\frac{N+k}{2} \mapsto \frac{N+k}{2} + \frac{N_f}{4}$ if $-m > B$, and $\frac{N+k}{2} \mapsto \frac{N+k}{2} - \frac{N_f}{4}$ if $-m < A$.

In the intermediate phase, characterized by $A < -m < B$, we must take into account two possibilities, namely $A < -m < B - A$ and $B - A < -m < B$. The final results in the two sub-cases are

$$\log \langle W_{\mathsf{A}_K} \rangle^{(\mathrm{II})} \Big|_{-m<B-A} = \begin{cases} rBK & 0 \leq K < \frac{N+k}{2} - \frac{N_f}{4} \\ r\left[(A-m)\left(K - \frac{N+k}{2} + \frac{N_f}{4}\right) + B\left(\frac{N+k}{2} - \frac{N_f}{4}\right)\right] & \frac{N+k}{2} - \frac{N_f}{4} \leq K < \frac{N+k}{2} + \frac{N_f}{4} \\ r\left[A\left(K - \frac{N+k}{2} + \frac{N_f}{4}\right) + B\left(\frac{N+k}{2} - \frac{N_f}{4}\right) - m\frac{N_f}{2}\right] & K \geq \frac{N+k}{2} + \frac{N_f}{4}, \end{cases}$$

$$\log \langle W_{\mathsf{A}_K} \rangle^{(\mathrm{II})} \Big|_{-m>B-A} = \begin{cases} r(A-m)K & 0 \leq K < \frac{N_f}{2} \\ r\left[(A-m)\frac{N_f}{2} + B\left(K - \frac{N_f}{2}\right)\right] & \frac{N_f}{2} \leq K < \frac{N+k}{2} + \frac{N_f}{4} \\ r\left[A\left(K - \frac{N+k}{2} + \frac{N_f}{4}\right) + B\left(\frac{N+k}{2} - \frac{N_f}{4}\right) - m\frac{N_f}{2}\right] & K \geq \frac{N+k}{2} + \frac{N_f}{4}, \end{cases}$$

where the superscript $^{(\mathrm{II})}$ means that we have computed the Wilson loop vev in the intermediate phase. Condition (6.2.12) guarantees that the inequalities are always well posed.

The Wilson loop is continuous but not differentiable function of m at both critical loci. It is also a continuous but not differentiable function of the scaling parameter κ in every phase.

The study of the expectation value of a Wilson loop in a large antisymmetric representation A_K for any number of mass scales can be addressed by the method presented here, specializing the argument of Section 6.2.4 to a given F and analyzing the various sub-cases in each phase.

6.4 Phases of $SU(N)$ theories

$SU(N)$ gauge theories descend from (twisted) compactifications of $6d$ SCFTs on a circle of radius β, with the Yang–Mills deformation $h \propto \beta^{-1}$.

The analysis is closely related to that of the $U(N)$ theory, but we now reintroduce the Lagrange multiplier $\tilde{\xi}$. At the end, we will set it to its physical value, determined by the requirement

$$\int_A^B \phi\rho(\phi)\mathrm{d}\phi = 0. \tag{6.4.1}$$

The details for finding the explicit solution $\rho(\phi)$ in each $SU(N)$ theory are exactly as in the corresponding $U(N)$ theory analyzed in Section 6.3, except that the endpoints A, B will carry an additional dependence on $\tilde{\xi}$. Notice that c_A and c_B are the same in the $U(N)$ and $SU(N)$ theory, and for this reason we will not explicitly discuss them throughout this section.

The generic solution $\rho(\phi)$ is again given by (6.2.15), and the constraint (6.4.1) reads

$$c_A A + c_B B - \sum_{\alpha=1}^{F} c_\alpha m_\alpha = 0. \tag{6.4.2}$$

This is meant as an equation fixing $\tilde{\xi}$ through the dependence of A and B on it.

A remark is in order to clarify the role of the multiplier $\tilde{\xi}$. A power counting in the integral representation (6.2.3) of the partition function suggests that the difference between $SU(N)$ and $U(N)$ is sub-leading in a $\frac{1}{N}$ expansion. Indeed, only the ratio $\tilde{\xi} \propto \frac{\xi}{N}$ enters the SPE (6.2.14), showing that ungauging an Abelian factor $U(1) \subset U(N)$ gives a next-to-leading order correction. In this work we do not go beyond the leading order at large N, and adopt the approach of [247] scaling the Lagrange multiplier in a 't Hooft-like way, with $\tilde{\xi}$ fixed at large N, to keep track of the tracelessness condition at large N. Stated more formally, we work in the direct limit Lie algebra $\mathfrak{u}(\infty)$ and restrict to the traceless subspace.

From the geometric engineering viewpoint, the 't Hooft limit (6.2.10) blows up the volume of a curve $\mathbb{P}^1_0$ belonging to the base $\mathscr{B}$ of the elliptic fibration $\widetilde{X} \to \mathscr{B}$. Imposing the same scaling for $\tilde{\xi}$ corresponds to scale the metric on the base $\mathscr{B}$ in such a way that the volume of a different curve $\mathbb{P}^1_* \subset \mathscr{B}$, transverse to $\mathbb{P}^1_0$, grows linearly with $\mathrm{vol}\left(\mathbb{P}^1_0\right)$. Essentially, this procedure amounts to keep track of the difference between ALF and ALE metrics on the Calabi–Yau threefold.

6.4.1 Pure gauge theory

The first theory we consider is the pure Yang–Mills-Chern–Simons gauge theory without matter. We compute A and B when the multiplier $\tilde{\xi}$ is taken into account in the SPE, and then impose (6.4.2) and solve for $\tilde{\xi}$. This gives

$$\tilde{\xi} = -\frac{t}{\lambda^2(t^2-1)}.$$

Plugging this value back in A and B we obtain

$$A = -\frac{t}{\lambda} \mp \frac{t^2}{\lambda(t-1)}, \qquad B = -\frac{t}{\lambda} \pm \frac{t^2}{\lambda(t+1)},$$

with signs chosen consistently depending on $\mathrm{sgn}(\lambda)$. There is no crucial difference between $SU(N)$ and $U(N)$ pure gauge theories, except for the details in determining A and B. In the present case, the tracelessness condition implies $A < 0$ and $B > 0$ $\forall \lambda^{-1} \in \mathbb{R}$.

6.4.2 One mass scale

The first example which includes matter is the $F = 1$ theory. As in Subsection 6.3.2 we discuss first the $|\lambda| \to \infty$ case, and then reintroduce a finite Yang–Mills term.

Infinite Yang–Mills 't Hooft coupling

We begin with the analysis of the $SU(N)$ theory with all hypermultiplets of equal mass and no Yang–Mills term. Solving for A, B and imposing (6.4.2) we find

$$\tilde{\xi} = \begin{cases} -\frac{\zeta m^2(t(\zeta-2t)+2)}{(\zeta^2-4)t^2+4\zeta t+4} & m < 0 \\ -\frac{\zeta m^2(t(\zeta+2t)-2)}{(\zeta^2-4)t^2-4\zeta t+4} & m > 0. \end{cases}$$

Plugging this back into A and B in both phases gives the explicit solution reported in Appendix 6.B.2, equation (6.B.15). These expressions are much simpler than the ones obtained in the $U(N)$ theory. We find a third order phase transition at $m = 0$ but, in contrast to the $U(N)$ theory, the solution is non-trivial for all values of $\frac{1}{t}$ in the window (6.2.12) on both sides of the critical wall. The phase structure is represented in Figure 6.13.

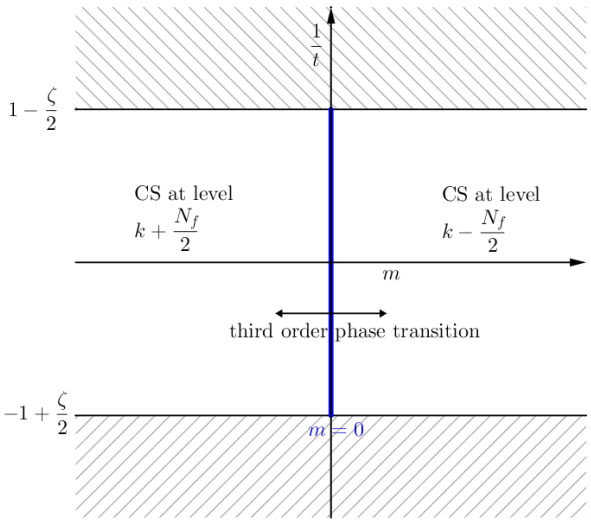


Figure 6.13. Phase diagram of the $SU(N)$ theory with N_f hypermultiplets all of mass m at $\lambda \to \pm\infty$. Across the blue wall $m_{\rm cr} = 0$ the theory undergoes a third order phase transition.

Finite Yang–Mills 't Hooft coupling

At finite Yang–Mills 't Hooft coupling $|\lambda| < \infty$, we solve the SPE and impose the constraint (6.4.2), which gives

$$\tilde{\xi} = \begin{cases} \frac{\zeta\lambda^2\left(-m^2\right)(t(\zeta-2t)+2)+2\zeta\lambda mt(\zeta t+2)+2t(\zeta t+2)}{\lambda^2((\zeta-2)t+2)((\zeta+2)t+2)} & m < m_{\rm cr,1} \\ \frac{t(-m\zeta\lambda(\zeta\lambda m-2(\zeta-2)t)-4)}{\lambda^2((\zeta-2)^2t^2-4)} & m_{\rm cr,1} < m < m_{\rm cr,2} \\ \frac{\zeta\lambda^2\left(-m^2\right)(t(\zeta+2t)-2)+2\zeta\lambda mt(\zeta t-2)-2t(\zeta t-2)}{\lambda^2((\zeta-2)t-2)((\zeta+2)t-2)} & m > m_{\rm cr,2}. \end{cases}$$

The endpoints of the support, reported in (6.B.16), are uniquely determined and take an especially simple form. The two critical surfaces are

$$m_{\rm cr,1}(t,\lambda,\zeta) = \left[\lambda\left(1+\frac{1}{t}\right)\right]^{-1}$$
$$m_{\rm cr,2}(t,\lambda,\zeta) = \left[-\lambda\left(1-\frac{1}{t}\right)\right]^{-1}$$

which, remarkably, are independent of ζ. These solutions have been obtained under the assumption $A < B$, which is self-consistent only for $\lambda < 0$, in agreement with the analysis of the $U(N)$ theory. A major difference with the $U(N)$ theory is that the condition (6.4.2) has introduced an explicit dependence on m in A and B in the intermediate phase. This dependence is necessary to balance the average $\int_A^B \phi\rho(\phi)\mathrm{d}\phi$ as the δ-function at $-m$ is moved inside the interval $[A,B]$.

The free energy is explicitly given by

$$\mathcal{F}_{\mathbb{S}^5}\,|_{m<m_{\rm cr,1}} = \frac{2\zeta\lambda^3m^3+2\zeta^2\lambda^3m^3t-t^2(\zeta\lambda m(\lambda m(3\zeta+2\lambda m)+6)+4)}{3\lambda^3((\zeta-2)t+2)((\zeta+2)t+2)} \tag{6.4.3a}$$

$$\mathcal{F}_{\mathbb{S}^5}\,|_{m_{\rm cr,1}<m<m_{\rm cr,2}} = \frac{2\zeta\lambda^3m^3-(\zeta-2)t^3\left(3\zeta\lambda^2m^2+2\right)+2\zeta\lambda mt^2\left((\zeta-1)\lambda^2m^2+3\right)-6\zeta\lambda^2m^2t}{3\lambda^3\left((\zeta-2)^2t^3-4t\right)} \tag{6.4.3b}$$

in the first and second phase respectively, and by (6.4.3a) with $\zeta \mapsto -\zeta$ in the third phase. Taking derivatives, we obtain

$$\left.\frac{\partial \mathcal{F}_{\mathbb{S}^5}}{\partial m}\right|_{m\uparrow m_{\mathrm{cr},1}} = -\frac{2\zeta t^3}{\lambda^2(t+1)^2((\zeta-2)t+2)} = \left.\frac{\partial \mathcal{F}_{\mathbb{S}^5}}{\partial m}\right|_{m\downarrow m_{\mathrm{cr},1}}$$
$$\left.\frac{\partial^2 \mathcal{F}_{\mathbb{S}^5}}{\partial m^2}\right|_{m\uparrow m_{\mathrm{cr},1}} = -\frac{2\zeta(t-1)t}{\lambda(t+1)((\zeta-2)t+2)} = \left.\frac{\partial^2 \mathcal{F}_{\mathbb{S}^5}}{\partial m^2}\right|_{m\downarrow m_{\mathrm{cr},1}}$$

yielding a third order phase transition. We conclude that the phase diagram of the $SU(N)$ theory is similar to that of the corresponding $U(N)$ theory, but with different critical loci and the order of the transitions is increased from second to third.

Large Chern–Simons 't Hooft coupling

For $\lambda < 0$ and $|t| \to \infty$ the solution is given in (6.B.2). The system has three phases, separated by third order transitions at the critical curves

$$m_{\mathrm{cr},1} = \lambda^{-1} = -m_{\mathrm{cr},2}.$$

In contrast to Subsection 6.3.2, the support of the eigenvalue density does not collapse sending $|\lambda| \to \infty$.

6.4.3 Two opposite mass scales

We consider the $SU(N)$ theory with $F = 2$ opposite mass scales $(-m, m)$ and equal number of hypermultiplets per mass, $\zeta_1 = \zeta_2 \equiv \zeta$. This is the traceless counterpart of the analysis carried out in Subsection 6.3.3. The SPE reads

$$-\int_A^B \mathrm{d}\psi\rho(\psi)\,(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi) = \frac{1}{t}\phi^2 + \frac{2}{\lambda}\phi + \tilde{\xi} - \frac{\zeta}{2}\Big[(\phi+m)^2\,\mathrm{sgn}\,(\phi+m) + (\phi-m)^2\,\mathrm{sgn}\,(\phi-m)\Big]. \tag{6.4.4}$$

Infinite Yang–Mills 't Hooft coupling

To begin with, we remove the Yang–Mills term sending $|\lambda| \to \infty$, thus the unique mass scale remaining in the problem in m. Solving (6.4.4) for A and B as functions of $\tilde{\xi}$ and imposing (6.4.2) we get

$$\tilde{\xi} = -\frac{4\zeta^2 m^2 t}{t^2-1} \tag{6.4.5}$$

in both phases. A and B are given in (6.B.18). We find a third order phase transition at $m = 0$, consistent with the general arguments presented so far.

Finite Yang–Mills 't Hooft coupling

Without loss of generality we restrict the analysis to $m > 0$, thanks to the $\mathbb{Z}_2$ symmetry exchanging the two sets of hypermultiplets.

Reintroducing a finite Yang–Mills term, $|\lambda| < \infty$, we solve (6.4.4) in analogy with the $U(N)$ theory of Section 6.3.3. The solution in each phase is reported in equation (6.B.19). The Lagrange

multiplier $\tilde{\xi}$ takes the values

$$\tilde{\xi} = \begin{cases} -\frac{t\left(\frac{1}{\lambda}-\zeta m\right)^2}{t^2-1} & m > m_{\mathrm{cr},1} \\ \frac{\zeta\lambda^2\left(-m^2\right)+\zeta(t-(\zeta-1)\lambda m t)^2+t(2-2\zeta\lambda m)}{2\lambda^2(t+1)((\zeta-1)t+1)} & m_{\mathrm{cr},2} < m < m_{\mathrm{cr},1} \text{ and } t\lambda < 0 \\ \frac{\zeta\lambda^2 m^2-\zeta(t-(\zeta-1)\lambda m t)^2+t(2-2\zeta\lambda m)}{2\lambda^2(t-1)((\zeta-1)t-1)} & m_{\mathrm{cr},2} < m < m_{\mathrm{cr},1} \text{ and } t\lambda > 0 \\ -\frac{t}{\lambda^2((\zeta-1)^2t^2-1)} & m < m_{\mathrm{cr},2}. \end{cases}$$

The analysis is carried out as for the corresponding unitary theory. The critical loci are

$$m_{\mathrm{cr},1}(t,\lambda,\zeta) = \begin{cases} \frac{t}{\lambda(\zeta t-t+1)} & t\lambda < 0 \\ \frac{t}{\lambda(\zeta t-t-1)} & t\lambda > 0, \end{cases}$$

$$m_{\mathrm{cr},2}(t,\lambda,\zeta) = \begin{cases} -\frac{t}{\lambda-\zeta\lambda t+\lambda t} & t\lambda < 0 \\ \frac{t}{\lambda+(\zeta-1)\lambda t} & t\lambda > 0. \end{cases}$$

The phase diagram is qualitatively analogous to that of the corresponding $U(N)$ theory, but the expressions for A and B, as well as the critical loci, are much simpler, as shown in (6.B.19). The free energy is directly evaluated in each phase, giving

$$\mathcal{F}_{\mathbb{S}^5}\big|_{m>m_{\mathrm{cr},1}} = \frac{\zeta\lambda^3m^3-\zeta\lambda m t^2\left(\lambda m\left(\left(\zeta^2+1\right)\lambda m-3\zeta\right)+3\right)+t^2}{3\lambda^3\left(t^2-1\right)}$$

$$\mathcal{F}_{\mathbb{S}^5}\big|_{m<m_{\mathrm{cr},1}} = \begin{cases} \frac{1}{6(t+1)}\left(\frac{3\zeta m^2(\zeta t+t+1)}{\lambda}-\frac{\zeta m^3(\zeta t+t+1)^2}{t}-\frac{3\zeta m t}{\lambda^2}+\frac{(\zeta-2)t^2}{\lambda^3((\zeta-1)t+1)}\right) & t\lambda<0 \\ \frac{3\zeta\lambda^2m^2t((\zeta-1)t-1)(\zeta t+t-1)-\zeta\lambda^3m^3((\zeta-1)t-1)(\zeta t+t-1)^2+3\zeta\lambda m t^2(-\zeta t+t+1)+(\zeta-2)t^3}{6\lambda^3(t-1)t((\zeta-1)t-1)} & t\lambda>0 \end{cases}$$

$$\mathcal{F}_{\mathbb{S}^5}\big|_{m<m_{\mathrm{cr},2}} = \frac{\zeta\lambda^2m^2(3-4\zeta\lambda m)-(\zeta-1)^3t^4\left((\zeta-1)\zeta\lambda^2m^2(4\zeta\lambda m-3)+1\right)+(\zeta-1)t^2\left(2(\zeta-1)\zeta\lambda^2m^2(4\zeta\lambda m-3)+5\right)}{3\lambda^3\left((\zeta-1)^2t^2-1\right)^2}.$$

Differentiating, a third order phase transition is found.

6.4.4 Wilson loops

From Subsection 6.2.4, the vev of a half-BPS Wilson loop in the fundamental representation is

$$\langle W_{\mathsf{F}}\rangle = c_A e^{2\pi r A} + c_B e^{2\pi r B} + \sum_{\alpha=1}^{F} c_\alpha e^{-2\pi r m_\alpha}$$

and its continuity at the critical surfaces follows from the continuity of A and B and the associated jump of c_A or c_B by $\frac{\zeta}{2}$.

When the gauge group is $SU(N)$, we can exploit the additional constraint (6.4.2) to prove that the derivative of $\langle W_{\mathsf{F}}\rangle$ is continuous, too. We show this for the phase with $m_\alpha \notin [A,B]$ $\forall\alpha = 1,\dots,F$, being the extension to any other phase straightforward.

Differentiate equation (6.4.2) together with (6.2.16b) with respect to a given m on both sides of the critical wall, and use the resulting expressions to get rid of the derivatives $\frac{\partial A}{\partial m}$ and $\frac{\partial B}{\partial m}$ in the formula for the derivative of the Wilson loop vev. This gives

$$\frac{1}{2\pi r}\frac{\partial}{\partial m}\langle W_{\mathsf{F}}\rangle\Big|_{m<m_{\mathrm{cr}}} = \frac{\zeta}{4}\tilde{s}\left(e^{2\pi r B}-e^{2\pi r A}\right)$$

$$\frac{1}{2\pi r}\frac{\partial}{\partial m}\langle W_{\mathsf{F}}\rangle\Big|_{m>m_{\mathrm{cr}}} = \frac{\zeta}{4}\left(e^{2\pi r B}+e^{2\pi r A}\right)-\frac{\zeta}{2}e^{-2\pi r m}$$

where we have followed the notation of Subsection 6.2.3 and introduced the auxiliary variable $\tilde{s}$, which is -1 if $-m > B$ and $+1$ if $-m < A$ in the first phase. Sending $m \to m_{\text{cr}}$, either $m = A$ or $m = B$ at the critical point, whence the continuity of $\frac{\partial}{\partial m}\langle W_{\mathsf{F}}\rangle$ follows.

We conclude that, in the $SU(N)$ theory, the vevs of Wilson loops in the fundamental representation are always differentiable. This is consistent with the explicit calculations, yielding third order phase transitions.

A similar reasoning can be applied to the expectation value of Wilson loops in large antisymmetric representations. The computations to determine $\langle W_{\mathsf{A}_K}\rangle$ follow closely those in Section 6.3.6. The additional constraints on the partial derivatives of A and B in the $SU(N)$ theory allow to show that the first derivative of the vev with respect to the mass is a continuous function.

6.4.5 Hypermultiplets in the symmetric representation

In this subsection we analyze the phase structure of $SU(N)$ theories with hypermultiplets in the symmetric representation.

Only symmetric hypermultiplet

Let us start with the simpler case of only a symmetric hypermultiplet of mass m, without Chern–Simons term. The SPE of this model is:

$$\begin{aligned}\frac{2}{\lambda}\phi = &-\int_A^B \mathrm{d}\psi\rho(\psi)\,(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi) \\ &+\frac{1}{4}\int_A^B \mathrm{d}\psi\rho(\psi)\left[(\phi+\psi+m)^2\,\mathrm{sgn}\,(\phi+\psi+m)+(\phi+\psi-m)^2\,\mathrm{sgn}\,(\phi-\psi-m)\right].\end{aligned} \tag{6.4.6}$$

We can take $m \geq 0$ without loss of generality. The symmetric ansatz

$$\rho(\phi) = \frac{1}{2}\delta(\phi+B)+\frac{1}{2}\delta(\phi-B) \tag{6.4.7}$$

solves (6.4.6) with

$$B = \begin{cases} \frac{m}{2}-\frac{1}{\lambda} & m \geq -\frac{2}{\lambda} \\ -\frac{2}{\lambda} & m \leq -\frac{2}{\lambda}. \end{cases}$$

As we are taking $m \geq 0$, the phase transition takes place at negative values of the Yang–Mills 't Hooft coupling. The Wilson loop in the fundamental representation acquires a vev

$$\langle W_{\mathsf{F}}\rangle = \begin{cases} \cosh\left(2\pi r\left(\frac{m}{2}-\frac{1}{\lambda}\right)\right) & m \geq -\frac{2}{\lambda} \\ \cosh\left(\frac{4\pi r}{\lambda}\right) & m \leq -\frac{2}{\lambda}. \end{cases}$$

We find a second order phase transition. As for the model with a single adjoint hypermultiplet of Subsection 6.2.5, at $m = 0$ we find a third order transition at the superconformal point $\frac{1}{\lambda} \to 0$, mirroring a flop transition in the dual Calabi–Yau geometry (see Appendix 6.A.1).

We can easily obtain a solution for the theory analytically continued to any real $n_{\mathsf{S}} < 2$. In that case the symmetric ansatz (6.4.7) solves the SPE with

$$B = \begin{cases} \frac{n_{\mathsf{S}}}{2}m-\frac{1}{\lambda} & m \geq -\frac{2}{\lambda(2-n_{\mathsf{S}})} \\ -\frac{2}{\lambda(2-n_{\mathsf{S}})} & m \leq -\frac{2}{\lambda(2-n_{\mathsf{S}})}. \end{cases}$$

The features of the phase diagram extend to this case.

Symmetric and fundamental hypermultiplets

We now consider $SU(N)$ theory with a massless symmetric hypermultiplet and two families of fundamental hypermultiplets with opposite masses $\pm m$, with equal Veneziano parameters $\zeta_1 = \zeta_2 \equiv \zeta$. That is, we introduce a massless symmetric or rank-two antisymmetric hypermultiplet in the model of Subsection 6.4.3. In absence of a Chern–Simons term, the Veneziano parameter is constrained by $\zeta \leq \frac{1}{2}$. The SPE is

$$\int_A^B \mathrm{d}\psi\rho(\psi)\left[-(\phi-\psi)^2\,\mathrm{sgn}\,(\phi-\psi)+\frac{1}{2}(\phi+\psi)^2\,\mathrm{sgn}\,(\phi+\psi)\right]= \\ \frac{2}{\lambda}\phi+\tilde{\xi}-\frac{\zeta}{2}\left[(\phi+m)^2\,\mathrm{sgn}\,(\phi+m)+(\phi-m)^2\,\mathrm{sgn}\,(\phi-m)\right]. \quad (6.4.8)$$

It is solved by a simple extension of the method in Subsection 6.2.3, with $\tilde{\xi}=0$. Let us begin with the case $\lambda<0$. Then, starting with the phase in which $m \notin \mathrm{supp}\rho$, we find that the symmetric ansatz (6.4.7) solves (6.4.8) with

$$B = 2\zeta m - \frac{2}{\lambda}, \qquad m > B.$$

This phase holds for $m > m_{\mathrm{cr}}$, with

$$m_{\mathrm{cr}}\,(\lambda<0,\zeta) = -\frac{2}{\lambda(1-2\zeta)},$$

which is positive. Beyond the critical point we find the solution

$$\rho(\phi) = \frac{1-2\zeta}{2}\left[\delta(\phi+B)+\delta(\phi-B)\right]+\zeta\left[\delta(\phi+m)+\delta(\phi-m)\right], \qquad B = -\frac{2}{\lambda(1-2\zeta)}.$$

For $\lambda>0$ the solution in the first phase is identical, but B becomes negative at $m=\frac{1}{\zeta\lambda}$, thus we should rename B and $-B$. The phase transition then takes place at

$$m_{\mathrm{cr}}\,(\lambda>0,\zeta) = \frac{2}{\lambda(1+2\zeta)} > 0,$$

which also equals $B^{(\mathrm{II})}$ computed in the second phase. Note that, consistently with the derivation for $\lambda>0$,

$$\frac{1}{\zeta\lambda} > \frac{2}{\lambda(1+2\zeta)}.$$

Computing the free energy, we find a third order phase transition.

The Wilson loop in the fundamental representation acquires a vev

$$\langle W_{\mathsf{F}}\rangle = \begin{cases} \cosh\left(2\pi r\left(2\zeta m-\frac{2}{\lambda}\right)\right) & m \geq m_{\mathrm{cr}}\,(\lambda<0,\zeta) \\ (1-2\zeta)\cosh\left(\frac{4\pi r}{\lambda(1-2\zeta)}\right)+2\zeta\cosh(2\pi r m) & m \leq m_{\mathrm{cr}}\,(\lambda<0,\zeta)\,, \end{cases}$$

whose derivative is a continuous but not differentiable function of m. This confirms that the phase transition is third order. The case of positive λ is analogous.

Spontaneous one-form symmetry breaking

5d $\mathcal{N}=1$ gauge theories with simple gauge group have a one-form symmetry associated to the centre of the group [255, 256]. It is $\mathbb{Z}_N$ for $SU(N)$ and $\mathbb{Z}_2$ for $USp(2N)$. This symmetry is compatible with matter in the adjoint or rank-two antisymmetric representation, whilst fundamental hypermultiplets break it explicitly [257].

It is argued in [257] that, for theories with adjoint or antisymmetric matter, the one-form symmetry is spontaneously broken. This expectation is confirmed by our results, in the regime considered, as signalled by $\langle W_{\mathsf{F}}\rangle$ following a perimeter law. Moreover we observe that transitions between two phases with spontaneously broken one-form symmetry are always second order (cf. subsections 6.2.5 and 6.4.5). Instead, whenever the one-form symmetry is explicitly broken from the onset, the phase transitions are third order.

We note a subtlety concerning the pure $SU(N)$ gauge theory at Chern–Simons level k of Subsection 6.4.1. The one-form symmetry should be restored at $k=0$, however, directly taking the limit $t\to\infty$ in $\langle W_{\mathsf{F}}\rangle$ is problematic. Instead, we write the result for large but finite N and tune $k\to 0$ first. The Wilson loop vev is then damped as

$$\langle W_{\mathsf{F}}\rangle \approx e^{-\frac{4\pi r|h|}{k}}\left[\frac{N-k}{2N}e^{\frac{2\pi r}{\lambda}\left(1+\frac{k}{N}\right)}+\frac{N+k}{2N}e^{\frac{2\pi r}{\lambda}\frac{k}{N}}\right]$$

for one sign of h, and a similar expression for the other sign. The term in bracket is finite in the $k\to 0$ limit with fixed N, thus we find agreement with [255, 256].

6.5 Phases of $USp(2N)$, $SO(2N)$ and $SO(2N+1)$ theories

In this section we study the large N phase structure of gauge theories with the other classical gauge groups: $USp(2N)$, $SO(2N)$ and $SO(2N+1)$.[16] In the large N setup of Section 6.2, the difference between $SO(2N)$ or $SO(2N+1)$ and $USp(2N)$ is sub-leading, thus it suffices to study the compact symplectic gauge group $USp(2N)$. The eigenvalues of the $\mathfrak{usp}(2N)$-valued adjoint scalar ϕ are

$$(\phi_1,\dots,\phi_N,-\phi_1,\dots,-\phi_N),$$

and we can take $\phi_a\geq 0$ for all $a=1,\dots,N$ without loss of generality. These groups do not admit a Chern–Simons term but have a $\mathbb{Z}_2$-valued theta parameter [220], which we set to zero.

Repeating the argument of Section 6.2 for $USp(2N)$ we arrive at the SPE:

$$\begin{aligned}-\int_A^B \mathrm{d}\psi\rho(\psi)\left[(\phi-\psi)^2\operatorname{sgn}(\phi-\psi)+(\phi+\psi)^2\operatorname{sgn}(\phi+\psi)\right]\\ =\frac{4}{\lambda}\phi-\sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{2}\left[(\phi+m_\alpha)^2\operatorname{sgn}(\phi+m_\alpha)+(\phi-m_\alpha)^2\operatorname{sgn}(\phi-m_\alpha)\right].\end{aligned}\tag{6.5.1}$$

The eigenvalue density is assumed to be supported on a single interval $[A,B]$ on the positive real axis. The contributions to the vector multiplet one-loop determinant from each pair of opposite eigenvalues are sub-leading at large N and do not appear in the SPE.

Convergence of the localized partition function in the large N limit requires

$$\sum_{\alpha=1}^{F}\zeta_\alpha\leq 2,\tag{6.5.2}$$

[16] For orthogonal groups the hypermultiplets are taken in the vector representation.

which matches the condition for the gauge theory to sit in the IR of a SCFT [251]. Taking three derivatives of (6.5.1) we find the solution

$$\rho(\phi) = c_A \delta(\phi - A) + c_B \delta(\phi - B) + \sum_{\alpha=1}^{F} \left[c_\alpha^- \delta(\phi + m_\alpha) + c_\alpha^+ \delta(\phi - m_\alpha) \right], \tag{6.5.3}$$

where the coefficients $c_\alpha^\pm$ are

$$c_\alpha^\pm = \begin{cases} \frac{\zeta_\alpha}{4} & \pm m_\alpha \in [A, B] \\ 0 & \text{otherwise} \end{cases}$$

for all $\alpha = 1, \dots, F$. Note that, as we are taking $0 < A < B$, at most one between c_α^- and c_α^+ is non-zero. The normalization condition (6.2.13) imposes

$$c_A + c_B + \sum_{\alpha=1}^{F} \left(c_\alpha^- + c_\alpha^+ \right) = 1. \tag{6.5.4}$$

To lighten the notation, let us define $\tilde{c}_\alpha = c_\alpha^- + c_\alpha^+$ and

$$\tilde{s}_\alpha^- = \begin{cases} -1 & -m_\alpha > B \\ 0 & A \leq -m_\alpha \leq B \\ +1 & -m_\alpha < A, \end{cases} \qquad \tilde{s}_\alpha^+ = \begin{cases} -1 & m_\alpha > B \\ 0 & A \leq m_\alpha \leq B \\ +1 & m_\alpha < A. \end{cases}$$

Plugging (6.5.3) back into the SPE (6.5.1) we find the three additional conditions

$$2c_A + \sum_{\alpha=1}^{F} \tilde{c}_\alpha \left(\tilde{s}_\alpha^- + \tilde{s}_\alpha^+ \right) = \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \left(\tilde{s}_\alpha^- + \tilde{s}_\alpha^+ \right) \tag{6.5.5a}$$

$$2c_B B + \sum_{\alpha=1}^{F} \tilde{c}_\alpha \left(\tilde{s}_\alpha^- - \tilde{s}_\alpha^+ \right) m_\alpha = -\frac{2}{\lambda} + \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \left(\tilde{s}_\alpha^- - \tilde{s}_\alpha^+ \right) m_\alpha \tag{6.5.5b}$$

$$2c_A A^2 + \sum_{\alpha=1}^{F} \tilde{c}_\alpha \left(\tilde{s}_\alpha^- + \tilde{s}_\alpha^+ \right) m_\alpha^2 = \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{2} \left(\tilde{s}_\alpha^- + \tilde{s}_\alpha^+ \right) m_\alpha^2. \tag{6.5.5c}$$

These three equations together with (6.5.4) determine c_A, c_B and the endpoints A and B. Inspection of the possible values of $\tilde{s}_\alpha^- + \tilde{s}_\alpha^+$ shows that the assumption of a δ-function supported at $\phi = A$ is redundant, because either $c_A = 0$ or the point A is merged with the singularities at $\phi = m_\alpha$ (or at $\phi = -m_\alpha$, depending on the sign of the mass). We therefore obtain the solution

$$\rho(\phi) = c_B \delta(\phi - B) + \sum_{\alpha=1}^{F} c_\alpha \delta(\phi - |m_\alpha|), \tag{6.5.6}$$

with coefficients

$$c_B = 1 - \sum_{\alpha=1}^{F} c_\alpha, \qquad c_\alpha = \begin{cases} \frac{\zeta_\alpha}{2} & -B < m_\alpha < B \\ 0 & \text{otherwise} \end{cases}$$

and endpoint

$$B = \left(1 - \sum_{\alpha=1}^{F} c_\alpha \right)^{-1} \left[-\frac{1}{\lambda} + \sum_{\alpha=1}^{F} \left(\frac{\zeta_\alpha}{2} - c_\alpha \right) m_\alpha \right].$$

In particular, B is independent of m_α when $-B < m_\alpha < B$. The full eigenvalue density $\rho_{USp(2N)}(\phi)$, that accounts for all the $2N$ eigenvalues, is symmetric under $\phi \mapsto -\phi$, and reads

$$\rho_{USp(2N)}(\phi) = \frac{\rho(\phi) + \rho(-\phi)}{2}.$$

The free energy computed using the solution (6.5.6) is

$$\begin{aligned}\mathcal{F}_{\mathbb{S}^5} = {} & \frac{4}{3}c_B^2 B^3 + \frac{c_B}{3}\sum_{\alpha=1}^{F} c_\alpha \left(|B-m_\alpha|^3 + |B+m_\alpha|^3\right) \\ & + \frac{1}{6}\sum_{\alpha=1}^{F} c_\alpha \sum_{\alpha'=1}^{F} c_{\alpha'} \left(|m_\alpha - m_{\alpha'}|^3 + |m_\alpha + m_{\alpha'}|^3\right) + \frac{2}{\lambda}\left(c_B B^2 + \sum_{\alpha=1}^{F} c_\alpha m_\alpha^2\right) \\ & - \sum_{\alpha=1}^{F} \frac{\zeta_\alpha}{6}\left[c_B\left(|B-m_\alpha|^3 + |B+m_\alpha|^3\right) + \sum_{\alpha'=1}^{F} c_{\alpha'}\left(|m_\alpha - m_{\alpha'}|^3 + |m_\alpha + m_{\alpha'}|^3\right)\right].\end{aligned}$$

Let us study the phase structure. We focus for clarity on the first phase transition, assuming that all masses are larger than B and a mass, which we take to be m_1, is decreased until it eventually crosses B. On one side of the wall we use $c_\alpha = 0$ for all $\alpha = 1, \dots, F$ and find

$$\mathcal{F}_{\mathbb{S}^5}^{(\mathrm{I})} = \frac{4}{3}\left(B^{(\mathrm{I})}\right)^3 + \frac{2}{\lambda}\left(B^{(\mathrm{I})}\right)^2 - \sum_{\alpha=1}^{F}\frac{\zeta_\alpha}{6}\left(|B^{(\mathrm{I})}-m_\alpha|^3 + |B^{(\mathrm{I})}+m_\alpha|^3\right),$$

while on the other side we use $c_1 = \frac{\zeta_1}{2}$ and $c_\alpha = 0$ for $\alpha = 2, \dots, F$ and get

$$\begin{aligned}\mathcal{F}_{\mathbb{S}^5}^{(\mathrm{II})} = {} & \frac{4}{3}\left(1-\frac{\zeta_1}{2}\right)^2\left(B^{(\mathrm{II})}\right)^3 + \frac{2}{\lambda}\left(\left(1-\frac{\zeta_1}{2}\right)\left(B^{(\mathrm{II})}\right)^2 + \frac{\zeta_1}{2}m_1^2\right) \\ & - \frac{\zeta_1^2}{24}\left(|B^{(\mathrm{II})}-m_1|^3 + |B^{(\mathrm{II})}+m_1|^3\right) - \sum_{\alpha=2}^{F}\frac{\zeta_\alpha}{6}\left(|B^{(\mathrm{II})}-m_\alpha|^3 + |B^{(\mathrm{II})}+m_\alpha|^3\right),\end{aligned}$$

where the superscripts $^{(\mathrm{I})}$ and $^{(\mathrm{II})}$ indicate the quantity evaluated in the corresponding phase. At the critical point, $B = m_1$ by definition, which guarantees the continuity of $\mathcal{F}_{\mathbb{S}^5}$. Taking derivatives and using

$$\frac{\partial B^{(\mathrm{I})}}{\partial m_1} = \frac{\zeta_1}{2}, \qquad \frac{\partial B^{(\mathrm{II})}}{\partial m_1} = 0,$$

we find that the first and second derivatives of the free energy are continuous, and the phase transition is third order.

We notice that the calculations of the free energy are akin to those in [208]. On a computational level, this stems from the symmetric form of the eigenvalues together with the lack of a Chern–Simons term. This is an incarnation of the fact that the $SU(N)$ and $USp(2N)$ gauge theories are UV-completed into the same SCFT, up to a shift in N_f that is invisible in the Veneziano limit.

The argument extends to theories with hypermultiplets in the adjoint or rank-two antisymmetric representation. The phase diagram is easily recovered from subsections 6.2.5 and 6.4.5.[17]

[17] $USp(2N)$ theories have been investigated from different angles in [248, 258–262].

6.A Calabi–Yau varieties and localization

As anticipated in Section 6.1, the Coulomb branches of the 5d $\mathcal{N} = 1$ gauge theories we are interested in can be built from resolutions of a singular Calabi–Yau threefold X. If X is realized as an elliptic fibration, the theory has special unitary gauge group [220, 263, 223, 224], while if X is realized as a $\mathbb{C}^*$-fibration the gauge group is unitary [228].

Appendix 6.A.1 embeds the results of the main text in the geometric framework and explains how to extract matrix models from Calabi–Yau geometries. Then, these facts are applied in Appendix 6.A.2 to match the reduction from $U(N)$ to $SU(N)$ gauge theories between geometry and partition function.

6.A.1 Geometric description

A 5d $\mathcal{N} = 1$ gauge theory $\mathscr{T}_X$ on its Coulomb branch can be read off from the geometry of a crepant resolution $\widetilde{X} \to X$ of a singular local Calabi–Yau threefold X.[18] Likewise, a 3d $\mathcal{N} = 2$ theory is obtained from crepant resolutions of singular local Calabi–Yau fourfolds. We focus on a threefold X.

Most of the theories we consider correspond to $\widetilde{X}$ containing compact divisors formed by N intersecting $\mathbb{P}^1$s fibered over a single holomorphic genus zero curve, that we denote $\mathbb{P}^1_0$. Besides, there are N_f divisors that are $\mathbb{P}^1$ fibrations over a non-compact curve inside $\widetilde{X}$. In the models of Subsection 6.2.5 the fibrations are instead over a genus one curve, while for quiver theories the compact divisors are fibered over a collection of intersecting $\mathbb{P}^1_{0,j}$, $j = 1, \dots, L$.

Kähler moduli $\{\phi_a\}$ of holomorphic curves that are Poincaré dual to compact divisors $\{S_a\}$ give rise to dynamical fields, while Kähler moduli $\{m_\alpha\}$ of curves Poincaré dual to non-compact divisors $\{D_\alpha\}$ give rise to background fields. The identification of the extended Coulomb branch with the extended Kähler cone $\mathscr{C}(X)$ of X stems from these relations.

The gauge theory is characterized by a Yang–Mills coupling $\frac{1}{g^2_{\text{YM}}} = \text{vol}\left(\mathbb{P}^1_0\right)$, thus the 't Hooft limit (6.2.10) increases the volume of $\mathbb{P}^1_0$ linearly with the number of compact divisors fibered over it. Besides, the Veneziano limit (6.2.11) corresponds to take the number of both compact and non-compact exceptional divisors in $\widetilde{X}$ large.

We have reviewed how to read off a gauge theory $\mathscr{T}_X$ from a resolution $\widetilde{X} \to X$. In turn, supersymmetric localization provides an explicit dictionary between the field content of a supersymmetric field theory and a matrix model representation of certain observables in such theory on a compact manifold. Therefore, a two-step procedure yields a map

$$\mathscr{C}(X) \ni \widetilde{X} \longleftrightarrow \mathscr{C}(X) \ni (\phi, m) \mapsto Z^{\mathscr{T}_X}_{\mathbb{S}^5}(\phi, m),$$

whose image is the measure in the matrix model on $\mathbb{S}^5$. The sphere partition function is the average over $\mathscr{C}_{\text{gauge}}$ of this quantity:

$$\mathcal{Z}^{\mathscr{T}_X}_{\mathbb{S}^5}(m) = \int_{\mathscr{C}_{\text{gauge}}} \mathrm{d}\phi \; Z^{\mathscr{T}_X}_{\mathbb{S}^5}(\phi, m).$$

In the above setup, these last steps define a dictionary whose entries include

$$\begin{aligned} \text{existence of } S_a &\implies \text{integrate over } \phi_a \\ \text{existence of } D_\alpha &\implies \text{hypermultiplet of mass } m_\alpha \\ \text{intersections} &\implies Z_{\text{class}}(\phi) Z^{\text{vec}}_{\text{1-loop}}(\phi) Z^{\text{hyp}}_{\text{1-loop}}(\phi). \end{aligned}$$

[18] A resolution is crepant if it preserves the canonical bundle.

Varying a Kähler parameter m_α, we have found a phase transition each time the corresponding volume crosses a threshold determined by $\frac{1}{N}\text{vol}\left(\mathbb{P}^1_0\right)$. Importantly, these transitions take place at strictly infinite rank of the gauge group, and differ in nature from the flop transitions among two birationally equivalent resolved geometries $\widetilde{X}$. An exception to this statement is discussed in Subsection 6.3.5.

6.A.2 Stückelberg mechanism from localization

In this Appendix we comment on the matrix model interpretation of the Stückelberg mechanism presented in [228] to pass from unitary to special unitary quiver gauge theories. The argument has been shown to hold for three-dimensional gauge theories [264].

The partition function on either $\mathbb{S}^3$ or $\mathbb{S}^5$ when Fayet–Iliopoulos parameters $\{\xi_j\}$ are turned on is written schematically as

$$\mathcal{Z}^{\mathscr{T}_X}_{\mathbb{S}^d}(m) = \frac{1}{\prod_j N_j!}\int_{\mathbb{R}^{\text{rank}(G)}} \mathrm{d}\phi\; e^{i4\pi r^2 \sum_{j=1}^{L}\xi_j\left(\sum_{a=1}^{N_j}\phi_{a,j}\right)}\; Z^{\mathscr{T}_X}_{\mathbb{S}^d}(\phi, m),$$

for gauge group $G = U(N_1)\times\cdots\times U(N_L)$. The parameters ξ_j are dual to the fibre of X. Therefore, according to the dictionary in Appendix 6.A.1, compactifying the $\mathbb{C}^*$-fibre we have to integrate over the scalars ξ_j, producing a δ-constraint at each gauge node. In 3d, this corresponds to gauging the $U(1)_j$ global symmetry at each node, in agreement with [264].

The situation is very similar in 5d, although we first have to address a subtlety with the integration contour. In Section 6.2 we have inserted ξ as a Lagrange multiplier, while now we want to treat it as the lowest component of a full-fledged dynamical Abelian vector multiplet. Hence, localization dictates to rotate its integration contour $\xi \mapsto i\xi$ [230], eventually producing the correct factor.

6.B Eigenvalue densities

The eigenvalue densities we have found in the main text have the generic form

$$\rho(\phi) = c_A\delta\left(\phi - A\right) + c_B\delta\left(\phi - B\right) + \sum_{\alpha=1}^{F} c_\alpha\delta\left(\phi + m_\alpha\right), \qquad \text{supp}\rho = [A, B].$$

In this appendix we collect the explicit expressions of the parameters c_A, c_B, and the endpoints A, B of the support. As explained in Subsection 6.2.3, $c_\alpha = \frac{\zeta_\alpha}{2}$ if $-m_\alpha \in [A, B]$ and $c_\alpha = 0$ otherwise. We do not report these coefficients below.

Recall that A and B are found solving quadratic equations. In all the subsequent expressions the correct choice of sign in front of each square root has already been made. For instance, $\sqrt{\lambda^2(\cdots)} = +|\lambda|\sqrt{(\cdots)}$ is understood.

6.B.1 Eigenvalue densities: $U(N)$ theories

In this appendix we collect the coefficients c_A, c_B and the endpoints A, B determining the eigenvalue density $\rho(\phi)$ in the various $U(N)$ gauge theories studied in Section 6.3.

$F = 1$, $|\lambda| \to \infty$

$$c_A = \begin{cases} \frac{(2-\zeta)t-2}{4t} & m < 0 \\ \frac{(2+\zeta)t-2}{4t} & m > 0, \end{cases} \qquad c_B = \begin{cases} \frac{(2+\zeta)t+2}{4t} & m < 0 \\ \frac{(2-\zeta)t+2}{4t} & m > 0, \end{cases} \tag{6.B.1a}$$

$$A=\begin{cases}\frac{m}{\zeta t+2}\left(\sqrt{2}\sqrt{\frac{\zeta t((\zeta+2)t+2)}{(\zeta-2)t+2}}-\zeta t\right) & m<0\\ \zeta mt-\frac{\sqrt{2}m\sqrt{\zeta(-t)((\zeta^2-4)t^2-4\zeta t+4)}}{(\zeta+2)t-2} & m>0,\end{cases}\tag{6.B.1b}$$

$$B=\begin{cases}\frac{\sqrt{2}m\sqrt{\zeta t((\zeta-2)t+2)((\zeta+2)t+2)}-\zeta mt((\zeta+2)t+2)}{(\zeta t+2)((\zeta+2)t+2)} & m<0\\ -\frac{m\left(\zeta t((\zeta-2)t-2)+\sqrt{2}\sqrt{\zeta(-t)((\zeta-2)t-2)((\zeta+2)t-2)}\right)}{((\zeta-2)t-2)(\zeta t-2)} & m>0.\end{cases}\tag{6.B.1c}$$

$F=1$, $|\lambda|<\infty$

$$c_A=\begin{cases}\frac{(2-\zeta)t-2}{4t} & m<m_{\rm cr,1}\\ \frac{(2-\zeta)t-2}{4t} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{(2+\zeta)t-2}{4t} & m>m_{\rm cr,2},\end{cases}\qquad c_B=\begin{cases}\frac{(2+\zeta)t+2}{4t} & m<m_{\rm cr,1}\\ \frac{(2-\zeta)t+2}{4t} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{(2-\zeta)t+2}{4t} & m>m_{\rm cr,2},\end{cases}\tag{6.B.2a}$$

$$A=\begin{cases}\frac{\sqrt{2}\sqrt{\lambda^2(-t)((\zeta-2)t+2)((\zeta+2)t+2)\left(2t(\zeta\lambda m+1)-\zeta\lambda^2m^2\right)}-\lambda t((\zeta-2)t+2)(\zeta\lambda m+2)}{\lambda^2((\zeta-2)t+2)(\zeta t+2)} & m<m_{\rm cr,1}\\ \frac{\sqrt{\lambda^2t^2\left((\zeta-2)^2t^2-4\right)}}{\lambda((\zeta-2)t+2)}-\frac{t}{\lambda} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{\sqrt{2}\sqrt{\lambda^2t((\zeta-2)t-2)((\zeta+2)t-2)\left(2t(\zeta\lambda m-1)-\zeta\lambda^2m^2\right)}+\lambda t((\zeta+2)t-2)(2-\zeta\lambda m)}{\lambda^2(\zeta t-2)((\zeta+2)t-2)} & m>m_{\rm cr,2},\end{cases}\tag{6.B.2b}$$

$$B=\begin{cases}-\frac{\sqrt{2}\sqrt{\lambda^2(-t)((\zeta-2)t+2)((\zeta+2)t+2)\left(2t(\zeta\lambda m+1)-\zeta\lambda^2m^2\right)}+\lambda t((\zeta+2)t+2)(\zeta\lambda m+2)}{\lambda^2(\zeta t+2)((\zeta+2)t+2)} & m<m_{\rm cr,1}\\ \frac{\sqrt{\lambda^2t^2\left((\zeta-2)^2t^2-4\right)}}{\lambda((\zeta-2)t-2)}-\frac{t}{\lambda} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ -\frac{\sqrt{2}\sqrt{\lambda^2t\left((\zeta^2-4)t^2-4\zeta t+4\right)\left(2t(\zeta\lambda m-1)-\zeta\lambda^2m^2\right)}+\lambda t((\zeta-2)t-2)(\zeta\lambda m-2)}{\lambda^2((\zeta-2)t-2)(\zeta t-2)} & m>m_{\rm cr,2}.\end{cases}\tag{6.B.2c}$$

$F=1$, $|t|\to\infty$

$$c_A=\begin{cases}\frac{2-\zeta}{4} & m<m_{\rm cr,1}\\ \frac{2-\zeta}{4} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{2+\zeta}{4} & m>m_{\rm cr,2},\end{cases}\qquad c_B=\begin{cases}\frac{2+\zeta}{4} & m<m_{\rm cr,1}\\ \frac{2-\zeta}{4} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{2-\zeta}{4} & m>m_{\rm cr,2},\end{cases}\tag{6.B.3a}$$

$$A=\begin{cases}\frac{2\left((2-\zeta)\lambda+\sqrt{(4-\zeta^2)\lambda^2(\zeta\lambda m+1)}\right)}{(\zeta-2)\zeta\lambda^2}-m & m<m_{\rm cr,1}\\ \frac{2}{(2-\zeta)\lambda} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ -\frac{2\left((\zeta+2)\lambda+\sqrt{(4-\zeta^2)\lambda^2(1-\zeta\lambda m)}\right)}{(-\zeta-2)\zeta\lambda^2}-m & m>m_{\rm cr,2},\end{cases}\tag{6.B.3b}$$

$$B=\begin{cases}-\frac{2\left((\zeta+2)\lambda+\sqrt{(4-\zeta^2)\lambda^2(\zeta\lambda m+1)}\right)}{\zeta(\zeta+2)\lambda^2}-m & m<m_{\rm cr,1}\\ -\frac{2}{(2-\zeta)\lambda} & m_{\rm cr,1}<m<m_{\rm cr,2}\\ \frac{2\left((2-\zeta)\lambda+\sqrt{(4-\zeta^2)\lambda^2(1-\zeta\lambda m)}\right)}{(2-\zeta)\zeta\lambda^2}-m & m>m_{\rm cr,2}.\end{cases}\tag{6.B.3c}$$

$F=2$, $|\lambda|\to\infty$. **Symmetric case**

If $t>\frac{1}{1-\zeta}$:

$$c_A=\frac{t-1}{2t},\qquad c_B=\frac{t+1}{2t},\tag{6.B.4a}$$

$$A=\begin{cases}\zeta mt\left(\frac{t+1}{\sqrt{t^2-1}}-1\right) & m<0\\ -\zeta mt\left(\frac{t+1}{\sqrt{t^2-1}}-1\right) & m>0,\end{cases}\qquad B=\begin{cases}\zeta mt\left(\frac{t-1}{\sqrt{t^2-1}}-1\right) & m<0\\ -\zeta mt\left(\frac{t-1}{\sqrt{t^2-1}}-1\right) & m>0.\end{cases}\tag{6.B.4b}$$

If $t < -\frac{1}{1-\zeta}$:

$$c_A = \frac{t-1}{2t}, \qquad c_B = \frac{t+1}{2t}, \tag{6.B.5a}$$

$$A = \begin{cases} -\frac{\zeta m t\left(\sqrt{t^2-1}+t-1\right)}{t-1} & m<0 \\ \frac{\zeta m t\left(\sqrt{t^2-1}+t-1\right)}{t-1} & m>0, \end{cases} \qquad B = \begin{cases} -\frac{\zeta m t\left(\sqrt{t^2-1}+t+1\right)}{t+1} & m<0 \\ \frac{\zeta m t\left(\sqrt{t^2-1}+t+1\right)}{t+1} & m>0. \end{cases} \tag{6.B.5b}$$

$F = 2$, $|\lambda| < \infty$. Symmetric case

$$c_A = \begin{cases} \frac{t-1}{2t} & m>m_{\mathrm{cr},1} \\ \frac{t(1-\zeta)-1}{2t} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t>0 \\ \frac{t-1}{2t} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t<0 \\ \frac{t(1-\zeta)-1}{2t} & m<m_{\mathrm{cr},2}, \end{cases} \qquad c_B = \begin{cases} \frac{t+1}{2t} & m>m_{\mathrm{cr},1} \\ \frac{t+1}{2t} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t>0 \\ \frac{t(1-\zeta)+1}{2t} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t<0 \\ \frac{t(1-\zeta)+1}{2t} & m<m_{\mathrm{cr},2}, \end{cases} \tag{6.B.6a}$$

$$A = \begin{cases} \frac{\sqrt{t^2(t^2-1)\left(\frac{1}{\lambda}-\zeta m\right)^2}}{1-t}+\zeta m t-\frac{t}{\lambda} & m>m_{\mathrm{cr},1} \\ \frac{\sqrt{2}\lambda\sqrt{\frac{t(t+1)((\zeta-1)t+1)\left(\zeta\lambda^2m^2+2t(\zeta\lambda m-1)\right)}{\lambda^2}}+(\zeta-1)t^2(\zeta\lambda m-2)+t(\zeta\lambda m-2)}{\lambda((\zeta-1)t+1)(\zeta t+2)} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{and } t>0 \\ -\frac{\sqrt{2}\sqrt{(t-1)t((\zeta-1)t-1)\left(2t(\zeta\lambda m-1)-\zeta\lambda^2m^2\right)}+(t-1)t(\zeta\lambda m-2)}{\lambda(t-1)(\zeta t-2)} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{and } t<0 \\ \sqrt{\frac{t^2\left((\zeta-1)^2t^2-1\right)}{\lambda^2}}\\ \frac{}{(\zeta-1)t+1}-\frac{t}{\lambda} & m<m_{\mathrm{cr},2}, \end{cases} \tag{6.B.6b}$$

$$B = \begin{cases} -\frac{\sqrt{t^2(t^2-1)\left(\frac{1}{\lambda}-\zeta m\right)^2}}{t+1}+\zeta m t-\frac{t}{\lambda} & m>m_{\mathrm{cr},1} \\ \frac{\sqrt{2}\sqrt{t(t+1)((\zeta-1)t+1)\left(\zeta\lambda^2m^2+2t(\zeta\lambda m-1)\right)}+t(t+1)(\zeta\lambda m-2)}{\lambda(t+1)(\zeta t+2)} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{and } t>0 \\ \frac{\sqrt{2}\sqrt{(t-1)t((\zeta-1)t-1)\left(2t(\zeta\lambda m-1)-\zeta\lambda^2m^2\right)}-t((\zeta-1)t-1)(\zeta\lambda m-2)}{\lambda((\zeta-1)t-1)(\zeta t-2)} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{and } t<0 \\ -\frac{\sqrt{t^2\left((\zeta-1)^2t^2-1\right)}}{(\zeta-1)t-1}+t\\ -\frac{}{\lambda} & m<m_{\mathrm{cr},2}. \end{cases} \tag{6.B.6c}$$

$F = 2$, $|\lambda| \to \infty$. Generic case

We impose $m_2 = m$ and $m_1 = -\frac{\zeta_2}{\zeta_1}m$.

$$c_A = \begin{cases} \frac{t(2-\zeta_1+\zeta_2)-2}{4t} & m<0 \\ \frac{t(2+\zeta_1-\zeta_2)-2}{4t} & m>0, \end{cases} \qquad c_B = \begin{cases} \frac{t(2+\zeta_1-\zeta_2)+2}{4t} & m<0 \\ \frac{t(2-\zeta_1+\zeta_2)+2}{4t} & m>0, \end{cases} \tag{6.B.7a}$$

$$A = \begin{cases} \frac{m\left(2\zeta_1\zeta_2t((\zeta_1-\zeta_2-2)t+2)+\sqrt{\zeta_1\zeta_2(-t)((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)\left(\zeta_1^2t+2\zeta_1(\zeta_2t+1)+\zeta_2(\zeta_2t-2)\right)}\right)}{\zeta_1((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2)t+2)} & m<0 \\ -\frac{m\left(2\zeta_1\zeta_2t((-\zeta_1+\zeta_2-2)t+2)+\sqrt{\zeta_1\zeta_2(-t)((-\zeta_1+\zeta_2-2)t+2)((-\zeta_1+\zeta_2+2)t+2)\left(\zeta_1^2t+2\zeta_1(\zeta_2t-1)+\zeta_2(\zeta_2t+2)\right)}\right)}{\zeta_1(\zeta_1(-t)+\zeta_2t+2)((-\zeta_1+\zeta_2-2)t+2)} & m>0, \end{cases} \tag{6.B.7b}$$

$$B = \begin{cases} \frac{2\zeta_1\zeta_2mt((\zeta_1-\zeta_2+2)t+2)-m\sqrt{\zeta_1\zeta_2(-t)((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)\left(\zeta_1^2t+2\zeta_1(\zeta_2t+1)+\zeta_2(\zeta_2t-2)\right)}}{\zeta_1((\zeta_1-\zeta_2)t+2)((\zeta_1-\zeta_2+2)t+2)} & m<0 \\ \frac{m\left(-2\zeta_2t-\frac{\sqrt{\zeta_1\zeta_2(-t)((-\zeta_1+\zeta_2-2)t+2)((-\zeta_1+\zeta_2+2)t+2)\left(\zeta_1^2t+2\zeta_1(\zeta_2t-1)+\zeta_2(\zeta_2t+2)\right)}}{\zeta_1(\zeta_1t-\zeta_2t-2(t+1))}\right)}{\zeta_1(-t)+\zeta_2t+2} & m>0. \end{cases} \tag{6.B.7c}$$

$F = 2$, $|\lambda| \to \infty$. Generic case revisited

When $-m_1 > B, -m_2 < A$,

$$c_A = \frac{2-\zeta_1+\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2+\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.8a}$$

$$A = \frac{\sqrt{-t((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)\left(-2\zeta_1 m_1^2+2\zeta_2 m_2^2+\zeta_1\zeta_2(m_1-m_2)^2 t\right)}-t(\zeta_1 m_1-\zeta_2 m_2)((\zeta_1-\zeta_2-2)t+2)}{((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2)t+2)}, \tag{6.B.8b}$$

$$B = -\frac{t(\zeta_1 m_1-\zeta_2 m_2)((\zeta_1-\zeta_2+2)t+2)+\sqrt{-t((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)\left(-2\zeta_1 m_1^2+2\zeta_2 m_2^2+\zeta_1\zeta_2(m_1-m_2)^2 t\right)}}{((\zeta_1-\zeta_2)t+2)((\zeta_1-\zeta_2+2)t+2)}. \tag{6.B.8c}$$

When $-m_2 < A < -m_1 < B$,

$$c_A = \frac{2-\zeta_1+\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2-\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.9a}$$

$$A = \frac{\zeta_2 m_2 t((-\zeta_1+\zeta_2+2)t-2)-\sqrt{2}m_2\sqrt{\zeta_2 t((\zeta_1-\zeta_2-2)t+2)((\zeta_1+\zeta_2-2)t-2)}}{((\zeta_1-\zeta_2-2)t+2)(\zeta_2 t-2)}, \tag{6.B.9b}$$

$$B = \frac{m_2\left(\zeta_2(-t)-\frac{\sqrt{2}\sqrt{\zeta_2 t((\zeta_1-\zeta_2-2)t+2)((\zeta_1+\zeta_2-2)t-2)}}{(\zeta_1+\zeta_2-2)t-2}\right)}{\zeta_2 t-2}. \tag{6.B.9c}$$

$F = 2$, $-\infty < \lambda < 0$. Generic case

When $A < -m_1, -m_2 < B$,

$$c_A = \frac{2-\zeta_1-\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2-\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.10a}$$

$$A = \frac{t}{\lambda}\left(\sqrt{1-\frac{4}{(\zeta_1+\zeta_2)\,t-2t+2}}-1\right), \qquad B = \frac{t}{\lambda}\left(\sqrt{\frac{4}{(\zeta_1+\zeta_2)\,t-2(t+1)}+1}-1\right). \tag{6.B.10b}$$

When $-m_2 < A$ and $A < -m_1 < B$,

$$c_A = \frac{2-\zeta_1+\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2-\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.11a}$$

$$A = \frac{\zeta_2(-\lambda)m_2 t+\frac{\sqrt{2}\sqrt{-t(\zeta_1 t-\zeta_2 t-2t+2)((\zeta_1+\zeta_2)t-2(t+1))(\zeta_2\lambda m_2(2t-\lambda m_2)-2t)}}{\zeta_1 t-\zeta_2 t-2t+2}+2t}{\lambda(\zeta_2 t-2)}, \tag{6.B.11b}$$

$$B = \frac{t((\zeta_1+\zeta_2-2)t-2)(2-\zeta_2\lambda m_2)+\sqrt{2}\sqrt{-t((\zeta_1-\zeta_2-2)t+2)((\zeta_1+\zeta_2-2)t-2)(\zeta_2\lambda m_2(2t-\lambda m_2)-2t)}}{\lambda(\zeta_2 t-2)((\zeta_1+\zeta_2-2)t-2)}. \tag{6.B.11c}$$

When $-m_1 > B$ and $A < -m_2 < B$,

$$c_A = \frac{2-\zeta_1-\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2+\zeta_1-\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.12a}$$

$$A = \frac{1}{\lambda^2(\zeta_1 t+2)}\left[\frac{\sqrt{2}\sqrt{\lambda^2 t((\zeta_1-\zeta_2+2)t+2)((\zeta_1+\zeta_2-2)t+2)(\zeta_1\lambda m_1(\lambda m_1-2t)-2t)}}{(\zeta_1+\zeta_2-2)t+2}-\lambda t(\zeta_1\lambda m_1+2)\right], \tag{6.B.12b}$$

$$B = -\frac{1}{\lambda^2(\zeta_1 t+2)}\left[\frac{\sqrt{2}\sqrt{\lambda^2 t((\zeta_1-\zeta_2+2)t+2)((\zeta_1+\zeta_2-2)t+2)(\zeta_1\lambda m_1(\lambda m_1-2t)-2t)}}{(\zeta_1-\zeta_2+2)t+2}+\lambda t(\zeta_1\lambda m_1+2)\right]. \tag{6.B.12c}$$

When $-m_1 > B$ and $-m_2 < A$,

$$c_A = \frac{2-\zeta_1-\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2+\zeta_1+\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.13a}$$

$$A = \big[\sqrt{\lambda^2(-t)((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)(-2\zeta_1\lambda m_1(\lambda m_1-2t)+\zeta_2\lambda(\zeta_1\lambda(m_1-m_2)^2t+2m_2(\lambda m_2-2t))+4t)} \\ -\lambda t((\zeta_1-\zeta_2-2)t+2)(\zeta_1\lambda m_1-\zeta_2\lambda m_2+2)\big]\,\tfrac{1}{\lambda^2((\zeta_1-\zeta_2-2)t+2)(\zeta_1 t-\zeta_2 t+2)}, \tag{6.B.13b}$$

$$B = -\big[\sqrt{\lambda^2(-t)((\zeta_1-\zeta_2-2)t+2)((\zeta_1-\zeta_2+2)t+2)(-2\zeta_1\lambda m_1(\lambda m_1-2t)+\zeta_2\lambda(\zeta_1\lambda(m_1-m_2)^2t+2m_2(\lambda m_2-2t))+4t)} \\ +\lambda t((\zeta_1-\zeta_2+2)t+2)(\zeta_1\lambda m_1-\zeta_2\lambda m_2+2)\big]\,\tfrac{1}{\lambda^2((\zeta_1-\zeta_2+2)t+2)(\zeta_1 t-\zeta_2 t+2)}. \tag{6.B.13c}$$

When $-m_1 < A$ and $-m_2 > B$,

$$c_A = \frac{2-\zeta_1-\zeta_2}{4} - \frac{1}{2t}, \qquad c_B = \frac{2+\zeta_1+\zeta_2}{4} + \frac{1}{2t}, \tag{6.B.14a}$$

$$A = \big[\sqrt{\lambda^2 t((\zeta_1+\zeta_2-2)t-2)((\zeta_1+\zeta_2+2)t-2)(-2\zeta_1\lambda m_1(\lambda m_1-2t)+\zeta_2\lambda(\zeta_1\lambda(m_1-m_2)^2t-2m_2(\lambda m_2-2t))-4t)} \\ -\lambda t((\zeta_1+\zeta_2+2)t-2)(\zeta_1\lambda m_1+\zeta_2\lambda m_2-2)\big]\,\tfrac{1}{\lambda^2((\zeta_1+\zeta_2)t-2)((\zeta_1+\zeta_2+2)t-2)}, \tag{6.B.14b}$$

$$B = \big[\sqrt{t((\zeta_1+\zeta_2-2)t-2)((\zeta_1+\zeta_2+2)t-2)(-2\zeta_1\lambda m_1(\lambda m_1-2t)+\zeta_2\lambda(\zeta_1\lambda(m_1-m_2)^2t-2m_2(\lambda m_2-2t))-4t)} \\ -t((\zeta_1+\zeta_2-2)t-2)(\zeta_1\lambda m_1+\zeta_2\lambda m_2-2)\big]\,\tfrac{1}{\lambda((\zeta_1+\zeta_2-2)t-2)((\zeta_1+\zeta_2)t-2)}. \tag{6.B.14c}$$

6.B.2 Eigenvalue densities: $SU(N)$ theories

In this appendix we collect the endpoints A, B determining the eigenvalue density $\rho(\phi)$ in the various $SU(N)$ gauge theories studied in Section 6.4. The coefficients c_A and c_B are equal to the ones in the corresponding $U(N)$ theory, and we do not report them as they already appear in Appendix 6.B.1.

$F = 1$, $|\lambda| \to \infty$

$$A = \begin{cases} -\frac{\zeta m t}{(\zeta-2)t+2} & m < 0 \\ -\frac{\zeta m t}{(\zeta+2)t-2} & m > 0, \end{cases} \qquad B = \begin{cases} -\frac{\zeta m t}{(\zeta+2)t+2} & m < 0 \\ \frac{\zeta m t}{2-(\zeta-2)t} & m > 0. \end{cases} \tag{6.B.15}$$

$F = 1$, $|\lambda| < \infty$

$$A = \begin{cases} -\frac{t(\zeta\lambda m+2)}{\lambda((\zeta-2)t+2)} & m < m_{\mathrm{cr},1} \\ -\frac{t(\zeta\lambda m+2)}{\lambda((\zeta-2)t+2)} & m_{\mathrm{cr},1} < m < m_{\mathrm{cr},2} \\ \frac{t(2-\zeta\lambda m)}{\lambda((\zeta+2)t-2)} & m > m_{\mathrm{cr},2}, \end{cases} \qquad B = \begin{cases} \frac{t(\zeta\lambda m-2)}{\lambda((\zeta+2)t+2)} & m < m_{\mathrm{cr},1} \\ \frac{t(\zeta\lambda m-2)}{\lambda((\zeta-2)t+2)} & m_{\mathrm{cr},1} < m < m_{\mathrm{cr},2} \\ \frac{t(2+\zeta\lambda m)}{\lambda((\zeta-2)t-2)} & m > m_{\mathrm{cr},2}. \end{cases} \tag{6.B.16}$$

$F = 1$, $|t| \to \infty$

$$A = \begin{cases} \frac{m\zeta\lambda+2}{(2-\zeta)\lambda} & m < \lambda^{-1} \\ \frac{2+\lambda m\zeta}{\lambda(2-\zeta)} & \lambda^{-1} < m < -\lambda^{-1} \\ \frac{2-m\zeta\lambda}{(2+\zeta)\lambda} & m > -\lambda^{-1}, \end{cases} \qquad B = \begin{cases} -\frac{m\zeta\lambda+2}{(2+\zeta)\lambda} & m < \lambda^{-1} \\ -\frac{2-\lambda m\zeta}{\lambda(2-\zeta)} & \lambda^{-1} < m < -\lambda^{-1} \\ \frac{m\zeta\lambda-2}{(2-\zeta)\lambda} & m > -\lambda^{-1}. \end{cases} \tag{6.B.17}$$

$F = 2$, $|\lambda| \to \infty$. **Symmetric case**

$$A = \begin{cases} \frac{2\zeta m t}{t-1} & m < 0 \\ -\frac{2\zeta m t}{t-1} & m > 0. \end{cases} \qquad B = \begin{cases} -\frac{2\zeta m t}{t+1} & m < 0 \\ \frac{2\zeta m t}{t+1} & m > 0. \end{cases} \tag{6.B.18}$$

$F = 2$, $|\lambda| < \infty$. **Symmetric case**

$$A = \begin{cases} \frac{t\left(\frac{1}{\lambda}-\zeta m\right)}{t-1} & m>m_{\mathrm{cr},1} \\ -\frac{t}{\lambda+(\zeta-1)\lambda t} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t\lambda<0 \\ \frac{t\left(\frac{1}{\lambda}-\zeta m\right)}{t-1} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t\lambda>0 \\ -\frac{t}{\lambda+(\zeta-1)\lambda t} & m<m_{\mathrm{cr},2}, \end{cases} \qquad B = \begin{cases} \frac{t(\zeta\lambda m-1)}{\lambda(t+1)} & m>m_{\mathrm{cr},1} \\ \frac{t(\zeta\lambda m-1)}{\lambda(t+1)} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t\lambda<0 \\ -\frac{t}{\lambda(1+t(1-\zeta))} & m_{\mathrm{cr},2}<m<m_{\mathrm{cr},1} \text{ and } t\lambda>0 \\ -\frac{t}{\lambda-\zeta\lambda t+\lambda t} & m<m_{\mathrm{cr},2}. \end{cases} \tag{6.B.19}$$

Chapter 7

On SQED and SQCD in three dimensions: phase transitions and integrability

7.1 Introduction to the chapter

The study of supersymmetric gauge theories in curved spacetime has been pushed forward considerably in the last decade due to the extension of the localization method of path integrals [56, 58]. By using localization, a much simpler integral representation of the observables of the gauge theories is achieved. In turn, these seemingly simple representations, in general of the matrix model type, contain a wealth of information. First, they are very useful for asymptotic analysis and, in the large N 't Hooft limit, have predicted phase transitions in the theory [198, 62, 102, 3]. Secondly, in many cases, especially for three dimensional theories, they are amenable to exact analytical solutions, even for finite N [62, 63]. Such exact evaluation, or the procedure leading to it, oftentimes may point towards a connection between the gauge theory and, for example, integrable systems [265].

All these aspects of the localization integral formulas will be exposed in what follows, as we will not only study finite and large N properties, together with large N phase transitions, but also give an integrable systems view of the gauge theory, by showing a connection with the hyperbolic Calogero–Moser system.

In what follows, we will consider $\mathcal{N} = 4$ theory on the 3d sphere $\mathbb{S}^3$, with gauge group $U(n)$ and an even number $N_f = 2N$ of massive chiral multiplets in the fundamental, N of them with mass m and N with mass $-m$, arranged into N hypermultiplets. We also insert a Fayet–Iliopoulos (FI) term. Localization [58, 96, 266] gives the integral representation of the partition function:

$$\mathcal{Z}_N^{U(n)} = \int_{\mathbb{R}^n} \mathrm{d}^n x \prod_{1\leq j<k\leq n} \left(2\sinh\frac{x_j - x_k}{2}\right)^2 \prod_{j=1}^{n} \frac{e^{\mathrm{i}\eta x_j}}{2^N \left[\cosh(x_j) + \cosh(m)\right]^N}, \tag{7.1.1}$$

where we set the radius of $\mathbb{S}^3$ to $1/2\pi$ and η is the FI parameter. We will eventually be interested in the limit in which the number of flavours $N_f = 2N$ is large, while the number of colours n is kept finite. Therefore, we consider $N_f = 2N \geq 2n$, so that the theory is "good" according to the Gaiotto–Witten classification [267], and the integral (7.1.1) is convergent.

The Abelian case $n = 1$ was studied in detail in [102]. In what follows, we will extend the results of [102], including $1/N$ corrections and the analysis of Wilson loops, as well as carrying over the study to non-Abelian theories, $n > 1$. In the simplest non-Abelian case $n = 2$ we will also compute $1/N$ corrections to the large N limit.

As a remark on notation, we stress that N and n have swapped meaning compared to the previous chapters.

7.2 SQED

7.2.1 Abelian theory at finite N

The partition function of the Abelian theory reads:

$$\mathcal{Z}_N^{U(1)} = 2^{-N} \int_{-\infty}^{+\infty} \mathrm{d}x e^{\mathrm{i}\eta x} \left[\cosh(x) + z\right]^{-N}, \tag{7.2.1}$$

where $z \equiv \cosh(m)$. The expression is significantly simpler than any non-Abelian case, since the one-loop determinant of the vector multiplet is trivial for $n = 1$. The partition function (7.2.1) can be computed exactly in terms of a hypergeometric function [102], as

$$\mathcal{Z}_N^{U(1)} = \frac{\sqrt{2\pi}}{2^N (1+z)^{N-\frac{1}{2}}} \frac{\Gamma(N+\mathrm{i}\eta)\Gamma(N-\mathrm{i}\eta)}{\Gamma(N)\Gamma\left(N+\frac{1}{2}\right)} {}_2F_1\left(\frac{1}{2}-\mathrm{i}\eta, \frac{1}{2}+\mathrm{i}\eta, N+\frac{1}{2}, \frac{1-z}{2}\right). \tag{7.2.2}$$

Using an Euler transformation for the hypergeometric [268, Ch. 2], we can rewrite (7.2.2) when $\eta \geq 1, m \geq 1$ as:

$$\mathcal{Z}_N^{U(1)} = \frac{e^{\mathrm{i}\eta m}}{2^N (\sinh(m))^N} \frac{\Gamma(N-\mathrm{i}\eta)\Gamma(\mathrm{i}\eta)}{\Gamma(N)} {}_2F_1\left(1-N, N, 1-\mathrm{i}\eta, -(e^{2m}-1)^{-1}\right) \; + \; (\mathrm{i}\eta \leftrightarrow -\mathrm{i}\eta).$$

This latter form is illustrative: since the first coefficient, $a = 1 - N$, is a non-positive integer, the hypergeometric series terminates and gives a polynomial of degree $N - 1$ in the variable $y \equiv -(e^{2m} - 1)^{-1}$. Moreover, in our case the second coefficient $b = N = 1 - a$, thus the hypergeometric function is actually an associated Legendre function of imaginary order [269, Eq. (15.9.21)]:

$${}_2F_1\left(1-N, N, 1-\mathrm{i}\eta, y\right) = \Gamma(1-\mathrm{i}\eta)\left(\frac{y}{1-y}\right)^{\frac{\mathrm{i}\eta}{2}} P_{N-1}^{\mathrm{i}\eta}(1-2y).$$

The partition function reads:

$$\mathcal{Z}_N^{U(1)} = \frac{\pi e^{-\frac{\pi\eta}{2}} \Gamma(N-\mathrm{i}\eta)}{2^N \mathrm{i} \sinh(\pi\eta) \sinh(m)^N \Gamma(N)} P_{N-1}^{\mathrm{i}\eta}(\coth(m)) \; + \; (\mathrm{i}\eta \leftrightarrow -\mathrm{i}\eta),$$

where we used the property $\Gamma(1-\mathrm{i}\eta)\Gamma(\mathrm{i}\eta) = \pi/\sin(\mathrm{i}\pi\eta)$.

We can represent the function (7.2.2) in yet another form, in terms of a conical function [270, 102]:

$$\mathcal{Z}_N^{U(1)} = \frac{\sqrt{2\pi}}{2^N (\sinh(m))^{N-\frac{1}{2}}} \frac{\Gamma(N+\mathrm{i}\eta)\Gamma(N-\mathrm{i}\eta)}{\Gamma(N)} P_{-\frac{1}{2}+\mathrm{i}\eta}^{\frac{1}{2}-N}(z),$$

where $P_{-\frac{1}{2}+\mathrm{i}\eta}^{\frac{1}{2}-N}(z)$ is an associated Legendre function of negative order and complex degree. This latter form is the most suitable to study the asymptotics for large mass. Indeed, when $m \to \infty$, $z = \cosh(m) \to \infty$ as well and we can use the approximation of [271]:

$$\begin{aligned} P_{-\frac{1}{2}+\mathrm{i}\eta}^{\frac{1}{2}-N}(z) &\approx \sqrt{\frac{2\pi}{z}} \frac{\sin\left(\eta \log(2z) + \theta_1 + \theta_2\right)}{\sinh(\pi\eta)|\Gamma(1+\mathrm{i}\eta)\Gamma(N+\mathrm{i}\eta)|} \\ &= \sqrt{\frac{2}{\pi z}} \frac{\sin\left(\eta \log(2z) + \theta_1 + \theta_2\right)}{\prod_{k=0}^{N-1} \sqrt{k^2+\eta^2}}, \end{aligned}$$

where $\theta_1 = \arg\Gamma(1+\mathrm{i}\eta)$ and $\theta_2 = \arg\Gamma(N-\mathrm{i}\eta)$, and in the second line we used elementary identities for the Γ function. Altogether, and approximating the hyperbolic functions for $m \to \infty$, we have:

$$\mathcal{Z}_N^{U(1)} \approx \frac{e^{-mN}\pi \prod_{k=1}^{N-1}\sqrt{k^2+\eta^2}}{2^{N-1}\Gamma(N)\sinh(\pi\eta)} \sin\left(\eta m + \theta_1 + \theta_2\right). \tag{7.2.3}$$

This approximation is in agreement with the large mass approximation found in [102, Eq. (8)] applying a different Euler transformation to (7.2.2), which led to:

$$\mathcal{Z}_N^{U(1)} \approx \frac{2\pi e^{-mN}}{\Gamma(N)\sinh(\pi\eta)} \Im\left(e^{\mathrm{i}m\eta}\prod_{k=1}^{N-1}(k-\mathrm{i}\eta)\right). \tag{7.2.4}$$

See Figure 7.1 for the match of expressions (7.2.3) and (7.2.4).

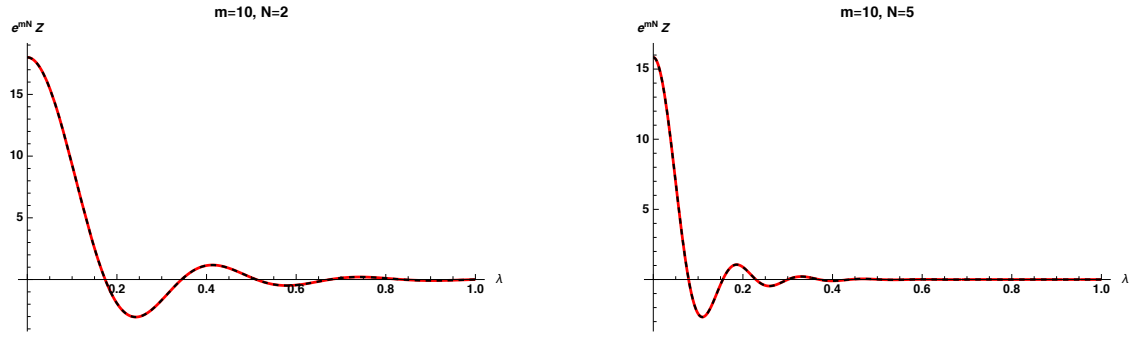


Figure 7.1. Approximation of $e^{mN}\mathcal{Z}_N^{U(1)}$ at large $m = 10$ as a function of $\lambda = \eta/N$, using (7.2.3) (red) and (7.2.4) (black, dashed), for $N = 2$ (left) and $N = 5$ (right).

The exact evaluation (7.2.2) of the partition function, or its equivalent representation as a conical function, relies on the hypothesis $\cosh(m) \geq 1$, thus on reality of the mass. However, the dependence of $\mathcal{Z}_N^{U(1)}$ on m should be holomorphic [96, 97]. For arbitrary complex masses the integral (7.2.1) can be evaluated by residue theorem [101], and we checked for many values of N that the result coincide with the prolongation of (7.2.2) to complex masses.

7.2.2 Integrability

The partition function satisfies the second-order differential equation [102]

$$\frac{\mathrm{d}^2\mathcal{Z}_N}{\mathrm{d}m^2} + 2N\coth(m)\frac{\mathrm{d}\mathcal{Z}_N}{\mathrm{d}m} + (\eta^2+N^2)\mathcal{Z}_N = 0, \tag{7.2.5}$$

which becomes the Schrödinger equation with a hyperbolic Pöschl–Teller potential, for the function $Z(m) = \sinh(m)^N\mathcal{Z}_N$ [102]. This quantum mechanical model has a discrete energy spectrum [272], and $Z(m)$ represents the wave function of a state with positive energy proportional to η^2. Furthermore, the fact that the potential appears with integer coefficient N implies that the wave function propagates without reflection.

The appearance of the quantum mechanical interpretation with a solvable Pöschl–Teller potential immediately suggests a possible role of the hyperbolic Calogero–Moser model, the celebrated integrable system, which can be seen as the many-body generalization of the quantum mechanical problem above. The Hamiltonian of the $A_{\widehat{N}-1}$ hyperbolic Calogero–Moser model is [273, 274]

$$H = \sum_{1\leq j<k\leq\widehat{N}}\left[-\hbar^2\partial_{x_j}\partial_{x_k} + \frac{g(g-\hbar)\mu^2}{4\sinh^2\left(\mu\left(x_j-x_k\right)/2\right)}\right], \tag{7.2.6}$$

and there exist $\widehat{N}-1$ additional independent partial differential operators H_l of order l. The simplest is the momentum operator

$$H_1 = -\mathrm{i}\hbar \sum_{j=1}^{\widehat{N}} \partial_{x_j} \tag{7.2.7}$$

whereas the others are made of correspondingly higher derivatives (and lower order terms as well). Here, $\widehat{N} = n+1$. Consider the two-particle case, the family is then the Hamiltonian and the momentum operator, (7.2.6) and (7.2.7).

Using recent work on the construction, by a recursive method, of the joint eigenfunctions of this integrable system [274], we show now that the Abelian theory above can be identified with this two-particle A_1 hyperbolic Calogero–Moser, where the coupling constant g in (7.2.6) will be identified with the half-number of flavours N. In particular, this two-particle interpretation follows from considering the function

$$\Psi_2(g;x,y) \equiv e^{iy_2(x_1+x_2)} \int_{-\infty}^{\infty} e^{\mathrm{i}(y_1-y_2)w} K_2(g;x,w)\mathrm{d}w,$$

with $x, y \in \mathbb{R}^2$and where the kernel is $(g>0)$

$$K_2(g;x,w) = \frac{\left[4\sinh^2(x_1-x_2)\right]^{g/2}}{\prod_{j=1}^{2}\left[2\cosh(w-x_j)\right]^g},$$

and is central in the recursion, taking the $\widehat{N}-1$ eigenfunction to the $\widehat{N}$ eigenfunction. The connection with the function $Z(m)$ defined above follows immediately from the identifications $g=N$, $x_1 = m/2 = -x_2$ and $(y_1-y_2)/2 = \eta$. It is shown in [274] that

$$\begin{aligned} H_1\Psi_2(x,y) &= (y_1+y_2)\Psi_2(x,y), \\ H\Psi_2(x,y) &= (y_1^2+y_2^2)\Psi_2(x,y). \end{aligned}$$

A different type of connection also exists relating the non-Abelian theory, with $\widehat{N}=N$, with the free case of the integrable system, given by $g=\hbar$ in (7.2.6). Using the customary adimensional coupling $\widehat{\lambda} \equiv g/\hbar = 1$, (7.2.6) is then the free N-body Hamiltonian. Thus, there is no identification here between g and number of flavours and is a very different relationship compared to the two-particle one. The integral representation given for $\Psi_N(\widehat{\lambda};x,y)$ [274] is then evaluated exactly for $\widehat{\lambda}=1$ and the explicit expression [274, Thm 3.1.] is the one for the partition function of the $T[SU(N)]$ linear quiver [275, 101, 276].

The relationship between the integral expressions in [274] and the well-known Heckman–Opdam hypergeometric functions [277], which are also relevant in [278, 279], is explained in [274]. By factorizing Ψ_N in two pieces, one describing the centre of mass, it is shown in [274] that the remaining piece is the A_{N-1} Heckman–Opdam hypergeometric function. In terms of two sets of N variables $(m_j,\zeta_j)_{j=1}^N$, this hypergeometric satisfies the condition $\sum_j m_j = 0 = \sum_j \zeta_j$, with $\zeta_j \in \mathbb{R}$ and complex m_j such that $|\Im(m_j-m_k)| < \pi$, cf. [274, Thm 7.1]. On the gauge theory side, those are exactly the constraints on the $T[SU(N)]$ theory [101], the first being the $SU(N)$ flavour symmetry and the latter arising from the redundancy of the N number of ζ_j variables, defined from the original $N-1$ FI parameters as $\zeta_j = \eta_j - \eta_{j+1}$. We underline that the partition function of the $T[SU(N)]$ quiver is evaluated for real masses and FI parameters but, by holomorphy, holds on the stripes $|\Im(m_j-m_k)| < \pi$, hence the identification is exact.

7.2.3 Abelian theory at large N

Sending $N \to \infty$ with $\lambda \equiv \eta/N$ fixed, the leading contribution to the partition function (7.2.1) comes from the saddle points of the action

$$S_1(x) = -\mathrm{i}\lambda x + \log\left[\cosh(x) + z\right], \tag{7.2.8}$$

which are given by the set $\mathscr{S} = \{x_s^{\pm} + \mathrm{i}2\pi k,\ k \in \mathbb{Z}\}$, with

$$x_s^{\pm} = \log\left(\frac{-\lambda z \pm \mathrm{i}\Delta}{\mathrm{i} + \lambda}\right), \tag{7.2.9}$$

where $\Delta \equiv \sqrt{1 - \lambda^2 \sinh(m)^2}$ and we recall that $z \equiv \cosh(m)$. The curve $\lambda \sinh(m) = 1$ determines a critical line in parameter space, along which the free energy $\mathcal{F} = -\frac{1}{N}\log \mathcal{Z}$ has a discontinuity in its second derivative. In the *sub-critical* phase $\lambda \sinh(m) < 1$, the leading contribution comes from x_s^+ and $k = 0$, while in the *super-critical* phase $\lambda \sinh(m) > 1$ both $x_s^{\pm}$ contribute, being complex conjugate and $S_1(x_s^-) = S_1(x_s^+)^*$.

Close to the saddle points $\bar{x} \in \mathscr{S}$, we can change variables $x = \bar{x} + t/\sqrt{N}$ and expand

$$S_1(x) = S_1(\bar{x}) + \frac{t^2 S_1''(\bar{x})}{2N} + \frac{t^3 S_1'''(\bar{x})}{6N^{\frac{3}{2}}} + \frac{t^4 S_1^{(iv)}(\bar{x})}{24N^2} + \dots$$

We now plug this expansion into (7.2.1) and keep the Gaussian part in t exponentiated, while expanding the rest of the exponential function. Elementary integration provides:

$$\mathcal{Z}^{U(1)} = 2^{-N}\sqrt{\frac{2\pi}{N}} \sum_{\bar{x}\in\mathscr{S}} \frac{e^{-NS_1(\bar{x})}}{\sqrt{S_1''(\bar{x})}} \left[1 + \frac{1}{24N}\left(\frac{5S_1'''(\bar{x})}{(S_1''(\bar{x}))^3} - \frac{3S_1^{(iv)}(\bar{x})}{(S_1''(\bar{x}))^2}\right) + \mathcal{O}(N^{-2})\right].$$

The relevant expressions for the derivatives of the action S_1 are reported in Appendix 7.A.1. When $\lambda \sinh(m) < 1$, only x_s^+ contributes, and we get:

$$\mathcal{Z}^{U(1)}_{\text{sub.}} = 2^{-N}\sqrt{\frac{2\pi}{N}} \frac{e^{-NS_1(x_s^+)}}{\sqrt{S_1''(x_s^+)}} \left[1 + \frac{1}{24N}\left(\frac{5S_1'''(x_s^+)}{(S_1''(x_s^+))^3} - \frac{3S_1^{(iv)}(x_s^+)}{(S_1''(x_s^+))^2}\right)\right] + \mathcal{O}(N^{-2}),$$

while in the supercritical phase $\lambda \sinh(m) > 1$ both $x_s^{\pm}$ must be taken into account, leading to:

$$\mathcal{Z}^{U(1)}_{\text{super.}} = 2\Re\left(\mathcal{Z}^{U(1)}_{\text{sub.}}\right) + \mathcal{O}(N^{-2}).$$

Dropping sub-leading corrections, one can evaluate $\mathcal{F}$ in both phases:

$$\mathcal{F}^{U(1)}_{\text{sub.}} = S_1(x_s^+), \quad \mathcal{F}^{U(1)}_{\text{super.}} = \Re\left(S_1(x_s^+)\right), \tag{7.2.10}$$

with discontinuous second derivative:

$$\frac{\partial^2 \mathcal{F}^{U(1)}_{\text{sub.}}}{\partial\lambda^2} - \frac{\partial^2 \mathcal{F}^{U(1)}_{\text{super.}}}{\partial\lambda^2} = \frac{z}{(1+\lambda^2)\Delta}.$$

Therefore, not only the susceptibility $\frac{\partial^2\mathcal{F}}{\partial\lambda^2}$ is discontinuous, but it is divergent as $(\lambda - \lambda_c)^{-\gamma_c}$, and we identify the critical exponent $\gamma_c = \frac{1}{2}$. The free energy yields analogous discontinuity with respect to the mass:

$$\frac{\partial^2 \mathcal{F}^{U(1)}_{\text{sub.}}}{\partial m^2} - \frac{\partial^2 \mathcal{F}^{U(1)}_{\text{super.}}}{\partial m^2} = \frac{z\Delta}{\sinh(m)^2} - \frac{\lambda z}{\Delta},$$

hence the critical exponent for the mass is again $\delta_c = \frac{1}{2}$.

In Figure 7.2 we present the convergence of the exact solution (7.2.2) and the large N expression (7.2.10) as N is increased.

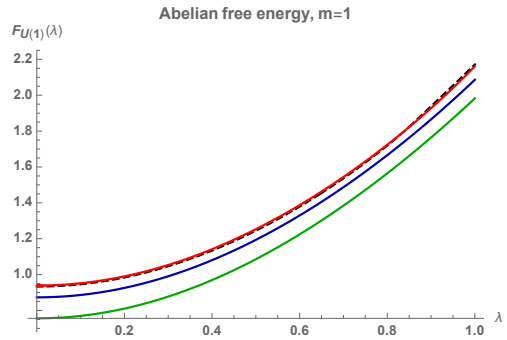


Figure 7.2. Exact solution of $\mathcal{F}^{U(1)}$ as a function of $\lambda = \eta/N$ at $m = 1$, for $N = 4, 7, 20$ (in green, blue, red, respectively) and large N expression (black, dashed).

7.2.4 Wilson loops

Irreducible complex representations of $U(1)$ are labelled by $\nu \in \mathbb{Z}$, thus Wilson loops can be written as $W_\nu = \mathrm{Tr}_\nu e^x = e^{\nu x}$ (recall that the radius of the three-sphere is $1/2\pi$), and their vacuum expectation value is:

$$\begin{aligned}\langle W_\nu \rangle &= \frac{1}{2^N \mathcal{Z}_N^{U(1)}} \int_{-\infty}^{+\infty} \mathrm{d}x \frac{e^{(\mathrm{i}\eta+\nu)x}}{[\cosh(x)+z]^N} \\ &= \frac{\Gamma(N+\nu+\mathrm{i}\eta)\Gamma(N-\nu-\mathrm{i}\eta)}{\Gamma(N+\mathrm{i}\eta)\Gamma\left(N-\mathrm{i}\eta\right)} \cdot \frac{{}_2F_1\left(\frac{1}{2}-\nu-\mathrm{i}\eta, \frac{1}{2}+\nu+\mathrm{i}\eta, N+\frac{1}{2}, \frac{1-z}{2}\right)}{{}_2F_1\left(\frac{1}{2}-\mathrm{i}\eta, \frac{1}{2}+\mathrm{i}\eta, N+\frac{1}{2}, \frac{1-z}{2}\right)},\end{aligned} \tag{7.2.11}$$

where we stress that the insertion of a Wilson loop is analogous to the complexification of the FI coupling. The integral representation (7.2.11) is well-defined as $\eta \to 0$ only for representations of size $|\nu| < N$: this is reflected in the poles of the Γ function at negative integers.

The quantum mechanical interpretation carries over for the Wilson loop without FI term, $\eta = 0$. In this case, $w_\nu \equiv \left[\sinh(m)^N \mathcal{Z}_N \langle W_\nu \rangle\right]_{\eta=0}$ satisfies the Schrödinger equation with Pöschl–Teller potential:

$$\left[\frac{\mathrm{d}^2}{\mathrm{d}m^2} - \frac{N(N-1)}{\sinh(m)^2}\right] w_\nu = \nu^2 w_\nu.$$

The latter equation describes the wave function of a bound state with energy proportional to ν^2, for integer $|\nu| < N$, which is indeed the case at hand [272].

For $\eta \neq 0$, however, the resulting potential acquires an imaginary part, seemingly spoiling unitarity of the evolution operator and producing a dissipation-like term in the probability conservation.

At large N with the size ν of the representation fixed, the Wilson loop can be approximated by the value of the integrand in (7.2.11) at the saddle points. Nevertheless, we can also consider the case of large representations, in which ν scales with N, so that $\kappa \equiv \nu/N$ is kept fixed as $N \to \infty$. Let us turn off the FI term for simplicity, $\eta = 0$, the saddle points of the action are given by:

$$\bar{x} = \log\left(\frac{\kappa\cosh(m) \pm \sqrt{1+\kappa^2\sinh(m)^2}}{1-\kappa}\right) + \mathrm{i}2\pi k,$$

with $k \in \mathbb{Z}$, that are real for every $-1 < \kappa < 1$.[19] Therefore, the Wilson loops without FI term do not experience phase transition. The limit with both η and ν scaling with N is discussed in Appendix 7.A.2.

[19] As $|\nu| < N$, the range of validity is $-1 < \kappa < 1$. x_s^+ is singular as $\kappa \to 1^-$, while x_s^- is singular as $\kappa \to -1^+$.

7.2.5 J_3 correlators

We can also consider other families of operators, besides Wilson loops. Higgs branch operators in 3d $\mathcal{N} = 4$ can be analyzed through localization techniques [280], and therefore represent a suitable choice for the present setting. In particular, we focus our attention on the gauge invariant, quadratic operator

$$J_3 = \frac{1}{N}\left[\widetilde{Q}_{+,j}Q_+^j - \widetilde{Q}_{-,j}Q_-^j\right],$$

where $(Q_{\pm,j}, \widetilde{Q}_{\pm,j})$, $j = 1, \dots, N$, are the hypermultiplets of mass $\pm m$, formed by chiral/antichiral pairs. The expectation value of this operator is

$$\langle J_3\rangle = \frac{1}{2N\mathcal{Z}_N}\frac{\mathrm{d}\mathcal{Z}_N}{\mathrm{d}m},$$

and correlation functions of J_3 are generated by higher derivatives.

The differential equation (7.2.5) satisfied by $\mathcal{Z}_N$ can be translated into a recursion relation for correlators of J_3:

$$\langle J_3 J_3\rangle = -\coth(m)\langle J_3\rangle - \frac{1}{4N}\left(1 + \frac{\eta^2}{N^2}\right).$$

Taking the first derivative of (7.2.5) gives $\frac{d^3\mathcal{Z}_N}{\mathrm{d}m^3}$ as a function of the first and second derivative of $\mathcal{Z}_N$, but the second order term can be eliminated using (7.2.5). Hence, we immediately obtain:

$$\langle J_3 J_3 J_3\rangle = \langle J_3\rangle\left[\frac{2N\cosh(m)^2 + 1}{2N\sinh(m)^2} - \frac{1}{4}\left(1 + \frac{\eta^2}{N^2}\right)\right] + \frac{1}{4}\left(1 + \frac{\eta^2}{N^2}\right).$$

One can take further derivatives and systematically plug (7.2.5) in the resulting expression. This allows to recursively compute k-point correlation functions of J_3: leveraging (7.2.5), the final result will be an expression only in terms of $\langle J_3\rangle$, hyperbolic functions of m and polynomials in $\left(1 + \eta^2/N^2\right)$.

7.3 SQCD

7.3.1 Non-Abelian theory: $SU(2)$

The simplest non-Abelian theory corresponds to the gauge group $SU(2)$. The partition function is again a single integral, but now the one-loop determinant of the vector multiplet contributes. Also, $SU(2)$ gauge theories do not admit a FI term, thus $\eta = 0$. The partition function is:

$$\mathcal{Z}_N^{SU(2)} = \int_{-\infty}^{+\infty}\mathrm{d}x\,\frac{\sinh(x)^2}{2^N\left[\cosh(x) + z\right]^N}.$$

Writing $\sinh(x)$ in terms of exponentials, we can see the $SU(2)$ partition function as a combination of expectation values of Wilson loops in the Abelian theory:

$$\mathcal{Z}_N^{SU(2)} = \left[\frac{\mathcal{Z}_N^{U(1)}}{2}\left(\langle W_2\rangle - 2 + \langle W_{-2}\rangle\right)\right]_{\eta=0},$$

with the expectation value $\langle W_r\rangle$ given in (7.2.11).

Due to the absence of FI term, the unique saddle point is $x_s = 0$, and the phase structure at large N is trivial.

7.3.2 Non-Abelian theory: $U(2)$

We now apply the same procedure to the $U(2)$ theory, i.e. two colours. Specialization of (7.1.1) for $n = 2$ gives:

$$\mathcal{Z}_N^{U(2)} = \int_{\mathbb{R}^2} \mathrm{d}x_1 \mathrm{d}x_2 \, \frac{e^{\mathrm{i}\eta(x_1+x_2)} \left(2\sinh\frac{x_1-x_2}{2}\right)^2}{2^{2N}\left[(\cosh(x_1)+z)\,(\cosh(x_2)+z)\right]^N}, \tag{7.3.1}$$

where, as above, $z \equiv \cosh(m)$. Through the equivalent representation of (7.3.1) as a determinant, one could write an exact solution

$$\mathcal{Z}_N^{U(2)} = 2! \det_{1\le j,k\le 2}[Z_{jk}],$$

with Z_{jk} entries of a 2×2 matrix formally given by (7.2.2) up to a shift in the FI coupling $\mathrm{i}\eta \mapsto \mathrm{i}\eta + j + k - 2$, $j,k \in \{1,2\}$. This equals the determinant of a matrix whose entry (j,k) is the expectation value, in the Abelian matrix model, of a Wilson loop in the irreducible representation labelled by $j+k-2$:

$$\mathcal{Z}_N^{U(2)} = 2\left(\mathcal{Z}_N^{U(1)}\right)^2 \left(\langle W_2\rangle - \langle W_1\rangle^2\right).$$

To study (7.3.1) in the limit in which the number of flavours N is large, we notice that the interaction between eigenvalues is sub-leading in $1/N$, thus the saddle points of the $U(2)$ theory are those of the action $S_1(x_1) + S_1(x_2)$:

$$\mathscr{S}^2 = \left\{(x_s^{\pm} + 2\pi k_1, x_s^{\pm} + 2\pi k_2),\ k_{1,2} \in \mathbb{Z}\right\}.$$

We proceed as in the Abelian case: we change variables $x_{1,2} = \bar{x}_{1,2} + t_{1,2}/\sqrt{N}$ and expand both the action and the hyperbolic interaction around the saddle point $(\bar{x}_1, \bar{x}_2)$. Expanding up to $\mathcal{O}(N^{-1})$ and integrating we obtain, for the sub-critical phase:

$$\mathcal{Z}_{\text{sub.}}^{U(2)} = \frac{\pi}{2^{2(N-1)}N^2} \frac{e^{-2NS_1(x_s^+)}}{(S_1''(x_s^+))^2}\left[1 + \frac{1}{2N}\left(\frac{1}{S_1''(x_s^+)} + \frac{17\,(S_1'''(x_s^+))^2}{6(S_1''(x_s^+))^3} - \frac{3S_1^{(iv)}(x_s^+)}{2(S_1''(x_s^+))^2}\right)\right],$$

while the expression in the super-critical phase $\lambda\sinh(m) > 1$ is a sum of four pieces, and is reported in Appendix 7.A.3.

Dropping $1/N$ corrections, the free energy is simply $\mathcal{F}^{U(2)} = 2\mathcal{F}^{U(1)}$, in particular the phase transition is second order with the same critical exponent $\gamma_c = \frac{1}{2}$. In Figure 7.3 we show how the exact solution approaches the large N expression as N is increased.

We study the most general non-Abelian case in Appendix 7.B, and only report here the main result. The free energy at large N of the $U(n)$ theory is n times the free energy of the Abelian theory:

$$\mathcal{F}^{U(n)} = n\mathcal{F}^{U(1)}.$$

7.A Actions, partition functions and Wilson loops

This appendix collects lengthy expressions and technical details that have been omitted from the main.

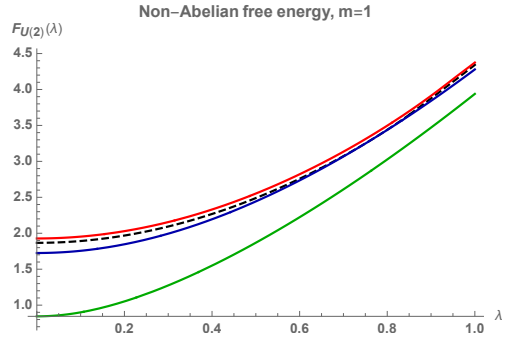


Figure 7.3. Exact solution from determinants of $\mathcal{F}^{U(2)}$ as a function of $\lambda = \eta/N$ at $m = 1$, for $N = 4, 7, 100$ (in green, blue, red, respectively) and large N expression (black, dashed).

7.A.1 Derivatives of the action S_1

Here we present the full expressions for the derivatives of the action in the Abelian theory, evaluated at the saddle point $\bar{x} = x_s^+$. In what follows, we denote $z \equiv \cosh(m)$, $\xi \equiv \lambda \sinh(m)$ and $\Delta \equiv \sqrt{1 - \lambda^2 \sinh(m)^2}$.

$$
\begin{aligned}
S_1(x_s^+) &= \log\left(\frac{\Delta z + \mathrm{i}\lambda \sinh(m)^2 + 1}{(\mathrm{i}+\lambda)(\lambda z - \mathrm{i}\Delta)}\right) - \mathrm{i}\lambda \log\left(\frac{-\lambda z + \mathrm{i}\Delta}{\mathrm{i}+\lambda}\right), \\
S_1''(x_s^+) &= \lambda^2\left[1 + \frac{z\Delta - 1}{\xi^2}\right], \\
S_1'''(x_s^+) &= \frac{\lambda(1-\mathrm{i}\lambda)}{4\sinh(m)^2(\lambda\sinh(m)^2 - \mathrm{i}z\Delta - \mathrm{i})^2}\left[2\xi - 6\xi z\Delta + 2\left(4\xi^2 - 3\right)\xi\cosh(2m) + 8\mathrm{i}\xi^2\sinh(m)^3\right. \\
&\quad \left. + 2\xi\cosh(3m)\Delta + 7\mathrm{i}\sinh(m) - 8i\xi^2\sinh(2m)\Delta + 2i\sinh(2m)\Delta - \mathrm{i}\sinh(3m)\right], \\
S_1^{(iv)}(x_s^+) &= -6\lambda^4 + \frac{23\lambda^2}{2\sinh^2 m} - \frac{\Delta\left(48\lambda^2\sinh(m)^2 - 23\right)z + 2\left(7\lambda^2\sinh(m)^2 - 4\right)\cosh(2m) + \Delta\cosh(3m) - 16}{4\sinh^4 m}.
\end{aligned}
$$

The values of the derivatives of S_1 when evaluated at x_s^- are immediately obtained through the relations:

$$
S_1(x_s^-) = (S_1(x_s^+))^*, \quad S_1''(x_s^-) = (S_1''(x_s^+))^*, \quad S_1'''(x_s^-) = -(S_1'''(x_s^+))^*, \quad S_1^{(iv)}(x_s^-) = (S_1^{(iv)}(x_s^+))^*
$$

7.A.2 Multiple scaling limit of Wilson loops

The large N limit of (7.2.11) whit $\lambda \equiv \eta/N$ and $\kappa \equiv \nu/N$ fixed, with $0 \leq \kappa < 1$, is obtained from the contributions of the saddle points:

$$
\bar{x} = \log\left(\frac{-\left(\lambda z + L\sin\frac{\theta}{2}\right) + \mathrm{i}\left(\kappa z + L\cos\frac{\theta}{2}\right)}{\lambda + \mathrm{i}(1-\kappa)}\right) + \mathrm{i}2\pi k, \quad k \in \mathbb{Z},
$$

where we defined L and θ as:

$$
\begin{aligned}
L \equiv L(\lambda,\kappa) &= \sqrt{1 + (\lambda^2+\kappa^2)^2\sinh(m)^4 - 2(\lambda^2-\kappa^2)\sinh(m)^2}, \\
\sin\theta = \frac{2\lambda\kappa\sinh(m)^2}{L}, \;\; \cos\theta &= \frac{1 - (\lambda^2-\kappa^2)\sinh(m)^2}{L}.
\end{aligned}
$$

Those saddle points are in general complex, and there is no critical surface in parameter space signalling a phase transition. The sub-critical phase of the case $\nu = 0$ now corresponds to the

system living in the surface in the (λ, κ, m) space determined by the equation:

$$\left(\lambda z + L \sin\frac{\theta}{2}\right)^2 + \left(\kappa z + L\cos\frac{\theta}{2}\right)^2 = \lambda^2 + (1-\kappa)^2,$$

while the rest of the 3d parameter space is qualitatively analogous to the super-critical phase of the partition function.

This is the expected behaviour, as we now explain. Indeed, as noticed above, decorating the theory with a Wilson loop has the net effect of complexifying the FI parameter. Therefore, on any fixed-m slice no interface is expected to separate different phases, but instead a phase transition would correspond to jump on a higher-codimensional locus in parameter space.

7.A.3 Partition function in the super-critical phase for $n = 2$

The non-Abelian theory with $n = 2$ has four relevant saddle points, obtained from the combinations $(\bar{x}_1, \bar{x}_2) = (x_s^{\pm}, x_s^{\pm})$. In the sub-critical phase, only (x_s^+, x_s^+) contributes, but in the super-critical phase all four saddle points are to be taken into account, and the partition function is therefore the sum of four pieces:

$$\mathcal{Z}^{U(2)}_{\text{super.}} = \mathcal{Z}(x_s^+, x_s^+) + \mathcal{Z}(x_s^+, x_s^-) + \mathcal{Z}(x_s^-, x_s^+) + \mathcal{Z}(x_s^-, x_s^-).$$

Taking advantage of the relations of Appendix 7.A.1, one immediately finds:

$$\mathcal{Z}(x_s^+, x_s^+) + \mathcal{Z}(x_s^-, x_s^-) = \mathcal{Z}^{U(2)}_{\text{sub.}} + c.c.,$$

at order $\mathcal{O}(N^{-1})$. The sum of the other two contributions is:

$$\begin{aligned}\mathcal{Z}(x_s^+, x_s^-) + \mathcal{Z}(x_s^-, x_s^+) = \frac{\pi e^{-2N\Re S_1(x_s^+)}}{2^{2(N-1)}N^2}\Bigg\{ &\frac{2\Re S_1''(x_s^+)}{|S_1''(x_s^+)|^3} \\ &+ \frac{1}{N}\Bigg[\frac{(\Re S_1''(x_s^+))^2}{|S_1''(x_s^+)|^5} - \frac{\Re\left((S_1''(x_s^+))^2\left(S_1^{(iv)}(x_s^+)\right)^*\left(5S_1''(x_s^+) + (S_1''(x_s^+))^*\right)\right)}{4|S_1''(x_s^+)|^7} + \\ &+ \frac{5\Re\left((S_1''(x_s^+))^3\left((S_1'''(x_s^+))^*\right)^2\left(7S_1''(x_s^+) + (S_1''(x_s^+))^*\right)\right) - 6|S_1''(x_s^+)|^4|S_1'''(x_s^+)|^2}{12|S_1''(x_s^+)|^9}\Bigg]\Bigg\}.\end{aligned}$$

7.B Non-Abelian theory: the general case

The same procedure applied in Section 7.3.2 for the case of $U(2)$ holds in principle for any $U(n)$ theory, i.e. arbitrary number of colours, as long as n is kept fixed in the large N limit. At finite N, one has the determinantal representation:

$$\mathcal{Z}_N^{U(n)} = N! \det_{1\le j,k\le N}[Z_{jk}] = N!\left(\mathcal{Z}^{U(1)}\right)^n \det_{1\le j,k\le N}\langle W_{j+k-2}\rangle.$$

Here we compute the large N limit of the partition function (7.1.1) of the $U(n)$ theory, and the $1/N$ corrections might be obtained in the same fashion as for the $U(2)$ case. The key observation is that, for every n, the interaction among eigenvalues is sub-leading as $N \to \infty$, and therefore the set of saddle points of the $U(n)$ theory is given by n copies of the set $\mathscr{S}$ of the Abelian theory. Another simplification arises from the observation that, at leading order in $1/N$, the determinant is linearized:

$$\prod_{1\le j<k\le n}\left(2\sinh\frac{x_j - x_k}{2}\right)^2 = \prod_{1\le j<k\le n}\frac{(t_j - t_k)^2}{N} + \mathcal{O}(N^{-2}).$$

Consequently, at large N the partition function $\mathcal{Z}_N^{U(n)}$ converges to:

$$\mathcal{Z}^{U(n)}_{\text{sub.}} = \frac{e^{-nNS_1(x_s^+)}}{2^{nN}N^{\frac{n^2}{2}}}\mathcal{Z}_{\text{GUE}}(S_1''(x_s^+)) = \frac{(2\pi)^{\frac{n}{2}}e^{-nNS_1(x_s^+)}}{2^{nN}N^{\frac{n^2}{2}}(S_1''(x_s^+))^{\frac{n^2}{2}}}G(n+2),$$

when $\lambda\sinh(m) < 1$, where $\mathcal{Z}_{\text{GUE}}(g)$ denotes the partition function of a Gaussian ensemble with coefficient g in the exponent, and $G(n+2) = \prod_{k=0}^{n}(k!)$ is the Barnes G-function. In the super-critical phase, $\mathcal{Z}^{U(n)}_{\text{super.}}$ is a sum over all possible combinations $(\bar{x}_1, \ldots, \bar{x}_n) = (x_s^\pm, \ldots, x_s^\pm)$. It is formally given by:

$$\mathcal{Z}^{U(n)}_{\text{super.}} = \frac{(2\pi)^{\frac{n}{2}}}{2^{nN}N^{\frac{n^2}{2}}}\sum_{(\bar{x}_1,\ldots,\bar{x}_n)\in\mathscr{S}^n}\prod_{j=1}^{n}\frac{e^{-NS_1(\bar{x}_j)}}{(S_1''(\bar{x}_j))^{n-\frac{1}{2}}}P_n\left(S_1''(\bar{x}_1),\ldots,S_1''(\bar{x}_n)\right),$$

with $P_n(s_1,\ldots,s_n)$ a symmetric polynomial of degree $n(n-1)/2$ in n variables, subject to the additional constraint:

$$P_n(s,\ldots,s) = G(n+2)s^{n(n-1)/2}.$$

For example, in the $U(3)$ theory it is:

$$P_3(s_1,s_2,s_3) = 3\left(s_1^2s_2 + s_1^2s_3 + s_1s_2^2 + s_1s_3^2 + s_2^2s_3 + s_2s_3^2 - 2s_1s_2s_3\right),$$

and for $U(4)$ it is:

$$\begin{aligned}P_4(s_1,s_2,s_3,s_4) = 9\big\{&5s_2s_3s_4\left[s_2(s_3-s_4)^2 + s_2^2(s_3+s_4) + s_3s_4(s_3+s_4)\right]\\
&+ s_1^3\left[5s_2^2(s_3+s_4) + 5s_3s_4(s_3+s_4) + s_2(5s_3^2 - 18s_3s_4 + 5s_4^2)\right]\\
&+ s_1^2\left[5s_3(s_3-s_4)^2s_4 + 5s_2^3(s_3+s_4) - 2s_2^2(5s_3^2 - 2s_3s_4 + 5s_4^2) + s_2(s_3+s_4)(5s_3^2 - s_3s_4 + 5s_4^2)\right]\\
&+ s_1\left[5s_3^2s_4^2(s_3+s_4) + s_2^3(5s_3^2 - 18s_3s_4 + 5s_4^2) + s_2^2(s_3+s_4)(5s_3^2 - s_3s_4 + 5s_4^2)\right.\\
&\left.\quad -2s_2s_3s_4(9s_3^2 - 2s_3s_4 + 9s_4^2)\right]\big\}.\end{aligned}$$

The expression may be further simplified, using the fact that every combination $(\bar{x}_1,\ldots,\bar{x}_n)$ with a fixed number l of entries equal to x_s^+, and the remaining $n-l$ equal to x_s^-, give the same contribution, independently on the position the $x_s^\pm$ appear. We obtain:

$$\mathcal{Z}^{U(n)}_{\text{super.}} = \frac{(2\pi)^{\frac{n}{2}}}{2^{nN}N^{\frac{n^2}{2}}}\sum_{l=0}^{n}\frac{e^{-NlS_1(x_s^+)-N(n-l)S_1(x_s^-)}}{(S_1''(x_s^+))^{(n-\frac{1}{2})l}(S_1''(x_s^-))^{(n-\frac{1}{2})(n-l)}}\binom{n}{l}P_n(\underbrace{s,\ldots,s}_{l},\underbrace{s^*,\ldots,s^*}_{n-l}),$$

where for shortness we denoted $s \equiv S_1''(x_s^+)$ and used $S_1''(x_s^-) = S_1''(x_s^+)^* \equiv s^*$ from Appendix 7.A.1.

To find the free energy, we reason as in [102] for the Abelian case. We write:

$$\begin{aligned}\mathcal{Z}^{U(n)}_{\text{super.}} &\propto \sum_{l=0}^{n}\exp\left[-NlS_1(x_s^+) - N(n-l)S_1(x_s^-) + \ldots\right]\\
&= \exp\left[-nN\Re\left(S_1(x_s^+)\right) + \log\left(1 + \sum_{l=0}^{n}\cos\left(lN\Im\left(S_1(x_s^+)\right)\right)\right) + \ldots\right],\end{aligned}$$

where the dots contain sub-leading terms at large N, and arrive at a closed formula for the free energy in the arbitrary $U(n)$ case:

$$\mathcal{F}^{U(n)} = n\mathcal{F}^{U(1)}.$$

This simple relation is indeed the expected one at leading order in N: on the Coulomb branch, the difference between a $U(1)^n$ theory and a $U(n)$ theory is encoded only in the one-loop determinant of the gauge W-bosons, which is sub-leading at large N with fixed n.

7.C General R-charges

The partition function of the $U(1)$ $\mathcal{N}=2$ theory with N chiral multiplets of mass m and N chiral multiplets of mass $-m$ with arbitrary R-charge q, and coupled to a FI background, is [96, 266]:

$$\mathcal{Z}^{U(1)}_{N,q} = \int_{-\infty}^{\infty} \mathrm{d}x \exp\left\{ \mathrm{i}\eta x + N\left[\ell\left(1-q+\frac{\mathrm{i}(x+m)}{2\pi}\right)+\ell\left(1-q-\frac{\mathrm{i}(x+m)}{2\pi}\right)\right] \right.$$
$$\left. +N\left[\ell\left(1-q+\frac{\mathrm{i}(x-m)}{2\pi}\right)+\ell\left(1-q-\frac{\mathrm{i}(x-m)}{2\pi}\right)\right]\right\},$$

where we recall that the theory is put on a three-sphere of radius $1/2\pi$. Here, $e^{\ell(z)}$ is the double sine function, defined as [96]:

$$\ell(u) = -u\log\left(1-e^{\mathrm{i}2\pi u}\right)+\frac{\mathrm{i}\pi}{2}u^2+\frac{\mathrm{i}}{2\pi}\mathrm{Li}_2\left(e^{\mathrm{i}2\pi u}\right)-\frac{\mathrm{i}\pi}{12}, \qquad u\in\mathbb{C}\,.$$

This function has logarithmic singularities when $q\in\mathbb{Z}$, or, more in general, when $\Im\left(\widetilde{m}\right)+2\pi(1-q)\in 2\pi\,\mathbb{Z}$, where $\widetilde{m}$ denotes a complexified mass parameter. Nevertheless, the partition function does not develop singularities, and in fact is holomorphic in $\widetilde{m}$, as the divergences cancel. This can be seen, for instance, from the identity

$$\ell\left(1-q-\frac{\mathrm{i}(x-m)}{2\pi}\right)\overset{\text{reg.}}{=}-\ell\left(1-q+\frac{\mathrm{i}(x-\widetilde{m})}{2\pi}\right), \qquad \widetilde{m}=m-\mathrm{i}4\pi(1-q),$$

where the equality is exact for the infinite product representation of the one-loop determinants and extends to the function ℓ through regularization by ζ-function.

The derivative of the double sine function satisfies the simple property:

$$\frac{\mathrm{d}\ell}{\mathrm{d}u} = -\pi u\cot(\pi u).$$

Therefore, in the large N limit, we arrive at the saddle point equation:

$$\frac{\frac{(x+m)}{2\pi}\sin\left(2\pi(1-q)\right)-\mathrm{i}(1-q)\sinh(x+m)}{\cosh(x+m)-\cos\left(2\pi(1-q)\right)}+\frac{\frac{(x-m)}{2\pi}\sin\left(2\pi(1-q)\right)-\mathrm{i}(1-q)\sinh(x-m)}{\cosh(x-m)-\cos\left(2\pi(1-q)\right)}=\lambda. \tag{7.C.1}$$

It is a simple exercise to see that, setting $q=\frac{1}{2}$, one recovers the saddle points of the $\mathcal{N}=4$ theory. When q is half-integer, $q\in\frac{1}{2}\mathbb{Z}$, the trigonometric functions take simple values and we can solve the saddle point equation exactly, obtaining by simple modification of the results in [102].

- If $q=1$, the action is pure imaginary, already at finite N, and admits no saddle point.
- If $q\in\frac{1}{2}+\mathbb{Z}$, the saddle point equation reduces to:

  $$\frac{\sinh(x)}{\cosh(x)+z}=\frac{\mathrm{i}\lambda}{2(1-q)},$$

  and the large N behaviour is identical to the case $q=\frac{1}{2}$ upon scaling $\lambda\mapsto\frac{\lambda}{2(1-q)}$.
- If $q\in\mathbb{Z}\setminus\{1\}$, the saddle point equation simplifies into:

  $$\frac{\sinh(x)}{\cosh(x)-z}=\frac{\mathrm{i}\lambda}{2(1-q)},$$

  and the phase structure at large N is identical to the case $q=\frac{1}{2}$, up to scaling $\lambda\mapsto\frac{\lambda}{2(1-q)}$ and replacing $z\mapsto -z$ everywhere. The critical line is $\lambda\sinh(m)=2|1-q|$.

We find out that, for $q = 1$, the action admits no saddle point. Here, we study what happens close to that point, for real $q = 1 - \varepsilon$. We assume ε small and approximate the expression at $\mathcal{O}(\varepsilon)$. From (7.C.1) we get:

$$\frac{\sinh(x)}{\cosh(x) - \cosh(m)} + \mathrm{i}\frac{x(\cosh(x)\cosh(m) - 1) + m\sinh(x)\sinh(m)}{(\cosh(x) - \cosh(m))^2} = \frac{\mathrm{i}\lambda}{2\varepsilon}.$$

The equation is still transcendental, but we can find an approximate solution in the large mass limit:

$$\sinh(x) \approx \frac{\lambda e^m}{2m\varepsilon} \implies x \approx \log\left[\frac{\lambda e^m}{2m\varepsilon}\left(1 + \sqrt{1 + \frac{4m^2\varepsilon^2 e^{-2m}}{\lambda^2}}\right)\right] \approx m + \log\frac{\lambda}{m\varepsilon}. \tag{7.C.2}$$

In Figure 7.4 we compare this expression at large m with a numerical solution to the saddle point equation.

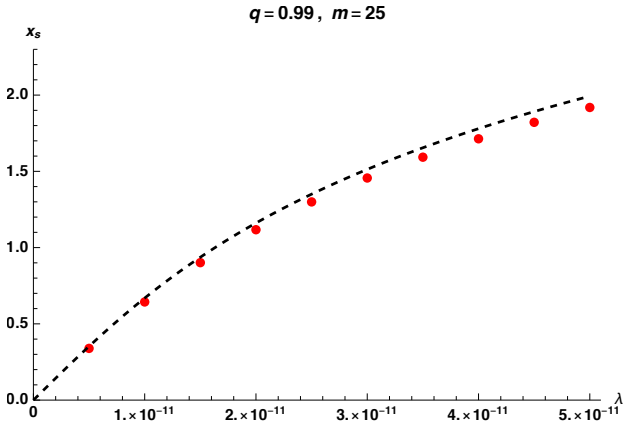


Figure 7.4. Comparison of the numerical solution of the saddle point equation (7.C.1) (red dots) and expression (7.C.2) (black, dashed line), for $q = 0.99$ and $m = 25$.

7.C.1 Comments on squashed geometry

If the supersymmetric $\mathcal{N} = 2$ theory is put on a squashed three-sphere, the partition function is obtained replacing the double sine functions by [281]

$$\ell\left(1 - q + \frac{\mathrm{i}\sigma}{2\pi}\right)\ell\left(1 - q - \frac{\mathrm{i}\sigma}{2\pi}\right) \mapsto \ell_b\left(\frac{1}{2}\left(b + \frac{1}{b}\right)(1 - q) + \frac{\mathrm{i}\sigma}{2\pi}\right)\ell_b\left(\frac{1}{2}\left(b + \frac{1}{b}\right)(1 - q) - \frac{\mathrm{i}\sigma}{2\pi}\right),$$

for $\sigma = x \pm m$, where $b = \sqrt{r_1/r_2}$ is the squashing parameter, and the average radius is $\sqrt{r_1 r_2} = 1/2\pi$. We now take advantage of the remarkable property of the double sine function:

$$\exp\left\{\ell_b\left(\frac{b}{2} + \frac{\mathrm{i}\sigma}{2\pi}\right)\ell_b\left(\frac{b}{2} - \frac{\mathrm{i}\sigma}{2\pi}\right)\right\} = \frac{1}{2\cosh\left(\frac{b\sigma}{2}\right)},$$

which holds for every real non-negative b, and for the round case $b = 1$ provides the partition function (7.2.1). Therefore, for hypermultiplets with R-charge $0 < q < 1$, we may tune the geometry of the manifold so that $(b + b^{-1})(1 - q) = b$, that is we may squash the sphere as

$$b = \sqrt{\frac{1 - q}{q}},$$

and the partition function reads:

$$\mathcal{Z}^{U(1)}_{N,\text{squash}} = \int_{-\infty}^{\infty} \mathrm{d}x \frac{e^{\mathrm{i}\eta x}}{2^N \left[\cosh(bx) + \cosh(bm)\right]^N}.$$

Notice that this procedure would also affect the one-loop determinant of the vector multiplet, but this is irrelevant in the Abelian theory, being such determinant trivial. Also, we see that the symmetry $q \leftrightarrow 1-q$ at the matrix model level is translated into a symmetry $b \leftrightarrow b^{-1}$ in the geometry. We therefore obtain a simple relation between the partition function of the $\mathcal{N}=2$ theory with arbitrary R-charge $0<q<1$ posed on a suitably squashed sphere and the $\mathcal{N}=4$ theory with R-charge $q=\frac{1}{2}$ on the round $\mathbb{S}^3$:

$$\mathcal{Z}^{U(1)}_{N,\text{squashed}}(m,\eta,q) = \frac{1}{b}\mathcal{Z}^{U(1)}_{N,\text{round}}\left(bm, \frac{\eta}{b}, q=\frac{1}{2}\right), \qquad b=\sqrt{\frac{1-q}{q}}.$$

As a byproduct, this equivalence holds for the $U(2)$ theory in the large N approximation. In fact, the squashing would modify:

$$\left(\sinh\frac{x_1-x_2}{2}\right)^2 \mapsto \left(\sinh\frac{b(x_1-x_2)}{2}\right)\left(\sinh\frac{x_1-x_2}{2b}\right),$$

and, as we have seen, the determinant is linearized at first order in $1/N$, producing cancellation of the b-dependence.

Chapter 8

Exact equivalences and phase transitions in unitary matrix models

8.1 Introduction to the chapter

Random matrix theory [39] has developed enormously, especially in the last two decades, attracting attention from researchers in a multitude of different fields [40, 41, 282, 283]. Indeed, one of its most exciting aspects is the inherent interdisciplinary nature of random matrices. Among the myriad of connections with other mathematical and physical areas, we have the important relationship with the study of random partitions, which in turn is deeply linked with problems in statistical mechanics.

In this chapter, we will see that random matrix ensembles which are related in different ways, either via direct mapping or emerging as two random matrix descriptions of the same object in the study of random partitions, have the same analytical evaluations for fixed values of the parameters yet, in some cases, the consideration of the large N limit leads to a different phase structure.

Originally, we started by noting that the analysis of probabilities given by a finite ensemble with weight of the Meixner type, for example as in [284], could equally be studied with a random unitary matrix model, with a distribution of eigenvalues supported on the unit circle. Indeed, we could check, as we shall see, that the corresponding unitary matrix model gives the same results. Furthermore, the equivalence holds for the calculation of correlations near the edge of the eigenvalue density, in a double-scaling limit. However, when studying the more general ensemble with Meixner weight and a hard wall [285] and the corresponding generalized unitary matrix model, while there is again an equivalence for fixed values of the parameter, a different asymptotic behaviour appears, with phase transitions of second order instead of third order. This discrepancy is a consequence of the complexification of the potential of the more general unitary random matrix ensembles, studied in Section 8.3.4 below.

The technical reason behind the apparent paradox lies in the choice of integration contour, as detailed at the end of Section 8.3.6. From this perspective, the results of this chapter could be phrased as the study of phase transitions when the potential is complex but we insist on keeping the eigenvalues on the unit circle. Contour deformations will be systematically analyzed in the second part of the next chapter.

We further elaborate on this whole notion by studying another equivalence, with continuous random matrix ensembles on the real line, via direct mapping.

8.2 Random partitions and unitary matrix models

Let $\lambda = (\lambda_1, \lambda_2, \dots)$, $\lambda_1 \geq \lambda_2 \geq \dots$, be a partition, and $\mathfrak{s}_\lambda(\underline{t})$ the Schur polynomial associated to it [90], evaluated at $\underline{t} = (t_1, t_2, \dots)$. The Schur measure is the probability measure on the space of partitions defined as [286]

$$\mathrm{Prob}_{(\underline{t},\underline{t}')}(\lambda) = \frac{1}{\mathcal{Z}_{\mathrm{Schur}}(\underline{t},\underline{t}')}\mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}'), \tag{8.2.1}$$

with normalization constant the inverse of the partition function

$$\mathcal{Z}_{\mathrm{Schur}}(\underline{t},\underline{t}') = \sum_\lambda \mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}') = \prod_{j,k}\left(1 - t_i t'_j\right)^{-1},$$

where last equality is the Cauchy identity. Here we are denoting $\mathrm{Prob}_{(\underline{t},\underline{t}')}$ the probability taken with respect to a given choice of parameters $(\underline{t},\underline{t}')$. Define the set

$$\mathfrak{S}(\lambda) := \left\{\lambda_j - j + \frac{1}{2},\ j = 1, 2, \dots\right\} \subset \mathbb{Z} + \frac{1}{2},$$

which encodes the shape of the partition λ. For a given subset $\mathfrak{X} \subset \mathbb{Z} + \frac{1}{2}$, the probability

$$\mathrm{Prob}_{(\underline{t},\underline{t}')}(\mathfrak{X}) = \frac{1}{\mathcal{Z}_{\mathrm{Schur}}(\underline{t},\underline{t}')}\sum_{\lambda|\mathfrak{S}(\lambda)\supset\mathfrak{X}}\mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}')$$

that, picking a random partition λ according to the Schur measure we get $\mathfrak{X} \subset \mathfrak{S}(\lambda)$, has a determinantal representation. Such representation is the determinant of a known correlation kernel admitting an integral representation [286] (see also [287, Sec. 4]).

In the present work, we are interested in a special case of Schur measure, which is a particular sub-case of what is known in the literature as z-measure [288, 289]. For any given $n \in \mathbb{N}$ the z-measure is a probability measure on the partitions $\{\lambda\}$ of n (thus with $|\lambda| = n$) depending on two parameters z and z'. A probability measure over all partitions $\{\lambda\}$ (with $|\lambda|$ arbitrary) is obtained passing to the grand canonical ensemble, summing over n. In defining the grand canonical ensemble, every $n \in \mathbb{N}$ is weighted with a negative binomial distribution [288].

We choose the parameters in (8.2.1) to be

$$\underline{t} = (\underbrace{t, \cdots, t}_{N_1 \text{ times}}, 0, \dots), \quad \underline{t}' = (\underbrace{t, \cdots, t}_{N_2 \text{ times}}, 0, \dots), \qquad 0 < t < 1, \tag{8.2.2}$$

so that the Schur measure (8.2.1) becomes a particular instance of z-measure with $z = N_1, z' = N_2 \in \mathbb{N}$, and the parameter of the binomial distribution being t^2. This choice of z, z' is the degenerate case, i.e., non-negative measure, and positive definite when restricted to partitions of length at most N_1. The probabilities

$$\mathrm{Prob}_t(\mathfrak{X}) = \left(1 - t^2\right)^{N_1 N_2}\sum_{\lambda|\mathfrak{S}(\lambda)\supset\mathfrak{X}}\mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}'),$$

where by Prob_t we understand the probability taken with specialization of parameters (8.2.2), admit a determinantal representation in terms of the so-called hypergeometric kernel [288]. The properties of the hypergeometric function were used in [288] to show that, when $z = N_1, z' = N_2$ are integers, the hypergeometric kernel becomes proportional to the Meixner kernel. For a collection of results about the z-measure, see the survey [290], and in particular [290, Prop. 6.1] for the connection with the Meixner ensemble.

In the present work we are interested in the quantities

$$\mathcal{Z}_H(t) = \sum_{\lambda} \mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}') = \sum_{\lambda} (\dim \lambda)^2 \, t^{2|\lambda|} = \left(1-t^2\right)^{-N_1 N_2} \tag{8.2.3}$$

and (the sum is over all λ with $\lambda_1 \leq K$)

$$\mathcal{Z}_E(t) = \sum_{\lambda|\lambda_1 \leq K} \mathfrak{s}_\lambda(\underline{t})\mathfrak{s}_\lambda(\underline{t}') = \sum_{\lambda|\lambda_1 \leq K} (\dim \lambda)^2 \, t^{2|\lambda|}, \tag{8.2.4}$$

with choice of parameter understood to be as in (8.2.2), $|\lambda| = \sum_j \lambda_j$ is the size of the partition λ, corresponding to the total number of boxes in its diagram, and $\dim \lambda$ is the dimension of the irreducible representation of the symmetric group $S_{|\lambda|}$ labelled by λ. The meaning of the subscripts H and E will be clear in a moment. Notice that, due to the property $\mathfrak{s}_\lambda(\underline{t}) = 0$ if length$(\lambda) >$ length$(\underline{t})$, the sums are effectively truncated to partitions of length at most N_1. Furthermore, the ratio

$$\frac{\mathcal{Z}_E(t)}{\mathcal{Z}_H(t)} = \mathrm{Prob}_t\left(\lambda_1 \leq K\right)$$

is the probability that, picking a random partition λ with probability distribution as described above, the length of its rows is at most K.

The z-measure induces a determinantal point process on $\mathbb{Z}+\frac{1}{2}$, thus the correlation functions have determinantal form

$$\mathrm{Prob}_t\left(\mathfrak{X} \subset \mathfrak{S}(\lambda)\right) = \det\left(\mathcal{K}\right)_{\mathfrak{X}},$$

for $\mathfrak{X} \subset \mathbb{Z}+\frac{1}{2}$, where $\mathcal{K}$ is the operator whose kernel, the function $\mathcal{K}(x,y)$ on $\left(\mathbb{Z}+\frac{1}{2}\right)\times\left(\mathbb{Z}+\frac{1}{2}\right)$, is the hypergeometric kernel. Therefore we have a Fredholm determinant representation of $\mathcal{Z}_E(t)$:

$$\begin{aligned}
\mathcal{Z}_E(t) &= (1-t^2)^{-N_1N_2}\mathrm{Prob}_t\left(\lambda_1 \leq K\right) \\
&= (1-t^2)^{-N_1N_2}\mathrm{Prob}_t\left(\left\{x \in \mathbb{Z}+\frac{1}{2} \mid x \leq K-\frac{1}{2}\right\} \cap \mathfrak{S}(\lambda) = \emptyset\right) \\
&= (1-t^2)^{-N_1N_2}\det\left(\mathbf{1}-\mathcal{K}\right)_{\mathbb{Z}_{\geq K}+\frac{1}{2}},
\end{aligned}$$

where $\mathbf{1}$ in the last line is the identity operator. Using the hook-length formula

$$\dim \lambda = \prod_{j<k} \frac{\lambda_j - \lambda_k - j + k}{k-j}$$

for the dimension in (8.2.4) and changing variables $h_j = \lambda_j - j + N$, we arrive at the expression

$$\begin{aligned}
\mathcal{Z}_E(t) = &\frac{t^{-N_1(N_1-1)}\left(\Gamma(N_2-N_1+1)\right)^{N_1}}{N_1!G(N_1+1)G(N_2+1)} \\
&\times \sum_{h_1=0}^{N_1+K-1} \cdots \sum_{h_{N_1}=0}^{N_1+K-1} \prod_{1\leq j<k\leq N_1} (h_j-h_k)^2 \prod_{j=1}^{N_1} \binom{N_2-N_1+h_j}{h_j} t^{2h_j},
\end{aligned} \tag{8.2.5}$$

where we also used the symmetry in the h_j variables to remove the restriction to the Weyl chamber $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{N_1}$. $G(\cdot)$ is the Barnes G-function [291], which, for integer values of the argument is $G(n) = \prod_{j=0}^{n-2} j!$. The expression (8.2.5) is a Meixner ensemble with summation restricted to $0 \leq h_j \leq N_1+K-1$ and coincides with the partition function of the dimer model studied in [284, 285],[20] and has also appeared in [293]. In the Coulomb gas picture, the restriction in the

[20] The dimer model is associated with the random tiling of an Aztec diamond with a square [284] or a rectangle [285] cut off. See e.g. [292] for an overview on dimers, random tilings and random matrices.

summation range corresponds to a hard wall for the charges (i.e. an infinite barrier) placed at $N_1 + K - 1$. We assumed $N_1 \leq N_2$, but a completely analogous expression can be easily obtained in the converse case.

Let us introduce the generating functions of, respectively, the complete homogeneous polynomials $\{\mathfrak{h}_k\}$ and the elementary symmetric polynomials $\{\mathfrak{e}_k\}$, specialized at $\underline{t}$:

$$H(z;\underline{t}) = \sum_{k=0}^{\infty} \mathfrak{h}_k(\underline{t}) z^k = \prod_k (1 - t_k z)^{-1} ,$$
$$E(z;\underline{t}) = \sum_{k=0}^{\infty} \mathfrak{e}_k(\underline{t}) z^k = \prod_k (1 + t_k z) .$$

The partition functions (8.2.3)-(8.2.4) admit a determinantal representation in terms of determinants of $K \times K$ Toeplitz matrices, with symbol, respectively

$$\sigma_H(z,t) = H(z;\underline{t}) H(z^{-1};\underline{t}') = (1 - tz)^{-N_1} \left(1 - tz^{-1}\right)^{-N_2} ,$$
$$\sigma_E(z,t) = E(z;\underline{t}) E(z^{-1};\underline{t}') = (1 + tz)^{N_1} \left(1 + tz^{-1}\right)^{N_2} .$$

See for example [294, 295] for the explicit derivation (see also [283] for an extensive account of Toeplitz determinants and their properties). In turn, using Andréief's identity [296, 297] we have that these Toeplitz determinants admit a representation as unitary matrix integrals

$$\mathcal{Z}_H(t) = \frac{1}{K!} \int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K \varphi}{(2\pi)^K} \prod_{1 \leq j < k \leq K} |e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}| \prod_{j=1}^{K} \left(1 - te^{\mathrm{i}\varphi_j}\right)^{-N_1} \left(1 - te^{-\mathrm{i}\varphi_j}\right)^{-N_2} , \tag{8.2.6}$$

$$\mathcal{Z}_E(t) = \frac{1}{K!} \int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K \varphi}{(2\pi)^K} \prod_{1 \leq j < k \leq K} |e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}| \prod_{j=1}^{K} \left(1 + te^{\mathrm{i}\varphi_j}\right)^{N_1} \left(1 + te^{-\mathrm{i}\varphi_j}\right)^{N_2} . \tag{8.2.7}$$

We will refer to (8.2.6) and (8.2.7) as the H-model and the E-model, respectively. Therefore we have two equivalent matrix model descriptions of the quantity $\mathcal{Z}_E(t)$, or, equivalently, of the probability $\mathrm{Prob}_t(\lambda_1 \leq K)$: either as a discrete matrix model on the bounded subset $\{0, 1, \ldots, N_1 + K - 1\} \subset \mathbb{Z}$, or as a continuous matrix model on the unit circle.

We stress that the equivalence between these two representations does not rely on a direct map, but rather on a two-step procedure relating the two matrix model formulations of $\mathcal{Z}_E(t)$ to the same Toeplitz determinant. As a consequence, the quantity $(N_1 - N_2)$, which has the meaning of a deformation parameter, plays different roles in the two pictures. This will be reflected in the mismatch of the phase structure when $N_1, N_2 \to \infty$ with $N_1 - N_2 \neq 0$.

8.3 Random matrix ensembles on the unit circle

In the present section, we consider the asymptotic behaviour of the unitary matrix models defined in (8.2.6) and (8.2.7), when the rank K is large and N_1, N_2 scale with K. Before that, we comment on the already known aspects of the exact solvability of some of the models above.

8.3.1 Exact evaluation

Following [298], the authors of [285] gave an explicit evaluation of the discrete matrix model (8.2.5) at the limit value $t = 1$. The exact formula of [285, Prop. 3.1] was obtained thanks to the fact that at the limit value $t = 1$ the Meixner ensemble with a hard wall becomes a Hahn ensemble.

In [285]	s	r	q	α	R	Q
In [298]	N	$N+t$	$M-N=[\gamma N]-N$	q	$\omega+1$	$\gamma-1$
Here	N_1	N_1+K	N_2-N_1	t^2	$1+\gamma^{-1}$	$2v/(1-v)$

Table 8.1: Dictionary between the notation in [285], in [298] and the present work.

On the unitary matrix model side, at $t=1$ (or more generally $|t|=1$) the symbol σ_E develops a Fisher-Hartwig singularity, and it can be evaluated exactly thanks to a formula by Böttcher and Silbermann [299] (see also [300] for another proof):

$$\mathcal{Z}_E(t) = \frac{G(N_1+1)G(N_2+1)G(N_1+N_2+K+1)}{G(N_1+N_2+1)G(N_1+K+1)G(N_2+K+1)} G(K+1). \tag{8.3.1}$$

This provides an exact check of the equivalence. We will provide a third independent derivation of this result later in Section 8.4.1. For ease of the reader, we report in table 8.1 the correspondence between the notation of [284, 285], that in [298] and ours.

Note that, if we choose the symmetric model $N_1 = N_2 \equiv N$ and modify the symbol by inserting a monomial factor z^s, with $s \in \mathbb{Z}$, the exact evaluation of the corresponding matrix model at $t=1$ is again given by the formula of [299, 300], thanks to the simple identity

$$z^s \sigma_E(z, t=1) = z^s \left(1+z\right)^{\beta} \left(1+z^{-1}\right)^{\beta} = \left(1+z\right)^{\beta+s} \left(1+z^{-1}\right)^{\beta-s}, \tag{8.3.2}$$

when $z = e^{\mathrm{i}\varphi}$, $-\pi \le \varphi \le \pi$. The effect of the monomial insertion is to shift the Fourier coefficients of the symbol $\sigma_E(z;1)$ by s: the k^{th} Fourier coefficient of the symbol $z^s\sigma_E(z;1)$ is the $(k+s)^{\text{th}}$ coefficient of $\sigma_E(z;1)$. Therefore, we can evaluate exactly the partition function of the unitary ensemble with weight (8.3.2), and it is again given by formula (8.3.1), with $N_1 = N+s, N_2 = N-s$.

It is worth mentioning that there are a wealth of analytical results on Toeplitz banded matrices, whose determinant is given by the E-models above [301].

We just quote here that for (8.2.7) with $N_1 = N_2 = 1$ we have the determinant of a symmetric tridiagonal Toeplitz matrix, known to be equal to a Chebyshev polynomial of the second type:

$$\mathcal{Z}_E(t)|_{N_1=N_2=1} = t^K U_K\left(\frac{1+t^2}{2t}\right) = \sum_{n=0}^{K} t^{2n}.$$

8.3.2 Unitary matrix models

Comparing the definition (8.2.4) of the E-model with (8.2.3), one sees that its $K \to \infty$ limit, with fixed N_1, N_2 coincides with the H-model:

$$\lim_{K\to\infty} \mathcal{Z}_E(t) = \mathcal{Z}_H(t).$$

In the scaling limit with N_1, N_2 growing together with K, it was found in [284, 285] that the presence of hard walls in the discrete matrix model (8.2.5) triggers a third order phase transition. For the case $N_1 = N_2$, we will prove the phase transition from the point of view of the unitary matrix model. This is done in Section 8.3.3 and reproduces an early result of Baik [302]. For the general case $N_1 \neq N_2$, however, the potential of the unitary matrix model is complex-valued, and the large K asymptotic becomes more involved. This topic is analyzed in Section 8.3.4.

As the asymptotic behaviour at large K does not depend on N_1, N_2 being integers, we consider the matrix models arising from Toeplitz determinants with more general symbols

$$\sigma_H(z;t) = \left(1-tz\right)^{-\beta_1} \left(1-tz^{-1}\right)^{-\beta_2},$$
$$\sigma_E(z;t) = \left(1+tz\right)^{\beta_1} \left(1+tz^{-1}\right)^{\beta_2},$$

and, without loss of generality, we assume $0 \leq \beta_1 \leq \beta_2$. We also introduce the notation

$$\beta_1 = \beta(1-v), \quad \beta_2 = \beta(1+v), \tag{8.3.3}$$

where $\beta = (\beta_1 + \beta_2)/2$ is the average power and $0 \leq v \leq 1$ measures the asymmetry in $z \leftrightarrow z^{-1}$. We then take the $K \to \infty$ limit with[21]

$$\gamma := \frac{\beta}{K} \quad \text{fixed.} \tag{8.3.4}$$

As customary, when studying the large rank behaviour of matrix models, we introduce the density of eigenvalues

$$\rho(\varphi) = \frac{1}{K}\sum_{j=1}^{K} \delta(\varphi - \varphi_j), \tag{8.3.5}$$

which at large K becomes a continuous function of φ, with compact support and normalized so that

$$\int_{-\pi}^{\pi} \mathrm{d}\varphi \rho(\varphi) = 1.$$

In each of the cases considered below, we will find the eigenvalue density ρ and use it to evaluate the free energy in this limit, defined as:

$$\mathcal{F} := -\frac{1}{K^2}\log \mathcal{Z}.$$

Our calculations are based on standard saddle point techniques, and we omit them from the main text and refer to the Appendices 8.A and 8.B. We will show that all the models undergo a phase transition when a gap opens in the support of the eigenvalue density, as schematized in Figure 8.1.

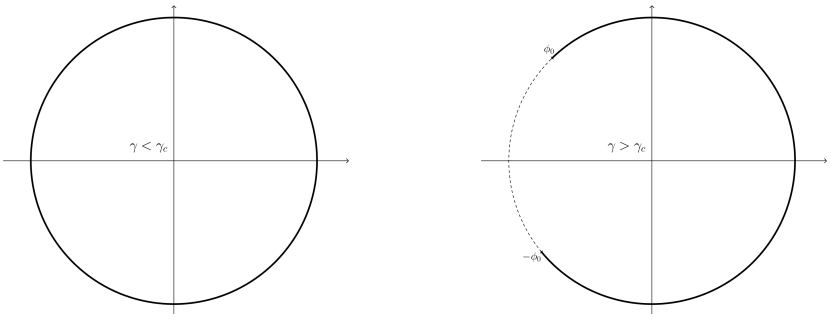


Figure 8.1. Increasing the coupling γ, the support of the eigenvalue density ρ develops a gap, signalling a phase transition.

8.3.3 Phase transition: symmetric case

We first focus on the symmetric case $\beta_1 = \beta_2 \equiv \beta$, while the analysis of the more general case $\beta_1 \neq \beta_2$ is undertaken later in Section 8.3.4.

[21] This is the inverse of the usual 't Hooft coupling as defined in gauge theories, but here we adopt to the notation of [183, 302]. Consistently, "weak" and "strong" coupling will refer to the values of γ.

Hence, the matrix models we analyze are:

$$\mathcal{Z}^{\text{sym.}}_{\text{u},H} = \frac{1}{K!}\int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K\varphi}{(2\pi)^K} \prod_{1\le j<k\le K} \left|e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}\right|^2 \prod_{j=1}^{K} \left[(1-te^{\mathrm{i}\varphi_j})(1-te^{-\mathrm{i}\varphi_j})\right]^{-\beta}, \tag{8.3.6}$$

$$\mathcal{Z}^{\text{sym.}}_{\text{u},E} = \frac{1}{K!}\int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K\varphi}{(2\pi)^K} \prod_{1\le j<k\le K} \left|e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}\right|^2 \prod_{j=1}^{K} \left[(1+te^{\mathrm{i}\varphi_j})(1+te^{-\mathrm{i}\varphi_j})\right]^{\beta}, \tag{8.3.7}$$

and we recall that $\mathcal{Z}^{\text{sym.}}_{\text{u},H}$ admits an exact solution through the Cauchy identity, whilst $\mathcal{Z}^{\text{sym.}}_{\text{u},E}$ does not. Nevertheless we have $\lim_{K\to\infty} \mathcal{Z}^{\text{sym.}}_{\text{u},E} = \mathcal{Z}^{\text{sym.}}_{\text{u},H}$.

In [302], Baik proved that the first system, described by the partition function (8.3.6), and which we will call for simplicity the H-model, undergoes a phase transition at large K. We prove that the second system (8.3.7), which we call E-model, undergoes the same phase transition. We prove it solving a singular integral equation in Appendix 8.A, but in fact the result may also be directly obtained from [302], with minor changes.

The H-model

Consider the matrix integral in (8.3.6), and take the limit $K \to \infty$ with β scaling as in (8.3.4). The leading contribution to $\mathcal{Z}^{\text{sym.}}_{\text{u},H}$ comes from the solution to the system of saddle point equations that, with the help of the eigenvalue density ρ_H as defined in (8.3.5), can be rewritten as a single singular integral equation:

$$-\mathrm{i}\gamma t\left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right] = \mathrm{P}\int \mathrm{d}\vartheta \rho_H(\vartheta)\cot\left(\frac{\varphi-\vartheta}{2}\right), \tag{8.3.8}$$

where the symbol $\mathrm{P}\int$ means principal value of the integral, and ρ_H is the eigenvalue density for the specific model considered presently. The details of the solution to (8.3.8) are spelled in Appendices 8.A.1 and 8.A.2. Two phases exists, separated by the critical curve [302]

$$\gamma = \frac{1+t}{2t} =: \gamma_{c,H}(t).$$

The eigenvalue density, plotted in Figure 8.2 for various t and γ, reads:

$$\rho_H(\varphi) = \begin{cases} \frac{1}{2\pi}\left[1 + 2\gamma t\left(\frac{\cos\varphi - t}{(1-t)^2+4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\right], & \gamma \le \gamma_{c,H}(t), \\ \frac{2(\gamma-1)t}{\pi}\left(\frac{\cos\frac{\varphi}{2}}{(1-t)^2+4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2}, & \gamma > \gamma_{c,H}(t) \end{cases} \tag{8.3.9}$$

and allows to evaluate the free energy $\mathcal{F}^{\text{sym.}}_{\text{u},H}$, obtaining:

$$\mathcal{F}^{\text{sym.}}_{\text{u},H} = \begin{cases} -\gamma^2\log(1-t^2), & \gamma \le \gamma_{c,H}(t), \\ -(2\gamma-1)\log(1-t) - \frac{1}{2}\log t + \mathcal{C}_H(\gamma), & \gamma > \gamma_{c,H}(t), \end{cases} \tag{8.3.10}$$

where $\mathcal{C}_H(\gamma)$ is t-independent. See Appendices 8.B.1 and 8.B.2 for the calculation of the free energy. From the latter expression one sees that $\frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t}$ and $\frac{\mathrm{d}^2\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t^2}$ are continuous functions for all values of γ, while

$$\lim_{\gamma\uparrow\gamma_{c,H}} \frac{\mathrm{d}^3\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t^3} - \lim_{\gamma\downarrow\gamma_{c,H}} \frac{\mathrm{d}^3\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t^3} = \frac{1}{t^3(1-t^2)},$$

and therefore the system undergoes a third order phase transition at the critical curve $\gamma_{c,H}(t) = \frac{1+t}{2t}$.

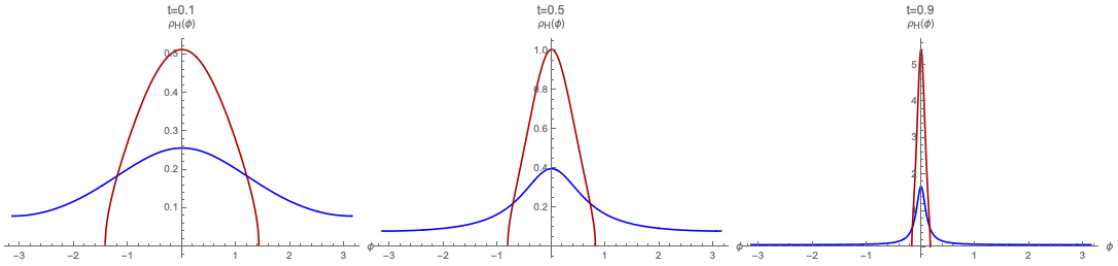


Figure 8.2. Eigenvalue density $\rho_H(\phi)$. The blue curve is at $\gamma = \frac{1}{2}\gamma_{c,H}(t)$ and the red curve is at $\gamma = 2\gamma_{c,H}(t)$, for $t = 0.1$ (left), 0.5 (center), 0.9 (right).

The E-model

We now turn to the second matrix model, defined in (8.3.7). The leading contribution in the large K limit, with scaling (8.3.4), is obtained solving the saddle point equation

$$-\mathrm{i}\gamma t\left[\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}}-\frac{e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}}\right]=\mathrm{P}\int \mathrm{d}\vartheta\rho_E(\vartheta)\cot\left(\frac{\varphi-\vartheta}{2}\right). \tag{8.3.11}$$

We solve this singular integral equation in Appendices 8.A.3 and 8.A.4. From direct comparison of the integral representation of $\mathcal{Z}^{\mathrm{sym.}}_{\mathrm{u},H}$ and $\mathcal{Z}^{\mathrm{sym.}}_{\mathrm{u},E}$ in (8.3.6) and (8.3.7), one would expect that, the solution to the second model is related to the solution to the first model by

$$(\gamma,t)\mapsto(-\gamma,-t).$$

Direct calculations prove that this is true and, in particular, the system undergoes a phase transition along the critical curve

$$\gamma=\frac{1-t}{2t}=:\gamma_{c,E}(t).$$

The eigenvalue density in the E-model, plotted in Figure 8.3 for different values of t and γ, is:

$$\rho_E(\varphi)=\begin{cases}\frac{1}{2\pi}\left[1+2\gamma t\left(\frac{\cos\varphi+t}{(1+t)^2-4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\right], & \gamma\le\gamma_{c,E}(t),\\ \frac{2(\gamma+1)t}{\pi}\left(\frac{\cos\frac{\varphi}{2}}{(1+t)^2-4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2-\left(\sin\frac{\varphi}{2}\right)^2}, & \gamma>\gamma_{c,E}(t),\end{cases}$$

which allows to compute the free energy $\mathcal{F}^{\mathrm{sym.}}_{\mathrm{u},E}$ at large K (see Appendices 8.B.1 and 8.B.2), and the final result is:

$$\mathcal{F}^{\mathrm{sym.}}_{\mathrm{u},E}=\begin{cases}-\gamma^2\log(1-t^2), & \gamma\le\gamma_{c,E}(t),\\ (2\gamma+1)\log(1+t)-\frac{1}{2}\log t+\mathcal{C}_E(\gamma), & \gamma>\gamma_{c,E}(t),\end{cases} \tag{8.3.12}$$

where $\mathcal{C}_E(\gamma)$ is t-independent. Taking derivatives, one finds again that first and second derivatives are continuous functions, while

$$\lim_{\gamma\uparrow\gamma_{c,E}}\frac{\mathrm{d}^3\mathcal{F}^{\mathrm{sym.}}_{\mathrm{u},E}}{\mathrm{d}t^3}-\lim_{\gamma\downarrow\gamma_{c,E}}\frac{\mathrm{d}^3\mathcal{F}^{\mathrm{sym.}}_{\mathrm{u},E}}{\mathrm{d}t^3}=\frac{1}{t^3(1-t^2)},$$

thus the phase transition is of third order also in this case.

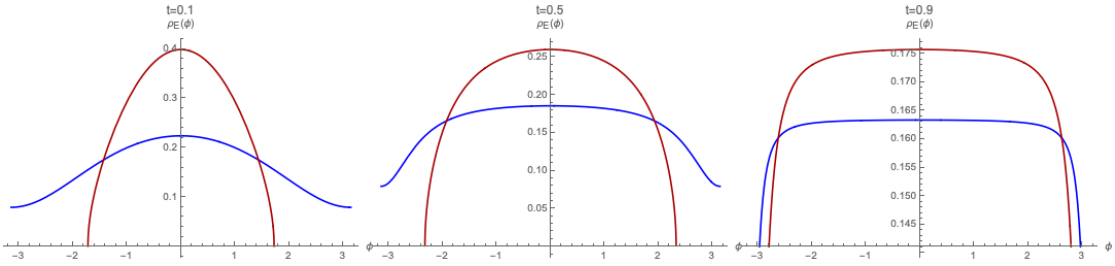


Figure 8.3. Eigenvalue density $\rho_E(\phi)$. The blue curve is at $\gamma = \frac{1}{2}\gamma_{c,E}(t)$ and the red curve is at $\gamma = 2\gamma_{c,E}(t)$, for $t = 0.1$ (left), 0.5 (center), 0.9 (right).

Double-scaling limit

We have shown that the free energy, and consequently the phase structure and the critical curve, of the matrix model with symmetric weight of E type at large K and large N agrees with the equivalent descriptions as a Meixner ensemble with a hard wall studied in [284]. In fact, we find an even stronger agreement between the two pictures, as also the correlations among eigenvalues near the edge of the distributions in a double-scaling limit match. For the unitary ensemble, the double-scaling limit is described in [302], along the lines of [183]. On the other hand, in the discrete ensemble, if the critical region is approached from the small coupling phase, the hard wall is not active and the results of [298] directly apply. The double-scaling is the same in both cases, and leads to the Tracy–Widom law [303]. A third consistency check comes from the asymptotics of the hypergeometric kernel: standard computations using a steepest descend method (see for example [287, Sec. 4] for an introduction) show that the result in the double-scaling limit matches the Tracy–Widom behaviour.

8.3.4 Phase transition: general case

While in the previous Section 8.3.3 we focused on the analysis of two matrix models that are symmetric under $z \leftrightarrow z^{-1}$, we now allow the more general situation $\beta_1 \neq \beta_2$, in which the symmetry is lost.

The matrix models we consider here are

$$\mathcal{Z}^{\text{gen.}}_{\text{u},H}(t) = \frac{1}{K!}\int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K\varphi}{(2\pi)^K} \prod_{1\le j<k\le K} |e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}| \prod_{j=1}^{K} \left(1 - te^{\mathrm{i}\varphi_j}\right)^{-\beta_1} \left(1 - te^{-\mathrm{i}\varphi_j}\right)^{-\beta_2}, \tag{8.3.13}$$

$$\mathcal{Z}^{\text{gen.}}_{\text{u},E}(t) = \frac{1}{K!}\int_{[-\pi,\pi]^K} \frac{\mathrm{d}^K\varphi}{(2\pi)^K} \prod_{1\le j<k\le K} |e^{\mathrm{i}\varphi_j} - e^{\mathrm{i}\varphi_k}| \prod_{j=1}^{K} \left(1 + te^{\mathrm{i}\varphi_j}\right)^{\beta_1} \left(1 + te^{-\mathrm{i}\varphi_j}\right)^{\beta_2}. \tag{8.3.14}$$

The weight functions are complex valued, thus we expect the eigenvalue densities at large K to be complex-valued functions. We rewrite (β_1, β_2) in terms of the parameters (β, v) defined in (8.3.3), and consider the limit $K \to \infty$ with scaling of β as introduced in (8.3.4).

The H-model

We first focus on the behaviour of the partition function (8.3.13) in the limit described above, and show how the discussion of the symmetric case in Section 8.3.3 is modified. Details of the calculations can be retrieved in Appendices 8.A.7 and 8.A.8 for weak and strong coupling respectively.

The leading contributions at large K may be encoded in the eigenvalue density ρ_H which solves

the integral equation

$$-\mathrm{i}\gamma t\left[\frac{(1-v)e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}}-\frac{(1+v)e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right]=\mathrm{P}\int \mathrm{d}\vartheta\rho_H(\vartheta)\cot\left(\frac{\varphi-\vartheta}{2}\right). \tag{8.3.15}$$

The asymmetry parameter v complexifies the left hand side of the latter equation, and the resulting eigenvalue density is complex (see Appendices 8.A.7 and 8.A.8):

$$\rho_H(\varphi)=\begin{cases}\frac{1}{2\pi}\left\{1+2\gamma t\left(\frac{\cos\varphi-t-\mathrm{i}v\sin(\varphi)}{(1-t)^2+4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\right\}, & \gamma\leq\gamma_{c,H},\\ \frac{2t(\gamma-1)}{\pi(1-t)}\left[\frac{(1-t)\cos\left(\frac{\varphi}{2}\right)-\mathrm{i}v(1+t)\sin\left(\frac{\varphi}{2}\right)}{(1-t)^2+4t\left(\sin\left(\frac{\varphi}{2}\right)\right)^2}\right]\sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2-\left(\sin\left(\frac{\varphi}{2}\right)\right)^2}, & \gamma>\gamma_{c,H},\end{cases}$$

with critical value $\gamma_{c,H}=\frac{1+t}{2t}$, the same as in the symmetric case. The corresponding free energy $\mathcal{F}^{\text{gen.}}_{\text{u},H}=-K^{-2}\log\mathcal{Z}^{\text{gen.}}_{\text{u},H}$, computed in Appendix 8.B.3, can be written as:

$$\mathcal{F}^{\text{gen.}}_{\text{u},H}=\mathcal{F}^{\text{sym.}}_{\text{u},H}+v^2\Delta\mathcal{F}_H,\qquad \Delta\mathcal{F}_H=\begin{cases}-\gamma^2\log(1-t^2), & \gamma\leq\gamma_{c,H}(t),\\ \log(1+t)-\frac{1}{2}\log t, & \gamma>\gamma_{c,H}(t).\end{cases} \tag{8.3.16}$$

The introduction of the asymmetry reduces the order of the phase transition from third to second, with:

$$\lim_{\gamma\uparrow\gamma_{c,H}}\frac{\mathrm{d}^2\mathcal{F}^{\text{gen.}}_{\text{u},H}}{\mathrm{d}t^2}-\lim_{\gamma\downarrow\gamma_{c,H}}\frac{\mathrm{d}^2\mathcal{F}^{\text{gen.}}_{\text{u},H}}{\mathrm{d}t^2}=-\frac{v^2}{t^2(1-t)},$$

The E-model

Consider now the second matrix model, the E-model (8.3.14), and take its scaled limit as in (8.3.3)-(8.3.4). The saddle point equation reads

$$-\mathrm{i}\gamma t\left((1-v)\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}}-\frac{(1+v)e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}}\right)=\mathrm{P}\int \mathrm{d}\vartheta\rho_E(\vartheta)\cot\frac{\varphi-\vartheta}{2}. \tag{8.3.17}$$

The eigenvalue density is complexified by the presence of the asymmetry parameter v (see Appendices 8.A.9 and 8.A.10):

$$\rho_E(\varphi)=\begin{cases}\frac{1}{2\pi}\left[1+2\gamma t\left(\frac{\cos(\varphi)+t-\mathrm{i}v\sin(\varphi)}{1+t^2+2t\cos(\varphi)}\right)\right], & \gamma\leq\gamma_{c,E},\\ \frac{2t(\gamma+1)}{\pi(1+t)}\left[\frac{(1+t)\cos\left(\frac{\varphi}{2}\right)-\mathrm{i}v(1-t)\sin\left(\frac{\varphi}{2}\right)}{(1+t)^2-4t\left(\sin\left(\frac{\varphi}{2}\right)\right)^2}\right]\sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2-\left(\sin\left(\frac{\varphi}{2}\right)\right)^2}, & \gamma>\gamma_{c,E},\end{cases}$$

with the same critical value as for the symmetric case: $\gamma_{c,E}=\frac{1-t}{2t}$. The free energy is (see Appendix 8.B.3):

$$\mathcal{F}^{\text{gen.}}_{\text{u},E}=\mathcal{F}^{\text{sym.}}_{\text{u},E}+v^2\Delta\mathcal{F}_E,\qquad \Delta\mathcal{F}_E=\begin{cases}-\gamma^2\log(1-t^2), & \gamma\leq\gamma_{c,E}(t),\\ \log(1-t)-\frac{1}{2}\log t, & \gamma>\gamma_{c,E}(t).\end{cases} \tag{8.3.18}$$

The first derivative of the free energy is continuous, but the second derivative is not:

$$\lim_{\gamma\uparrow\gamma_{c,E}}\frac{\mathrm{d}^2\mathcal{F}^{\text{sym.}}_{\text{u},E}}{\mathrm{d}t^2}-\lim_{\gamma\downarrow\gamma_{c,E}}\frac{\mathrm{d}^2\mathcal{F}^{\text{sym.}}_{\text{u},E}}{\mathrm{d}t^2}=-\frac{v^2}{t^2(1+t)},$$

thus the phase transition is second order. We plot the first derivative of free energy as a function of γ, at different values of t, in Figure 8.4.

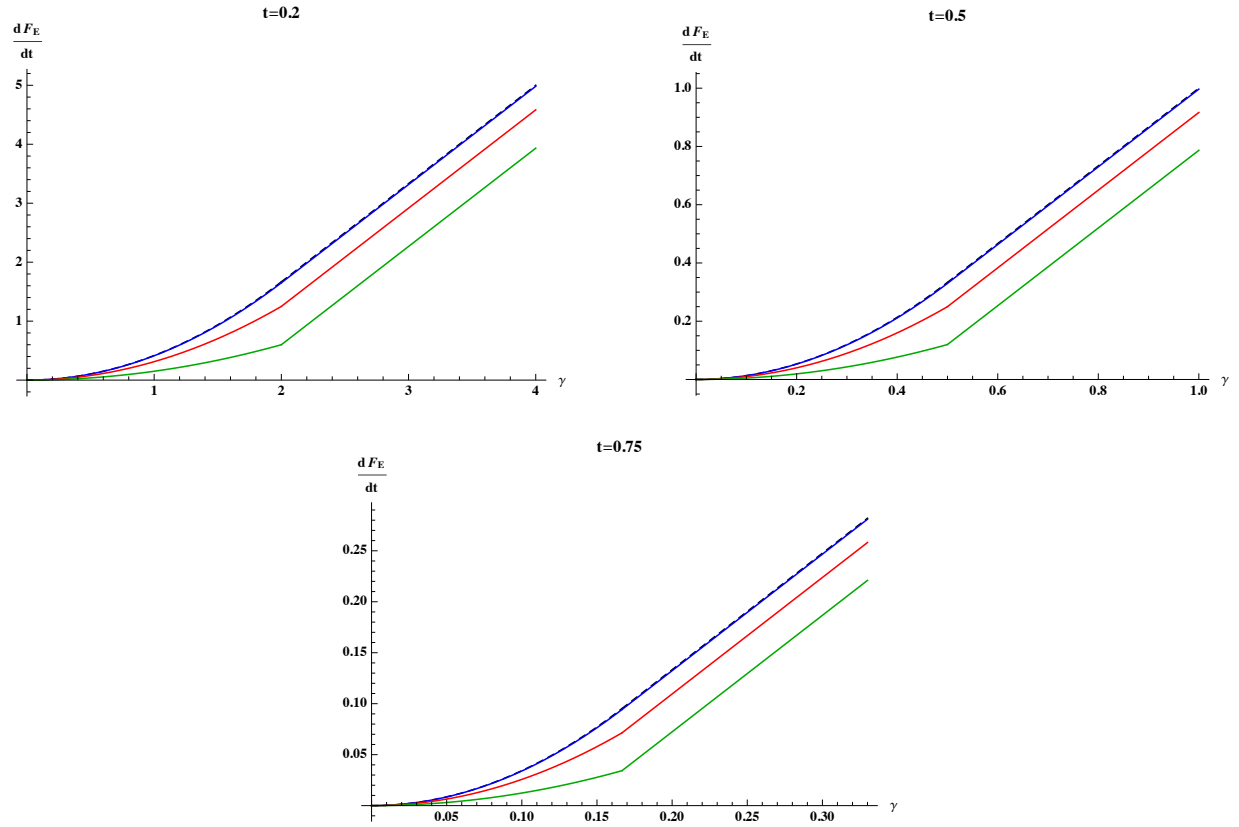


Figure 8.4. Planar limit of $\frac{\mathrm{d}\mathcal{F}_E}{\mathrm{d}t}$ as a function of γ, for $t = 0.2$ (up left), $t = 0.5$ (up right), $t = 0.75$ (down). In each plot, there appear $\frac{\mathrm{d}\mathcal{F}_E}{\mathrm{d}t}$ for different values of the asymmetry parameter: $v = 0.1$ (blue), $v = 0.5$ (red), and $v = 0.75$ (green). The dashed back line is the symmetric case $v = 0$. The point at which the curve is continuous but with discontinuous derivative becomes more and more visible as v is increased.

8.3.5 The Gross–Witten–Wadia limit

Consider the two matrix models of Section 8.3.3, defined in equations (8.3.6) and (8.3.7). The potential of those models can be written as:

$$V_{E/H}^{\text{sym.}}(z) = \pm\beta \left[\log\left(1 \pm tz\right) + \log\left(1 \pm tz^{-1}\right)\right],$$

with $+$ sign for the E and $-$ for the H. We send $t \to 0$ and $\beta \to \infty$, keeping their product $\beta_{\text{GWW}} := t\beta$ fixed. This gives:

$$V_{E/H}^{\text{sym.}}(z) \to \beta_{\text{GWW}} \left(z + z^{-1}\right),$$

which is the potential of the Gross–Witten–Wadia (GWW) matrix model [190, 192]. Note that the limit is the same for both the E- and the H-model. In the more general, non-symmetric case, with potential

$$V_{E/H}^{\text{gen.}}(z) = \pm\left[\beta_1 \log\left(1 \pm tz\right) + \beta_2 \log\left(1 \pm tz^{-1}\right)\right],$$

we pass from (β_1, β_2) to (β, v) as prescribed in (8.3.3), and define $\beta_{\text{GWW}} := t\beta$. The same limit as above gives:

$$V_{E/H}^{\text{gen.}}(z) \to \beta_{\text{GWW}} \left[\left(z + z^{-1}\right) - v\left(z - z^{-1}\right)\right],$$

which is the potential of the Gross–Witten–Wadia model with a topological θ-term [304, 305]. Since, by construction, we are in the regime $|v| < 1$, the GWW phase transition is still present, as discussed in [305].

Within the setting laid down in Section 8.2, the limit above with $\beta = N \in \mathbb{N}$ of the normalization of the z-measure leads to the normalization of the Poissonized Plancherel measure on partitions [306, 290]. This is consistent with what we mentioned above, as the Poissonized Plancherel measure is mapped to the GWW unitary matrix model [307]. On the other side, the Poissonized Plancherel measure is the $t \to 1$ limit of the Meixner ensemble [308] as well.

8.3.6 Phase transition: from third to second order

For $0 < |v| < 1$, the potential of the unitary matrix models (8.3.13)-(8.3.14) becomes complex-valued. We have studied the large K limit and showed that the model undergoes a second order phase transition, which becomes third order turning off the asymmetry parameter, $v \to 0$. Nevertheless, the phase transition in the GWW model remains third order if the potential is modified into

$$\cos\varphi - \mathrm{i}v\sin\varphi.$$

An alternative approach to face the matrix models with complexified potential is to analytically continue them in the sense of [309]. We relax the condition $|z|^2 = 1$ and deform the integration cycle $\mathbb{S}^1 \rightsquigarrow \mathcal{C}_v$ in the complex plane so that the potential remains real along $\mathcal{C}_v$ (see Figure 8.5). Of course, for the deformation to be smooth, $\mathcal{C}_v$ should be a smooth Jordan curve, whose shape depends on v and which is the unit circle $\mathbb{S}^1$ when $v = 0$.

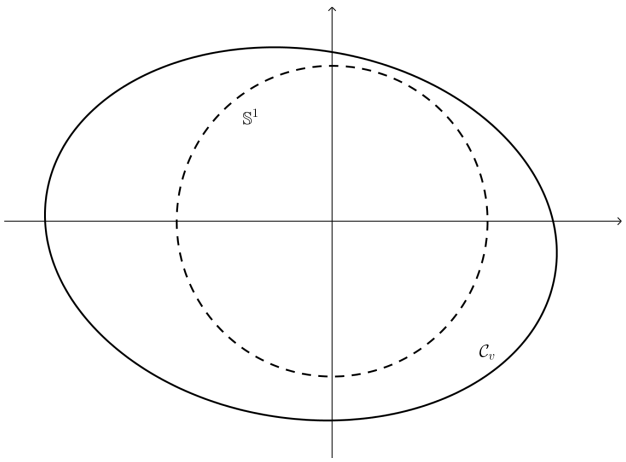


Figure 8.5. We deform the integration contour from the unit circle to a Jordan curve along which the potential is real.

As a warm up, we apply the analytic continuation to the Gross–Witten–Wadia model [190, 192]: we seek a contour $\mathcal{C}_v$ such that

$$\left[\left(z + z^{-1}\right) - v\left(z - z^{-1}\right)\right] \in \mathbb{R}$$

for all $z \in \mathcal{C}_v$. We find that $\mathcal{C}_v$ is simply a circle of radius $\sqrt{(1+v)/(1-v)}$. Moreover, the potential is

$$V(z) = 2\sqrt{1-v^2}\cos\varphi, \qquad z \in \mathcal{C}_v,\ \varphi = \mathrm{Arg}\ z.$$

Hence, not only it is real-valued, but we also recover the original, symmetric GWW model sitting on a rescaled circle, see Figure 8.6. Furthermore, $\mathcal{C}_v$ is sent to infinity or shrinks to a point as $v \to 1$ or $v \to -1$ respectively, and we cannot prolong beyond these values. This provides a new perspective on the result of [304, 305], where a drastic difference in the behaviour was observed crossing from $0 < |v| < 1$ to $|v| > 1$.

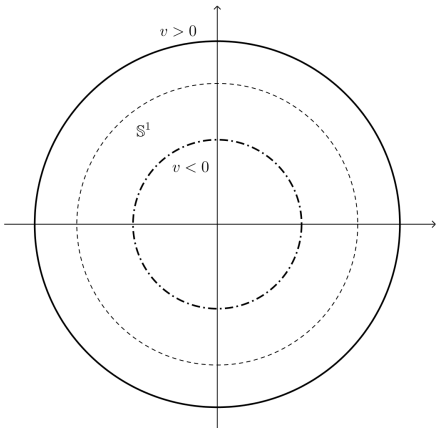


Figure 8.6. The deformed contour for the complexified Gross–Witten–Wadia model is a new circle, with bigger ($0 < v < 1$) or smaller ($-1 < v < 0$) radius.

For a generic Laurent polynomial potential

$$(1-v)\sum_{k=1}^{n} c_k z^k + (1+v)\sum_{k=1}^{n} c_k z^{-k}, \qquad c_k \in \mathbb{R},\ k = 1, \dots, n,$$

the smooth integration cycle $\mathcal{C}_v$ is determined as the locus in $\mathbb{C}$ that solves the algebraic equation

$$\Im\left(z^n V(z)\right) = 0, \qquad \forall z \in \mathcal{C}_v.$$

Stated more formally, provided the potential is a Laurent polynomial, we determine a suitable contour $\mathcal{C}_v$ as a rational algebraic curve of genus zero. If the curve has disconnected components, we simply retain as integration cycle the component homotopic to $\mathbb{S}^1$. Under the assumption that the coefficients $\{c_k\}_{k=1,\dots,n}$ are generic, this guarantees that the locus $\mathcal{C}_v$ will be a small deformation of $\mathbb{S}^1$ if $V(z)$ is deformed by a small $v \neq 0$. See, for example, [310] for a textbook reference on affine algebraic curves.

However, the potentials of the matrix models with weights of E and H type are not polynomials but logarithms. The form of the potential is precisely the reason for the discontinuous behaviour as a function of $|N_1 - N_2|$, because then imposing

$$\Im V^{\text{gen.}}_{E/H}(z) = 0, \qquad z \in \mathcal{C}_v, \tag{8.3.19}$$

does not define an algebraic curve embedded in $\mathbb{C}$. For every real $v \neq 0$ we find that the unique contour solving (8.3.19) is a half real line together with an open segment (the details depend on whether we consider the E- or the H-model), which is not a smooth deformation of $\mathbb{S}^1$.

It is perhaps instructive to look at the problem from the converse perspective. In general, the existence of a Jordan curve $\mathcal{C}_{\underline{v}} \subset \mathbb{C}$, depending on a collection of parameters $\underline{v}$, along which the potential is real-valued is not guaranteed. For non-polynomial potential $V(z)$, it requires to place the model at certain special points of the parameter space. For the weights of E and H type, this dictates $v = 0$, equivalently $N_1 = N_2$.

Phase transition and universality

To sum up our conclusions in one sentence, the unitary matrix models with symmetric E and H weight undergo a third order GWW phase transition, but their more general, non-symmetric extensions have complex potentials and therefore fall out of the GWW universality class. This

ought to be contrasted with what happens in the equivalent description as a discrete ensemble with a hard wall. In that case, increasing the parameter γ from zero, the system undergoes a third order phase transitions at the value $\gamma = \gamma_{c,E}$ when the hard wall becomes active. This holds both in the symmetric and the more general setting: they belong to the same universality class. See [160, 311] for detailed discussion on the universality of the third order phase transition in presence of a hard wall. We also stress that the discrete topology further constrains the eigenvalue density, but the condition is satisfied for all values of the parameters [284, 285]. Thus, the discrete nature of the ensemble plays no role in determining the phase structure of this model.

It is worth mentioning that the symmetric H-model has also been studied in the context of supersymmetric gauge theories, both with unitary [312] and orthogonal and symplectic symmetry [313]. The third order phase transition was observed also in the latter case [314]. The modification $N_1 \neq N_2$ in [312, Eq. (2.18)] would correspond to a different number of chiral and anti-chiral matter fields: it would be interesting to interpret the change of order of the phase transition from second to third in terms of an anomaly of four-dimensional SQCD.[22]

Final remarks on the phase transition

To wrap up, let us highlight the effects that led to a change in the order of the transition.

First of all, from (8.3.2) we infer that, at any $0 < t < 1$, the generalized models include a term which is a refinement of the insertion $(\det U)^{\pm\beta v}$. We therefore expect the phase discrepancy to be akin to the discussion in Section 4.2, even though the refinement prevents a direct identification.

The origin of the additional term in the free energy, that produces a discontinuity in the second derivative, can be traced back to two facts.

- We are taking a 't Hooft scaling for $|N_1 - N_2|$, thus enhancing the asymmetry $z \leftrightarrow z^{-1}$ to contribute at leading order at large K.
- We are imposing that the eigenvalues live on the unit circle. This might not be a natural choice when the potential of a matrix model is complex. To stick to this choice we had to accept a complexified density of eigenvalues, which nevertheless yielded a real-valued free energy.

We could have not followed the second point and, instead, looked for a contour in $\mathbb{C}$ along which the density of eigenvalues is real-valued. This would completely change the phase diagrams and the critical curves.

8.4 Cauchy ensembles and Romanovski orthogonal polynomials

We analyse the matrix models $\mathcal{Z}^{\text{sym.}}_{\text{u},H}$ and $\mathcal{Z}^{\text{sym.}}_{\text{u},E}$ defined in (8.3.6) and (8.3.7), respectively, and consider a change of topology, in which we remove a single point from the unit circle, to obtain $\mathbb{S}^1 \setminus \{\infty\} \cong \mathbb{R}$. We expect the asymptotic behavior of these models at large K to correspond to the "gapped" phase of the unitary matrix model (see Figure 8.1), which appears at strong coupling. We pass from angular variables to the real line using the stereographic projection [40, Sec. 2.5]:

$$e^{\mathrm{i}\varphi} = \frac{1 + \mathrm{i}x}{1 - \mathrm{i}x}, \qquad -\pi < \varphi < \pi, \ x \in \mathbb{R},$$

as sketched in Figure 8.7.

[22] As in Chapter 6, it could be that a change in the order of the transition signals the ill behaviour of the gauge theory. We do not have a conclusive answer. Notice that, as in Chapter 6, we are again enhancing the additional term in the potential by scaling it linearly with N in the 't Hooft limit.

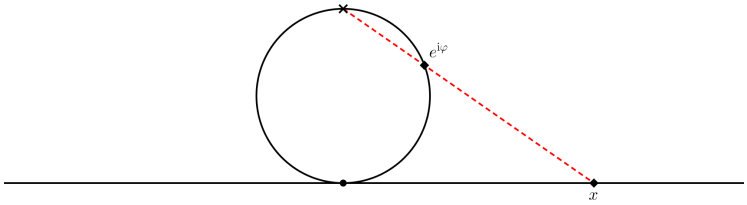


Figure 8.7. The 1d stereographic projection.

The Vandermonde determinant is mapped to:

$$\prod_{1\le j<k\le K}\left|e^{\mathrm{i}\varphi_j}-e^{\mathrm{i}\varphi_k}\right|^2\mathrm{d}\varphi_1\ldots\mathrm{d}\varphi_K=2^{K^2}\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}\frac{1}{(1+x_j^2)^K}\mathrm{d}x_1\ldots\mathrm{d}x_K,$$

and, with the corresponding transformation of the symbols, the resulting matrix models are:

$$\mathcal{Z}^{\text{sym.}}_{H,\text{stereo.}}=\frac{2^{K^2}}{K!(2\pi)^K}\int_{\mathbb{R}^K}\mathrm{d}^Kx\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}\frac{(1+x_j^2)^{\beta-K}}{\left[(1-t)^2+x_j^2(1+t)^2\right]^\beta},\tag{8.4.1}$$

$$\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}=\frac{2^{K^2}}{K!(2\pi)^K}\int_{\mathbb{R}^K}\mathrm{d}^Kx\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}\frac{\left[(1+t)^2+x_j^2(1-t)^2\right]^\beta}{(1+x_j^2)^{\beta+K}}.\tag{8.4.2}$$

Likewise, we can also use the stereographic projection in the more general, non-symmetric, matrix models (8.3.13) and (8.3.14). We obtain:

$$\begin{aligned}\mathcal{Z}^{\text{gen.}}_{H,\text{stereo.}}=&\frac{2^{K^2}}{K!(2\pi)^K}\int_{\mathbb{R}^K}\mathrm{d}^Kx\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}(1+x_j^2)^{\beta-K}\\&\times\left[\left(\frac{1}{(1-t)^2+x_j^2(1+t)^2}\right)\left(\frac{(1-t)-2\mathrm{i}tx_j+x_j^2(1+t)}{(1-t)+2\mathrm{i}tx_j+x_j^2(1+t)}\right)^v\right]^\beta,\end{aligned}\tag{8.4.3}$$

$$\begin{aligned}\mathcal{Z}^{\text{gen.}}_{E,\text{stereo.}}=&\frac{2^{K^2}}{K!(2\pi)^K}\int_{\mathbb{R}^K}\mathrm{d}^Kx\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}\frac{1}{(1+x_j^2)^{\beta+K}}\\&\times\left[\left((1+t)^2+x_j^2(1-t)^2\right)\left(\frac{(1+t)-2\mathrm{i}tx_j+x_j^2(1-t)}{(1+t)+2\mathrm{i}tx_j+x_j^2(1-t)}\right)^v\right]^\beta.\end{aligned}\tag{8.4.4}$$

where we used the redefinition of parameters $\beta_1=\beta(1-v)$ and $\beta_2=\beta(1+v)$.

8.4.1 Exact evaluation

In the limit case $t=1$, the partition function of the symmetric E-model, after stereographic projection, takes the simple form:

$$\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}(t=1)=\frac{2^{K^2+2K\beta}}{K!(2\pi)^K}\int_{\mathbb{R}^K}\mathrm{d}^Kx\prod_{1\le j<k\le K}(x_j-x_k)^2\prod_{j=1}^{K}\left(1+x_j^2\right)^{-K-\beta}.\tag{8.4.5}$$

This random matrix ensemble has been studied as a Cauchy ensemble [315], Lorentz ensemble [316], and in this form it is a particular case of the classical ensemble with weight [40, 317]

$$\sigma(x)=(1-\mathrm{i}x)^{-\alpha_1}(1+\mathrm{i}x)^{-\alpha_2},\qquad\alpha_1+\alpha_2>1,\ x\in\mathbb{R}.$$

This weight function satisfies the Pearson equation and hence the associated random matrix ensemble is classical [317, 318, 40], although in many references the listing of classical ensembles appears restricted to Hermite, Laguerre and Jacobi. See e.g. [318] for the expanded list of possible classical weights. The associated polynomials go under many names, including pseudo-Jacobi [319–323] due to their (non-trivial) relationship with Jacobi polynomials, and also appear in [324]. A proper name seems Romanovski polynomials $\left\{R_n^{(\alpha_1,\alpha_2)}\right\}$ given [325, 319], see [326] for a review. We evaluate now the matrix integral (8.4.5) using Romanovski polynomials. They satisfy (the dependence on (α_1, α_2) is understood):

$$\frac{1}{2\pi}\int_{\mathbb{R}} \mathrm{d}x R_m R_n (1-\mathrm{i}x)^{-\alpha_1}(1+\mathrm{i}x)^{-\alpha_2} = h_n \delta_{mn}$$

where we have chosen a normalization such that the polynomials R_n are monic, and their norm squared is [327]

$$h_n = 2^{-\alpha_1-\alpha_2+1}\left[(-1)^{n+1}\frac{\Gamma(-n+\alpha_1+\alpha_2)}{n!(2n-\alpha_1-\alpha_2+1)\Gamma(-n+\alpha_1)\Gamma(-n+\alpha_2)}\right]\left[\frac{n!2^n\Gamma(-2n+\alpha_1+\alpha_2)}{\Gamma(-n+\alpha_1+\alpha_2)}\right]^2$$

where the second square bracket is the change of normalization to obtain monic Romanovski polynomials from the normalization of [327]. Therefore, using

$$\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}(t=1) = \left(\frac{2^{K^2+2K\beta}}{K!}\right) K! \prod_{n=0}^{K-1} h_n,$$

with h_n specialized to the case of interest $\alpha_1 = \alpha_2 = K+\beta$, we obtain

$$\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}(t=1) = G(K+1)\frac{G(\beta+1)^2 G(K+2\beta+1)}{G(K+\beta+1)^2 G(2\beta+1)},$$

where the Barnes G-function [291] is identified using

$$\prod_{n=0}^{K-1}\Gamma(K+\beta-n) = \frac{G(K+\beta+1)}{G(\beta+1)}.$$

The result is indeed the same as (8.3.1). Therefore, the $\mathcal{Z}_E(t=1)$ is computed in three different ways: as a discrete ensemble on a finite set, as a unitary ensemble, and as the Cauchy ensemble on the real line. The tools used in each of the three approaches are, respectively: the Hahn polynomials [285], the Toeplitz determinant with symbol with a pure Fisher–Hartwig singularity [299], and the Romanovski polynomials, as represented in Figure 8.8.

Asymmetric Toeplitz matrix

A much studied symbol in spectral analysis of Toeplitz matrices [328, 329] is

$$\sigma(z) = z^s(1+tz)^\beta(1+tz^{-1})^\beta, \tag{8.4.6}$$

for integer $s \in \mathbb{Z}$. As discussed in Section 8.3.1, this has the effect of shifting the Fourier coefficients of the symbol by s, thus it shifts the diagonals of the Toeplitz matrix upward (if $s > 0$) or downward ($s < 0$) by s and the Toeplitz banded matrix associated to our model becomes asymmetric, making it more similar to the generalized model in terms of this asymmetry of the associated Toeplitz matrix. The stereographic projection of this extra term is:

$$e^{\mathrm{i}s\varphi} = \left(\frac{1+\mathrm{i}x}{1-\mathrm{i}x}\right)^s = \exp\left(s\log\frac{1+\mathrm{i}x}{1-\mathrm{i}x}\right) = e^{2s\arctan x}.$$

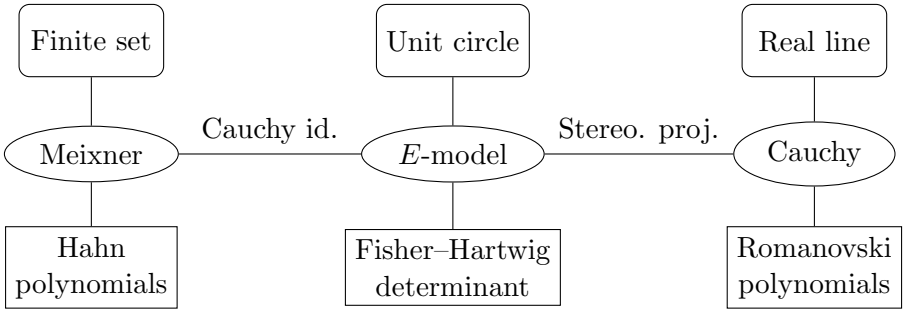


Figure 8.8. Relationships between matrix ensembles: the supports (above), the weight functions (middle) and the corresponding tools providing an exact solution at $t = 1$ (below).

For $t = 1$, the resulting weight function is

$$\sigma(x) = (1 - \mathrm{i}x)^{-K-\beta-s}(1 + \mathrm{i}x)^{-K-\beta+s}. \tag{8.4.7}$$

This is the stereographic projection of the weight introduced in (8.3.2), and it is still of the Romanovski form (see [317]) with the identification

$$\alpha_1 = K + \beta + s, \quad \alpha_2 = K + \beta - s.$$

We can therefore give explicit evaluation of the determinant of the Toeplitz banded matrix with symbol (8.4.7) using again Romanovski polynomials:

$$\det T_K\left(\sigma\left(x\right)\right) = \left(\frac{K!}{2^{K^2-2\beta K}}\right) G(K+1) \frac{G(\beta+s+1)G(\beta-s+1)}{G(K+\beta+s+1)G(K+\beta-s+1)} \frac{G(K+2\beta+1)}{G(2\beta+1)}. \tag{8.4.8}$$

Note that, if we consider the random matrix ensemble with weight (8.4.7) and keep the normalization such that it coincides with the stereographic projection of the unitary ensemble with weight (8.3.2), discussed in Section 8.3.1, the factor in bracket in (8.4.8) cancels exactly.

8.4.2 Large K limit

We now focus on $\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}$, cf. (8.4.2), and study its large K limit, with $\gamma = \beta/K$ fixed. The saddle point equation is:

$$\mathrm{P}\int \mathrm{d}y \frac{\rho(y)}{x-y} = (1+\gamma)\frac{x}{1+x^2} - \gamma\frac{(1-t)^2 x}{(1+t)^2 + x^2(1-t)^2}. \tag{8.4.9}$$

The parity symmetry of the matrix model guarantees that we can look for a symmetric solution with $\mathrm{supp}\rho = [-A, A]$, $A > 0$. We expect that A will be back-projected to $e^{\mathrm{i}\phi_0}$ of Section 8.3.3 undoing the stereographic projection.

We report the details of the computations in Appendix 8.C.1. We arrive at the eigenvalue density:

$$\rho(x) = \frac{\sqrt{A^2 - x^2}}{\pi}\left[\frac{(1+\gamma)}{\sqrt{A^2+1}(x^2+1)} - \frac{\gamma t_0}{\sqrt{A^2+t_0^2}(x^2+t_0^2)}\right],$$

with the boundary A fixed by normalization:

$$A^2 = \frac{(2\gamma+1)(1+t)^2}{(2\gamma+1-t)(2\gamma-1+t)}.$$

This solution is well defined as long as

$$\gamma > \frac{1-t}{2t},$$

and sending $A \to \infty$, which would be back-projected to $\phi_0 \to \pi$, corresponds to the limit $\gamma \downarrow \gamma_{c,E} = \frac{1-t}{2t}$. This matches the analysis on the circle.

We can do the same with the more general matrix model $\mathcal{Z}^{\text{gen.}}_{E,\text{stereo.}}$ in (8.4.4). In this case, as happened on the circle, turning on the asymmetry parameter v complexifies the eigenvalue density, which becomes:

$$\rho_{\text{stereo.}}(x) = \frac{\sqrt{A^2-x^2}}{\pi}\left[\frac{(1+\gamma)+\mathrm{i}v\gamma x}{\sqrt{A^2+1}(x^2+1)} - \frac{\gamma t_0 + \mathrm{i}v\gamma x}{\sqrt{A^2+t_0^2}(x^2+t_0^2)}\right],$$

with same value of A as above. The derivation of the result is given in Appendix 8.C.2. For completeness, we also report the details of the large K analysis of the matrix model with weight (8.4.6), corresponding to the determinant of an asymmetric Toeplitz matrix, in Appendix 8.C.3.

8.A Solution to saddle point equations on the unit circle

Here we present the calculations to solve the saddle point equations obtained in the scaled large size limit of the unitary matrix models discussed in Section 8.3.

8.A.1 Solution for $\mathcal{Z}^{\text{sym.}}_{\text{u},H}$ at weak coupling

We start with the solution of the saddle point equation (8.3.8), which we rewrite here for completeness:

$$-\mathrm{i}\gamma t\left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right] = \mathrm{P}\int \mathrm{d}\vartheta \rho_H(\vartheta)\cot\left(\frac{\varphi-\vartheta}{2}\right). \tag{8.A.1}$$

The solution was originally obtained by Baik in [302], and the present derivation via saddle point equations was discussed in [11].

We make the ansatz $\text{supp}\rho = [-\pi,\pi]$. On the right hand side of (8.A.1) we use the expansion (for $\vartheta \neq \varphi$):

$$\cot\left(\frac{\varphi-\vartheta}{2}\right) = 2\sum_{n=1}^{\infty}\left[\sin(n\varphi)\cos(n\vartheta) - \cos(n\varphi)\sin(n\vartheta)\right],$$

thus the integration extracts the Fourier coefficients of the eigenvalue density:

$$\mathrm{P}\int \mathrm{d}\vartheta\rho(\vartheta)\cot\left(\frac{\varphi-\vartheta}{2}\right) = 2\pi\sum_{n=1}^{\infty}\left[a_n\sin(n\varphi) - b_n\cos(n\varphi)\right],$$

where

$$a_n = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{d}\vartheta\rho_H(\vartheta)\cos(n\vartheta), \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi}\mathrm{d}\vartheta\rho_H(\vartheta)\sin(n\vartheta).$$

For the left hand side we use:

$$-\mathrm{i}\gamma t\left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right] = -(\mathrm{i}\gamma)\frac{2\mathrm{i}t\sin(\varphi)}{1+t^2-2t\cos(\varphi)} = 2\gamma t\sin(\varphi)\sum_{n=0}^{\infty}U_n(\cos(\varphi))t^n,$$

where in the second equality we recognized the generating function of Chebyshev polynomials of second kind, $U_n(x)$. Exploiting the basic property $U_n(\cos\varphi) = \frac{\sin(n+1)\varphi}{\sin\varphi}$, equation (8.A.1) is rewritten as:

$$2\gamma\sum_{n=1}^{\infty}\sin(n\varphi)t^n = 2\pi\sum_{n=1}^{\infty}\left[a_n\sin(n\varphi) - b_n\cos(n\varphi)\right],$$

and we immediately get:

$$a_n = \frac{\gamma}{\pi} t^n, \qquad b_n = 0, \tag{8.A.2}$$

for all $n = 1, 2, \ldots$. The only yet undetermined coefficient a_0 is fixed by normalization to $a_0 = 1/(2\pi)$. Putting all together:

$$\rho_H(\varphi) = \frac{1}{2\pi}\left[1 + 2\gamma \sum_{n=1}^{\infty} \cos(n\varphi) t^n\right],$$

which can be further simplified recognizing the generating function of Chebyshev polynomials of first kind $T_n(x)$, using the property $T_n(\cos(\varphi)) = \cos(n\varphi)$. We finally arrive at:

$$\rho_H(\varphi) = \frac{1}{2\pi}\left[1 + 2\gamma t \left(\frac{\cos\varphi - t}{1 + t^2 - 2t\cos\varphi}\right)\right]. \tag{8.A.3}$$

The minima of this function are located at $\varphi = \pm\pi$, and imposing the condition $\rho(\varphi) \geq 0$ for all $-\pi \leq \varphi \leq \pi$, we see that the present solution holds in the regime

$$0 \leq \gamma \leq \frac{1+t}{2t} =: \gamma_{c,H}.$$

Above the critical value $\gamma_{c,H}$, solution (8.A.3) ceases to be valid and we must drop the assumption $\mathrm{supp}\rho = [-\pi, \pi]$ and look for a different solution.

8.A.2 Solution for $\mathcal{Z}^{\mathrm{sym.}}_{\mathrm{u},H}$ at strong coupling

In this appendix we present the calculations to solve the saddle point equation (8.3.8) at strong coupling. The procedure follows [330, App. B] (see also [11, Sec. 4]). We look for a one-cut solution with $\mathrm{supp}\rho = [-\phi_0, \phi_0]$, $0 < \phi_0 < \pi$.

We introduce the following complex function, named resolvent:

$$\omega(z) = \int_{\mathcal{L}} \frac{\mathrm{d}u}{\mathrm{i}u} \frac{z+u}{z-u} \psi_H(u), \qquad z \in \mathbb{C}\setminus\mathcal{L}.$$

Here, $\mathcal{L} \subset \mathbb{S}^1$ is the arc of the unit circle from $e^{-\mathrm{i}\phi_0}$ to $e^{\mathrm{i}\phi_0}$, oriented anti-clockwise, and the complex function $\psi_H(u)$ is the continuation of $\rho_H(\varphi)$ to the complex plane, with $\psi_H(e^{\mathrm{i}\varphi}) = \rho_H(\varphi)$. The eigenvalue density is recovered from the resolvent through the relation

$$\omega_+(e^{\mathrm{i}\varphi}) - \omega_-(e^{\mathrm{i}\varphi}) = 4\pi\psi_H(e^{\mathrm{i}\varphi}),$$

with the subscript $+$ (resp. $-$) meaning the limit taken from outside (resp. inside) the unit circle.

Following standard methods, we introduce two auxiliary complex functions:

$$h(z) = \sqrt{(e^{\mathrm{i}\phi_0} - z)(e^{-\mathrm{i}\phi_0} - z)}, \qquad z \in \mathbb{C}, \tag{8.A.4}$$

and a function $\Phi(z)$, to be determined, such that

$$\omega(z) = h(z)\Phi(z), \qquad z \in \mathbb{C}\,. \tag{8.A.5}$$

The saddle point equation (8.3.8) becomes a discontinuity equation for $\Phi(z)$. We will use this discontinuity equation, together with the fact that, by definition, $\Phi(z)$ decays as $\sim 1/z$ as $|z| \to \infty$, to obtain an integral expression for $\Phi(z)$ (see [330, 11]). Explicitly:

$$\Phi(z) = \oint_{\mathcal{C}} \frac{\mathrm{d}u}{2\pi} \frac{W(u)}{(u-z)h(u)},$$

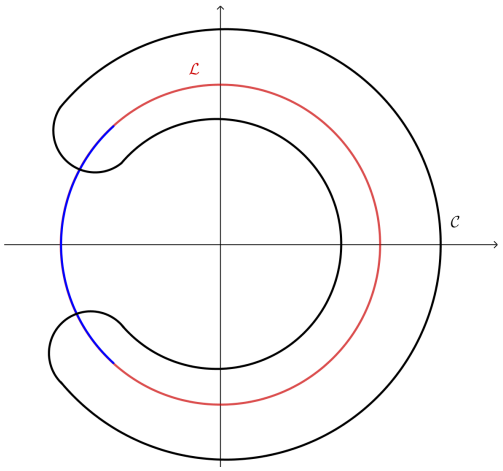


Figure 8.9. Integration contour in the complex plane. The red line is the cut $\mathcal{L}$, the black curve represents a choice of the contour $\mathcal{C}$. The blue arc on the unit circle, complementary to $\mathcal{L}$, is the gap, where the resolvent $\omega(z)$ is continuous.

where we denoted for shortness $W(u)$ the left hand side of (8.A.1), and the contour $\mathcal{C}$ encloses the branch cut $\mathcal{L}$, the arc along the unit circle, cf. Figure 8.9.

At this stage, we notice that the left hand side of (8.A.1) has poles but no branch cuts, so we can manipulate the integration contour $\mathcal{C}$ in a convenient way. We arrive at:

$$\Phi(z) = \mathcal{I}_1(z) + \mathcal{I}_2(z) + \mathcal{I}_3(z),$$

where the three contributions are:

$$\begin{aligned}
\mathcal{I}_1(z) &= -\frac{\gamma t}{h(z)}\left(\frac{z}{1-tz} - \frac{z^{-1}}{1-tz^{-1}}\right), \\
\mathcal{I}_2(z) &= \sum_{z_p} \mathrm{Res}_{u=z_p} \frac{-\gamma t}{h(u)(u-z)}\left(\frac{u}{1-tu} - \frac{u^{-1}}{1-tu^{-1}}\right), \\
\mathcal{I}_3(z) &= \lim_{R\to\infty} \oint_{\mathcal{C}_R} \frac{\mathrm{d}u}{2\pi \mathrm{i}} \frac{\gamma t}{h(u)(u-z)}\left(\frac{u}{1-tu} - \frac{u^{-1}}{1-tu^{-1}}\right),
\end{aligned}$$

The sum in $\mathcal{I}_2$ runs over the poles $\{z_p\}$ of the derivative of the potential (that is, the poles on the left hand side of (8.A.1)), in this case $z_p = t^{\pm 1}$, and $\mathcal{C}_R$ is a large circle of radius R. We have $\mathcal{I}_3(z) = 0$ and $\mathcal{I}_1(z)h(z)$ yields no discontinuity, so it is irrelevant for the evaluation of $\psi(z)$. The unique relevant contribution comes from:

$$\begin{aligned}
\mathcal{I}_2(z) &= -\gamma t \left\{ \mathrm{Res}_{u=t} \frac{\left(\frac{u}{1-tu} - \frac{u^{-1}}{1-tu^{-1}}\right)}{h(u)(u-z)} + \mathrm{Res}_{u=1/t} \frac{\left(\frac{u}{1-tu} - \frac{u^{-1}}{1-tu^{-1}}\right)}{h(u)(u-z)} \right\} \\
&= \gamma \left[\frac{t}{h(t)(t-z)} + \frac{t^{-1}}{h(t^{-1})(t^{-1}-z)} \right] = \frac{-\gamma t}{h(t)}\left(\frac{1}{1-tz} + \frac{z^{-1}}{1-tz^{-1}}\right),
\end{aligned} \tag{8.A.6}$$

where for the last equality we used $h(t^{-1}) = -t^{-1}h(t)$. In general, on the real axis $h(x) > 0$ if $x > 1$ and $h(x) < 0$ if $x < 1$. Therefore, since $0 < t < 1$, we bare in mind that $-h(t) > 0$.

Plugging the expression for $\Phi(z)$ into $\omega(z)$ and taking its discontinuity along the arc $\mathcal{L}$, we arrive at:

$$\psi_H(e^{\mathrm{i}\varphi}) = -\frac{2\gamma t(1-t)}{\pi h(t)}\left(\frac{\cos\frac{\varphi}{2}}{(1-t)^2 + 4t\left(\sin\frac{\varphi}{2}\right)^2}\right)\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2}. \tag{8.A.7}$$

The boundary ϕ_0 of the support is fixed by normalization:

$$1 = \int_{-\phi_0}^{\phi_0} \mathrm{d}\rho_H(\varphi) = \gamma \left(\frac{1-t}{h(t)} + 1 \right),$$

which provides

$$h(t) = -\frac{\gamma(1-t)}{(\gamma-1)}$$

and hence, writing $h(t)$ explicitly, we obtain:

$$\left(\sin \frac{\phi_0}{2} \right)^2 = \frac{(1-t)^2(2\gamma-1)}{4t(\gamma-1)^2}.$$

Also, plugging the expression for $h(t)$ in (8.A.7), we arrive at the final expression

$$\rho_H(\varphi) = \frac{2(\gamma-1)t}{\pi} \left(\frac{\cos\frac{\varphi}{2}}{(1-t)^2 + 4t\left(\sin\frac{\varphi}{2}\right)^2} \right) \sqrt{\left(\sin\frac{\phi_0}{2} \right)^2 - \left(\sin\frac{\varphi}{2} \right)^2}.$$

8.A.3 Solution for $\mathcal{Z}^{\text{sym.}}_{\text{u},E}$ at weak coupling

In this appendix we solve the saddle point equation (8.3.11), which we rewrite here for completeness:

$$-\mathrm{i}\gamma t \left[\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}} \right] = \mathrm{P} \int \mathrm{d}\vartheta \rho_E(\vartheta) \cot\left(\frac{\varphi-\vartheta}{2} \right). \tag{8.A.8}$$

The analysis will follow closely Appendix 8.A.1 for small coupling and Appendix 8.A.2 for strong coupling, and we omit most of the details.

We again start with the ansatz $\mathrm{supp}\rho = [-\pi, \pi]$, and manipulate the right hand side of (8.A.8) exactly as we did in Appendix 8.A.1, and for the left hand side we use:

$$-\mathrm{i}\gamma t \left[\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}} \right] = -2\gamma t \sin\varphi \sum_{n=0}^{\infty} U_n(\cos\varphi)(-t)^n = 2\gamma \sum_{n=1}^{\infty} \sin(n\varphi)(-t)^n.$$

Then, (8.A.8) becomes

$$2\gamma \sum_{n=1}^{\infty} \sin(n\varphi)(-t)^n = -2\pi \sum_{n=1}^{\infty} \left[a_n \sin(n\varphi) - b_n \cos(n\varphi) \right],$$

with a_n, b_n the Fourier coefficients of $\rho_E(\varphi)$, and we immediately obtain:

$$\rho_E(\varphi) = \frac{1}{2\pi} \left[1 - 2\gamma \sum_{n=1}^{\infty} \cos(n\varphi)(-t)^n \right],$$

where, as usual, $a_0 = 1/(2\pi)$ is fixed by normalization. We recognize the generating function of Chebyshev polynomials of the first kind, using the standard identity $T_n(\cos(\varphi)) = \cos(n\varphi)$, and arrive at the final expression for the eigenvalue density

$$\rho_E(\varphi) = \frac{1}{2\pi} \left[1 + 2\gamma t \left(\frac{\cos\varphi + t}{1 + t^2 + 2t\cos\varphi} \right) \right].$$

The minima are placed at $\varphi = \pm\pi$, and the non-negativity condition $\rho_E \geq 0$ holds as long as:

$$\gamma \leq \frac{1-t}{2t} =: \gamma_{c,E}.$$

We see that, as expected by naïve comparison of the integral representation of the matrix models (8.3.6) and (8.3.7), the eigenvalue densities ρ_H and ρ_E, as well as the respective critical points, are related through $(\gamma, t) \leftrightarrow (-\gamma, -t)$.

8.A.4 Solution for $\mathcal{Z}^{\text{sym.}}_{\text{u},E}$ at strong coupling

The procedure at strong coupling is as in Appendix 8.A.2, and we avoid the technical details here. We assume a one-cut solution supported on $[-\phi_0, \phi_0]$, $0 < \phi_0 < \pi$, and introduce the resolvent

$$\omega(z) = \int_{\mathcal{L}} \frac{\mathrm{d}u}{\mathrm{i}u} \frac{z+u}{z-u} \psi_E(u), \qquad z \in \mathbb{C} \setminus \mathcal{L},$$

with integration contour $\mathcal{L}$ meant to be the arc along the unit circle connecting $e^{-\mathrm{i}\phi_0}$ to $e^{\mathrm{i}\phi_0}$, and the function $\psi_E(u)$ being the continuation of ρ_E in $\mathbb{C}$, with $\psi(e^{\mathrm{i}\varphi}) = \rho_E(\varphi)$. We introduce, as in Appendix 8.A.2, the complex functions $h(z)$ and $\Phi(z)$, and the saddle point equation (8.A.8) becomes a discontinuity equation for $\Phi(z)$. The only relevant contribution (for the computation of ρ_E) to $\Phi(z)$ is:

$$\begin{aligned}\mathcal{I}_2(z) &= -\gamma t \left\{ \mathrm{Res}_{u=-t} \frac{\left(\frac{u}{1+tu} - \frac{u^{-1}}{1+tu^{-1}}\right)}{h(u)(u-z)} + \mathrm{Res}_{u=-1/t} \frac{\left(\frac{u}{1+tu} - \frac{u^{-1}}{1+tu^{-1}}\right)}{h(u)(u-z)} \right\} \\ &= -\gamma \left[\frac{t}{h(-t)(t+z)} + \frac{t^{-1}}{h(-t^{-1})(t^{-1}+z)} \right] = \frac{-\gamma t}{h(-t)} \left(\frac{1}{1+tz} + \frac{z^{-1}}{1+tz^{-1}} \right),\end{aligned}$$

where for the last equality we used $h(-t^{-1}) = t^{-1} h(-t)$, with both understood to be negative. We get:

$$\rho_E(\varphi) = -\frac{2\gamma t(1+t)}{\pi h(-t)} \left(\frac{\cos\frac{\varphi}{2}}{(1+t)^2 - 4t\left(\sin\frac{\varphi}{2}\right)^2} \right) \sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2}, \tag{8.A.9}$$

and $\sin(\phi_0/2)$ is fixed by the normalization:

$$1 = \int_{-\phi_0}^{\phi_0} \rho_E(\varphi) \mathrm{d}\varphi = -\gamma \left(\frac{1+t}{h(-t)} + 1 \right).$$

Thus,

$$h(-t) = -\frac{\gamma(1+t)}{\gamma+1} \quad \Longrightarrow \quad \left(\sin\frac{\phi_0}{2}\right)^2 = \frac{(1+t)^2(2\gamma+1)}{4t(\gamma+1)^2}.$$

Combining this latter expression with (8.A.9), we finally arrive at:

$$\rho_E(\varphi) = \frac{2}{\pi}(\gamma+1)t \left[\frac{\cos\left(\frac{\varphi}{2}\right)}{(1+t)^2 - 4t\left(\sin\left(\frac{\varphi}{2}\right)\right)^2} \right] \sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2},$$

supported in $[-\phi_0, \phi_0]$.

8.A.5 Solution for $\mathcal{Z}^{\text{mod.}}_{\text{u},H}$ and $\mathcal{Z}^{\text{mod.}}_{\text{u},E}$ at weak coupling

In this appendix we solve in the weak coupling regime the saddle point equations, which we write below, for the matrix models with symmetric weight multiplied by a monomial,

$$\sigma_{E \text{ or } H}(z) \to z^s \sigma_{E \text{ or } H}(z).$$

The procedure is essentially as in Appendices 8.A.1 and 8.A.3.

The first saddle point equation, for the large K scaling limit of $\mathcal{Z}^{\text{mod.}}_{\text{u},H}$, is:

$$-\mathrm{i}\gamma t \left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}} \right] - \mathrm{i}\alpha_1 e^{\mathrm{i}\varphi} + \mathrm{i}\alpha_2 e^{-\mathrm{i}\varphi} = \mathrm{P} \int \mathrm{d}\vartheta \rho^{\text{mod.}}_H(\vartheta) \cot\left(\frac{\varphi-\vartheta}{2}\right). \tag{8.A.10}$$

We begin with the ansatz $\text{supp}\rho_H^{\text{mod.}} = [-\pi, \pi]$, and follow the same steps as in Appendix 8.A.1. We see that the unique difference with the case without the extra insertion in the potential, is in the ± 1 Fourier coefficients of the eigenvalue density, and we get:

$$\rho_H^{\text{mod.}}(\varphi) = \frac{1}{2\pi}\left[1 + 2\gamma t\left(\frac{\cos\varphi - t}{1 + t^2 - 2t\cos\varphi}\right) + \alpha_+\cos(\varphi) - \mathrm{i}\alpha_-\sin(\varphi)\right],$$

where we defined

$$\alpha_\pm = \alpha_2 \pm \alpha_1.$$

We notice that the modification of the potential leads to a complex-valued eigenvalue density, with imaginary part proportional to the parameter α_- which measures the asymmetry between $z \leftrightarrow z^{-1}$. This solution holds for

$$\gamma \leq \frac{1+t}{2t}\left(1 - \alpha_+\right),$$

in particular, when considering monomial insertions generated by the potential studied here, we will eventually set $\alpha_\pm = 0$, and the critical value is unchanged in that case.

The second saddle point equation of interest, that of $\mathcal{Z}_{\text{u},E}^{\text{mod.}}$, is:

$$-\mathrm{i}\gamma t\left[\frac{e^{\mathrm{i}\varphi}}{1 + te^{\mathrm{i}\varphi}} - \frac{e^{-\mathrm{i}\varphi}}{1 + te^{-\mathrm{i}\varphi}}\right] - \mathrm{i}\alpha_1 e^{\mathrm{i}\varphi} + \mathrm{i}\alpha_2 e^{-\mathrm{i}\varphi} = \mathrm{P}\int \mathrm{d}\vartheta \rho_E^{\text{mod.}}(\vartheta)\cot\left(\frac{\varphi - \vartheta}{2}\right). \tag{8.A.11}$$

Assuming $\text{supp}\rho_E^{\text{mod.}} = [-\pi, \pi]$, we see that the unique difference with the case discussed in Appendix 8.A.3 are the ± 1 Fourier coefficients of the eigenvalue density $\rho_E^{\text{mod.}}$. The standard procedure gives:

$$\rho_E^{\text{mod.}}(\varphi) = \frac{1}{2\pi}\left[1 + 2\gamma t\left(\frac{\cos\varphi + t}{1 + t^2 + 2t\cos\varphi}\right) + \alpha_+\cos(\varphi) - \mathrm{i}\alpha_-\sin(\varphi)\right],$$

which extends up to

$$\gamma \leq \frac{1-t}{2t}\left(1 - \alpha_+\right).$$

8.A.6 Solution for $\mathcal{Z}_{\text{u},H}^{\text{mod.}}$ and $\mathcal{Z}_{\text{u},E}^{\text{mod.}}$ at strong coupling

In this appendix we solve equations (8.A.10) and (8.A.11) in the strong coupling phase. We therefore drop the assumption that the eigenvalue densities are supported o the whole circle, but still assume a symmetric support.

We first consider (8.A.10), and follow the same steps as in Appendix 8.A.2. The different with the case analyzed there is that now, the resolvent receives two additional contributions: one from the residue at $u = 0$, given by the term $\alpha_2 u^{-1}$, and one from the contribution of $\alpha_1 u$ to the integration along the large circle:

$$\mathcal{I}_3(z) = \lim_{R\to\infty}\oint_{\mathcal{C}_R}\frac{\mathrm{d}u}{2\pi\mathrm{i}}\frac{\alpha_1 u + \cdots}{h(u)(u - z)} = \alpha_1,$$

where the dots include terms which are irrelevant when $|u| \to \infty$. Altogether, we have that $\rho_H^{\text{mod.}}(\varphi)$ equals the one with $\alpha_1 = 0 = \alpha_2$, plus those two extra contributions:

$$-\frac{\mathrm{i}h(e^{\mathrm{i}\varphi})}{\pi}(\alpha_1 + \alpha_2),$$

and we finally arrive at:

$$\rho_H^{\text{mod.}}(\varphi) = \frac{1}{\pi}\left[-\frac{2\gamma t(1-t)}{h(t)}\left(\frac{\cos\frac{\varphi}{2}}{(1-t)^2 + 4t\left(\sin\frac{\varphi}{2}\right)^2}\right) + \alpha_+\cos\frac{\varphi}{2} - \mathrm{i}\alpha_-\sin\frac{\varphi}{2}\right]\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2}.$$

The normalization condition implies:

$$1 = \int_{-\phi_0}^{\phi_0} \mathrm{d}\rho_H^{\mathrm{mod.}}(\varphi) = \gamma \left(\frac{1-t}{h(t)} + 1 \right) + \alpha_+ \left(\sin \frac{\phi_0}{2} \right)^2 ,$$

thus now the solution for $\left(\sin \frac{\phi_0}{2}\right)^2$ depends on α_+.

The strong coupling solution $\rho_H^{\mathrm{mod.}}$ is derived analogously.

8.A.7 Solution for $\mathcal{Z}_{\mathrm{u},H}^{\mathrm{gen.}}$ at weak coupling

This appendix is dedicated to the solution, in the weak coupling regime, to the saddle point equation (8.3.15), which we rewrite here for convenience:

$$-\mathrm{i}\gamma t \left[\frac{(1-v)e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{(1+v)e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}} \right] = \mathrm{P} \int \mathrm{d}\vartheta \rho_H(\vartheta) \cot \left(\frac{\varphi - \vartheta}{2} \right) . \tag{8.A.12}$$

The procedure is similar to the one adopted in Appendix 8.A.1, but now the parameter v makes the left hand side complex-valued. Therefore, we expand the eigenvalue density in the exponential form:

$$\rho_H(\vartheta) = \sum_{n \in \mathbb{Z}} \rho_{H,n} e^{\mathrm{i}n\vartheta},$$

while we also use:[23]

$$\cot \left(\frac{\varphi - \vartheta}{2} \right) = -\mathrm{i} \sum_{n=1}^{\infty} \left[e^{\mathrm{i}n(\varphi - \vartheta)} - e^{-\mathrm{i}n(\varphi - \vartheta)} \right] .$$

The right hand side of (8.A.12) then becomes:

$$\mathrm{P} \int \mathrm{d}\vartheta \rho_H(\vartheta) \cot \left(\frac{\varphi - \vartheta}{2} \right) = -2\pi\mathrm{i} \sum_{n=1}^{\infty} \left[\rho_{H,n} e^{\mathrm{i}n\varphi} - \rho_{H,-n} e^{-\mathrm{i}n\varphi} \right] .$$

For the left hand side we can expand the geometric series and obtain:

$$-\mathrm{i}\gamma t \left[\frac{(1-v)e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} - \frac{(1+v)e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}} \right] = -\mathrm{i}\gamma \sum_{n=1}^{\infty} t^n \left[(1-v)e^{\mathrm{i}n\varphi} - (1+v)e^{-\mathrm{i}n\varphi} \right] .$$

Putting all together, and taking into account the normalization condition that fixes ρ_0, we arrive at:

$$\rho_H(\varphi) = \frac{1}{2\pi} \left\{ 1 + 2\gamma \sum_{n=1}^{\infty} t^n \left[\cos(n\varphi) - \mathrm{i}v \sin(n\varphi) \right] \right\} .$$

We see that the parameter v, which controls the asymmetry, introduces an imaginary part in the eigenvalue density. Since $-1 < v < 1$, the relevant minima are those of the real part, thus the critical value is the same as in the symmetric case:

$$\gamma_{c,H} = \frac{1+t}{2t},$$

and above this value the solution ceases to be valid.

[23] The expression holds for $\varphi \neq \vartheta$, which is guaranteed by the principal value of the integral.

8.A.8 Solution for $\mathcal{Z}^{\text{gen.}}_{\text{u},H}$ at strong coupling

When $\gamma > \gamma_{c,H}$ we have to look for a different solution to the saddle point equation (8.A.12). At strong coupling, the procedure is exactly the same as in Appendix 8.A.2, since we were already working in the complex plane, so the complexification of the left hand side of (8.A.12) for $v \neq 0$ does not alter the strategy. We assume a gapped one-cut solution supported on $[-\phi_0, \phi_0]$ and introduce the resolvent

$$\omega(z) = \int_{\mathcal{L}} \frac{\mathrm{d}u}{\mathrm{i}u} \frac{z+u}{z-u} \psi(u), \qquad z \in \mathbb{C} \setminus \mathcal{L}, \ \mathcal{L} \subset \mathbb{S}^1.$$

Everything goes trough exactly as in Appendix 8.A.2, except for the factors of $(1 \pm v)$ appearing on the left hand side of (8.A.12). One easily finds:

$$\psi(e^{\mathrm{i}\varphi}) = -\gamma \frac{t}{\pi h(t)} \left[\frac{(1-v)e^{\mathrm{i}\varphi/2}}{1-te^{\mathrm{i}\varphi}} + \frac{(1+v)e^{-\mathrm{i}\varphi/2}}{1-te^{-\mathrm{i}\varphi}} \right] \sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2 - \left(\sin\left(\frac{\varphi}{2}\right)\right)^2}.$$

The normalization fixes $\sin(\phi_0/2)^2$, through:

$$1 = \int_{-\phi_0}^{\phi_0} \mathrm{d}\varphi \rho_H(\varphi) = \gamma \left(\frac{1-t}{h(t)} + 1 \right).$$

Notice that, due to the parity of the integral, the result is independent of v, in particular it is equal to the symmetric case ($v = 0$). The final expression for the eigenvalue density:

$$\rho_H(\varphi) = \frac{2t(\gamma-1)}{\pi(1-t)} \left[\frac{(1-t)\cos\left(\frac{\varphi}{2}\right) - \mathrm{i}v(1+t)\sin\left(\frac{\varphi}{2}\right)}{(1-t)^2 + 4t\left(\sin\left(\frac{\varphi}{2}\right)\right)^2} \right] \sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2 - \left(\sin\left(\frac{\varphi}{2}\right)\right)^2},$$

with $\sin(\phi_0/2)$ the same as in [302] and Appendix 8.A.2, that is

$$\left(\sin\frac{\phi_0}{2}\right)^2 = \frac{(1-t)^2(2\gamma-1)}{4t(\gamma-1)^2}.$$

8.A.9 Solution for $\mathcal{Z}^{\text{gen.}}_{\text{u},E}$ at weak coupling

Here we solve the large K limit of the second matrix model, cf. (8.3.14), in the general case. The saddle point equation (8.3.17) is:

$$-\mathrm{i}\gamma t \left(\frac{(1-v)e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}} - \frac{(1+v)e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}} \right) = \mathrm{P} \int \mathrm{d}\vartheta \rho_E(\vartheta) \cot\frac{\varphi-\vartheta}{2}. \tag{8.A.13}$$

The procedure is as in Appendix 8.A.7. The right hand side reads:

$$-2\pi\mathrm{i} \sum_{n=1}^{\infty} \left[\rho_{E,n} e^{\mathrm{i}n\varphi} - \rho_{E,-n} e^{-\mathrm{i}n\varphi} \right],$$

while, expanding the geometric series, the left hand side becomes:

$$\mathrm{i}\gamma \sum_{n=1}^{\infty} (-t)^n \left[(1-v)e^{\mathrm{i}n\varphi} - (1+v)e^{-\mathrm{i}n\varphi} \right].$$

This gives:

$$\rho_E(\varphi) = \frac{1}{2\pi} \left[1 + 2\gamma t \left(\frac{\cos(\varphi) + t - \mathrm{i}v\sin(\varphi)}{1+t^2+2t\cos(\varphi)} \right) \right].$$

Again, the validity of this solution extends up to:

$$\gamma_{c,E} = \frac{1-t}{2t},$$

as in the symmetric case.

8.A.10 Solution for $\mathcal{Z}_{\mathrm{u},E}^{\mathrm{gen.}}$ at strong coupling

When $\gamma > \gamma_{c,E}$, we drop the assumption $\mathrm{supp}\rho_E = [-\pi, \pi]$ and look for a one-cut solution with a gap. The procedure is clear from the previous Appendices, and the result is:

$$\rho_E(\varphi) = -\frac{\gamma t}{\pi h(-t)}\left[\frac{(1-v)e^{\mathrm{i}\varphi/2}}{1+te^{\mathrm{i}\varphi}} + \frac{(1+v)e^{-\mathrm{i}\varphi/2}}{1+te^{-\mathrm{i}\varphi}}\right]\sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2 - \left(\sin\left(\frac{\varphi}{2}\right)\right)^2},$$

which, after imposing normalization and rewriting in terms of trigonometric functions, gives:

$$\rho_E(\varphi) = \frac{2t(\gamma+1)}{\pi(1+t)}\left[\frac{(1+t)\cos\left(\frac{\varphi}{2}\right) - \mathrm{i}v(1-t)\sin\left(\frac{\varphi}{2}\right)}{(1+t)^2 - 4t\left(\sin\left(\frac{\varphi}{2}\right)\right)^2}\right]\sqrt{\left(\sin\left(\frac{\phi_0}{2}\right)\right)^2 - \left(\sin\left(\frac{\varphi}{2}\right)\right)^2},$$

supported on

$$\mathrm{supp}\rho_E = [-\phi_0, \phi_0], \quad \left(\sin\left(\frac{\phi_0}{2}\right)\right)^2 = \frac{(1+t)^2(2\gamma+1)}{4t(\gamma+1)}.$$

8.B Free energies

This appendix contains the calculations to obtain the free energy of the unitary matrix models considered in the main text.

8.B.1 Free energy for $\mathcal{Z}_{\mathrm{u},H}^{\mathrm{sym.}}$ and $\mathcal{Z}_{\mathrm{u},E}^{\mathrm{sym.}}$ at weak coupling

The simplest way to obtain the free energy is to evaluate its derivative with respect to the parameter t and then integrate. In the weak coupling phase, $0 \leq \gamma \leq \gamma_c$, we can use the boundary condition $\mathcal{Z}_{\mathrm{u}}(\gamma = 0) = 1$, which follows immediately from the normalization of the Haar measure on $U(N)$. At strong coupling we use the continuity of $\log \mathcal{Z}_{\mathrm{u}}(\gamma)$ at $\gamma = \gamma_c$.

For what concerns $\mathcal{F}_{\mathrm{u},H}^{\mathrm{sym.}}$, at weak coupling we have:

$$\frac{\mathrm{d}\mathcal{F}_{\mathrm{u},H}^{\mathrm{sym.}}}{\mathrm{d}t}(\gamma \leq \gamma_{c,H}) = \gamma \int \mathrm{d}\varphi \rho_H(\varphi)\left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} + \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right] = \frac{2\gamma^2 t}{1-t^2}.$$

Integrating with boundary condition $\mathcal{Z}_{\mathrm{u}}^H(\gamma = 0) = 1$ gives

$$\mathcal{F}_{\mathrm{u},H}^{\mathrm{sym.}}(\gamma \leq \gamma_{c,H}) = -\gamma^2 \log(1-t^2). \tag{8.B.1}$$

For $\mathcal{F}_{\mathrm{u},E}^{\mathrm{sym.}}$ at weak coupling we have:

$$\frac{\mathrm{d}\mathcal{F}_{\mathrm{u},E}^{\mathrm{sym.}}}{\mathrm{d}t}(\gamma \leq \gamma_{c,E}) = \gamma \int \mathrm{d}\varphi \rho_E(\varphi)\left[\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}} + \frac{e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}}\right] = \frac{2\gamma^2 t}{1-t^2},$$

which, after integration, gives

$$\mathcal{F}_{\mathrm{u},E}^{\mathrm{sym.}}(\gamma \leq \gamma_{c,E}) = -\gamma^2 \log(1-t^2). \tag{8.B.2}$$

In particular, we see that the free energies of the two models are equal in the weak coupling phase. This is expected, since, in the weak coupling phase, the free energy equals the result provided by Szegő theorem, i.e. the limit without scaling (see e.g. [11] for a proof of this statement), and it is well known that the limits of the models (8.3.6) and (8.3.7) without scaling are equal.

8.B.2 Free energy for $\mathcal{Z}^{\text{sym.}}_{\text{u},H}$ and $\mathcal{Z}^{\text{sym.}}_{\text{u},E}$ at strong coupling

We now pass to the strong coupling phase, and use the form of the eigenvalue densities ρ_H and ρ_E at $\gamma > \gamma_c$.

Starting with $\mathcal{F}^{\text{sym.}}_{\text{u},H}$, we have:

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t}(\gamma > \gamma_{c,H}) &= \gamma \int \mathrm{d}\varphi \rho_H(\varphi) \left[\frac{e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}} + \frac{e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right] \\
&= \frac{4\gamma(\gamma-1)t}{\pi} \int_{-\phi_0}^{\phi_0} \mathrm{d}\varphi \frac{\cos\frac{\varphi}{2}\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2 - \left(\sin\frac{\varphi}{2}\right)^2}}{(1-t)^2 + 4t\left(\sin\frac{\varphi}{2}\right)^2} \left[\frac{\cos\varphi - t}{(1-t)^2 + 4t\left(\sin\frac{\varphi}{2}\right)^2}\right] \\
&= \frac{\gamma(\gamma-1)}{\pi t} \int_{-1}^{1} \mathrm{d}y \frac{\sqrt{1-y^2}\left(\frac{2t(\gamma-1)^2}{(1-t)(2\gamma-1)} - y^2\right)}{\left[\frac{(\gamma-1)^2}{(2\gamma-1)} + y^2\right]^2} \\
&= -\frac{1+t-4\gamma t}{2t(1-t)},
\end{aligned}$$

where we used the change of variables $y = \frac{\sin(\varphi/2)}{\sin(\phi_0/2)}$ and used the explicit form of $\sin(\phi_0/2)^2$. Immediate integration gives:

$$\mathcal{F}^{\text{sym.}}_{\text{u},H}(\gamma > \gamma_{c,H}) = -(2\gamma-1)\log(1-t) - \frac{1}{2}\log(t) + \mathcal{C}_H(\gamma),$$

with $\mathcal{C}_H(\gamma)$ a t-independent integration constant fixed by continuity at $\gamma = \gamma_{c,H}$.

For $\mathcal{F}^{\text{sym.}}_{\text{u},E}$ the calculations are almost the same, except for the change of sign in front of all factors of t and γ. We get:

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},E}}{\mathrm{d}t}(\gamma > \gamma_{c,E}) &= \gamma \int \mathrm{d}\varphi \rho_E(\varphi) \left[\frac{e^{\mathrm{i}\varphi}}{1+te^{\mathrm{i}\varphi}} + \frac{e^{-\mathrm{i}\varphi}}{1+te^{-\mathrm{i}\varphi}}\right] \\
&= \frac{\gamma(\gamma+1)}{\pi t} \int_{-1}^{1} \mathrm{d}y \frac{\sqrt{1-y^2}\left(\frac{2t(\gamma+1)^2}{(1+t)(2\gamma+1)} - y^2\right)}{\left[\frac{(\gamma+1)^2}{(2\gamma+1)} - y^2\right]^2} \\
&= -\frac{1-t-4\gamma t}{2t(1+t)}.
\end{aligned}$$

Notice that the final line here could not be inferred from the final line of $\mathcal{F}^{\text{sym.}}_{\text{u},H}$ with reversed signs. After integration we obtain the result:

$$\mathcal{F}^{\text{sym.}}_{\text{u},E}(\gamma > \gamma_{c,E}) = (2\gamma+1)\log(1+t) - \frac{1}{2}\log(t) + \mathcal{C}_E(\gamma), \tag{8.B.3}$$

where $\mathcal{C}_E(\gamma)$ is t-independent and fixed by continuity at the critical value $\gamma_{c,E}$.

8.B.3 Free energy for $\mathcal{Z}^{\text{gen.}}_{\text{u},H}$ and $\mathcal{Z}^{\text{gen.}}_{\text{u},E}$

The free energies for the matrix models (8.3.13) and (8.3.14) are computed at weak coupling in exactly the same manner as in Appendix 8.B.1. It is easy to check that, thanks to the parity of the integral, the free energies receive contribution from the terms proportional to v but they remain real. Direct computations give:

$$\frac{\mathrm{d}\mathcal{F}^{\text{gen.}}_{\text{u},H}}{\mathrm{d}t}(\gamma < \gamma_{c,H}) = \frac{2\gamma^2 t}{1-t^2}(1-v^2) = \frac{\mathrm{d}\mathcal{F}^{\text{gen.}}_{\text{u},E}}{\mathrm{d}t}(\gamma < \gamma_{c,E}).$$

At strong coupling we get:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{F}^{\text{gen.}}_{\text{u},H}}{\mathrm{d}t}(\gamma>\gamma_{c,H}) &= \gamma\int_{-\phi_0}^{\phi_0}\mathrm{d}\varphi\rho_H(\varphi)\left[\frac{(1-v)e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}}+\frac{(1+v)e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right]\\ &= \frac{4\gamma(\gamma-1)t}{\pi(1-t)}\int_{-\phi_0}^{\phi_0}\mathrm{d}\varphi\frac{\left((1-t)\cos\frac{\varphi}{2}-\mathrm{i}v(1+t)\sin\frac{\varphi}{2}\right)}{\left[(1-t)^2+4t\left(\sin\frac{\varphi}{2}\right)^2\right]^2}\\ &\times\sqrt{\left(\sin\frac{\phi_0}{2}\right)^2-\left(\sin\frac{\varphi}{2}\right)^2}\,(\cos\varphi-t-\mathrm{i}v\sin\varphi)\\ &= \frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t}(\gamma>\gamma_{c,H})-v^2\frac{16\gamma(\gamma-1)t(1+t)}{\pi(1-t)}\int_{-x_0}^{x_0}\mathrm{d}x\frac{x^2\sqrt{x_0^2-x^2}}{[(1-t)^2+4tx^2]^2}\\ &= \frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},H}}{\mathrm{d}t}(\gamma>\gamma_{c,H})-\frac{v^2(1+t)}{2t(1-t)},\end{aligned}$$

where in the second line we changed variables $x=\sin(\varphi/2)$, with boundary at $x_0=\sin(\phi_0/2)$, and in the last line we used the explicit form of x_0^2 to simplify the resulting expression. Notice that the term proportional to v^2 does not depend explicitly on γ.

Following the same steps, the free energy $\mathcal{F}^{\text{gen.}}_{\text{u},E}$ is computed at strong coupling as:

$$\begin{aligned}\frac{\mathrm{d}\mathcal{F}^{\text{gen.}}_{\text{u},E}}{\mathrm{d}t}(\gamma>\gamma_{c,E}) &= \gamma\int_{-\phi_0}^{\phi_0}\mathrm{d}\varphi\rho_H(\varphi)\left[\frac{(1-v)e^{\mathrm{i}\varphi}}{1-te^{\mathrm{i}\varphi}}+\frac{(1+v)e^{-\mathrm{i}\varphi}}{1-te^{-\mathrm{i}\varphi}}\right]\\ &= \frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},E}}{\mathrm{d}t}(\gamma>\gamma_{c,E})-v^2\frac{16\gamma(\gamma+1)t(1-t)}{\pi(1+t)}\int_{-x_0}^{x_0}\mathrm{d}x\frac{x^2\sqrt{x_0^2-x^2}}{[(1+t)^2-4tx^2]^2}\\ &= \frac{\mathrm{d}\mathcal{F}^{\text{sym.}}_{\text{u},E}}{\mathrm{d}t}(\gamma>\gamma_{c,E})-\frac{v^2(1-t)}{2t(1+t)}.\end{aligned}$$

8.C Solution to the saddle point equations on the real line

Here we consider the solution to the saddle point equation obtained from the large N limit of the matrix model after stereographic projection on the real line, as described in Section 8.4.2. We focus on $\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}$ and $\mathcal{Z}^{\text{gen.}}_{E,\text{stereo.}}$, and the solutions for the H-models can be obtained in an analogous way.

8.C.1 Solution for $\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}$

We consider first the matrix model $\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}$ defined in (8.4.2). In the limit in which the number of variables, K, is large, the leading contribution comes from the eigenvalue density $\rho_{\text{stereo.}}$ that solves the saddle point equation (8.4.9), which we report here for convenience:

$$\mathrm{P}\int\mathrm{d}y\frac{\rho_{\text{stereo.}}(y)}{x-y}=W(x),\qquad x\in\mathbb{R},\tag{8.C.1}$$

with

$$W(x)=-\frac{\gamma x}{x^2+t_0^2}+\frac{(\gamma+1)x}{x^2+1},\qquad t_0:=\frac{1+t}{1-t}.$$

We proceed following standard methods for the analysis of Hermitian matrix models at large K, as reviewed for example in [27]. We introduce the resolvent

$$\omega(z)=\int_{\mathcal{L}}\mathrm{d}y\frac{\rho_{\text{stereo.}}(y)}{z-y},\qquad z\in\mathbb{C}\setminus\mathcal{L},$$

where now, assuming a solution with symmetric support $[-A, A]$, the path $\mathcal{L}$ is the segment $[-A, A]$ on the real line. The eigenvalue density is recovered as the jump of the resolvent along $\mathcal{L}$:

$$\omega_+(z) - \omega_-(z) = -2\pi\mathrm{i}\rho_{\text{stereo.}}(x), \qquad x \in [-A, A],$$

with $\pm$ meaning the limit taken approaching the real line from the upper (resp. lower) half plane.

A solution for the resolvent is the following: write

$$\omega(z) = h(z)\Phi(z), \quad h(z) = \sqrt{(-A-z)(A-z)},$$

and the saddle point equation (8.C.1) becomes a discontinuity equation for $\Phi(z)$, with solution:

$$\Phi(z) = \oint_{\mathcal{C}} \frac{\mathrm{d}u}{2\pi\mathrm{i}} \frac{W(u)}{(z-u)h(u)},$$

where the contour $\mathcal{C}$ is a closed curve surrounding the cut $\mathcal{L}$, as in Figure 8.10.

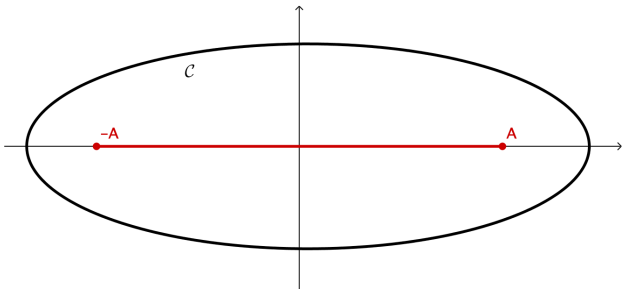


Figure 8.10. Integration contour in the complex plane. The red line is the cut $\mathcal{L}$, the black curve represents a choice of contour $\mathcal{C}$.

$W(u)$ is a meromorphic function, with simple poles at:

$$u = z_p, \quad z_p \in \{\pm\mathrm{i},\ \pm\mathrm{i}t_0\}\,.$$

We can thus deform the contour, to avoid the branch cut of the square root, and pick the poles of the integrand. This leaves a residual integral along an infinitely large circle. This latter contribution vanishes, since $W(u) \sim 1/u$ at large $|u|$. The pole at $u = z$ generates the regular part of $\omega(z)$, thus yields no contribution to the eigenvalue density. Therefore, the relevant contributions to the ρ arise from the poles of $W(u)$ in the complex plane. For the calculations, one has to be careful with the signs in front of the square roots, according to the definition of $h(z)$. With our conventions,

$$h(\mathrm{i}) = -\mathrm{i}\sqrt{A^2+1} = -h(-\mathrm{i}),$$
$$h(\mathrm{i}t_0) = -\mathrm{i}\sqrt{A^2+t_0^2} = -h(-\mathrm{i}t_0).$$

After simple computations, one arrives to:

$$\sum_{z_p} \mathrm{Res}_{u=z_p} \frac{W(u)}{(z-u)\sqrt{u^2-A^2}} = \frac{\gamma t_0}{(z^2+t_0^2)\sqrt{A^2+t_0^2}} - \frac{(\gamma+1)}{(z^2+1)\sqrt{A^2+1}}.$$

The eigenvalue density is obtained this function multiplied by the jump of $h(z)$:

$$\rho_{\text{stereo.}}(x) = \frac{\sqrt{A^2-x^2}}{\pi}\left[\frac{(1+\gamma)}{\sqrt{A^2+1}(x^2+1)} - \frac{\gamma t_0}{\sqrt{A^2+t_0^2}(x^2+t_0^2)}\right].$$

The value of the boundary A can be fixed by normalization, or equivalently looking at the large z behaviour of the resolvent $\omega(z)$. From the definition, one has $\omega(z) \sim 1/z$, as $z \to \infty$, while from the explicit evaluation, we have:

$$\begin{aligned}\omega(z) &= W(z) + \sqrt{z^2 - A^2}\left(\frac{\gamma t_0}{(z^2+t_0^2)\sqrt{A^2+t_0^2}} - \frac{(\gamma+1)}{(z^2+1)\sqrt{A^2+1}}\right) \\ &= \frac{1}{z}\left(1 + \frac{\gamma t_0}{\sqrt{A^2+t_0^2}} - \frac{\gamma+1}{\sqrt{A^2+1}}\right) + \mathcal{O}(z^{-2}).\end{aligned}$$

This provides the solution (recall that $t_0 = (1+t)/(1-t)$)

$$A^2 = \frac{(2\gamma+1)(1+t)^2}{(2\gamma+1-t)(2\gamma-1+t)}.$$

This is a positive quantity, and thus $\rho_{\text{stereo.}}$ is supported on the real line, as long as

$$\gamma > \frac{1-t}{2t}.$$

This is consistent with our general analysis: due to the change of topology, the stereographic projection should only provide the gapped phase of the unitary matrix model. We also notice that, undoing the stereographic projection, we obtain:

$$e^{\mathrm{i}\phi_0} = \frac{1+\mathrm{i}A}{1-\mathrm{i}A} \quad \Longrightarrow \quad \left(\sin\frac{\phi_0}{2}\right)^2 = \frac{(2\gamma+1)(1+t)^2}{4t(\gamma+1)^2},$$

thus the boundary we obtain on the real line is in fact the stereographic projection of the boundary of the model on the unit circle, as expected.

8.C.2 Solution for $\mathcal{Z}^{\text{gen.}}_{E,\text{stereo.}}$

In this Appendix, we study the large K limit of the matrix model $\mathcal{Z}^{\text{gen.}}_{E,\text{stereo.}}$ defined in (8.4.4). The saddle point equation is:

$$\mathrm{P}\int \mathrm{d}y \frac{\rho_{\text{stereo.}}(y)}{x-y} = W(x), \qquad x \in \mathbb{R}, \tag{8.C.2}$$

with $W(x)$ in the present, more general case given by:

$$W(x) = -\frac{\gamma x}{x^2+t_0^2} + \frac{(\gamma+1)x}{x^2+1} + 4\mathrm{i}v\gamma\frac{t}{1-t}\frac{x^2-t_0}{(x^2+1)(x^2+t_0^2)}.$$

From this expression, it is clear that the asymmetry in the integrand introduces an imaginary part in the eigenvalue density, but does not introduce new poles of $W(u)$ in the complex plane.

We proceed as in the previous Appendix, and find:

$$\omega(z) = W(z) + h(z)\sum_{z_p}\mathrm{Res}_{u=z_p}\frac{W(u)}{(z-u)\sqrt{u^2-A^2}},$$

where now the residues at $u = \pm\mathrm{i}, \pm\mathrm{i}t_0$ yield an extra term, proportional to the parameter v. After some simple calculations, we see that this extra contribution is:

$$2\mathrm{i}v\gamma\left(\frac{z}{(z^2+t_0^2)\sqrt{A^2+t_0^2}} - \frac{z}{(z^2+1)\sqrt{A^2+1}}\right).$$

The final expression for the eigenvalue density is:

$$\rho_{\text{stereo.}}(x) = \frac{\sqrt{A^2-x^2}}{\pi}\left[\frac{(1+\gamma)+\mathrm{i}v\gamma x}{\sqrt{A^2+1}(x^2+1)} - \frac{\gamma t_0 + \mathrm{i}v\gamma x}{\sqrt{A^2+t_0^2}(x^2+t_0^2)}\right],$$

Imposing normalization, from the parity of the integral we obtain the same value of A as above.

8.C.3 Solution for $\mathcal{Z}^{\text{sym.}}_{E,\text{stereo.}}$ with modified symbol

We now solve the saddle point equation (8.C.1) with, on the right hand side, the modified function

$$W(x) = -\frac{\gamma x}{x^2+t_0^2} + \frac{(\gamma+1)x-2b}{x^2+1}.$$

The procedure follows closely that of Appendix 8.C.1, but now we do not assume symmetry of the support for $\rho_{\text{stereo.}}$, and let

$$\text{supp}\rho_{\text{stereo.}} = [-A, B].$$

We therefore modify the definition of the auxiliary function

$$h(z) = \sqrt{(-A-z)(B-z)}.$$

For later convenience, we also introduce the functions

$$\begin{aligned}\tilde{h}(x) &= \sqrt{|x^2+AB+\mathrm{i}x(B-A)|} = \left(x^4+A^2B^2+x^2(A^2+B^2)\right)^{\frac{1}{4}},\\ \tilde{\theta}(x) &= \frac{1}{2}\arg\left(x^2+AB+\mathrm{i}x(B-A)\right) = \frac{1}{2}\arctan\frac{x(B-A)}{x^2+AB},\end{aligned} \tag{8.C.3}$$

for $x \in \mathbb{R}$. Note also that $\tilde{h}(-x) = \tilde{h}(x)$ and $\tilde{\theta}(-x) = -\tilde{\theta}(x)$. Following Appendix 8.C.1, we introduce the resolvent $\omega(z)$ and arrive at:

$$\omega(z) = W(z) + h(z)\sum_{z_p}\text{Res}_{u=z_p}\frac{W(u)}{(z-u)h(u)},$$

with poles at $z_p = \pm\mathrm{i}, \pm\mathrm{i}t_0$. Using

$$h(\mathrm{i}) = -\mathrm{i}\tilde{h}(1)e^{\mathrm{i}\tilde{\theta}(1)}, \quad h(-\mathrm{i}) = \mathrm{i}\tilde{h}(1)e^{-\mathrm{i}\tilde{\theta}(1)},$$

and similarly for $\pm\mathrm{i}t_0$, we arrive at:

$$\begin{aligned}\sum_{z_p}\text{Res}_{u=z_p}\frac{W(u)}{(z-u)h(u)} = &\frac{\gamma\left[t_0\cos\tilde{\theta}(t_0) - z\sin\tilde{\theta}(t_0)\right]}{\tilde{h}(t_0)(z^2+t_0^2)}\\ &-\frac{\left[(\gamma+1)\cos\tilde{\theta}(1)+2b\sin\tilde{\theta}(1)\right]+z\left[2b\cos\tilde{\theta}(1)-(\gamma+1)\sin\tilde{\theta}(1)\right]}{\tilde{h}(1)(z^2+1)},\end{aligned}$$

and the eigenvalue density is:

$$\begin{aligned}\rho_{\text{stereo.}}(x) = \frac{\sqrt{(x-A)(B-x)}}{\pi}\Bigg[&\frac{\gamma\left[t_0\cos\tilde{\theta}(t_0) - z\sin\tilde{\theta}(t_0)\right]}{\tilde{h}(t_0)(z^2+t_0^2)}\\ &-\frac{\left[(\gamma+1)\cos\tilde{\theta}(1)+2b\sin\tilde{\theta}(1)\right]+z\left[2b\cos\tilde{\theta}(1)-(\gamma+1)\sin\tilde{\theta}(1)\right]}{\tilde{h}(1)(z^2+1)}\Bigg]\end{aligned}$$

for $x \in [-A, B]$. The boundaries of the support are fixed by imposing the condition $\omega(z) \sim 1/z$ at large z. Using

$$\cos 2\tilde{\theta}(x) = \frac{x^2+AB}{\tilde{h}(x)^2},$$

order z^0 in the expansion of $\omega(z)$ gives the constraint

$$\frac{\gamma}{2\tilde{h}(t_0)}\left[\frac{t_0^2+AB}{\tilde{h}(t_0)^2}-1\right]-\frac{1}{2\tilde{h}(1)}\left[\left(\frac{1+AB}{\tilde{h}(1)^2}\right)(2b-\gamma-1)+2b+\gamma+1\right]=0,$$

while from order z^1, and using the previous expression to simplify the equation, we get the constraint

$$\frac{\gamma t_0}{2\tilde{h}(t_0)}\left[\frac{t_0^2+AB}{\tilde{h}(t_0)^2}+1\right]-\frac{1}{2\tilde{h}(1)}\left[\left(\frac{1+AB}{\tilde{h}(1)^2}\right)(2b+\gamma+1)-2b+\gamma+1\right]=0.$$

Chapter 9

Phases of unitary matrix models and their meromorphic deformations

9.1 Introduction to the chapter

The study of the critical properties of models of random matrices has become a widely popular and interdisciplinary subject in this century, in great part due to the vast scope of fields where such models appear naturally and play a prominent role [40, 41, 283].

A paradigmatic example of this phenomenon could very well be the so-called Gross–Witten–Wadia (GWW) model [190–192]. Originally proposed in the study of gauge theory, it is ubiquitous and pivotal in many other areas, such as combinatorics, representation theory and spectral theory [40, 41, 283].

With this fact in mind, in this chapter we will study unitary matrix models, starting with a generalized form of the Gross–Witten–Wadia model. A possible interpretation is as a one-plaquette model of two-dimensional lattice QCD with fermionic or bosonic quarks, which equivalently corresponds to a massive deformation of a model introduced by Minahan [331, 332].

We will show that, while quite simple, this model retains several features of a sensible quantum field theory in the continuum. In turn, its simplicity allows us to exploit standard techniques from random matrix theory to characterize the theory at large N and suggests more general and deeper problems to consider. Some of them, we will already tackle here, by discussing at length the case of meromorphic deformations of unitary matrix model, as we explain below.

Before that, a rich phase diagram will be obtained and analyzed in detail. Phase transitions such as the ones we obtain in our analysis are relevant to the study of deconfinement transitions in QCD models in four dimensions [333, 334] and in black holes physics [335–337]. More recently, this type of phase transitions has been argued to describe the critical behaviour of models exhibiting partial deconfinement [338–342].

We introduce the model in what follows, in Section 9.2, which includes a discussion on interpretations of the model, notation, relationship with other systems and an introductory discussion of its mathematical properties, including exact evaluations without scaling limits.

Then, the main results of the chapter are presented and organized as follows: there are three main contributions, as far as new results are concerned. In Section 9.3, we fully characterize the rich phase structure of the unitary matrix model. In Section 9.4, we study Wilson loops in the same setting of Section 9.3 and discuss at length the physical interpretation of the phase transitions, including the role of instanton contributions.

Finally, in Section 9.5 we study, in a general framework and going beyond the specific model studied in the previous sections, the case where the integration contour is deformed in $\mathbb{C}^*$ away from the unit circle. In spite of the vast body of results on random matrix ensembles, holomorphic

matrix models [343] are arguably understudied.

The aim of Section 9.5 is to adapt the results on holomorphic matrix models to unitary matrix models. We do so for a very general set of unitary matrix models and, only as an illustrative example, we discuss the particular case of the holomorphic GWW matrix model. Non-traditional tools in this area, such as Hasse diagrams, are introduced to fully understand the meromorphic models.

9.2 The model

In this section we present the model and study some of its exact features at finite N.

9.2.1 The model and its interpretations

Consider a one-plaquette model of two-dimensional lattice gauge theory [344, 22] with gauge group $U(N)$ and K pairs of real fields, that can be either bosonic or fermionic. Each pair is called a flavour. We encode the choice of matter fields in the binary variable

$$\epsilon = \begin{cases} +1 & \text{fermions,} \\ -1 & \text{bosons.} \end{cases}$$

We must impose (anti-)periodic boundary conditions, so the discrete space-time is effectively reduced to a point with a loop attached to it, along which the gauge connection travels. Let $m_f > 0$ be the mass of the f^{th} flavour, and introduce the notation

$$\mu_f^{(\epsilon)} = 1 + \sqrt{\epsilon} m_f .$$

The partition function of this theory has the matrix model representation [331]

$$\mathcal{Z}^{\epsilon,K}_{U(N)}(\lambda) = \int_{U(N)} \mathrm{d}U \ \prod_{f=1}^{K} \left[\det\left(\mu_f^{(\epsilon)} - U\right) \det\left(\mu_f^{(\epsilon)} - U\right)^{\dagger} \right]^{\epsilon} \ e^{\frac{N}{\lambda}\left(\operatorname{Tr} U + \operatorname{Tr} U^{\dagger}\right)}$$

where $\lambda \equiv N g_{\text{YM}}^2$ is the 't Hooft coupling for the bare gauge coupling g_{YM}, $U \in U(N)$ is the plaquette gauge variable and $\mathrm{d}U$ is the normalized Haar measure on $U(N)$. Particularizing to the degenerate case with all equal masses, $\mu_f = \mu \ \forall f = 1, \dots, K$, the partition function becomes:

$$\begin{aligned} \mathcal{Z}^{\epsilon,K}_{U(N)}(\lambda) &= \int_{U(N)} \mathrm{d}U \ \left[\det(\mu - U) \det(\mu - U)^{\dagger} \right]^{\epsilon K} \ e^{\frac{N}{\lambda}\left(\operatorname{Tr} U + \operatorname{Tr} U^{\dagger}\right)} \\ &= \oint_{\mathbb{T}^N} \prod_{1 \le j < k \le N} |z_j - z_k|^2 \ \prod_{j=1}^{N} \left[(\mu - z_j)\left(\bar{\mu} - z_j^{-1}\right) \right]^{\epsilon K} \ e^{\frac{N}{\lambda}\left(z_j + z_j^{-1}\right)} \ \frac{\mathrm{d}z_j}{2\pi \mathrm{i} z_j} . \end{aligned} \tag{9.2.1}$$

The second line is written in terms of the eigenvalues $z_j \in \mathbb{T}$ of $U \in U(N)$.

The matrix model (9.2.1) is a massive deformation of the model studied in [345, 346]. Besides, (9.2.1) is a generalization of the celebrated Gross–Witten–Wadia (GWW) model [190, 192] by determinant insertions, and reduces to it for $K = 0$ or $\mu \to \infty$. For $\lambda^{-1} = 0$ the model (9.2.1) is a particular case of a correlator of characteristic polynomials in a Circular Unitary ensemble (CUE) [347], a fundamental object in random matrix theory with many applications [348, 349]. Likewise, for this value of $\lambda^{-1} = 0$, it generalizes a matrix model that describes non-intersecting random walks [302], and has also appeared in gauge theory, for instance in [350] and later on in somewhat disguised forms.

Integrable systems interpretation

It is worth mentioning a possible interpretation of (9.2.1), from the point of view of integrable systems [351]. The study of the so-called Schur flow [352, 353], analogous to Toda flows but on the unit circle, precisely entails the generalization of a given weight function of the matrix model by multiplication by the GWW weight function.

This induces a flow that has many implications. For example, the recurrence coefficients of the polynomials, orthogonal with regards to the weight function of the matrix model, satisfy the non-linear Ablowitz–Ladik equation, with the parameter $\frac{1}{g_{\mathrm{YM}}^2} = \frac{N}{\lambda}$ interpreted as time.

Because of this and since our analysis is in a planar limit and centered around the matrix model, the results obtained are not obviously transferable into this integrable systems and spectral theory language [354].

Gauge theory interpretation

We consider (9.2.1) as a toy model for lattice two-dimensional QCD, although we will comment on some of the many other interpretations of the model. For example, (9.2.1) can be regarded as an effective description of two-dimensional QCD on a small spatial circle [350]. In fact, compactifying the spatial direction generates a mass gap for all but the zero-modes. Taking the small circumference limit, we are left with an effective theory with integration only over the gauge and matter zero-modes.

It was observed in [345] that the massless theory with fermions, that is $\epsilon = +1$ and $\mu = 1$, shows a Fisher–Hartwig (FH) singularity. The theory with bosons, on the contrary, yields a singular matrix model in the massless case. Here we recognize the singularities encountered in [345] as the remnants of the IR singularities due to massless fields, and resolve them via mass deformation.

On the parameter μ

Notice that there is a slight difference in the definition of μ between the fermionic and the bosonic theory in (9.2.1). In the first case $\mu > 1$ is real, while in the second case $\mu \in \mathbb{C}$ with $|\mu| > 1$. It is easy to see that the phase of $\mu \in \mathbb{C}$ can be reabsorbed in a rotation of the integration contour $\mathbb{T}$, and we henceforth restrict our attention to a real $\mu > 1$ in both cases, with the understanding that for bosons μ really means $|\mu|$.

Besides, treating μ as a real variable with this caveat in mind, the integrand in (9.2.1) is analytic in $\mu > 1$, and therefore the results for any other $\mu \in \mathbb{C}$ with $|\mu| > 1$ are obtained from analytic continuation of our results. In particular, we could not attain negative values of μ moving along the real line, because we would cross the FH singularity. Nevertheless, it is possible to take a path from $\mu > 1$ to any $\mu' < -1$ that runs in the complex plane outside the unit disk.

Remark. The independence of the partition function on $\arg \mu$ is the $U(1)$ freedom to choose the origin of $\mathbb{T}$, and is the incarnation of the residual diagonal $U(1) \subset U(N)$ gauge symmetry. Our choice $\arg \mu \in 2\pi\mathbb{Z}$ fixes this residual gauge freedom.

Notation

We introduce the notation

$$Y = \frac{1}{\lambda}, \qquad \tau = \epsilon \frac{K}{N}$$

for, respectively, the inverse of the 't Hooft coupling and a real Veneziano parameter, whose sign carries information on the type of fields we consider.

The parameter space of the theory is

$$\mathfrak{M} = \left\{(\mu, \tau, Y) \in (1, \infty) \times \mathbb{R}^2\right\}.$$

Remark. For the role of the mass as a regulator (for the FH singularities on the mathematical side, for the IR singularities on the field theory side), we do not expect the continuation from $\mathfrak{M}$ to the sheet $\{\mu = 1\} \times \mathbb{R}^2$, studied in [345, 346], to be analytic.

9.2.2 Exact finite N evaluations

We now present various analytical results for the matrix model (9.2.1).

It is possible to evaluate the partition function of any unitary matrix model at finite N via the Heine–Szegő identity, that for (9.2.1) gives

$$\mathcal{Z}^{\epsilon,K}_{U(N)}(\lambda, \mu) = N! \det_{1\le j,k\le N} [\mathbf{Z}_{jk}], \tag{9.2.2}$$

with

$$\mathbf{Z}_{jk} = \sum_{p=0}^{K} \frac{K!}{p!(K-p)!} (-\mu)^p (1+\mu^2)^{K-p} \frac{\mathrm{d}^p}{\mathrm{d}x^p} I_{k-j}(2x)\Big|_{x=\frac{N}{\lambda}}$$

where $I_k(x)$ is the modified Bessel function. We have simplified the expression assuming $\epsilon = +1$, although a similar determinant expression exists for $\epsilon = -1$ as well. Formula (9.2.2) allows to efficiently compute $\mathcal{Z}^{\epsilon,K}_{U(N)}$ exactly for fixed N and K, cf. [7, App. A1].

The partition function for generic masses, encoded in the parameters $(\mu_1, \dots, \mu_K)$, can also be related to the expectation value of Wilson loops in arbitrary representations in the pure GWW model, thanks to the Cauchy identity (see Chapter 2). For the fermionic theory we write

$$\begin{aligned}
\mathcal{Z}^{+1,K}_{U(N)} &= \left(\prod_{f=1}^{K} \mu_f^2\right) \oint_{\mathbb{T}^N} \prod_{1\le j<k\le N} |z_j - z_k|^2 \prod_{j=1}^{N} \left[\prod_{f=1}^{K} \left(1 - \frac{z_j}{\mu_f}\right)\left(1 - \frac{\bar{z}_j}{\mu_f}\right)\right] e^{\frac{N}{\lambda}(z_j+\bar{z}_j)} \frac{\mathrm{d}z_j}{2\pi \mathrm{i} z_j} \\
&= \left(\prod_{f=1}^{K} \mu_f^2\right) \sum_{\mathcal{R}_1} \sum_{\mathcal{R}_2} \mathfrak{s}_{\mathcal{R}_1'}(-\mu_1^{-1}, \dots, -\mu_K^{-1}) \mathfrak{s}_{\mathcal{R}_2'}(-\mu_1^{-1}, \dots, -\mu_K^{-1}) \\
&\times \oint_{\mathbb{T}^N} \prod_{1\le j<k\le N} |z_j - z_k|^2 \, \mathfrak{s}_{\mathcal{R}_1}(z_1, \dots, z_N) \mathfrak{s}_{\mathcal{R}_1}(\bar{z}_1, \dots, \bar{z}_N) \prod_{j=1}^{N} e^{\frac{N}{\lambda}(z_j+\bar{z}_j)} \frac{\mathrm{d}z_j}{2\pi \mathrm{i} z_j}
\end{aligned}$$

where the sum runs over Young diagrams $\mathcal{R}$ of length at most N and the first row at most K, $\mathcal{R}'$ is the conjugate diagram, and $\mathfrak{s}_{\mathcal{R}}$ is the corresponding Schur polynomial, that is, the character of the irreducible representation associated to the diagram $\mathcal{R}$. In the last line we identify the correlator of two Wilson loops in the pure GWW model,

$$\frac{\mathcal{Z}^{+1,K}_{U(N)}}{\mathcal{Z}^{(\mathrm{GWW})}_{U(N)}} = \left(\prod_{f=1}^{K} \mu_f^2\right) \sum_{\mathcal{R}_1} \sum_{\mathcal{R}_2} \mathfrak{s}_{\mathcal{R}_1'}(-\mu_1^{-1}, \dots, -\mu_K^{-1}) \mathfrak{s}_{\mathcal{R}_2'}(-\mu_1^{-1}, \dots, -\mu_K^{-1}) \, \langle \mathcal{W}_{\mathcal{R}_1} \overline{\mathcal{W}}_{\mathcal{R}_2} \rangle^{(\mathrm{GWW})}, \tag{9.2.3}$$

with $\overline{\mathcal{W}}$ meaning that the Wilson loop involves conjugated variables. We can in principle further expand the product of the two Schur polynomials with the Littlewood–Richardson rule, but this would entail inverting the variables $\bar{z}_1, \dots, \bar{z}_N$ in the second Schur polynomial, as in [1].

Expression (9.2.3) is suggestive but not very useful as it is, because the vacuum expectation value (vev) of a Wilson loop in a generic representation is not known in closed form. Nevertheless, we can go deeper in the character expansion thanks to the formula [355]

$$\exp\left(\frac{N}{\lambda}\operatorname{Tr}U\right)=\sum_{\mathcal{R}}\left(\frac{N}{\lambda}\right)^{|\mathcal{R}|}\dim\mathcal{R}\left(\prod_{j=1}^{N}\frac{(N-j)!}{(N-j+\mathcal{R}_j)!}\right)\mathfrak{s}_{\mathcal{R}}(z_1,\dots,z_N).$$

Using this equation and its conjugate and applying twice the Littlewood–Richardson rule we get

$$\begin{aligned}\mathcal{Z}^{+1,K}_{U(N)}=&\left(\prod_{f=1}^{K}\mu_f^2\right)\sum_{\mathcal{R}_1,\mathcal{R}_2,\mathcal{R}_3,\mathcal{R}_4}\mathfrak{s}_{\mathcal{R}_1'}(-\mu_1^{-1},\dots,-\mu_K^{-1})\mathfrak{s}_{\mathcal{R}_2'}(-\mu_1^{-1},\dots,-\mu_K^{-1})\\&\times\left(\frac{N}{\lambda}\right)^{|\mathcal{R}_3|+|\mathcal{R}_4|}\dim\mathcal{R}_3\dim\mathcal{R}_4\left[\prod_{j=1}^{N}\frac{((N-j)!)^2}{(N-j+\mathcal{R}_{3,j})!(N-j+\mathcal{R}_{4,j})!}\right]c_{13;24}\end{aligned}$$

where $c_{13;24}\equiv\sum_{\mathcal{R}}c^{\mathcal{R}}_{\mathcal{R}_1\mathcal{R}_3}c^{\mathcal{R}}_{\mathcal{R}_2\mathcal{R}_4}$, with $c^{\mathcal{R}}_{\mathcal{R}_j\mathcal{R}_k}$ the Littlewood–Richardson coefficients, and we have used the orthogonality of the Schur polynomials. The bosonic model $\mathcal{Z}^{-1,K}_{U(N)}$ admits a closely related expression, with the partitions $\mathcal{R}_1$ and $\mathcal{R}_2$ instead of their conjugate in the first line and dropping the restriction on the first rows.

Finally, there is an additional, simpler although more formal closed form expression for $\mathcal{Z}^{\epsilon,K}_{U(N)}$ [351]

$$\mathcal{Z}^{\epsilon,K}_{U(N)}=\sum_{\mathcal{R}}\left(\mathfrak{s}_{\mathcal{R}}\left(NY-\epsilon\frac{K}{\mu},-\epsilon\frac{K}{\mu^2},-\epsilon\frac{K}{\mu^3},\dots\right)\right)^2$$

where now the Schur functions must be interpreted as characters of $U(\infty)$, and the sum runs over Young diagrams $\mathcal{R}$ with at most N rows.

9.3 Phase structure

This section is dedicated to the large N analysis of the matrix model (9.2.1) and the determination of its phase diagram.

Write the partition function (9.2.1) as

$$\mathcal{Z}_{U(N)}=\oint_{\mathbb{T}^N}e^{-N^2S_{\text{eff}}(z_1,\dots,z_N)}\prod_{j=1}^{N}\frac{\mathrm{d}z_j}{2\pi\mathrm{i}z_j}\tag{9.3.1}$$

where the effective action S_{eff} is the sum of a potential and the Coulomb interaction between eigenvalues:

$$\begin{aligned}S_{\text{eff}}\left(z_1,\dots,z_N\right)&=\frac{1}{N}\sum_{j=1}^{N}V_{\text{eff}}(z_j)+\frac{1}{N^2}\sum_{j=1}^{N}\sum_{k\neq j}V_{\text{int}}\left(z_j,z_k\right)\\V_{\text{eff}}\left(e^{\mathrm{i}\theta}\right)&=-\frac{2}{\lambda}\cos\theta-\tau\log\left(1+\mu^2-2\mu\cos\theta\right)\\V_{\text{int}}\left(e^{\mathrm{i}\theta},e^{\mathrm{i}\varphi}\right)&=-\log 2\sin\left(\frac{\theta-\varphi}{2}\right).\end{aligned}$$

writing $z\in\mathbb{T}$ as $z=e^{\mathrm{i}\theta}$, $-\pi<\theta\leq\pi$.

The potential $V_{\text{eff}}\left(e^{\mathrm{i}\theta}\right)$ admits isolated minima at each point in $\mathfrak{M}$. Besides, there exists a surface $\{\lambda = \lambda_*(\mu, \lambda)\} \subset \mathfrak{M}$ at which it passes from a single-well to a double-well profile. Explicitly, these two regimes are separated by

$$\lambda_*(\mu, \tau) = \frac{(\mu - 1)^2}{\tau\mu},$$

and the potential develops stationary points at $\theta = \pm\theta_*$ with

$$\tan\theta_* = \pm\frac{\sqrt{\mu^2\left(2 - \lambda^2\tau^2\right) + 2\lambda\mu^3\tau + 2\lambda\mu\tau - \mu^4 - 1}}{-\lambda\mu\tau + \mu^2 + 1}.$$

We stress that λ_* is not a critical value of the model (9.2.1).

The potential is plotted in Figure 9.1 ($\tau > 0$) and in Figure 9.2 ($\tau < 0$). Clearly, the two figures have the same shape but upside down. Nonetheless, it is important to distinguish the role of minima to understand the phase structure.

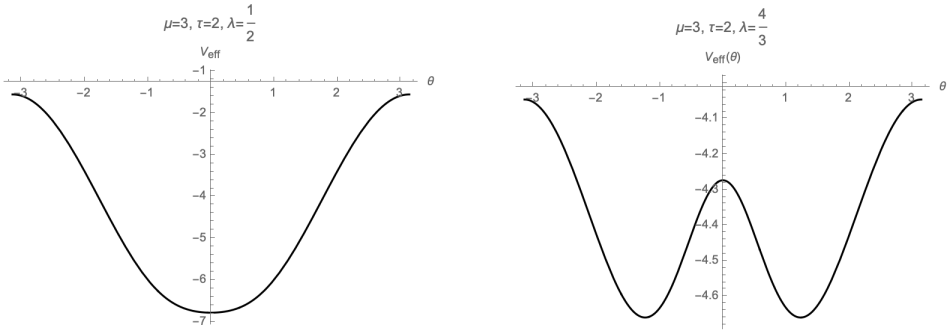


Figure 9.1. $V_{\text{eff}}(e^{\mathrm{i}\theta})$ at $\mu = 3$ and $\tau = 2$. Left: $\lambda = \frac{1}{2}$. Right: $\lambda = \frac{4}{3}$.

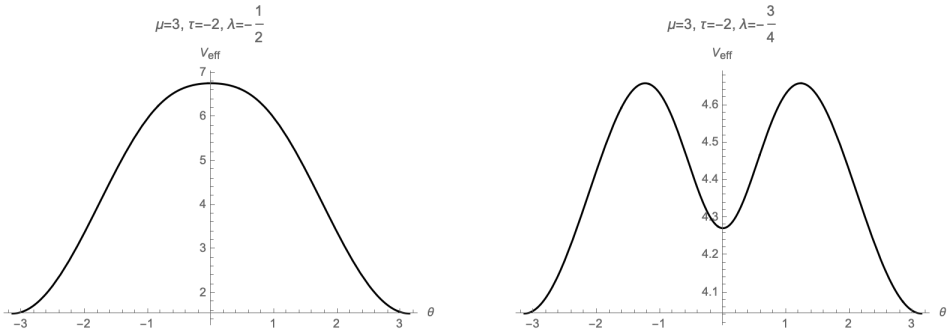


Figure 9.2. $V_{\text{eff}}(e^{\mathrm{i}\theta})$ at $\mu = 3$ and $\tau = -2$. Left: $\lambda = -\frac{1}{2}$. Right: $\lambda = -\frac{4}{3}$.

Now that we have set the ground, we are ready to discuss the large N limit of the model (9.2.1).

9.3.1 Large N

We now take the large N 't Hooft and Veneziano limit of the partition function (9.3.1). This means that we consider the planar limit with both λ and τ fixed. The leading contributions to the integral at large N come from the saddle points of the effective action:

$$\frac{\partial S_{\text{eff}}}{\partial\theta_j} = 0 \qquad j = 1, \ldots, N. \tag{9.3.2}$$

Introducing the eigenvalue density

$$\rho(\theta) = \frac{2\pi}{N} \sum_{j=1}^{N} \delta\left(e^{\mathrm{i}\theta} - e^{\mathrm{i}\theta_j}\right),$$

with normalization chosen so that

$$\int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \rho(\theta) = 1, \tag{9.3.3}$$

we can collect the system (9.3.2) of N coupled equations in a single singular integral equation at large N. The saddle point equation then reads

$$\mathrm{P} \int_{-\pi}^{\pi} \frac{\mathrm{d}\varphi}{2\pi} \rho(\varphi) \ \cot\left(\frac{\theta - \varphi}{2}\right) = 2Y \sin\theta - \frac{2\mu\tau \sin\theta}{1 + \mu^2 - 2\mu\cos\theta}. \tag{9.3.4}$$

The solution to (9.3.4) must satisfy the non-negativity constraint

$$\rho(\theta) \geq 0, \qquad -\pi < \theta \leq \pi, \tag{9.3.5}$$

that follows from the compactness of the integration domain.

Ungapped solution: Phase 0

We begin assuming $\rho(\theta)$ is supported on the whole circle, $-\pi < \theta \leq \pi$. We exploit $\mu > 1$ to obtain the solution [356]

$$\rho_0(\theta) = 1 + 2Y\cos\theta - 2\tau \frac{\mu\cos\theta - 1}{1 + \mu^2 - 2\mu\cos\theta}. \tag{9.3.6}$$

The derivation is standard [190, 6], thus we omit it. We call Phase 0 the region of $\mathfrak{M}$ for which the solution (9.3.6) is valid.

Critical loci

In those regions of $\mathfrak{M}$ for which the solution (9.3.6) violates the constraint (9.3.5), we should drop the assumption $\mathrm{supp}\rho = (-\pi, \pi]$ and look for a new solution, whose support has one or more gaps on the unit circle. The arcs on which ρ is supported are called cuts.

We find a phase transition with a gap opening at $\theta = \pm\pi$ at the critical surface

$$Y_{\mathrm{cr,a}} = \frac{1}{2} + \frac{\tau}{\mu + 1}. \tag{9.3.7}$$

Another phase transition, with a gap opening at $\theta = 0$, takes place at the critical surface

$$Y_{\mathrm{cr,b}} = -\frac{1}{2} + \frac{\tau}{\mu - 1}. \tag{9.3.8}$$

Besides, there exists a multi-critical point at the value $\tau = \tau_{\mathrm{cr},+}(\mu)$ at which $\rho_0(\pm\pi) = 0 = \rho_0(0)$, determined as the unique point at which $Y_{\mathrm{cr,a}}$ and $Y_{\mathrm{cr,b}}$ meet:

$$\tau_{\mathrm{cr},+}(\mu) = \frac{\mu^2 - 1}{2}. \tag{9.3.9}$$

Examples of the limiting cases of $\rho_0(\theta)$ are shown in Figure 9.3.

Besides the two critical surfaces just described, looking at $\rho_0(\theta)$ for negative Y and τ we also find values at which it attains zero value at two distinct, symmetric points in the interior of $(-\pi, \pi)$, as in Figure 9.4. We expect a new phase transition into a two-cut solution.

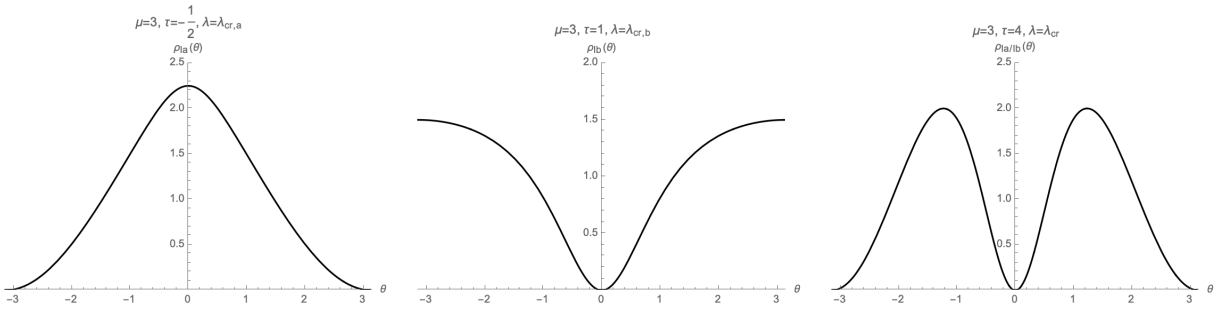


Figure 9.3. $\rho_0(\theta)$ at the transition point. Left: $\mu = 3$, $\tau = -\frac{1}{2}$ and $\lambda = \lambda_{\mathrm{cr,a}}$. Centre: $\mu = 3$, $\tau = 1$ and $\lambda = \lambda_{\mathrm{cr,b}}$. Right: $\mu = 3$, $\tau = 4$ and $\lambda = \frac{2}{3}$.

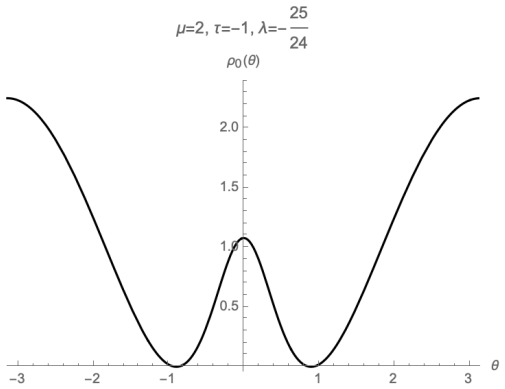


Figure 9.4. $\rho_0(\theta)$ at the transition point. $\mu = 2$, $\tau = -1$ and $\lambda = -\frac{25}{24}$. The support of ρ will break in two disjoint cuts beyond this critical value.

One-cut solution: Phase Ia

We now solve Equation (9.3.4) dropping the assumption that $\rho(\theta)$ is supported on the whole $\mathbb{T}$, and replace it by the assumption that the support is an arc $\Gamma \subset \mathbb{T}$. The derivation is standard and we relegate it to Appendix 9.A.

Introduce the trace of the resolvent in the large N limit,

$$\omega(z) = \int_\Gamma \frac{\mathrm{d}w}{2\pi \mathrm{i} w} \varrho(w) \frac{z+w}{z-w}, \qquad z \in \mathbb{C} \setminus \Gamma. \tag{9.3.10}$$

We adopt the standard notation

$$\omega_\pm(e^{\mathrm{i}\theta}) \equiv \lim_{\varepsilon \to 0^+} \omega(z = (1 \pm \varepsilon) e^{\mathrm{i}\theta}).$$

Then

$$\omega_+(e^{\mathrm{i}\theta}) - \omega_-(e^{\mathrm{i}\theta}) = 2\varrho(e^{\mathrm{i}\theta}), \qquad e^{\mathrm{i}\theta} \in \Gamma.$$

We find (see Appendix 9.A for the details)

$$\omega_{\mathrm{Ia}}(z) = -\mathrm{i}W(z) + \sqrt{(e^{\mathrm{i}\theta_0} - z)\,(e^{-\mathrm{i}\theta_0} - z)} \left[Y \left(1 + \frac{1}{z}\right) - \frac{\tau}{\sqrt{1 + \mu^2 - 2\mu\cos\theta_0}} \left(\frac{\mu}{z - \mu} - \frac{1}{z - \mu^{-1}} \right) \right].$$

The first term is regular and, taking the discontinuity at $z = e^{\mathrm{i}\theta} \in \Gamma$, we arrive at

$$\rho_{\mathrm{Ia}}(\theta) = 2\cos\frac{\theta}{2} \cdot \sqrt{2\cos\theta - 2\cos\theta_0} \cdot \left[Y - \frac{\tau\mu(\mu-1)}{\sqrt{1 + \mu^2 - 2\mu\cos\theta_0}\,(1 + \mu^2 - 2\mu\cos\theta)} \right] \tag{9.3.11}$$

The angle θ_0 is fixed by normalization:

$$Y\,(1-y_0)+\tau\left(\frac{\mu-1}{\sqrt{1+\mu^2-2\mu y_0}}-1\right)=1. \tag{9.3.12}$$

where $y_0 := \cos\theta_0$. Equation (9.3.12) admits a unique real solution, thus the problem is completely determined.

One-cut solution: Phase Ib

The solution above has been derived assuming that Γ is an arc along $\mathbb{T}$ joining $e^{-\mathrm{i}\theta_0}$ to $e^{\mathrm{i}\theta_0}$ running counter-clockwise, thus the gap has opened around $\theta=\pi$. For the gap opening at $\theta=0$, the procedure is identical, but now Γ is an arc from $\tilde\theta_0>0$ to $2\pi-\tilde\theta_0$. The procedure of Appendix 9.A leads us to

$$\rho_{\mathrm{Ib}}(\theta)=2\left|\sin\frac{\theta}{2}\right|\sqrt{2\cos\tilde\theta_0-2\cos\theta}\left[-Y+\frac{\tau\mu(\mu+1)}{\sqrt{1+\mu^2-2\mu\cos\tilde\theta_0}\,(\mu^2+1-2\mu\cos\theta)}\right]$$

which is non-negative definite. There is, however, a more direct route to get the correct answer. Looking back at the matrix model (9.2.1) we can chose a different parametrization $0\le\theta<2\pi$, and the solution with the gap opening at $\theta=0$ is recovered from the solution (9.3.11) in Phase Ia upon replacement $Y\mapsto -Y$, $\mu\mapsto-\mu$ and eventually $\theta+\pi\mapsto\theta$.

In conclusion, we have two different phases with a one-cut solution, as expected: one for $Y>Y_{\mathrm{cr,a}}(\tau,\mu)$, that we have called Phase Ia, and one for $Y<Y_{\mathrm{cr,b}}(\tau,\mu)$, that we have called Phase Ib.

Two-cut solution: Phase II

We have seen that at $\tau=\tau_{\mathrm{cr},+}=(\mu^2-1)/2$ the critical surfaces $Y=Y_{\mathrm{cr,a}}$ and $Y=Y_{\mathrm{cr,b}}$ meet. Thus, we expect a new phase characterized by a two-cut solution in the region

$$\left\{(\mu,\tau,Y)\ :\ \mu>1,\tau>\frac{\mu^2-1}{2},Y_{\mathrm{cr,a}}<Y<Y_{\mathrm{cr,b}}\right\}\subset\mathfrak{M}.$$

with gaps around $\theta=0$ and $\theta=\pm\pi$, and eigenvalue density supported on

$$\mathrm{supp}\rho_{\mathrm{II}}=\Gamma\cong\Gamma_{\mathrm{u}}\sqcup\Gamma_{\mathrm{d}}:=\left\{e^{\mathrm{i}\varphi}\in\mathbb{T}\ :\ \tilde\theta_0\le\theta\le\theta_0\right\}\sqcup\left\{e^{\mathrm{i}\varphi}\in\mathbb{T}\ :\ -\theta_0\le\theta\le-\tilde\theta_0\right\}.$$

That is, Γ is the union of two disjoint arcs, Γ_{u} and Γ_{d}, as in Figure 9.5.

To determine $\rho_{\mathrm{II}}(\theta)$ it is simpler to adopt a different strategy, detailed in Section 9.3.4 below.

Two-cut solution: Phase III

The fact that the potential $V_{\mathrm{eff}}(e^{\mathrm{i}\theta})$ develops a double well for negative Y and τ in a given range hints at the existence of a two-cut solution in that region of $\mathfrak{M}$, with the eigenvalues sitting around the two minima. This observation is corroborated looking at the shape of $\rho_{\mathrm{Ia}}(\theta)$ and $\rho_{\mathrm{Ib}}(\theta)$ in the negative quadrant, where they become negative in $Y_{\mathrm{cr,b}}<Y<Y_{\mathrm{cr,a}}$ for τ below a certain threshold.

We find a transition from Phase 0 to a two-cut phase in

$$Y_{\mathrm{cr,c}+}<Y<Y_{\mathrm{cr,a}}\quad\text{and}\quad Y_{\mathrm{cr,b}}<Y<Y_{\mathrm{cr,c}-}$$

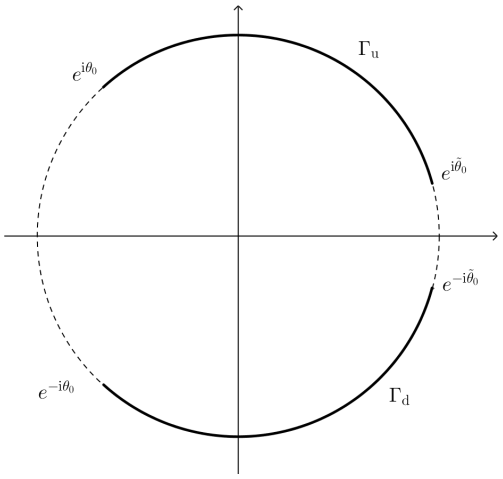


Figure 9.5. The two-cut support Γ in Phase II.

where the critical surfaces $Y = Y_{\rm cr,c\pm}$ are given by

$$Y_{\rm cr,c\pm}(\tau,\mu) = \frac{\mu}{(\mu^2+1)^2}\left[\mu^2(\tau-1) - 3\tau - 1 \pm 2\sqrt{-\tau\left[2\tau(\mu^2-1)+(\mu^4-1)\right]}\right].$$

The two curves $Y_{\rm cr,\pm}$ form an ellipse in each (τ, Y)-leaf of $\mathfrak{M}$ at fixed μ, with the physical critical curve being the first branch of the ellipse encountered when decreasing Y from 0.

In this phase, that we call Phase III, the eigenvalues distribute along a contour Γ which consists of two cuts, with gaps opening around $\pm\theta_*$, see Figure 9.6.

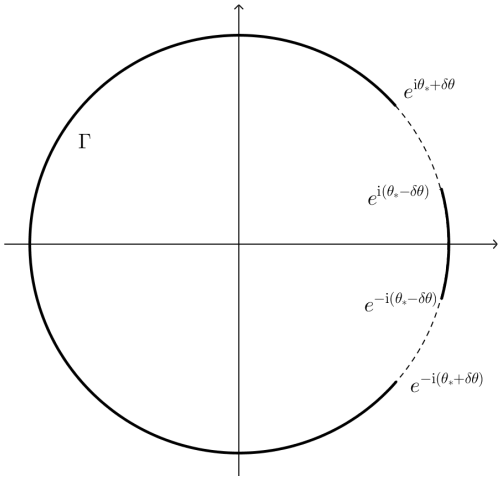


Figure 9.6. The two-cut support Γ in Phase III.

The eigenvalue density is

$$\rho_{\rm III}(\theta) = 2\sqrt{\left[\cos(\theta_*-\delta\theta)-\cos\theta\right]\left[\cos(\theta_*+\delta\theta)-\cos\theta\right]} \times\left[-Y + \frac{\tau\mu(\mu+1)(\mu-1)}{\sqrt{(\mu^2+1-2\mu\cos(\theta_*-\delta\theta))\,(\mu^2+1-2\mu\cos(\theta_*+\delta\theta))\,[\mu^2+1-2\mu\cos\theta]}}\right]. \tag{9.3.13}$$

Note that the argument of the outer square root is non-negative definite. The value of θ_* is known explicitly, as obtained from Phase 0, and the dependence of $\delta\theta$ on the parameters is fixed

by normalization. Equivalently, we can fix $\cos(\theta_* + \delta\theta)$ and $\cos(\theta_* - \delta\theta)$ comparing the large z behaviour of $\omega(z)$ computed in this phase with its definition.

For multi-cut solutions, the dependence on the number of eigenvalues filling each cut should be taken into account when computing physical observables [357]. We analyze the role of the filling fractions in Appendix 9.B: the upshot is that our conclusions are unaltered, both in phase II and III, although for different reasons.

9.3.2 Phase diagram

Putting all the information together, the following phase diagram emerges.

0) When both Y and τ are small, Phase 0 holds, with the eigenvalues spread on the whole circle.

Ia) When $Y > Y_{\mathrm{cr,a}}$ the system is in a new phase, Phase Ia, with a one-cut solution gapped around $\theta = \pm\pi$.

Ib) Likewise when $Y < Y_{\mathrm{cr,b}}$ the system is in Phase Ib, with a one-cut solution gapped around $\theta = 0$.

II) At $\tau > \frac{\mu^2-1}{2}$ the two critical surfaces cross each other. In the region $Y_{\mathrm{cr,a}} < Y < Y_{\mathrm{cr,b}}$ the system is in Phase II, a two-cut solution with density of eigenvalues gapped both around $\theta = 0$ and $\theta = \pi$.

III) The system develops a new two-cut phase, Phase III, in the region $Y_{\mathrm{cr,b}} < Y < Y_{\mathrm{cr,a}}$ and also bounded by an arc of ellipse determined by $Y_{\mathrm{cr,c}\pm}$. The density of eigenvalues is gapped around $\theta = \pm\theta_*$, with $\theta_* \to \pi$ as $Y \to Y_{\mathrm{cr,a}}$ and $\theta_* \to 0$ as $Y \to Y_{\mathrm{cr,b}}$.

See Figure 9.7 for a slice of $\mathfrak{M}$ at fixed μ.

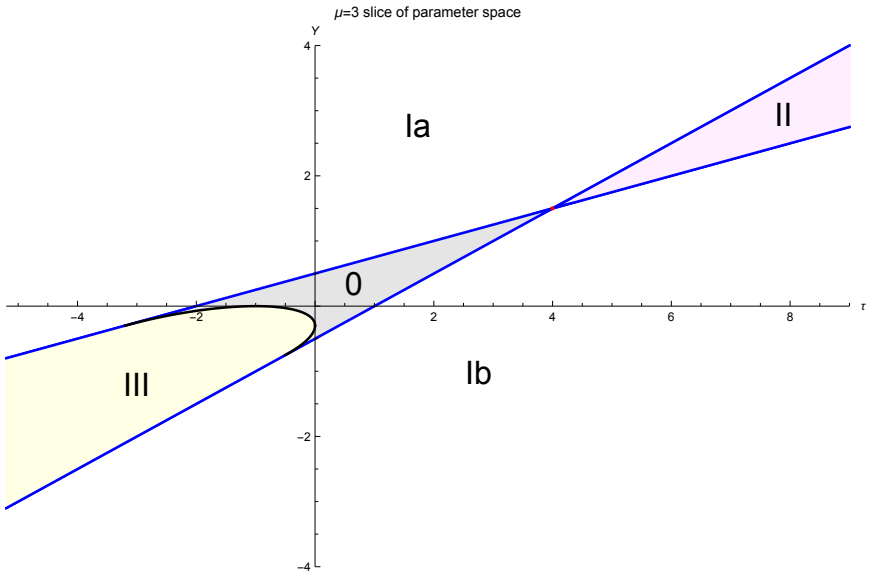


Figure 9.7. Phase diagram of the model in the (τ, Y) plane, at $\mu = 3$. The blue straight lines are $Y = Y_{\mathrm{cr,a}}$ and $Y = Y_{\mathrm{cr,b}}$, the black curve is $Y = Y_{\mathrm{cr,c}\pm}$, the red dot is the multi-critical point at $\tau = \frac{\mu^2-1}{2}$. The gray shaded region in the ungapped phase, Phase 0. The other light shaded regions are the two-cut phases, Phase II and III.

Taking the massless limit $\mu \to 1^+$, the critical surface $Y_{\text{cr,b}}$ is rotated onto the vertical axis. Using the analytic dependence on μ, we can also reach $\mu \to -1^-$ by first going to the negative real axis walking through $\mathbb{C}$ outside of the unit disk and then taking the limit $|\mu| \to 1^+$. In that case, it is $Y_{\text{cr,a}}$ that is rotated onto the vertical axis.

9.3.3 Free energy and massless theory

Before delving in the analysis of Wilson loop vevs in the next section, we comment on the free energy of the model, defined as

$$\mathcal{F} = \frac{1}{N^2} \log \mathcal{Z}.$$

The free energy in Phase 0 is easily obtained, and corresponds to the analytic continuation of Szegő's strong limit theorem in the bulk of the 't Hooft parameter space [11]. It takes the value

$$\mathcal{F}_0 = Y^2 - 2Y\frac{\tau}{\mu} - \tau^2 \log\left(1 - \frac{1}{\mu^2}\right). \tag{9.3.14}$$

It is clearly separated into three contributions: pure gauge (Y^2), matter only ($\propto \tau^2$) and the interaction. At strong coupling $\lambda \to \infty$ ($Y \to 0$) we are left with a matter contribution which counts gauge singlets: indeed, the integral over the gauge group projects onto gauge invariant states.

Massless theory

As we have stressed, a core assumption of our analysis is $|\mu| > 1$, and the massless limit $|\mu| \to 1^+$ can only be taken at the end. Due to the non-analyticity for $\mu \in \mathbb{T}$, the resulting model will differ from a model with massless matter [345, 346]. A consequence of this non-analyticity is the spontaneous chiral symmetry breaking, that we will discuss in Section 9.4.4. On the other hand, it is well known that the large N limit and the massless limit do not commute.

In the naïve $|\mu| \to 1$ limit, the free energy in (9.3.14) has a logarithmic divergence in the matter contribution. Setting instead $|\mu| = 1$ from the beginning, and $\arg \mu = \tilde{\theta}$, $0 < \tilde{\theta} \leq 2\pi$, the partition function acquires a FH singularity and the large N limit cannot be understood by standard methods. We use known results on Toeplitz determinants to derive the free energy in Phase 0 in the massless theory [358]:

$$\mathcal{F}_0\left(\mu = e^{\mathrm{i}\tilde{\theta}}\right) = Y^2 - 2Y\tau\cos\tilde{\theta} - 2\tau^2 \log\left|2\sin\frac{\tilde{\theta}}{2}\right| + \tau^2 \log N.$$

That is, the contribution from matter fields has an additional factor of $\log N$ and dominates at large N. Remarkably, this matches the logarithmic divergence of the naïve massless limit of (9.3.14). The result is in fact much more general [358] and directly extends to the case of various Veneziano parameters $\tau_1, \ldots, \tau_n$ associated to different $\mu_1, \ldots, \mu_n$ that approach the unit circle from outside at different angles $\tilde{\theta}_1, \ldots, \tilde{\theta}_n$.

9.3.4 Stereographic projection

To better understand Phase II and the transition from a one-cut to a two-cut phase, we map the model onto the real line and study the resulting Hermitian matrix model at large N. It can be interpreted as a massive deformation of the model in [359].

We conformally map the unit circle on the real line through the stereographic projection, see Figure 9.8. The drawback of the stereographic map is that it introduces a puncture on the circle

at $\theta = \pm\pi$: this has no effect at finite N, but the Hermitian matrix model will fail to reproduce Phase 0 of the unitary matrix model because of this change in topology [6]. Phase 0 and its associated transitions are well understood from the unitary matrix model side, and we use the conformally mapped model as yet another way to gain further insight into the one-cut to two-cut transition.

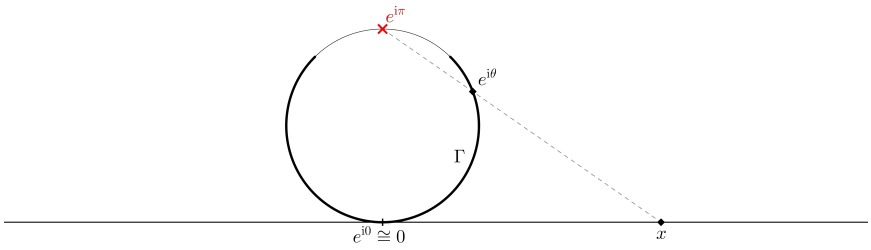


Figure 9.8. The stereographic projection. The red cross is the puncture on the circle, the ticker line is the cut Γ.

Our choice of coordinates is consistent with Phase Ia on the circle, but Phase Ib is easily retrieved rotating $\mathbb{T}$ by $e^{i\pi}$, so that the puncture is placed at $\theta = 0$. The Hermitian matrix model is

$$\mathcal{Z}^{\mathrm{p.}}_{U(N)} = (\mu-1)^{2K} \int_{\mathbb{R}^N} \prod_{1\le j<k\le N} (x_j - x_k)^2 \prod_{j=1}^{N} \frac{(1+\eta^2 x_j^2)^K}{(1+x_j^2)^{K+N}}\, e^{2NY\left(\frac{1-x_j^2}{1+x_j^2}\right)}\, \frac{\mathrm{d}x_j}{2\pi} \tag{9.3.15}$$

with the superscript p. as notation to remind that it comes from the projection of the model (9.2.1). We have adopted the shorthand notation

$$\eta := \frac{\mu+1}{\mu-1}, \qquad 1 < \eta < \infty.$$

Remark. Thanks to the mapping of the Vandermonde determinant form the unit circle to the real line that shifts $\tau \mapsto \tau + 1$ in the denominator of the integrand in (9.3.15), the stability issues pointed out in [359] do not arise here. We can thus safely allow $\tau < 0$ without spoiling the convergence of the matrix model.

Phase I

The saddle point equation for the Hermitian matrix model (9.3.15) is

$$\mathrm{P}\int \frac{\mathrm{d}y}{2\pi} \frac{\rho^{\mathrm{p.}}(y)}{x-y} = x\left[\frac{4Y}{(x^2+1)^2} + \frac{\tau+1}{1+x^2} - \frac{\eta^2\tau}{1+\eta^2x^2}\right].$$

The solution is found by standard large N techniques [27]. Using a one-cut ansatz for the density $\rho^{\mathrm{p.}}(x)$ supported on $[-A, A] \subset \mathbb{R}$ we find

$$\rho^{\mathrm{p.}}_{\mathrm{I}}(x) = 2\sqrt{A^2-x^2}\left[\frac{\tau+1}{\sqrt{1+A^2}(1+x^2)} - \frac{\tau\eta^2}{\sqrt{1+\eta^2A^2}(1+\eta^2x^2)} - 2Y\frac{A^2\left(x^2-1\right)-2}{(1+A^2)^{\frac{3}{2}}(1+x^2)^2}\right]. \tag{9.3.16}$$

The value of A is fixed by normalization:

$$\int_{-A}^{A} \frac{\mathrm{d}x}{2\pi}\rho^{\mathrm{p.}}_{\mathrm{I}}(x) = 1 \quad \Longrightarrow \quad \frac{\tau+1}{\sqrt{1+A^2}} - \frac{\tau}{\sqrt{1+\eta^2A^2}} + 2Y\frac{A^2}{(1+A^2)^{\frac{3}{2}}} = 0.$$

As a cross-check, turning off the mass deformation, $\mu \to 1$, sends $\eta \to \infty$ and we recover the eigenvalue density found in [359]. Besides, sending $A^2 \to \infty$ and expanding at leading order in $\frac{1}{A}$ the normalization becomes the consistency condition

$$Y = \frac{1}{2} + \frac{\tau}{\mu + 1},$$

correctly reproducing the critical surface $Y_{\mathrm{cr,a}}$ in the limit in which A is back-projected to $e^{\mathrm{i}\pi}$. We stress that, requiring that $\rho^{\mathrm{p}\cdot}(x)\mathrm{d}x$ descends from a measure on $\mathbb{T}$, the non-negativity constraint $\rho(x) \geq 0$ must be imposed.

Looking at $\rho_{\mathrm{I}}^{\mathrm{p}\cdot}(0)$, we find that the critical point is fixed by the condition

$$2Y\left(\frac{2+A^2}{1+A^2}\right) + \tau + 1 - \tau\eta^2\sqrt{\frac{1+A^2}{1+\eta^2A^2}} = 0.$$

Phase II

From the result above as well as from the analysis of the unitary matrix model, we find a phase transition to a two-cut solution, with a gap opening at $x = 0$. The new phase is the conformal image of Phase II of the unitary matrix model.

We look for a new eigenvalue density, supported on $[-A, -B] \cup [B, A]$. The result is

$$\rho_{\mathrm{II}}^{\mathrm{p}\cdot}(x) = 2\sqrt{(A^2 - x^2)(x^2 - B^2)}\ |x| \left[-2Y\frac{x^2(A^2+B^2+2) + 3(A^2+B^2) + 2A^2B^2 + 4}{[(1+A^2)(1+B^2)]^{\frac{3}{2}}\,(1+x^2)^2}\right.$$
$$\left. - \frac{\tau+1}{\sqrt{(1+A^2)(1+B^2)}(1+x^2)} + \frac{\tau\eta^4}{\sqrt{(1+\eta^2A^2)(1+\eta^2B^2)}(1+\eta^2x^2)}\right]. \tag{9.3.17}$$

The parameters A and B are fixed by normalization,

$$-Y\frac{(A^2-B^2)^2}{[(1+A^2)(1+B^2)]^{\frac{3}{2}}} - (\tau+1)\left(\frac{A^2+B^2+2}{\sqrt{(1+A^2)(1+B^2)}} - 1\right) + \tau\left(\frac{\eta^2A^2+\eta^2B^2+2}{\sqrt{(1+\eta^2A^2)(1+\eta^2B^2)}} - 1\right) = 1,$$

and by an additional self-consistency condition on $\omega(z)$,

$$2Y\frac{A^2+B^2+2}{(1+A^2)(1+B^2)} + \tau + 1 - \tau\eta^2\sqrt{\frac{(1+A^2)(1+B^2)}{(1+\eta^2A^2)(1+\eta^2B^2)}} = 0,$$

which reproduces the criticality condition for $B \to 0$.

This is not the end of the story for Phase II. Indeed, fluctuations in the number of eigenvalues in each cut may contribute at leading order in the evaluation of observables [357]. However, for the symmetric two-cut solution we find out that this is not the case, as proved in Appendix 9.B.

9.4 Wilson loops and instantons

We continue the investigation of the features of the phase transitions and establish their order by evaluating the vacuum expectation value (vev) of the Wilson loop in the fundamental representation. Moreover, we further discuss the different physics of the various transitions by looking at the different contributions by instantons.

9.4.1 Wilson loops

Wilson loops are order operators in gauge theories that, for simple connected gauge group, describe the holonomy of the gauge connection around a closed path. For our one-plaquette model, we consider the Wilson loop in the fundamental representation wrapping the plaquette, and compute its vev. It is given by

$$\langle \mathcal{W} \rangle = \left\langle \frac{1}{2N} \operatorname{Tr} U + \frac{1}{2N} \operatorname{Tr} U^\dagger \right\rangle = \left\langle \frac{1}{N} \sum_{j=1}^{N} \cos\theta_j \right\rangle$$

with the average taken in the unitary ensemble (9.2.1). We use the eigenvalue density at large N found in each phase to evaluate the Wilson loop.

Wilson loops: Generalities

From the matrix model (9.2.1) we immediately get the relation

$$\langle \mathcal{W} \rangle = \frac{1}{2N} \frac{1}{\mathcal{Z}} \frac{\partial}{\partial (NY)} \mathcal{Z} = \frac{1}{2} \frac{\partial \mathcal{F}}{\partial Y}.$$

Therefore, all the information about the order of the transition can be extracted from the Wilson loop vev. This is precisely what we expect from an order parameter, and follows from the Wilson loop belonging to the class of order operators of QCD_2.

Being $\rho(\theta)$ continuous on the whole $\mathfrak{M}$, the Wilson loop vevs are continuous as well, implying that every phase transition we find must be at least second order.

Wilson loops: Evaluation

We focus now on the Wilson loop vev at large N. In the ungapped phase we find

$$\langle \mathcal{W} \rangle_0 = \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \rho_0(\theta)\, e^{\mathrm{i}\theta} = Y - \frac{\tau}{\mu}. \tag{9.4.1}$$

This reproduces the GWW result as $\tau \to 0$, but also as $\mu \to \infty$, as expected when the matter becomes non-dynamical. For a Wilson loop winding $k > 1$ times around the plaquette, either in clockwise or anti-clockwise direction, we get

$$\langle \mathcal{W}^k \rangle_0 = -\frac{\tau}{\mu^k}.$$

In Phase Ia the Wilson loop vev is

$$\begin{aligned} \langle \mathcal{W} \rangle_{\mathrm{Ia}} = \int_{-\pi}^{\pi} \frac{\mathrm{d}\theta}{2\pi} \rho_{\mathrm{Ia}}(\theta)\, e^{\mathrm{i}\theta} &= \frac{2}{\pi} \int_{y_0}^{1} \mathrm{d}y\; y \sqrt{\frac{y-y_0}{1-y}} \left[Y - \frac{\tau\mu(\mu-1)}{\sqrt{1+\mu^2-2\mu y_0}(1+\mu^2-2\mu y)} \right] \\ &= Y \frac{(1-y_0)(3+y_0)}{4} - \frac{\tau}{2\mu} \left[\mu^2 + 1 + \frac{1+\mu(\mu-1)y_0 - \mu^3}{\sqrt{1+\mu^2-2\mu y_0}} \right] \end{aligned} \tag{9.4.2}$$

where we have used the change of variables $y = \cos\theta$, with $y_0 = \cos\theta_0$. The value of y_0 as a function of the gauge theory parameters is known from (9.3.12).

The Wilson loop vev in Phase Ib is obtained likewise,

$$\langle \mathcal{W} \rangle_{\mathrm{Ib}} = Y \frac{(1+\tilde{y}_0)(3-\tilde{y}_0)}{4} - \frac{\tau}{2\mu} \left[\mu^2 + 1 + \frac{\mu(\mu+1)\tilde{y}_0 - \mu^3 - 1}{\sqrt{1+\mu^2-2\mu\tilde{y}_0}} \right], \tag{9.4.3}$$

where $\tilde{y}_0 = \cos\tilde{\theta}_0$.

To study the derivative of $\langle\mathcal{W}\rangle_{\mathrm{Ia}}$ and establish the order of the phase transition, it suffices to notice that

$$\frac{\mathrm{d}}{\mathrm{d}Y}\langle\mathcal{W}\rangle_{\mathrm{Ia}} = 1 + \left[\frac{Y}{4}(-2y_0 - 2) + \frac{\tau}{2}\cdot\frac{(\mu-1)\mu(y_0+1)}{(1+\mu^2-2\mu y_0)^{3/2}}\right]\frac{\partial y_0}{\partial Y}.$$

This implies

$$\lim_{y_0\to -1}\frac{\mathrm{d}}{\mathrm{d}Y}\langle\mathcal{W}\rangle_{\mathrm{Ia}} = 1$$

which matches the derivative of $\langle\mathcal{W}\rangle_0$. The computations are identical for the transition between Phase 0 and Phase Ib. Taking a further derivative, $\frac{\mathrm{d}^2}{\mathrm{d}Y^2}\langle\mathcal{W}\rangle$ vanishes identically in Phase 0, but does not vanish at the critical loci when computed in Phases Ia and Ib.

We conclude that the Wilson loop vev is an order parameter of class C^1 at the critical surfaces $Y = Y_{\mathrm{cr,a}}(\tau,\mu)$ and $Y = Y_{\mathrm{cr,b}}(\tau,\mu)$, thus the system shows a pair of third order phase transitions. In particular, both the GWW transition [190, 192] and the transition in [302] are special points on the critical locus of the present model.

Crossing from a one-cut to a two-cut phase, the first derivative of the Wilson loop is not protected. Indeed, in Phase II the derivative of the Wilson loop vev has the schematic form

$$\frac{\mathrm{d}}{\mathrm{d}Y}\langle\mathcal{W}\rangle_{\mathrm{II}} = \int_{y_0}^{\tilde{y}_0} y\frac{\partial}{\partial Y}f(y,y_0,\tilde{y}_0)\mathrm{d}y + \frac{\partial y_0}{\partial Y}\int_{y_0}^{\tilde{y}_0} y\frac{\partial}{\partial y_0}f(y,y_0,\tilde{y}_0)\mathrm{d}y + \frac{\partial\tilde{y}_0}{\partial Y}\int_{y_0}^{\tilde{y}_0} y\frac{\partial}{\partial\tilde{y}_0}f(y,y_0,\tilde{y}_0)\mathrm{d}y, \tag{9.4.4}$$

with the first term coming from the derivative of the explicit dependence on Y, and the other two from the dependence on Y through y_0 and $\tilde{y}_0$. The integrand evaluated at the endpoint vanishes, hence those contributions do not appear.

In (9.4.4), $f(y,y_0,\tilde{y}_0)$ is known explicitly from Section 9.3.1,

$$f(y,y_0,\tilde{y}_0) = \frac{2}{\pi}\sqrt{\frac{(y-y_0)(\tilde{y}_0-y)}{(1+y)(1-y)}}\left[-Y + \frac{\tau\mu(\mu+1)(\mu-1)}{\sqrt{(1+\mu^2-2\mu y_0)(1+\mu^2-2\mu\tilde{y}_0)}(1+\mu^2-2\mu y)}\right],$$

but the difficulty comes from the only implicit knowledge of the dependence of $y, \tilde{y}_0$ on Y.

Passing from Phase II to Phase Ia, the first term in (9.4.4) matches continuously with the corresponding expression in Phase Ia, as $\tilde{y}_0 \to 1$. The integral in the second summand in (9.4.4) also agrees with the corresponding contribution in Phase Ia at $\tilde{y}_0 \to 1$. Both facts follow from

$$\lim_{\tilde{y}_0\to 1} y_0|_{\mathrm{II}} = y_0|_{\mathrm{Ia}}.$$

Moreover, the symmetries of the integrand allow to combine the third term in (9.4.4) with the second term, in a simpler expression. Moreover, the symmetric form of the equations fixing $y_0, \tilde{y}_0$ can be used to show that

$$\frac{\partial\tilde{y}_0}{\partial Y} = \left.\frac{\partial y_0}{\partial Y}\right|_{\tilde{y}_0\leftrightarrow y_0}.$$

By this we mean that the expressions on the two sides agree upon exchanging all $\tilde{y}_0$ with y_0.

Due to the complicated dependence on the parameters, the derivatives of the boundaries $y_0, \tilde{y}_0$ are not continuous at the transition point. The differentiability of $\langle\mathcal{W}\rangle$ above followed by the vanishing of the term multiplying such derivatives. This does not happen for the transition from a two-cut to a one-cut phase. Therefore, the sum of the second and third terms in (9.4.4) gives an obstruction to the differentiability of $\langle\mathcal{W}\rangle$, so we expect a second order transition. In a sense, the obstruction arises from taking a limit that breaks explicitly the $y_0 \leftrightarrow \tilde{y}_0$ symmetry of Phase II.

The proof is very similar for the transition from Phase II to Phase Ib or from Phase III to either Phase Ia or Ib.

The argument fails at the critical surface $Y_{\rm cr,c}$ and at the multi-critical point at which $Y_{\rm cr,a} = Y_{\rm cr,b}$. Indeed, when passing directly from Phase 0 to a two-cut phase, the simplifications that arise from closing both gaps simultaneously imply that the Wilson loop vev is C^1. This is consistent with the observation of the previous paragraph, as these transitions preserve the $\mathbb{Z}_2$-symmetry of the two-cut phase.

9.4.2 Phase structure and remarks

Summing up the results extracted from the analysis of Wilson loop vev, we find that

- the transition from Phase 0 to any other phase is third order, but
- the transition from a one-cut to a two-cut phase is likely to be second order.

We emphasize that we have not ultimately established the order of the transitions of second type, but we have gathered evidence that it should be second order.

In the rest of this subsection we comment on various aspects of the phase structure we uncovered, insisting on the role of the second order phase transitions. The outcome of the analysis substantiates the belief that the order should be second in the transitions among gapped phases.

Remark. As obtained in the previous subsection, the second order discontinuities are finite jumps, not divergences. The correlation lengths remain finite at each transition. These finite discontinuities vanish in the limit $|\mu| \to 1$.

Metastability

While the third order transitions we find are a continuation of the GWW transition in $\mathfrak{M}$, it is worth to further comment on the second order transitions we obtain. The phase transition to a two-cut solution happens slightly beyond the values of τ where the potential develops a double-well structure. The proposal in [360] states that a second order transition can be associated with tunneling from a metastable vacuum to a stable one. Our analysis confirms that picture in the one-plaquette model we consider. In Subsection 9.4.5 we study instanton effects, expanding this discussion leading to a further refined distinction between second and third order phase transitions in this model, from the instanton point of view.

Critical behaviour

It is worthwhile to notice that the phase diagram in Figure 9.7 resembles that in [361], where a unitary matrix model with potential $Y_1 \cos(\theta) + Y_2 \cos(2\theta)$ was analyzed.

The critical behaviour close to a transition to a two-cut phase in our model differs from that found in similar models in the literature, for matrix models with potentials of the form $\sum_{n=1}^{K} Y_n \cos(n\theta)$. This is so because the potential in (9.2.1) includes both a polynomial and a logarithmic part, requiring different scaling approaching the critical regime from the two-cut phase. This distinction, however, fades away approaching the multicritical point.

Double-scaling limit

The statements above can be refined exploiting the double-scaling limit.

In particular, we can zoom in the critical regime, tuning Y towards a transition to the ungapped phase. In the double-scaling limit, the dynamics is governed by Painlevé II equation. The proof follows from [183, 302] with minimal variations. An alternative proof can be given using orthogonal

polynomials [182]. We have checked explicitly that, in the double-scaling limit, the problem reduces to the analogous one for the pure GWW model.

For the transition from a one-cut to a two-cut phase, however, there is no double scaling that gives Painlevé II.

9.4.3 Continuum limit and β-function

The β-function of the theory, as a function of the 't Hooft coupling λ, can be computed using the chain rule through [190]

$$\beta(\lambda) = 2\lambda^2 \frac{\langle \mathcal{W} \rangle \log \langle \mathcal{W} \rangle}{\frac{\partial}{\partial Y} \langle \mathcal{W} \rangle}. \tag{9.4.5}$$

This quantity can be used to test whether our model reproduces the expected features of QCD_2 in the continuum limit. The fixed points of the RG flow, that capture the continuum physics, are given by $\beta(\lambda) = 0$, which, from (9.4.5), can only happen at $\langle \mathcal{W} \rangle = 0$ or at $\langle \mathcal{W} \rangle = 1$.

Direct computations in Phases 0, Ia, Ib, show that only the solution to $\langle \mathcal{W} \rangle = 0$ is consistent, while the solution to $\langle \mathcal{W} \rangle = 1$ always falls out of the phase in which it has been computed, and thus should be discarded. It is a nice consistency check that the solution to be discarded is precisely the one that would violate Elitzur's theorem [362], and the one to be retained is in agreement with the confining nature of QCD_2 [190].

The continuum limit of a lattice theory consists in sending the lattice spacing to zero while approaching a critical curve [363]. In particular, this requires $|\mu| \to 1$.

Taking the continuum limit from Phase 0, approaching either $Y_{\mathrm{cr,a}}$ or $Y_{\mathrm{cr,b}}$, we find that the unique consistent solution is $Y = 0$ (i.e. $\lambda = \infty$). This is physically meaningful for a toy model of QCD_2: the theory flows to a strongly interacting theory in the deep infrared.

Taking the continuum limit from Phase Ia close to the transition to Phase II, we find a trivial solution with $\lambda = 0 = \tau$, describing a theory of free gauge bosons without matter. The continuum limit approaching the critical surface between Phase Ib and Phase II, instead, yields a non-trivial fixed point at

$$\lambda = \frac{1}{Y} \approx 121.4.$$

Remark. The existence of a continuum theory is not established by our analysis, because correlation lengths remain finite. While this has no effect in our model, which consists of a single plaquette, it may (and most likely shall) wash away the fixed point in the continuum limit of a more realistic lattice model.

9.4.4 Chiral symmetry breaking

Let us focus now on the model with fermionic matter. The fermion two-point function is by definition

$$\langle \bar{\psi}_f \psi_f \rangle = -\frac{1}{N\mathcal{Z}} \frac{\partial}{\partial \mu_f} \mathcal{Z} = -\frac{\partial}{\partial \mu_f} \frac{1}{N} \log \mathcal{Z}.$$

Due to our degenerate choice of masses, we can only compute the average over flavours of such quantity:

$$\left\langle \frac{1}{K} \sum_{f=1}^{K} \bar{\psi}_f \psi_f \right\rangle = \frac{1}{\tau} \frac{\partial \mathcal{F}}{\partial \mu}.$$

In Phase 0 we find

$$\frac{1}{\tau} \frac{\partial \mathcal{F}_0}{\partial \mu} = \frac{2}{\mu^2} \left(Y + \frac{\tau \mu}{\mu^2 - 1} \right).$$

This quantity diverges as $\mu \to 1$, therefore we expect the chiral symmetry to be spontaneously broken in the continuum, consistently with the analysis of the β-function in Phase 0.

In Phases Ia and Ib, we can move along $Y = 0$ and study the behaviour of $\langle \frac{1}{K} \sum_f \bar{\psi}_f \psi_f \rangle$ on that subspace of $\mathfrak{M}$. The result is read off directly from [6]:

$$\frac{1}{\tau} \frac{\partial \mathcal{F}_{\mathrm{Ia}}}{\partial \mu}\bigg|_{Y=0} = -\frac{\mu + 1 + 4\tau}{2\tau\mu(\mu - 1)}$$

in Phase Ia, which is non-vanishing and continuous at the transition point.

In Phase Ib we get

$$\frac{1}{\tau} \frac{\partial \mathcal{F}_{\mathrm{Ib}}}{\partial \mu}\bigg|_{Y=0} = -\frac{1 - \mu + 4\tau}{2\tau\mu(\mu + 1)},$$

again non-vanishing and continuous at the transition point, and goes to 1 in the $\mu \to 1^+$ limit. This latter result, in turn, hints at a transition to a free theory: the free energy of a theory of K free flavours of mass m goes as $\mathcal{F}_{\mathrm{free}} \propto Km$, whence $\langle \frac{1}{K} \sum_f \bar{\psi}_f \psi_f \rangle|_{\mathrm{free}} = 1$. Note that this computation has been done at infinite gauge 't Hooft coupling, which has the physical meaning of governing the theory in the deep infrared.

To sum up, we have observed that the phase transition from Phase Ib to Phase 0 is accompanied with spontaneous chiral symmetry breaking.

9.4.5 Instantons in unitary matrix models

We discuss non-perturbative effects in the unitary matrix model, coming from unstable saddle point configurations [364].

An instanton configuration is characterized by a collection of integers $\{N_0, N_1, \dots\}$ with $\sum_k N_k = N$. For example, the d-instanton configuration is associated with the symmetry breaking pattern

$$U(N) \to U(N_0) \times U(N_1) \times \cdots \times U(N_d),$$

with the eigenvalues $z_j \in \mathbb{T}$ of $U \in U(N)$ grouped in d different sets, sitting at d different extrema of the potential. For the one-instanton configuration,

$$\mathcal{Z}_{U(N)}(\nu) = \sum_{\ell=0}^{N} e^{-\nu\ell} \mathcal{Z}_\ell$$

with $\mathcal{Z}_\ell$ the partition function of a $U(N-\ell) \times U(\ell)$ model, and we have turned on a chemical potential $\nu > 0$ for the instanton number.

The ℓ-sector leads to non-perturbative corrections to the free energy of the matrix model, of the form

$$e^{-N\ell S_{\mathrm{inst}}(\vec{\lambda})} f_\ell(\vec{\lambda})$$

where $\vec{\lambda}$ generically denotes the couplings of the theory, and the functions $\{f_\ell\}_\ell$ admit themselves a $\frac{1}{N}$ expansion.

Instanton effects and third order transitions

Let us consider our model (9.2.1) at large N and focus on the one-cut phase, in which the interpretation of instanton effects is more transparent. We discuss them in Phase Ia, being the corresponding analysis in Phase Ib completely analogous. Most of the details are just an extension of the thorough analysis in [364].

The contribution of an instanton excitation, obtained moving one eigenvalue from the minimum of V_{eff} to the maximum at $\theta = \pm\pi$ is found to be

$$\pi S_{\text{inst}} = 2Y\left[\sqrt{1-x_0^2} - x_0^2 \cosh^{-1}\left(\frac{1}{x_0}\right)\right] + \tau\left[\tanh^{-1}\left((\mu-1)\sqrt{\frac{1-x_0^2}{(\mu-1)^2+4\mu x_0^2}}\right) + \frac{\mu-1}{\sqrt{(\mu-1)^2+4\mu x_0^2}}\log\left(\frac{x_0}{1+\sqrt{1-x_0^2}}\right)\right]$$

where $\cosh^{-1}$ and $\tanh^{-1}$ are the inverse of the hyperbolic functions, and $x_0 = \sin\frac{\theta_0}{2}$. One of the results in [364] (already conjectured in [365]) is that the GWW transition is triggered by instantons. We see that the result carries over to the present model, as

$$\lim_{x_0\to 1} S_{\text{inst}} = 0$$

and the instanton excitations cease to be suppressed at the critical point when the gap closes.

Analogous conclusions are found if we go to Phase III, in which the effective potential has developed a double well, and consider the instanton configuration with a few eigenvalues taken to the local maximum at θ_*. Approximating close to the transition to Phase 0, we find

$$\pi S_{\text{inst}} = (\delta\theta)^2 \sin\theta_* \left[-Y + \frac{\tau\mu(\mu+1)}{(\mu-1)\left(1+\mu^2-2\mu\cos\theta_*\right)}\right] + \mathcal{O}\left((\delta\theta)^3\right).$$

At the critical surface, $\delta\theta \to 0$ and we find again that the third order transition is triggered by instantons.

Instanton effects and second order transitions

We now turn our attention to the analysis of instanton effects in the two-cut phase, starting from Phase III. We consider a single eigenvalue placed on the maximum of the potential, as sketched in Figure 9.9.

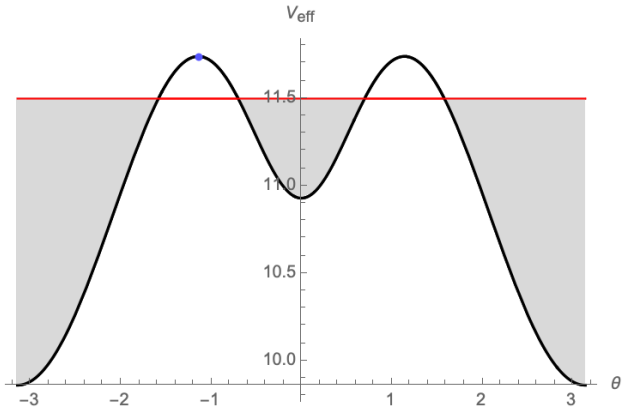


Figure 9.9. Instantons in Phase III. A single eigenvalue (blue dot) is moved on top of the maximum of V_{eff}, while all the others (gray sea) fill the minima.

We find that the instanton action is the sum of two pieces,

$$S_{\text{inst},L} = \int_{y_L}^{y_*} \frac{\mathrm{d}y}{\pi}\sqrt{\frac{(y-y_L)(y_R-y)}{1-y^2}}\left[-Y + \frac{\tau\mu(\mu^2-1)}{\sqrt{(1+\mu^2-2\mu y_L)(1+\mu^2-2\mu y_R)}[1+\mu^2-2\mu y]}\right],$$

$$S_{\text{inst},R} = \int_{y_*}^{y_R} \frac{\mathrm{d}y}{\pi}\sqrt{\frac{(y-y_L)(y_R-y)}{1-y^2}}\left[-Y + \frac{\tau\mu(\mu^2-1)}{\sqrt{(1+\mu^2-2\mu y_L)(1+\mu^2-2\mu y_R)}[1+\mu^2-2\mu y]}\right],$$

where $y_* = \cos\theta_*$ and $y_{L,R} = \cos(\theta_* \pm \delta\theta)$. The two are associated with the eigenvalue escaping from the left and right cut, respectively. There exists a third relevant quantity, namely the tunneling from one cut to the other,

$$S_{\text{tunnel}} = \int_{y_L}^{y_R} \frac{\mathrm{d}y}{\pi} \sqrt{\frac{(y-y_L)(y_R-y)}{1-y^2}} \left[-Y + \frac{\tau\mu(\mu^2-1)}{\sqrt{(1+\mu^2-2\mu y_L)(1+\mu^2-2\mu y_R)}[1+\mu^2-2\mu y]} \right].$$

All the three effects are non-perturbatively suppressed by a factor $e^{-NS_{\text{inst}}}$, with S_{inst} the corresponding action. The three contributions are still suppressed at the critical loci, although the tunneling term will coalesce with one of the other two.

The situation is slightly different in Phase II, where the two wells have equal depth, see Figure 9.10. In this case, S_{tunnel} is simply twice S_{inst}, and both go to zero as a gap closes. The phase

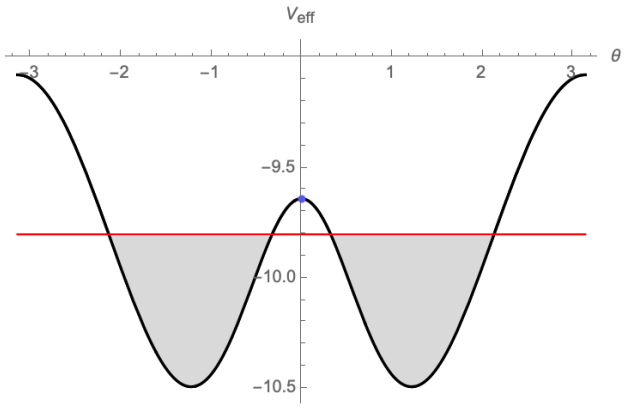


Figure 9.10. Instantons in Phase II. A single eigenvalue (blue dot) is moved on top of the local maximum of V_{eff} at $\theta = 0$, while all the others (gray sea) fill the minima.

transition takes place when the tunneling between the two cuts ceases to be suppressed in one direction (e.g. passing through $\theta = 0$ in Figure 9.10) but remains non-perturbative in the other direction (e.g. passing through $\theta = \pi$ in Figure 9.10).

The picture we infer is that the third order phase transitions are associated with releasing non-perturbative instabilities, while the second order transitions correspond to release only those instabilities in one direction. This is also in agreement with the proposal in [360, 342] relating second order phase transitions in GWW-type models to partial deconfinement.

9.5 Meromorphic deformation of unitary matrix models

We now depart from the model (9.2.1) with the aim of setting the stage for the study of meromorphic deformations of unitary matrix models, in which the integration contour is deformed in $\mathbb{C}^*$ and not bound to be the unit circle. This consists of an adaptation of the theory of holomorphic matrix models [343] to unitary matrix models and, as we shall show, is instrumental in understanding their phase diagram from new angles. This section can be read independently of the rest of the manuscript.

A unitary matrix model is characterized by a weight function $e^{-\frac{N}{\lambda}V(z)}$, with $V(z)$ admitting the expansion

$$V(z) = \sum_{n\geq 1} \left(\frac{t_n}{n} z^n + \frac{t_{-n}}{n} z^{-n} \right). \tag{9.5.1}$$

The function $e^{-\frac{N}{\lambda}V(z)}$ is singular at $z \in \{0,\infty\} \subset \mathbb{P}^1$ and possibly has other zeros and poles in $\mathbb{C}^* \cong \mathbb{P}^1 \setminus \{0,\infty\}$. The Vandermonde determinant appearing in a unitary matrix model is conveniently rewritten in meromorphic form

$$\prod_{1\le j<k\le N}(z_j-z_k)\left(\frac{1}{z_j}-\frac{1}{z_k}\right)=\left(\prod_{j=1}^{N}\frac{1}{z_j^{N-1}}\right)\prod_{1\le j<k\le N}(z_j-z_k)^2.$$

To deform a unitary matrix model, the integration cycle $\mathbb{T}^N$ is replaced by any half-dimensional cycle $\mathcal{C}_N$ in $(\mathbb{C}^*)^N$.

Definition. Let $N \in \mathbb{N}$, $\lambda \in \mathbb{C}^*$ and $V(z)$ as in (9.5.1). A meromorphic matrix model $\mathcal{Z}$ is the integral

$$\mathcal{Z}=\oint_{\mathcal{C}_N}\prod_{1\le j<k\le N}(z_j-z_k)^2\prod_{j=1}^{N}e^{-\frac{N}{\lambda}V(z_j)}\,\frac{\mathrm{d}z_j}{2\pi\mathrm{i}z_j^N}, \tag{9.5.2}$$

with integration contour

$$\mathcal{C}_N=\sum_\ell N_\ell\mathcal{C}_\ell,\qquad \sum_\ell N_\ell=N,\qquad [\mathcal{C}_\ell]=[\mathbb{T}]\in H_1(\mathbb{C}^*). \tag{9.5.3}$$

Condition (9.5.3) means that each $\mathcal{C}_\ell$ is homotopic to the unit circle in the holed plane $\mathbb{C}^*$. Dropping it, we may stretch $\mathcal{C}_\ell$ along any direction along which, asymptotically, $\Re\frac{1}{\lambda}V(z) > 0$. Eventually the one-cycle $\mathcal{C}_\ell$ pinches at $z=\infty \in \mathbb{P}^1$. To get an honest deformation of a unitary matrix model we do not allow this situation, otherwise we would fall back in the holomorphic deformation of a Hermitian matrix model.

The couplings $\{t_n\}$ in (9.5.1) are usually subject to reality conditions, such as

$$t_{-n}=\bar{t}_n,\qquad \forall n\ge 1,$$

and possibly other relations. We write the constraints collectively as $\vec{\Phi}(\{t_n\})=0$. Besides, a rescaling of all $\{t_n\}$ together can be reabsorbed in a redefinition of λ, hence the couplings $\{t_n\}$ are homogeneous coordinates on a projective space.

The matrix model (9.5.2) sets a natural stage to complexify the couplings. We denote by $\mathscr{T}$ the physical space of couplings, namely the collection of independent $\{t_n\}$ after imposing the constraints and modulo scaling. More formally,

$$\mathscr{T}=\left(\{t_n\in\mathbb{C},\ \forall n\neq 0\}/\mathbb{C}^*\right)\cap\left\{\vec{\Phi}(\{t_n\})=0\right\}, \tag{9.5.4}$$

with the $\mathbb{C}^*$-action being multiplication of all couplings by a non-vanishing constant. Whenever the constraints $\vec{\Phi}$ can be rewritten in homogeneous form, $\mathscr{T}$ is a projective variety.

9.5.1 Large N limit

We are interested in the large N limit of (9.5.2). Define the effective potential

$$W(z):=\frac{1}{\lambda}V(z)+\log z.$$

At large N, the eigenvalues will be gathered around the saddle points of $W(z)$ in $\mathbb{C}^*$,

$$W'(z_{\mathrm{sp},\ell})=0,\quad \ell=1,2,\dots,g+1.$$

Here we are assuming there is a finite number $g+1$ of saddle points z_{sp}. The integration contour $\mathcal{C}_N$ in (9.5.3) can be chosen in such a way that each $\mathcal{C}_\ell$ passes through $z_{\mathrm{sp},\ell}$. The integers N_ℓ in

(9.5.3) then count the number of eigenvalues around the ℓ^{th} saddle point $z_{\text{sp},\ell}$. At large N, the density of eigenvalues will vanish on $\mathcal{C}_\ell$ away from a compact interval $\Gamma_\ell \subset \mathcal{C}_\ell$, called a cut, with $z_{\text{sp},\ell} \in \Gamma_\ell$. Therefore

$$\text{supp}\rho = \bigcup_{\ell=1}^{g+1} \Gamma_\ell =: \Gamma,$$

and $\rho(z)$ is normalized.

Remark. The requirement that the integration cycle passes through all the $g+1$ saddle points does not fix it uniquely. The shape of each $\mathcal{C}_\ell$, and thus of the cuts Γ_ℓ at large N, can be homotopically deformed in an open neighbourhood of $z_{\text{sp},\ell}$, meaning that the matrix model (9.5.2) depends on (up to) g additional parameters.

In the large N limit, the eigenvalue density solves the saddle point equation

$$\int_\Gamma \frac{\mathrm{d}w}{2\pi} \frac{\rho(w)}{z-w} = \frac{1}{2} W'(z), \tag{9.5.5}$$

where $'$ means holomorphic derivative $\frac{\partial}{\partial z}$. Here, $\mathrm{d}w$ is a holomorphic differential on Γ and, given any parametrization $w : s \mapsto w(s) \in \Gamma$, $\mathrm{d}w = \dot{w}(s)\mathrm{d}s$ is understood, with $\mathrm{d}s$ the line element and $\dot{w}$ the derivative of the map $w : s \mapsto w(s)$.

Recall the definition of the trace of the resolvent $\omega(z)$ at large N:

$$\omega(z) := \int \mathrm{d}w \frac{\rho(w)}{z-w}, \qquad z \in \mathbb{P}^1 .$$

Equation (9.5.5) implies that $\omega(z)$ solves [27]

$$\omega(z)^2 - W'(z)\omega(z) + f(z) = 0, \tag{9.5.6}$$

where

$$f(z) = \int_\Gamma \frac{W'(z) - W'(w)}{z-w} \rho(w) \, \frac{\mathrm{d}w}{2\pi}.$$

Defining

$$y(z) = \omega(z) - \frac{1}{2} W'(z), \tag{9.5.7}$$

(9.5.6) becomes

$$y^2 - \left(\frac{1}{2} W'(z)\right)^2 + f(z) = 0. \tag{9.5.8}$$

This equation goes under the name of spectral curve. The steps from (9.5.5) to (9.5.8) are standard and have been applied to holomorphic matrix models since their early days [366] to establish a bridge between matrix models and geometric problems. The novel aspect of (9.5.8) compared to the literature is hidden in the form of W' and f, which in the present case are not ordinary polynomials but Laurent polynomials, or meromorphic functions.

We assume for now that $V(z)$ is a Laurent polynomial on $\mathbb{P}^1$, with singularities at $\{0, \infty\}$. The extension to a meromorphic weight function on $\mathbb{C}^*$ is worked out below. Write

$$V(z) = \sum_{n=-d_-+1}^{d_++1} \frac{t_n}{|n|} z^n \quad \Longrightarrow \quad W'(z) = \frac{1}{\lambda} \sum_{n=-d_-}^{d_+} \text{sgn}(n) t_{n+1} z^n + \frac{1}{z}, \tag{9.5.9}$$

where we assume $d_- \geq 2$ (otherwise we get back the known setting).

We need to introduce some notation. Define

$$g = d_+ + d_- - 1, \tag{9.5.10}$$

which agrees with the counting of saddle points above, and also $t_0 = -\lambda$ for later convenience. Besides, denote ρ_k the moments of the eigenvalue density,

$$\rho_k = \int_\Gamma w^k \rho(w) \mathrm{d}w, \qquad k \in \mathbb{Z}.$$

After some rewriting we get

$$W'(z) = \frac{1}{\lambda z^{d_-}} \sum_{n=0}^{g+1} z^n t_{n+1-d_-} \ \mathrm{sgn}(n-d_-), \tag{9.5.11}$$

$$f(z) = \frac{1}{\lambda z^{d_-}} \left[\sum_{n=0}^{d_--1} z^n \left(\sum_{k=-1}^{n-1} t_{k+2-d_-} \rho_{k-n} \right) + \sum_{n=d_-}^{g} z^n \left(\sum_{k=n}^{g} t_{k+2-d_-} \rho_{k-n} \right) \right], \tag{9.5.12}$$

where $\mathrm{sgn}(0) = +1$ by convention. The spectral curve takes the schematic form

$$y^2 = \frac{P(z)}{4\lambda^2 z^{2d_-}} \tag{9.5.13}$$

where $P(z)$ is a polynomial in z of degree $\deg(P) = 2g+2$, with coefficients read off from (9.5.11)-(9.5.12) and that depend on the parameters $\{t_n\}$, on λ and on the moments $\{\rho_k\}$.

A major difference with respect to the standard unitary matrix models is that $\{\rho_k\}$ are free complex parameters of the theory: they can be tuned deforming Γ, as discussed in Subsection 9.5.1. Recalling that both $z, \lambda \in \mathbb{C}^*$, it is possible to recast (9.5.13) in a more standard form $\hat{y} = P(z)$, describing an hyperelliptic complex curve of genus g [366].

Γ is the union of $g+1$ branch cuts stretched between pairs of roots of $P(z)$. The roots of $P(z)$ move inside $\mathbb{C}^*$ as the parameters are varied.[24] The coalescence of two roots produces a singularity of the curve (9.5.13) and corresponds, on the matrix model side, to a phase transitions from a $(g+1)$-cut to a g-cut phase, with either

- two cuts joining, or
- one cut collapsing.

The hyperelliptic curve (9.5.13) is fibered over the moduli space $\mathscr{M}$ of the model (9.5.2), defined as

$$\mathscr{M} = \mathbb{C}^* \times \mathscr{T} \times \mathbb{C}^g,$$

with $\mathbb{C}^*$ parametrized by λ, $\mathscr{T}$ defined in (9.5.4), and the last factor parametrized by the moments $\{\rho_k\}$. Note that one of the moments is fixed comparing (9.5.6) with the definition of $\omega(z)$ at $|z| \to \infty$.

Definition. A critical locus $\mathscr{C}$ is an irreducible component of the locus in $\mathscr{M}$ at which two roots of $P(z)$ coalesce.

[24] Without loss of generality, $\{0, \infty\} \subset \mathbb{P}^1$ are not roots of $P(z)$, because they would correspond to a "non-minimal" choice of $d_\pm$ in (9.5.9). They can be avoided simply defining $\hat{y} = 2y\lambda z^{d_- \pm m}$ with m the multiplicity of the root, and minus (resp. plus) sign if the root is $z = 0$ (resp. $z = \infty$).

The critical loci $\mathscr{C} \subset \mathscr{M}$ necessarily have positive complex codimension, and the hyperelliptic fibration is singular along them. Singularities in higher codimension, placed at the (self-)intersection of critical loci, correspond to multicritical points of the matrix model.

The theory of Abelian differentials provides a suitable framework to analyze the genus g hyperelliptic curve (9.5.13) [367]. At this stage, the analysis of the spectral curve works exactly as in the holomorphic deformation of Hermitian matrix models, thus we omit the details and refer to [367, 368].

Remark. We are now in the position to elaborate more on the remark in Subsection 9.5.1, from a point of view advocated in [369]. Let $\{A^\ell, B_\ell\}$ be a basis of one-cycles in the hyperelliptic curve. The A-cycles are chosen to go around the cuts Γ_ℓ. Therefore

$$\oint_{A^\ell} y(z)\mathrm{d}z = \oint_{A^\ell} \omega(z)\mathrm{d}z = \frac{N_\ell}{N} =: \xi_\ell.$$

The first equality follows from the definition (9.5.7) noting that $y(z)$ and $\omega(z)$ only differ by a regular term. Introducing chemical potentials for the filling fractions ξ_ℓ and extremizing the action with respect to these quantities gives their saddle point value as a function on $\mathscr{M}$. More precisely, one gets a set of equations analytic in the ratios $s_\ell := \frac{\xi_\ell}{\lambda}$ [369]. At this point, it is possible to invert the relations and express the moments $\{\rho_k\}$ in terms of the complex variables s_ℓ, keeping the latter as free parameters.

Note that only g out of the $g+1$ of both quantities are free.

The study of the phases of the model (9.5.2) leads to a stratification of the parameter space $\mathscr{M}$. We postpone the analysis to Section 9.5.4, discussing explicit models first.

Genus 0

The unique way to obtain a genus 0 spectral curve is from the holomorphic deformation of the CUE. In that case, the model has no couplings and, as opposed to $g \geq 1$, the additional condition derived from the definition of $\omega(z)$ is automatically fulfilled, leaving ρ_{-1} as unique, unconstrained parameter. Then, (9.5.13) describes a $\mathbb{P}^1$ fibered over $\mathbb{C}$. If we try to get a less trivial model by considering the insertion of $(\det U)^{\tau N}$, the consistency condition, which in $g \geq 1$ fixes one of the $\{\rho_k\}$, imposes $\tau = 0$.

9.5.2 Holomorphic GWW

We now put the machinery at work and revisit the phase structure of the holomorphic GWW model. The phase diagram of this model has been obtained in [370] for $\lambda \in \mathbb{R}$, while the behaviour at complex coupling has been partially analyzed in [371], although without fully exploiting the holomorphic deformation.

The GWW model has $t_{-1} = t_1 = 1$, and $t_{n\neq\pm1} = 0$, whence $d_+ = 0$, $d_- = 2$, $g = 1$ and only the moments ρ_{-2}, ρ_{-1} appear in (9.5.12). Fixing ρ_{-2} as a function of λ and ρ_{-1}, the spectral curve of the holomorphic GWW model is [370]

$$\hat{y}^2 = z^4 + 2\lambda z^3 + \left[(\rho_{-1}+1)\lambda^2 - 2\right] z^2 + 2\lambda z + 1. \tag{9.5.14}$$

It is an elliptic curve. Following the strategy outlined above, we think of (9.5.14) as an elliptic fibration over $\mathbb{C}^* \times \mathbb{C}$, with coordinates on the base λ and ρ_{-1}, and identify the phase transitions with singularities of the fibration.

The discriminant of (9.5.14) is

$$\Delta = \frac{\lambda^2}{4}\left(\lambda^2\rho_{-1} - 4\right)^2 \left(\lambda\left(\rho_{-1}+1\right) - 4\right)\left(\lambda\left(\rho_{-1}+1\right) + 4\right), \tag{9.5.15}$$

from which the critical loci are

$$\begin{aligned}\{\Delta = 0\} &= \mathscr{C}_1 \cup \mathscr{C}_{1'} \cup \mathscr{C}_2,\\ \mathscr{C}_1 &:= \left\{\rho_{-1} = -1 + \frac{4}{\lambda},\ \lambda \in \mathbb{C}^*\right\}, \quad \mathscr{C}_{1'} := \left\{\rho_{-1} = -1 - \frac{4}{\lambda},\ \lambda \in \mathbb{C}^*\right\},\\ \mathscr{C}_2 &:= \left\{\rho_{-1} = \frac{4}{\lambda^2},\ \lambda \in \mathbb{C}^*\right\}.\end{aligned}$$

In Kodaira's classification [372], $\mathscr{C}_1$ and $\mathscr{C}_{1'}$ are singularities of type I_1 and $\mathscr{C}_2$ is of type I_2. The GWW critical points $(\lambda, \rho_{-1}) = (\pm 2, 1)$ are singled out as the codimension-two singularities at which $\mathscr{C}_2$ intersects one of the other two critical curves. Besides, we recognize the elliptic curve (9.5.14) as the Seiberg–Witten curve of $\mathcal{N} = 2$ supersymmetric four-dimensional $SU(2)$ gauge theory with two flavours [373].

The original GWW transition is thus, from the perspective of the holomorphic deformation, one of the possible ways to approach the codimension-two singularity from a generic direction. The singularity at the multicritical points $(\lambda, \rho_{-1}) = (\pm 2, 1)$ is of Kodaira type III. The corresponding symmetry is A_1, which is precisely the symmetry of Painlevé II, that is known to control the GWW phase transition [182, 183]. If, instead, we approach the multicritical point not from a generic direction but moving along a critical locus, the singularity type is enhanced to I_0^*.

It is possible to allow $t_{-1} \neq t_1$. This corresponds to introduce a θ-term in the GWW lattice action, $\frac{\theta}{2\pi} = \frac{t_1 - t_{-1}}{2}$. The procedure goes through with only minor modifications, the unique difference being that the singularities $\mathscr{C}_1, \mathscr{C}_{1'}$ are placed at $\rho_{-1} = -\frac{1}{t_{-1}} \pm \frac{4}{\lambda\sqrt{t_{-1}}}$, so in particular the $\mathbb{Z}_2$-symmetry is preserved.

9.5.3 Meromorphic deformations

The formulation can be extended to include weight functions with zeros and poles in $\mathbb{C}^*$. For concreteness, we consider the illustrative example of our original model (9.2.1) at $\lambda^{-1} = 0$. In this case

$$\begin{aligned}W'(z) &= -\tau\left[\frac{1}{z-\mu} + \frac{1}{z-\mu^{-1}}\right] + \frac{1+\tau}{z},\\ f(z) &= \tau\left[\frac{\tilde{\rho}_{-1}(\mu)}{z-\mu} + \frac{\tilde{\rho}_{-1}(\mu^{-1})}{z-\mu^{-1}}\right] - \frac{(1+\tau)}{z}\rho_{-1}.\end{aligned}$$

In the second line, we have defined

$$\tilde{\rho}_{-1}(\mu) := \int_\Gamma \frac{\rho(w)}{w-\mu}\mathrm{d}w,$$

with, in particular, $\tilde{\rho}_{-1}(0) = \rho_{-1}$. Comparing the definition of $y(z)$ with the spectral curve at large $|z|$, we find a pair of consistency conditions, fixing $\tilde{\rho}_{-1}$ as a function of the other parameters,

$$\tilde{\rho}_{-1}(\mu) = -\frac{4\rho_{-1}(\tau+1) + \mu\tau(\tau+6) + \mu}{4\,(\mu^2-1)\,\tau}.$$

Note that the two conditions fix $\tilde{\rho}_{-1}(\mu^{\pm 1})$ independently, and the solutions are consistently mapped into each other under $\mu \leftrightarrow \mu^{-1}$. We get

$$y^2 = \frac{P(z)}{4z^2(z-\mu)^2(z-\mu^{-1})^2}, \tag{9.5.16}$$

with $P(z)$ a polynomial of degree 4. The spectral curve thus describes again en elliptic fibration over the moduli space $\mathbb{C}^* \times \{\mu \in \mathbb{P}^1 \ : \ 1 < |\mu| < \infty\} \times \mathbb{C}$, parametrized by (τ, μ, ρ_{-1}). The discriminant takes the form $\Delta = \mu^5(\mu+1)^2(\mu-1)^2(\tau+1)^4\widetilde{\Delta}$, the last term being a cumbersome polynomial of degree 6 in ρ_{-1}, degree 8 in τ and degree 10 in μ. The critical points of the undeformed model become higher-codimensional singularities, at which two roots of $P(z)$ collide. The collection of all critical loci in this model is

$$\mathscr{C}_4 \cup \mathscr{C}_2 \cup \mathscr{C}_{2'} \bigcup_{j=1}^{6} \mathscr{C}_{1_j},$$

with the subscript indicating the order of vanishing of Δ along the component $\mathscr{C}$. Taking what we have called the continuum limit in Section 9.4.3, that is, sending $\tau \to \tau_{\rm cr}(\mu)$ and then $\mu \to \pm 1$, with $\rho_{-1} = -\frac{\tau}{\mu}$ set to its undeformed value, yields a non-minimal singularity $\Delta \propto (\mu \pm 1)^{12}$.

9.5.4 Stratification of the moduli space

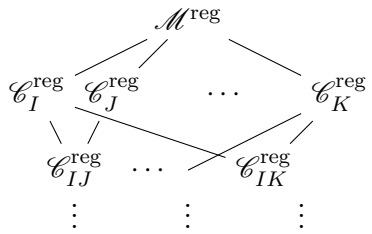
The critical loci and their intersections endow the parameter space $\mathscr{M}$ with additional structure.

The stratification of an algebraic variety $\mathscr{V}$ is a collection of open sets $\{\mathscr{V}_I\}$, with $\mathscr{V}_0$ a point and $\overline{\mathscr{V}}_{\rm max} = \mathscr{V}$, with a partial order given by the inclusion of the closures of $\{\mathscr{V}_I\}$. The parameter space $\mathscr{M}$ of the model (9.5.2) is the union of

$$\mathscr{M}^{\rm reg},\ \mathscr{C}_I^{\rm reg},\ \mathscr{C}_{IJ}^{\rm reg},\ \dots$$

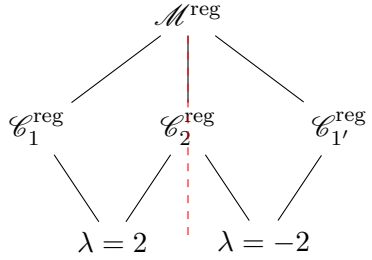
where the superscript means the regular part, and $\mathscr{C}_{IJ} = \mathscr{C}_I \cap \mathscr{C}_J$, and so on. The inclusion relations $\overline{\mathscr{C}_{IJ}^{\rm reg}} = \overline{\mathscr{C}_I} \cap \overline{\mathscr{C}_J} \subset \overline{\mathscr{C}_I}$ are obvious.

The partial order can be represented with the aid of a Hasse diagram:



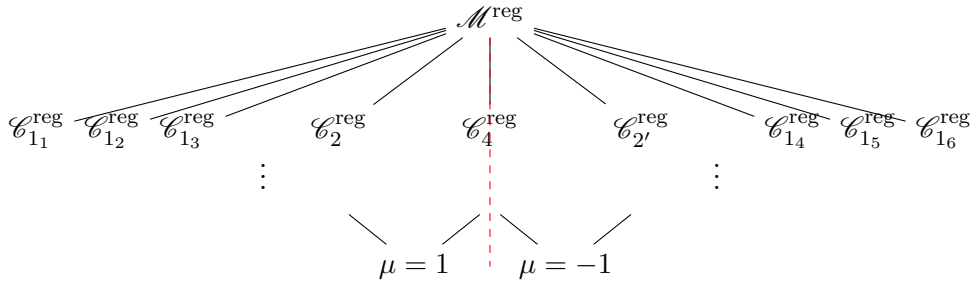
In general, this does not define a full-fledged stratification of $\mathscr{M}$ because multiple final points may exist. Nonetheless, whenever the potential (9.5.1) has a $\mathbb{Z}_2$-symmetry, the Hasse diagram inherits it. This $\mathbb{Z}_2$-symmetry acts as an automorphism of the Hasse diagram, which is mapped into itself under reflection along the vertical axis. By construction, the Hasse diagram resulting from folding the initial diagram via this $\mathbb{Z}_2$-symmetry determines a stratification of $\mathscr{M}/\mathbb{Z}_2$.

We draw the Hasse diagram of the holomorphic GWW model of Section 9.5.2:



(9.5.17)

The diagram of the meromorphic model of Section 9.5.3 is schematically



The vertical red, dashed line is there to emphasize the $\mathbb{Z}_2$ reflection symmetry. Folding the diagram along that line yields the stratification of $\mathscr{M}/\mathbb{Z}_2$.

Remark. The results of Section 9.5.2 with $t_{-1} \neq t_1$ show that, even for models in which a $\mathbb{Z}_2$ reflection symmetry is not manifest from the potential but emerges at large N, the $\mathbb{Z}_2$-folding yields a stratified moduli space.

Symplectic singularities

Recall that $H_1(\Sigma_g, \mathbb{C})$, the first homology group of a hyperelliptic curve Σ_g of genus g, is a symplectic space. The A- and B-cycles that we have implicitly used in the study of the spectral curve (9.5.8) can be chosen to be Darboux coordinates in $H_1(\Sigma_g, \mathbb{C})$. Moving along $\mathscr{M}^{\rm reg}$ corresponds to vary the symplectic structure without changing the topology of Σ_g. At the critical loci $\mathscr{C}_I$, however, either

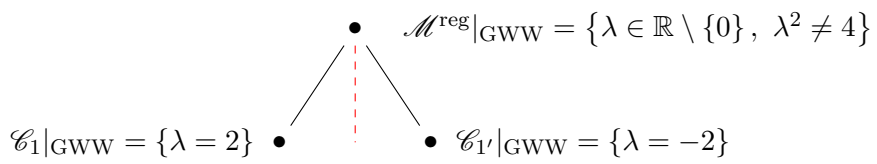
- a B-cycle collapses, or
- an A-cycle collapses.

Both situations correspond to a singularity of the symplectic form. Therefore, the analysis of the phase structure of the meromorphic matrix models can be rephrased in terms of symplectic foliations of symplectic singularities, in the sense of Kaledin [374].

The appearance of symplectic singularities is not entirely unexpected. The consideration of holomorphic matrix models in their large N limit lead to the Seiberg–Witten curves [373] of certain $\mathcal{N} = 2$ gauge theories [366]. The so-called Coulomb branch of these theories is a symplectic singularity and is stratified [375]. In fact, the use of Hasse diagrams in the present work was inspired by [376, 377].

As a final remark, we emphasize that the structure uncovered in this section is not specific of the meromorphic matrix models. The parameter spaces of unitary or Hermitian matrix models inherit it, as they can be realized as slices inside the parameter space of our meromorphic models.

As an example, the phase diagram of the GWW model is



(9.5.18)

It is found by fixing $\rho_{-1} = 1$, taking the slice $\lambda \in \mathbb{R} \setminus \{0\}$ and identifying the intersection of such subspace with the strata in (9.5.17).[25]

[25] Accidentally, this is precisely the Hasse diagram of the reduction to three dimensions of the $SU(2)$ theory with two flavours, captured by the holomorphic GWW of Section 9.5.2, cf. [377, Eq.(4.2)]. It should be stressed, however, that the strata in (9.5.18) are real, not hyperKähler.

9.A Large N limit: Gapped solutions

In this appendix we sketch the computation of $\omega(z)$, defined in (9.3.10), which allows to extract the eigenvalue density in the phases with one or more gaps. The procedure is standard and we follow closely [330, 11], glossing over many details. We work in Phase Ia, since all other phases are analyzed in similar fashion.

Introduce the function $\varrho(z)$ of complex variable $z \in \mathbb{C}$ such that $\varrho(e^{\mathrm{i}\theta}) = \rho(\theta)$ for $e^{\mathrm{i}\theta} \in \Gamma$. The saddle point equation (9.3.4) is rewritten as

$$\mathrm{P}\int_\Gamma \frac{\mathrm{d}w}{2\pi w}\varrho(w)\frac{z+w}{z-w} = W'(z) \tag{9.A.1}$$

where

$$W'(z) = -\mathrm{i}\left[Y\left(z-\frac{1}{z}\right) - \tau\left(1+\frac{\mu}{z-\mu}+\frac{\mu^{-1}}{z-\mu^{-1}}\right)\right].$$

Equation (9.A.1) is valid for $z \in \Gamma$, and is complemented by the normalization condition

$$\int_\Gamma \frac{\mathrm{d}w}{2\pi \mathrm{i} w}\varrho(w) = 1. \tag{9.A.2}$$

Recall that we have started with a $\mathbb{Z}_2$-symmetric system, invariant under $z \mapsto z^{-1}$ for $z \in \mathbb{T}$. We will thus find an eigenvalue density with symmetric support, and in particular $\partial\Gamma = \{e^{-\mathrm{i}\theta_0}, e^{\mathrm{i}\theta_0}\}$ in a one-cut phase. Then, depending on whether the gap opens at $\theta = \pi$ or $\theta = 0$, Γ will be the arc on the unit circle connecting $-\theta_0$ to θ_0 or θ_0 to $-\theta_0$, respectively, with orientation always taken counter-clockwise.

Recall from the definition (9.3.10) that

$$\omega_+(e^{\mathrm{i}\theta}) - \omega_-(e^{\mathrm{i}\theta}) = 2\varrho(e^{\mathrm{i}\theta}), \qquad e^{\mathrm{i}\theta} \in \Gamma.$$

In turn, from the definition of Cauchy principal value and (9.A.1) we immediately get

$$\omega_+(e^{\mathrm{i}\theta}) + \omega_-(e^{\mathrm{i}\theta}) = -2\mathrm{i}W'(e^{\mathrm{i}\theta}). \tag{9.A.3}$$

The normalization (9.A.2) and the definition (9.3.10) imply that $\omega(z) \to 1$ as $|z| \to \infty$. We have then reduced the problem of finding the eigenvalue density to the problem of determining the discontinuity of $\omega(z)$ along Γ, from the knowledge of its regular part and the boundary condition $\omega(z \to \infty) = 1$. It is standard procedure to reduce the problem (9.A.3) to a discontinuity equation for a new, auxiliary function $\Omega(z)$ related to $\omega(z)$ via

$$\omega(z) = \sqrt{(e^{\mathrm{i}\theta_0}-z)\,(e^{-\mathrm{i}\theta_0}-z)}\,\Omega(z). \tag{9.A.4}$$

We take the square root with positive value, but any potential ambiguity in the intermediate steps and definitions from now on, would drop out from the final answer.

Writing

$$\left[\sqrt{(e^{\mathrm{i}\theta_0}-z)\,(e^{-\mathrm{i}\theta_0}-z)}\right]_\pm = \left[\sqrt{z}\cdot\sqrt{2\cos\theta - 2\cos\theta_0}\right]_\pm = \mp e^{\mathrm{i}\theta/2}\sqrt{2\cos\theta-2\cos\theta_0}$$

for $z = e^{\mathrm{i}\theta} \in \Gamma$, we obtain from (9.A.3) the discontinuity equation for $\Omega(z)$:

$$\Omega_+(e^{\mathrm{i}\theta}) - \Omega_-(e^{\mathrm{i}\theta}) = 2e^{-\mathrm{i}\theta/2}\frac{\mathrm{i}W'(e^{\mathrm{i}\theta})}{\sqrt{2\cos\theta-2\cos\theta_0}}. \tag{9.A.5}$$

For a multi-cut phase, with

$$\Gamma \cong \{\theta_{0,-} \le \theta \le \theta_{0,+}\} \cup \{\theta_{1,-} \le \theta \le \theta_{1,+}\} \cup \cdots \cup \{\theta_{k,-} \le \theta \le \theta_{k,+}\}$$

the procedure is the same, but with $\Omega(z)$ defined via

$$\omega(z) = \sqrt{\prod_{j=0}^{k} \left(e^{\mathrm{i}\theta_{j,+}} - z\right)\left(e^{\mathrm{i}\theta_{j,-}} - z\right)}\,\Omega(z).$$

Let us now introduce a closed contour C_Γ which is a Jordan curve enclosing Γ but not z, and oriented counter-clockwise. See Figure 9.11 for the contour C_Γ in Phase Ia.

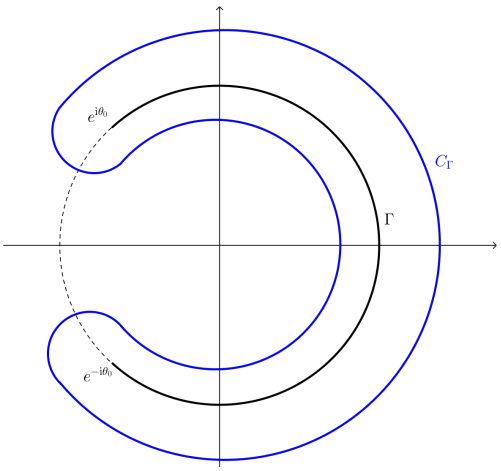


Figure 9.11. Contour C_Γ encircling the cut Γ.

From the definitions (9.3.10) and (9.A.4) it follows that $\Omega(z)$ falls off (at least) as $1/z$ at infinity. Then, for z lying in the exterior of C_Γ, Cauchy's theorem together with (9.A.5) implies

$$\Omega(z) = \oint_{C_\Gamma} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \frac{\mathrm{i}W'(w)}{(z-w)\sqrt{(e^{\mathrm{i}\theta_0}-w)\,(e^{-\mathrm{i}\theta_0}-w)}}$$

On the other hand, because $W'(w)$ is meromorphic we can deform the contour C_Γ into an infinitely large circle, picking the poles of the integrand. We find

$$\begin{aligned}\Omega(z) = &-\frac{\mathrm{i}W'(z)}{\sqrt{(e^{\mathrm{i}\theta_0}-z)\,(e^{-\mathrm{i}\theta_0}-z)}} - \oint_{C_\infty} \frac{\mathrm{d}w}{2\pi\mathrm{i}} \frac{\mathrm{i}W'(w)}{(z-w)\sqrt{(e^{\mathrm{i}\theta_0}-w)\,(e^{-\mathrm{i}\theta_0}-w)}} \\ &+ \sum_{z_p \in \{0,\mu,\mu^{-1}\}} \operatorname*{Res}_{w=z_p} \frac{\mathrm{i}W(w)}{(z-w)\sqrt{(e^{\mathrm{i}\theta_0}-w)\,(e^{-\mathrm{i}\theta_0}-w)}}\end{aligned} \tag{9.A.6}$$

where the first term is the residue at $w = z$, the second term is the remaining contour integral along a circle at infinity, which in our case simply contributes Y, and the last term includes the residues at the poles z_p of $W'(w)$.

In Phase Ia, explicit computation of each term leads to

$$\omega_{\mathrm{Ia}}(z) = -\mathrm{i}W'(z) + \sqrt{(e^{\mathrm{i}\theta_0}-z)\,(e^{-\mathrm{i}\theta_0}-z)}\left[Y\left(1+\frac{1}{z}\right) - \frac{\tau}{\sqrt{1+\mu^2-2\mu\cos\theta_0}}\left(\frac{\mu}{z-\mu} - \frac{1}{z-\mu^{-1}}\right)\right].$$

The solution in the other phases is found likewise.

9.B Filling fraction fluctuations

This appendix contains the analysis of the effect of taking into account fluctuations of the filling fractions around the equilibrium configuration.

For a generic matrix model in a two-cut phase, the dependence of the filling fractions on the parameters of the theory should be taken into account when computing physical observables [357].[26] Below we briefly review how this effect comes about, and study it for the model at hand. We start with Phase III, and look at Phase II projected onto the real line, as in Section 9.3.4.

9.B.1 Phase III

For the two-cut solution in Phase III, let N_L be the number of eigenvalues in the left arc around $\theta = \pi$, $0 \leq N_L \leq N$, and $N_R = N - N_L$ the number of eigenvalues in the right arc around $\theta = 0$. Let also $\xi = \frac{N_L}{N}$ and $1 - \xi = \frac{N_R}{N}$ denote the corresponding filling fractions.

The values of $y_L = \cos(\theta_* + \delta\theta)$ and $y_R = \cos(\theta_* - \delta\theta)$ can be fixed, as functions of ξ and of the other parameters, through the equations

$$\frac{2}{\pi}\int_{-1}^{y_L} \mathrm{d}y \sqrt{\frac{(y_R - y)(y_L - y)}{1 - y^2}} \left[-Y + \frac{\tau\mu(\mu^2 - 1)}{\sqrt{(1 + \mu^2 - 2\mu y_L)(1 + \mu^2 - 2\mu y_R)}(1 + \mu^2 - 2\mu y)} \right] = \xi,$$

$$\frac{2}{\pi}\int_{y_R}^{1} \mathrm{d}y \sqrt{\frac{(y - y_R)(y - y_L)}{1 - y^2}} \left[-Y + \frac{\tau\mu(\mu^2 - 1)}{\sqrt{(1 + \mu^2 - 2\mu y_L)(1 + \mu^2 - 2\mu y_R)}(1 + \mu^2 - 2\mu y)} \right] = 1 - \xi,$$

that come from the definition of ξ after changing variables $y = \cos\theta$. Then, the value of ξ is fixed by the equilibrium condition

$$\left.\frac{\mathrm{d}S_{\text{eff}}}{\mathrm{d}\xi}\right|_{\xi = \xi_{\text{sp}}} = 0. \tag{9.B.1}$$

For example, approximating close to the critical surface diving Phase III from Phase Ib, we find

$$\left.\frac{\partial \xi_{\text{sp}}}{\partial y_R}\right|_{y_R = 1} = \frac{\sqrt{2 - 2y_L}}{\pi}\left(\frac{1}{2} - \frac{\tau}{\mu - 1} + \frac{\tau\mu(\mu + 1)}{(\mu - 1)^3\sqrt{1 + \mu^2 - 2\mu y_L}} \right)$$

where we have also substituted $Y = Y_{\text{cr,b}}$. It has been shown in [357] that the quantum fluctuations around the saddle point ξ_{sp} contribute to the free energy a term of the form $-\frac{1}{N^2}\log\vartheta(N\xi_{\text{sp}})$, where $\vartheta(z)$ is the Jacobi theta function. See Appendix 9.B.2 below for more details and a very short review of the derivation. This is a sub-leading contribution to the free energy but, due to the dependence on $N\xi_{\text{sp}}$, each derivative generates a factor of N. Therefore, the ξ_{sp}-dependent part becomes of the same order as the leading order term when differentiating the Wilson loop vev, and must be taken into account. The relevant part of the derivative is

$$\sum_{y \in \{y_L, y_R\}} \left[\frac{\mathrm{d}}{\mathrm{d}z} \log\vartheta(z)|_{z = N\xi_{\text{sp}}} \right]^2 \left(\frac{\partial y}{\partial Y}\frac{\partial \xi_{\text{sp}}}{\partial y} \right)^2,$$

which yields a non-trivial contribution to the derivative of the Wilson loop vev in Phase III. However, when approaching the critical loci, $\xi \to 0$ if $Y \to Y_{\text{cr,a}}$ or $\xi \to 1$ if $Y \to Y_{\text{cr,b}}$, and the derivative of the theta function evaluated at an integer vanishes.

This shows that the effect of the filling fractions does not play any role in determining the order of the phase transition, despite being non-trivial in the bulk of Phase III.

[26] The original work [357] dealt with Hermitian matrix models, but the argument extends to the present setting.

9.B.2 Phase II

We now discuss the same effect in Phase II. It is more convenient and akin to the work [357] to do this in the alternative, Hermitian matrix model presentation of Section 9.3.4. The argument can be succinctly summarized as follows.

Consider a two-cut solution with support $\mathrm{supp}\rho_{\mathrm{II}}^{\mathrm{p.}} = \Gamma_L \cup \Gamma_R$, and denote by $\xi = \frac{N_L}{N}$ and $1-\xi = \frac{N_R}{N}$ the corresponding filling fractions, as above. The saddle point value ξ_{sp} of ξ is fixed by (9.B.1). Then, in the large N approximation, the partition function takes the form [357]

$$\mathcal{Z} = \sum_{N_L=0}^{N} e^{-N^2\mathcal{F}_{\mathrm{pert}} - \frac{N^2}{2}(\xi-\xi_{\mathrm{sp}})^2\,\partial_\xi^2 S_{\mathrm{eff}}\big|_{\xi=\xi_{\mathrm{sp}}} + \mathcal{O}((\xi-\xi_{\mathrm{sp}})^3)}$$

where $\mathcal{F}_{\mathrm{pert}}$ is the perturbative free energy to all orders in the $\frac{1}{N^2}$ expansion. This yields [357, 378]

$$-\frac{1}{N^2}\log\mathcal{Z} = \mathcal{F} + \frac{1}{N^2}\mathcal{F}^{\mathrm{nlo}} - \frac{1}{N^2}\log\vartheta(N\xi_{\mathrm{sp}}) + \frac{1}{2N^2}\log\left(2\pi\,\partial_\xi^2 S_{\mathrm{eff}}\big|_{\xi=\xi_{\mathrm{sp}}}\right) + \mathcal{O}(N^{-4}),$$

where $\mathcal{F}$ is the leading order or planar free energy, $\mathcal{F}^{\mathrm{nlo}}$ the next-to-leading order correction, and (after an implicit resummation) we have recognized the Jacobi theta function $\vartheta(N\xi_{\mathrm{sp}})$. The modular parameter of the theta function is $\mathrm{i}2\pi/(\partial_\xi^2 S_{\mathrm{eff}}(\xi_{\mathrm{sp}}))$, and the dependence on it is kept implicit in the notation.

For the case at hand, however, the effective action is an even function, the two wells have identical depth, and all the physical observables we consider preserve this property. We thus have $\xi_{\mathrm{sp}} = \frac{1}{2}$, independent of the parameters of the theory, and the effect we have just described will remain sub-leading [378]. This would not be the case for other type of physical observables that are not protected by the parity symmetry. See [357] for discussion and examples.

Part III

Deformations of two-dimensional Yang–Mills theories

Chapter 10

Prologue to Part III

10.1 Yang–Mills theory in two dimensions

Low dimensional quantum field theories have been proved for decades to be a very valuable source of exact results, providing numerous insights into quantum theory as well as showing many direct relationships with statistical mechanical systems, strongly correlated systems and integrable systems, just to name a few.

Yang–Mills theory, which accurately describes strong interactions in four dimensions, is exactly solvable in two dimensions [379, 380]. The solvability stems from the lack of propagating gauge degrees of freedom in two spacetime dimensions. Nonetheless, two-dimensional Yang–Mills theory is an inexhaustible source of inspiration and has received an enormous amount of attention over the years.

A chief reason for this widespread interest is that, despite its apparent simplicity, the theory is not trivial and retains topological and geometric information on the spacetime manifold. Moreover, a host of its deformations is still exactly solvable, thus providing a valuable laboratory to probe ideas and dualities that would be otherwise hard if not impossible to test non-perturbatively. A third motivation is the appearance of two-dimensional Yang–Mills as a protected subsector of four-dimensional supersymmetric gauge theories [381–383].

We review quantum Yang–Mills theory on a closed oriented Riemann surface Σ of genus h and area A [384], which is the main subject of Part III of the dissertation. We focus on gauge group $SU(N)$ for concreteness. This brief chapter is meant as a minimal introduction to the topic, and several details of q-deformed Yang–Mills will be given in due course in the following chapters.

The Yang–Mills Lagrangian density is $\mathrm{Tr} F \wedge *F$, where F is the curvature of a connection in a principal $SU(N)$-bundle (i.e. F is the curvature of the gauge field), and $*$ is the Hodge dual operator. We endow Σ with a symplectic form ω compatible with the Riemannian metric. The Yang–Mills action on Σ is then

$$S_{\mathrm{YM}} = -\frac{1}{2g_{\mathrm{YM}}^2}\int_\Sigma \mathrm{Tr} F * F. \tag{10.1.1}$$

A peculiarity of two dimensions in that, when exponentiated and inserted in a path integral, the action (10.1.1) is equivalent to

$$S_{\mathrm{YM}} = \int_\Sigma \mathrm{Tr}\left(-\mathrm{i}\phi F + \frac{g_{\mathrm{YM}}^2}{2}\phi^2\omega\right).$$

A lattice regularization of the functional integral and the self-reproducing property of the theory

on a plaquette led to the group theoretical expansion of the partition function [379, 380]

$$\mathcal{Z}[\Sigma] = \sum_R (\dim R)^{2-2h} \exp\left(-\frac{A g_{\text{YM}}^2}{2} C_2(R)\right), \tag{10.1.2}$$

where the sum runs over all isomorphism classes R of irreducible representations of the $SU(N)$ gauge group, $\dim R$ is the dimension of the representation R and $C_2(R)$ is the quadratic Casimir invariant of R. For $U(N)$ gauge theory, the expression is identical except that the sum runs over irreducible $U(N)$ representations, that are obtained from those of $SU(N)$ via $U(N) = SU(N) \times U(1)/\mathbb{Z}_N$. Stated differently, using the short exact sequence of groups

$$1 \longrightarrow SU(N) \longrightarrow U(N) \longrightarrow U(1) \longrightarrow 1$$

the irreducible $U(N)$ representations are obtained from those of $SU(N)$ summing over the $U(1)$ sector.

As expected from power counting of mass dimensions, the gauge coupling g_{YM}^2 only appears multiplying the parameter $A = \text{Area}(\Sigma)$, thus, in prevision of a 't Hooft large N limit, we redefine $A' = AN g_{\text{YM}}^2$ (and drop the prime henceforth).

The large N limit of (10.1.2) was studied in [193], showing the presence of a phase transition, known as Douglas–Kazakov (DK) phase transition. As was the case with the Gross–Witten–Wadia phase transition [190–192], it is of third order. Then, in [385], using the method of orthogonal polynomials, it was shown how the transition is triggered by instantons.

Chapter 11

Five-dimensional cohomological localization and squashed q-deformations of two-dimensional Yang–Mills theory

11.1 Introduction to the chapter

Supersymmetric gauge theories in five and six dimensions have undergone a surge of extensive and diverse investigations in recent years. They have played critical roles in the understanding of the strong coupling dynamics of quantum field theory, of M-theory where they can be engineered, and of various problems in geometry. Their compactifications generate many interesting theories in lower dimensions, which can be used to elucidate novel features of lower-dimensional quantum field theories from a geometric perspective.

Five-dimensional gauge theories with $\mathcal{N} = 1$ supersymmetry can be engineered from compactifications of M-theory on Calabi-Yau threefolds [220, 386, 387], which are related to the embedding of topological string theory into M-theory [388]. In six dimensions the worldvolume theory of coincident M5-branes is a six-dimensional $\mathcal{N} = (2, 0)$ superconformal field theory. Twisted compactification of this worldvolume theory on an n-punctured Riemann surface $\Sigma_{h,n}$ of genus h generically engineers an $\mathcal{N} = 2$ superconformal field theory of class $\mathcal{S}$ in four dimensions [389, 390], which upon further twisted compactification on a circle $\mathbb{S}^1$ in the Schur limit is conjecturally equivalent to two-dimensional q-deformed Yang–Mills theory on $\Sigma_{h,n}$ in the zero area limit [391]. Because of the lack of a Lagrangian description of this six-dimensional theory, its twisted partition function is most easily computed by dimensionally reducing on a circle of radius β and computing the twisted partition function of five-dimensional $\mathcal{N} = 2$ supersymmetric Yang–Mills theory with gauge coupling $g^2_{\mathrm{YM}} = 2\pi\,\beta$ [392, 393]; the duality with two-dimensional q-deformed Yang–Mills theory in this five-dimensional setting was checked by [394] using explicit supersymmetric localization techniques. From the four-dimensional perspective, the class $\mathcal{S}$ theory is then equivalent to a three-dimensional theory on a certain squashed deformation M_b of the three-dimensional part of the compactification in six dimensions. We mostly focus on the round three-sphere $\mathbb{S}^3$ and its squashed deformations $\mathbb{S}^3_b$.

These dualities have by now been extensively discussed. The purpose of the present chapter is to survey and investigate these correspondences from a new and more detailed perspective; we give extensive pointers to and comparisons with relevant previous works on the subject as we go along. The main technique which we exploit is cohomological localization, as pioneered by

Källén for three-dimensional theories in [137], and subsequently extended to five dimensions by Källén, Qiu and Zabzine in [230, 231]. This enables a unified treatment of supersymmetric gauge theories in three and five dimensions using solely topological techniques based on the Atiyah–Singer index theorem, which extends and simplifies previous treatments based on supersymmetric localization; this method is ultimately one of the main messages of the present work. Compared to previous work on the subject, we study full topologically twisted theories, instead of partially twisted theories, with which we further extend general results in the literature on cohomological localization. Our simplified treatment also unavoidably comes with some limitations, and we shall extensively discuss which backgrounds do not fit into our localization framework.

In three dimensions we provide different simplified derivations of some known scattered results, treated in a unified framework. For example, the index theory calculations of localization on $\mathbb{S}^3_b$ in [395] uses a procedure which is different from that of [137], and their Lie derivative appearing in the square of the supercharge is not along the Reeb vector field. We will explain this point thoroughly in this chapter. When the supersymmetry transformations of [395, App. B] are put into the cohomological form of [137], index theorem calculations can be applied to obtain the same results in a more economical way. Generally, our procedure of topological twisting rigorously justifies the localization calculations we employ at the field theory level; in particular, it justifies the usage of the index formula and related techniques from [230, 231].

In five dimensions the core of our work starts, where the reader versed in localization techniques may begin. We adapt the formalism of [230, 231] to reobtain some known results in a different way using a twisted gauge theory approach and also extend some general results (see in particular our localization formulas (11.4.4) and (11.4.6)). In particular, we rederive the results of [396, 244] through the Atiyah–Singer index theorem, first in the case of $\mathbb{S}^3 \times \Sigma_{h,0}$ (which for $h = 0$ is a relatively straightforward adaptation from the literature), and then extending it to $\mathbb{S}^3_b \times \Sigma_{h,0}$. We also describe the pushdown of these theories to the horizontal four-dimensional part of the geometry, which was justified in [244] only for $h = 0$, and is different from [231]. We study the resulting theory in detail and describe its precise relation to q-deformed Yang–Mills theory on $\Sigma_{h,0}$; we improve on various results in the literature (with some overlap with [244]), for example determining the two-dimensional theory at non-zero area, and pay more attention to the underlying matrix model, which has not been previously considered. For non-trivial squashing parameters $b \neq 1$, the resulting two-dimensional gauge theory is new and we refer to it as a squashed q-deformation of Yang–Mills theory on $\Sigma_{h,0}$. When $b^2 \in \mathbb{Q}$, we show that this theory is closely related to the q-deformations of Yang–Mills theory considered in [397] in a completely different setting, which in turn is related to three-dimensional Chern–Simons theory on general lens spaces $L(p, s)$.

The outline of the remainder of the current chapter is as follows. We have endeavoured throughout to give a relatively self-contained presentation, while glossing over some well-known technical aspects for which we refer to the pertinent literature; thus some of the earlier sections in the chapter are somewhat expository in nature. We begin in Section 11.2 by giving a more detailed introduction and background to the setting discussed briefly above, summarising the geometric settings and classifications, techniques and notation used.

Section 11.3 is dedicated to the three-dimensional case, wherein we review the ideas behind Källén's cohomological localization technique, and explain how they are modified on squashed manifolds. We discuss the construction of the cohomological gauge theory, the computation of the one-loop determinants for the vector multiplet on the two possible classes of Seifert three-manifolds admitting $\mathcal{N} = 2$ supersymmetry, and the computation of the hypermultiplet one-loop determinants. We describe several explicit applications of our localization formulas. Section 11.4 presents the five-dimensional analog of our considerations from Section 11.3. We describe the cohomological field theory and derive the one-loop determinants of the vector multiplet in the two classes of

Seifert five-manifolds admitting $\mathcal{N}=1$ supersymmetry. We describe explicit applications of our localization formulas, and also explain in detail the relationship with a four-dimensional theory. Section 11.5 explains the relation between our five-dimensional cohomological field theories and two-dimensional Yang–Mills theory. We derive the standard q-deformation of Yang–Mills theory on $\Sigma_{h,0}$ through a localization calculation on $\mathbb{S}^3 \times \Sigma_{h,0}$, and subsequently extend these considerations to the squashed deformations $\mathbb{S}^3_b \times \Sigma_{h,0}$ where we obtain a new two-parameter deformation. We investigate the matrix model in detail and derive a new correspondence with Chern–Simons gauge theory on lens spaces $L(p,s)$. We conclude by briefly addressing how to obtain more general deformations of two-dimensional Yang–Mills theory through localization calculations in higher dimensions.

Three appendices at the end of the chapter provide various technical details complementing some of the analysis in the main text. Appendix 11.A summarises our conventions and notation for spinors, which are adapted to treat the three-dimensional and five-dimensional cases, as well as the vector multiplets and matter hypermultiplets, in a unified way. Appendix 11.B provides mathematical details of the different types of squashings of spheres that preserve $\mathcal{N}=2$ supersymmetry in three dimensions. Appendix 11.C provides some mathematical details on Sasaki-Einstein manifolds, which preserve $\mathcal{N}=1$ supersymmetry in five dimensions, and we briefly review the formalism of [398] to better explain why our results appear to be so different from those of [398, 399].

11.2 Preliminaries on superconformal field theories and localization

In this preliminary section we collect the relevant background material that will be used throughout the current chapter. We begin with a discussion of superconformal field theories in six dimensions, which gives one of the primary motivations behind the present investigations. We discuss how certain compactifications of these theories on a Riemann surface suggest a duality between four-dimensional superconformal theories and the standard q-deformation of two-dimensional Yang–Mills theory. The purpose of this chapter is to investigate in detail how this duality is modified in the case where the three-dimensional part of the compactification is a squashed geometry, and we describe how cohomological localization techniques for superconformal field theories on Seifert manifolds will be applied to investigate the correspondence.

11.2.1 Squashed geometry and six-dimensional superconformal field theories

Consider the six-dimensional superconformal $\mathcal{N}=(2,0)$ theory with gauge group G of ADE type on a twisted compactification of the form

$$M_6 = \mathbb{S}^1 \times \mathbb{S}^3 \times \Sigma_h \ ,$$

where $\mathbb{S}^3$ is the standard round three-sphere and Σ_h is a compact oriented Riemann surface of genus h without boundaries. This setup dictates a remarkable duality: The correlators of a certain two-dimensional topological quantum field theory on Σ_h compute the partition function of a four-dimensional $\mathcal{N}=2$ field theory of class $\mathcal{S}$ on $\mathbb{S}^1 \times \mathbb{S}^3$, which is the superconformal index [400]

$$C_{\rm I}(u,q,t) = \mathrm{Tr}\,(-1)^{\mathsf{F}} \left(\frac{t}{u\,q}\right)^{\mathsf{r}} u^{\mathsf{J}_+}\, q^{\mathsf{J}_-}\, t^{\mathsf{R}},$$

where $\mathsf{J}_\pm$ are the rotation generators in the two orthogonal planes constructed from the Cartan generators of the Lorentz $SU(2)_{\rm L} \times SU(2)_{\rm R}$ isometry group of $\mathbb{S}^3$, the operator r is the $U(1)_r$

generator, and R is the $SU(2)_R$ generator of R-symmetries. The superconformal index of four-dimensional $\mathcal{N}=2$ superconformal field theories was originally introduced in [334, 401], and it is a highly non-trivial function of the three superconformal fugacities (u,q,t).

However, this duality is difficult to test because our current understanding of the six-dimensional $\mathcal{N}=(2,0)$ theory is rather incomplete. Instead, we dimensionally reduce over $\mathbb{S}^1$, which yields five-dimensional supersymmetric Yang–Mills theory. The Yang–Mills coupling in five dimensions has dimensions of length, $g_{\rm YM}^2 = 2\pi\,\beta$, where β is the radius of $\mathbb{S}^1$. The dimensional reduction over $\mathbb{S}^1$ of the supersymmetric partition function on $\mathbb{S}^1\times\mathbb{S}^3$ is achieved by assigning scaled chemical potentials to the fugacities according to

$$q = e^{2\pi{\rm i}\beta\,\epsilon_1}\ , \qquad t = e^{2\pi{\rm i}\beta\,\epsilon_2} \qquad \text{and} \qquad u = e^{2\pi{\rm i}\beta}\ ,$$

and taking the limit $\beta\to 0$. Then the four-dimensional index becomes a three-dimensional ellipsoidal partition function [402–404], i.e. the partition function on the squashed sphere $\mathbb{S}^3_b$ with squashing parameter

$$b = \sqrt{\frac{\epsilon_1}{\epsilon_2}}\ .$$

This deformation of the three-sphere of radius $r=\sqrt{\epsilon_1\,\epsilon_2}$ can be parametrized by the ellipsoid in $\mathbb{C}^2$ defined by

$$b^2\,|z_1|^2 + b^{-2}\,|z_2|^2 = r^2\ ,$$

which has isometry group $U(1)\times U(1)$.

Hence we consider supersymmetric Yang–Mills theory on the five-manifold

$$M_5 = \mathbb{S}^3_b\times\Sigma_h\ .$$

This theory sits in the infrared of a renormalization group flow triggered by a relevant deformation of a five-dimensional superconformal field theory in the ultraviolet [215]. The coupling to the ellipsoid $\mathbb{S}^3_b$ [281] described above, preserves four supercharges, and the theory can alternatively be taken to be a half-BPS descendant of the $\mathcal{N}=(1,0)$ superconformal field theory on $\mathbb{S}^1\times\mathbb{S}^3_b\times\Sigma_h$. On the other hand, to preserve eight supercharges we may use the squashed sphere $\mathbb{S}^3_b$ of [405] (see Appendix 11.B for details), for which the five-dimensional theory descends from twisted compactification of the $\mathcal{N}=(2,0)$ superconformal field theory on $\mathbb{S}^1\times\mathbb{S}^3\times\Sigma_h$ with round $\mathbb{S}^3$. The geometric meaning of the squashing parameter b is different in the two cases: $b>0$ for the ellipsoid, while $b\in\mathbb{C}$ with $|b|=1$ for the squashed sphere. From the point of view of the matrix ensemble we will find in Section 11.5.4, the natural choice would be $b>0$. In Section 11.5 we consider both cases. We will find that, regardless of what choice we make for $\mathbb{S}^3_b$, the partition function is the same and is a holomorphic function of b in the punctured complex plane $\mathbb{C}\setminus\{0\}\cong\mathbb{P}^1\setminus\{0,\infty\}$, so we may start with either the squashed sphere or the ellipsoid and then analytically continue the result.

11.2.2 Reductions to two and three dimensions

By a localization calculation over the squashed sphere $\mathbb{S}^3_b$, we will identify the partition function of the two-dimensional gauge theory dual on Σ_h of the $\mathcal{N}=2$ theory of class $\mathcal{S}$. We are exclusively interested in the slice of the superconformal fugacity space defined by the Schur limit $u=0$, $q=t$ of the superconformal index which is the Schur index

$$\mathcal{C}_{\rm I}(q) = {\rm Tr}\,(-1)^{\mathsf{F}}\,q^{\mathsf{J}_-+\mathsf{R}},$$

where the trace is now restricted to states with $U(1)_r$ charge $\mathsf{r}=\mathsf{J}_+$. In this case the index was computed in [391, 406] from a topological quantum field theory on the Riemann surface Σ_h, which

can be identified with the zero area limit $\mathrm{vol}(\Sigma_h) = 0$ of the usual q-deformed two-dimensional Yang–Mills theory; the duality with this two-dimensional gauge theory is confirmed in [394] by an explicit localization computation on $\mathbb{S}^3 \times \Sigma_h$ (i.e. for $b = 1$). The q-deformed Yang–Mills theory is not topological when $\mathrm{vol}(\Sigma_h) \neq 0$, but it still has a natural class $\mathcal{S}$ theory interpretation as the supersymmetric partition function of the $(2,0)$ theory on $\mathbb{S}^1 \times \mathbb{S}^3 \times \Sigma_h$ where the area of the ultraviolet curve Σ_h is kept finite [407]. We shall study this proposal explicitly via a localization calculation on the five-manifold $\mathbb{S}^3_b \times \Sigma_h$. The two-dimensional theory we find is a further deformation by the squashing parameter b, that we call 'squashed' q-deformed Yang–Mills theory. For later use, let us now briefly review the standard q-deformation of two-dimensional Yang–Mills theory.

Let $\mathfrak{g}$ be the Lie algebra of a connected Lie group G. Let $\triangle$ be the root system of $\mathfrak{g}$ and $\triangle_+$ the system of positive roots; similarly let $\Lambda \cong \mathbb{Z}^{\mathrm{rank}(G)}$ be the weight lattice of $\mathfrak{g}$ with dominant weights Λ_+. We fix an invariant bilinear form $(-,-)$ on $\mathfrak{g}$, usually the Killing form. Let

$$\delta = \frac{1}{2} \sum_{\alpha \in \triangle_+} \alpha$$

be the Weyl vector of $\mathfrak{g}$. The vector $2\,\delta$ is always a weight of $\mathfrak{g}$; if G is semi-simple, then δ is also a weight and we can identify it with an integer vector $\delta \in \mathbb{Z}^{\mathrm{rank}(G)}$.

The partition function for the q-deformation of Yang–Mills theory with gauge group G on a closed oriented Riemann surface Σ_h of genus $h \geq 0$ can be written as a generalization of the Migdal heat kernel expansion given by [408]

$$\mathcal{Z}_{h,p}(q) = \sum_{\lambda \in \Lambda_+} \dim_q(R_\lambda)^{2-2h}\, q^{\frac{p}{2}\,(\lambda+2\,\delta,\lambda)}\ , \tag{11.2.1}$$

where $p \in \mathbb{Z}$ is a discrete parameter and the sum runs over all isomorphism classes of irreducible unitary representations R_λ of G which are parametrized by dominant weights $\lambda \in \Lambda_+$. The deformation parameter

$$q = e^{-g_{\mathrm{str}}}$$

is identified with the coupling constant g_{str} in topological string theory. The quantum dimension of the representation R_λ labelled by $\lambda \in \Lambda_+$ is

$$\dim_q(R_\lambda) = \prod_{\alpha \in \triangle_+} \frac{\big[(\lambda+\delta,\alpha)\big]_q}{\big[(\delta,\alpha)\big]_q}, \tag{11.2.2}$$

where

$$[x]_q = \frac{q^{x/2} - q^{-x/2}}{q - q^{-1}} \tag{11.2.3}$$

for $x \in \mathbb{R}$ is a q-number. This theory is closely related to Chern–Simons theory on a principal $U(1)$-bundle of degree p over Σ_h [136].

For many computations it is useful to have an explicit expression for the partition function (11.2.1) in terms of highest weight variables. For this, we define shifted weights $\vec{k} \in \mathbb{Z}^{\mathrm{rank}(G)}$ by

$$\vec{k} = \lambda + \delta\ , \tag{11.2.4}$$

and use the Weyl reflection symmetry of the summand of the partition function (11.2.1) to remove the restriction to the fundamental chamber of the summation over $\vec{k}$. Up to overall normalization,

the partition function (11.2.1) can thus be written as

$$\mathcal{Z}_{h,p}(q) = \sum_{\vec{k}\in(\mathbb{Z}^{\mathrm{rank}(G)})_{\mathrm{reg}}} \Delta(g_{\mathrm{str}}\,\vec{k}\,)^{2-2h}\, e^{-\frac{p\,g_{\mathrm{str}}}{2}\,(\vec{k},\vec{k}\,)}\ , \tag{11.2.5}$$

where the Weyl determinant is given by

$$\Delta(\vec{x}\,) = \prod_{\alpha\in\triangle_+} 2\sinh\frac{(\alpha,\vec{x}\,)}{2}$$

for $\vec{x} = (x_1,\dots,x_{\mathrm{rank}(G)}) \in \mathbb{C}^{\mathrm{rank}(G)}$. The sum in (11.2.5) is restricted to those shifted weights where $\Delta(g_{\mathrm{str}}\,\vec{k}\,)$ is non-zero, i.e. $(\alpha,\vec{k}\,) \neq 0$ for $\alpha \in \triangle_+$.

When the gauge group is the unitary group $G = U(N)$, this two-dimensional gauge theory is conjecturally a non-perturbative completion of topological string theory on the local Calabi-Yau threefold which is the total space of the rank 2 holomorphic vector bundle

$$\mathcal{O}_{\Sigma_h}(p+2h-2)\oplus\mathcal{O}_{\Sigma_h}(-p) \longrightarrow \Sigma_h,$$

with N D4-branes wrapping the exceptional divisor $\mathcal{O}_{\Sigma_h}(-p)$ and D2-branes wrapping the base Σ_h [408]. In turn, for $h = 0$ the two-dimensional theory defines an analytical continuation of Chern–Simons gauge theory on the lens space $L(p,1)$ to the case where q is not a root of unity [409]. In Section 11.5 we shall find that five-dimensional cohomological localization over $\mathbb{S}^3_b \times \Sigma_h$ gives a squashed deformation of this theory at $p = 1$, which for genus $h = 0$ and rational values p/s of the squashing parameter b^2 is an analytical continuation of Chern–Simons theory on the more general lens spaces $L(p,s) \cong \mathbb{S}^3_b$.

The correspondence with three-dimensional field theories has also been studied from other perspectives. In [410], the dimensional reduction of six-dimensional theories on $\mathbb{S}^1 \times \mathbb{S}^3_b \times \Sigma_h$ is considered by reducing to five dimensions as above, compactifying on Σ_h, and then obtaining a three-dimensional theory on the squashed sphere; this enables a comparison of the two possible compactification paths: first along $\mathbb{S}^1$ and then on Σ_h to obtain a theory on $\mathbb{S}^3_b$, or first along Σ_h to obtain a four-dimensional theory of class $\mathcal{S}$ and then relating it by standard reasoning to the three-dimensional partition function. In [411–413] the five-dimensional theory on $M_3 \times \mathbb{S}^2$ is obtained from dimensional reduction of the six-dimensional superconformal field theory on $\mathbb{S}^1 \times M_3 \times \mathbb{S}^2$ for more general three-manifolds M_3. The resulting theory on M_3 is related to complex Chern–Simons theory. In these cases the theory is partially twisted along M_3, and supersymmetric localization on $\mathbb{S}^2$ is used to reduce to a twisted three-dimensional theory; this differs from the perspective of [394, 396], where the partial twist is along Σ_h and localization over $\mathbb{S}^3$ reduces the theory to two dimensions. In contrast, here we will consider the fully twisted theory on $\mathbb{S}^3_b \times \Sigma_h$. Finally, in [414], with squashed sphere of [405], the six-dimensional theory on $\mathbb{S}^3_b \times M_3$ is reduced along the Hopf fibre of $\mathbb{S}^3_b$, then twisted along M_3 and localized to reduce along $\mathbb{S}^2$; the resulting three-dimensional theory is the same as in [411, 412].

11.2.3 Basics of cohomological localization

In this chapter we use techniques based on localization theorems in equivariant cohomology, applied to supersymmetric quantum field theory, see e.g. [37, 38] for introductions to the subject; we will now briefly sketch the main ideas that will be used extensively in the remainder of this chapter. Supersymmetric localization is a technique which allows the reduction of a supersymmetry preserving Euclidean path integral to an integral over the smaller set of fixed points of a supercharge $\mathcal{Q}$. To compute the partition function, one adds a $\mathcal{Q}$-exact term $\mathcal{Q}V$ to the action S of the theory and computes the deformed partition function $\mathcal{Z}(t)$ defined by functional integration of

the Boltzmann weight $e^{-S-t\,\mathcal{Q}V}$ for $t\in\mathbb{R}$. $\mathcal{Z}(0)$ is the original partition function we wish to compute. Supersymmetry of the path integral then implies that $\mathcal{Z}(t)$ is formally independent of the parameter t, so that letting $t\to\infty$ and choosing the localizing term V to be positive semi-definite, the functional integration reduces to a localization calculation around the fixed points $\mathcal{Q}V=0$ of the supercharge $\mathcal{Q}$. In general the set of fixed points is a superspace, with odd coordinates associated with supersymmetric fermionic modes that have vanishing action in the localizing term in the bosonic background.

As in the rest of the dissertation, we consider Euclidean manifolds preserving rigid supersymmetry; we restrict them to non-trivial circle bundles $M_{2n+1}\to K_{2n}$, whose total spaces have odd dimension $2n+1$, in order to avoid dealing with fermionic fixed points. We shall derive the fixed point loci based on cohomological forms of the (BRST) supersymmetry transformations which are compatible with the $U(1)$-action on the circle bundle [137, 415]; this procedure is called topological twisting and the resulting theory is called a cohomological field theory (in the sense of equivariant cohomology). We shall also only consider localization on the Coulomb branch of the supersymmetric gauge theory, where the path integral is reduced to a finite-dimensional integral over a classical moduli space parameterized by scalars in vector multiplets, and holonomies and fluxes of gauge fields around non-contractible cycles in M_{2n+1} (possibly together with other continuous moduli).

The localization calculation amounts to computing a ratio of one-loop fluctuation determinants which is schematically given by [56]

$$h(\phi)=\frac{\det \mathrm{i}\mathsf{L}_\phi|_{\mathrm{coker}\,\mathsf{D}}}{\det \mathrm{i}\mathsf{L}_\phi|_{\ker\mathsf{D}}}\ ,$$

where D denotes differential operators entering the localizing terms V, and $\mathsf{L}_\phi=-\mathrm{i}\mathcal{Q}^2$ generates the geometric $U(1)$-action and gauge transformations parametrized by ϕ on the fields of the theory; the adjoint scalar ϕ is $\mathcal{Q}$-closed and does not have a fermionic partner. Then the equivariant cohomology in the localization of the supersymmetric gauge theory consists of gauge-invariant states on the base space K_{2n} of the circle bundle, together with an infinite tower of Kaluza-Klein modes on the $\mathbb{S}^1$ fibre. The effective contribution to this ratio from the zero modes which remain after cancellation between fermionic and bosonic states is computed using the Atiyah–Singer index theorem for transversally elliptic operators and the Atiyah–Bott localization formula in equivariant cohomology for the $U(1)$-action on M_{2n+1}. The schematic form of the localized partition function is then given by

$$Z(M_{2n+1})=\int_{\mathfrak{g}}\mathrm{d}\sigma\int_{\mathscr{M}_G^{\mathrm{BPS}}(M_{2n+1})}\mathrm{d}\mathbf{m}\ e^{-S_{\mathrm{cl}}(\mathbf{m};\sigma)}\ Z_{\mathrm{vec}}(M_{2n+1})\,Z_{\mathrm{hyp}}(M_{2n+1})\ ,$$

where $\mathfrak{g}$ is the Lie algebra of the gauge group G, and $\mathscr{M}_G^{\mathrm{BPS}}(M_{2n+1})$ is the BPS locus inside a moduli space of G-connections on M_{2n+1} parametrized by moduli $\mathbf{m}$. The action S_{cl} is the classical bosonic action, while Z_{vec} and Z_{hyp} are respectively the one-loop fluctuation determinants associated with the vector multiplet and the matter hypermultiplets of the supersymmetric gauge theory.

11.2.4 Localization of $\mathcal{N}=1$ gauge theories on Seifert manifolds

Our focus is mainly on Seifert manifolds which admit a free $U(1)$ action, so that they admit a $U(1)$ isometry. We shall comment where appropriate on the extension to more general principal $U(1)$-bundles over orbifolds, where the $U(1)$ action has fixed points.

Geometric setup

Let $M_{2n+1}\xrightarrow{\pi}K_{2n}$ be a circle bundle of degree p over a compact Kähler manifold (K_{2n},ω) of real dimension $2n$ with $[\omega]\in H^2(K_{2n},\mathbb{Z})$. The almost contact structure $\kappa\neq0$ on M_{2n+1} can be

chosen to be a connection one-form on this bundle which is locally written as

$$\kappa = \mathrm{d}\theta + p\,\pi^*(a)\ , \tag{11.2.6}$$

where $\theta \in [0, 2\pi\, r)$ is a local coordinate of the $\mathbb{S}^1$ fibre and a is a local symplectic potential for $\omega = \mathrm{d}a$. Then

$$\mathrm{d}\kappa = p\,\pi^*(\omega)\ . \tag{11.2.7}$$

In contrast to [230, 231], we will not assume κ to be a K-contact structure on M_{2n+1}. Instead, our interest will mainly focus on product manifolds $M_{2n+1} = M_{2n-1} \times \Sigma_h$, where M_{2n-1} is a compact contact manifold, but in general not K-contact, and Σ_h is a compact Riemann surface of genus h. Then $\omega = \omega_{K_{2n-2}} + \omega_{\Sigma_h}$ is the sum of the symplectic forms on the base K_{2n-2} of the Seifert fibration of M_{2n-1} and Σ_h, and $\kappa \wedge (\mathrm{d}\kappa)^{\wedge(n-1)} \neq 0$ is proportional to the volume form on M_{2n-1} induced by a metric compatible with the contact structure. The canonical volume form on the total space M_{2n+1} is

$$\mathrm{d}\Omega_{M_{2n+1}} = \frac{(-1)^n}{2^{n-1}\,(n-1)!}\ \kappa \wedge (\mathrm{d}\kappa)^{\wedge(n-1)} \wedge \omega_{\Sigma_h}\ . \tag{11.2.8}$$

The Reeb vector field ξ is defined by the duality contraction

$$\xi \lrcorner \kappa = 1$$

and the invariance condition

$$\mathcal{L}_\xi \kappa = \xi \lrcorner\, \mathrm{d}\kappa = 0\ ,$$

where $\mathcal{L}_\xi = \mathrm{d}\,\xi \lrcorner + \xi \lrcorner\, \mathrm{d}$ is the Lie derivative along ξ. It is the generator of the $U(1)$-action on M_{2n+1}, and in the coordinates (11.2.6) it assumes the form

$$\xi = \frac{\partial}{\partial\theta}\ .$$

A natural choice of $U(1)$-invariant metric on M_{2n+1} is given by

$$\mathrm{d}s^2_{M_{2n+1}} = \pi^*\big(\mathrm{d}s^2_{K_{2n}}\big) + \kappa \otimes \kappa\ ,$$

where $\mathrm{d}s^2_{K_{2n}}$ is the Kähler metric on K_{2n}. Any k-form α on M_{2n+1} can be decomposed using the projector $\kappa \wedge \xi \lrcorner$ into horizontal and vertical components as

$$\alpha = \alpha_H + \alpha_V := (1 - \kappa \wedge \xi \lrcorner)\alpha + \kappa \wedge \xi \lrcorner\, \alpha\ ,$$

where $\xi \lrcorner\, \alpha$ is the $k-1$-form component of α along the fibre direction.

The computation of the perturbative partition function of (twisted) $\mathcal{N} = 1$ supersymmetric Yang–Mills theory on M_{2n+1} is described in [230, 231], using equivariant localization techniques with respect to the $U(1)$ action on M_{2n+1} and the maximal torus of the gauge group G. The relevant computations typically involve the determinant of the kinetic operator

$$\mathsf{L}_\phi = \mathcal{L}_\xi + \mathcal{G}_\phi$$

acting on the tangent space to the space of fields, where $\mathcal{G}_\phi$ denotes the action by an element ϕ valued in the Cartan subalgebra of the Lie algebra $\mathfrak{g}$ of G; for fields in the vector multiplet of the supersymmetric gauge theory, $\mathcal{G}_\phi = \mathrm{ad}_\phi$ is the adjoint action. Here we assume momentarily that the localization locus consists of constant field configurations ϕ; the case of non-constant ϕ is discussed below. The operator L_ϕ acts with the same eigenvalue on both even and odd degrees in

the spaces $\Omega_H^{(0,\bullet)}(M_{2n+1},\mathfrak{g})$ of horizontal anti-holomorphic $\mathfrak{g}$-valued forms. The cancellation between bosonic and fermionic fluctuation determinants in the localized path integral is determined by the index of the Dolbeault complex of K_{2n} twisted by the line bundles $\mathscr{L}^{\otimes m}$ for $m\in\mathbb{Z}$, where $\mathscr{L}\to K_{2n}$ is the complex line bundle associated to the circle bundle $M_{2n+1}\xrightarrow{\pi}K_{2n}$ with first Chern class $c_1(\mathscr{L})=p\,[\omega]$. Denoting the corresponding twisted Dolbeault operators as $\bar\partial^{(m)}$, the Atiyah–Singer index theorem gives the index as

$$\operatorname{index}\bar\partial^{(m)}=\int_{K_{2n}}\operatorname{ch}\big(\mathscr{L}^{\otimes m}\big)\wedge\operatorname{Td}\big(T^{1,0}K_{2n}\big)\ ,$$

where $T^{1,0}K_{2n}$ is the holomorphic tangent bundle of K_{2n}, while ch and Td respectively denote the Chern character and the Todd class.

Geometry from rigid supersymmetry

We want to define a supersymmetric field theory on the backgrounds M_{2n+1} described above. We focus on $\mathcal{N}=1$ theories in five dimensions; a thorough discussion on localization in five-dimensional superconformal field theories can be found in [237]. We follow the approach of [416]: we add a supergravity multiplet in flat spacetime, and then take the rigid limit of the supergravity theory. This is done in two steps: in the first step we set to zero all fermion fields, and in the second step we set to zero the supersymmetry variations of the fermion fields. The equation obtained by imposing the vanishing of the gravitino variation is called the (generalized) Killing spinor equation. The vector multiplet and hypermultiplets of the gauge theory are then coupled to the background values of the supergravity multiplet, and the theory is effectively put on a curved spacetime. For the description of supersymmetric backgrounds in three dimensions we mainly follow [417, 418], and in five dimensions we follow [419, 420].

Let ε be a solution to the Killing spinor equation for M_{2n+1}. We use it to define the vector field v on M_{2n+1} as

$$v^\mu=\varepsilon^\dagger\,\Gamma^\mu\varepsilon\ , \tag{11.2.9}$$

where Γ^μ are the gamma-matrices in either three or five dimensions. The fact that ε satisfies the Killing spinor equation guarantees that v is a nowhere vanishing Killing vector field. In particular, its orbits foliate M_{2n+1}.

As explained in [417, Sec. 5] and in [418, Sec. 4], in three dimensions there are two possibilities:

(I) The orbits of v are closed. In this case M_3 is a Seifert manifold and v coincides with the Reeb vector field ξ of the $U(1)$ fibration of M_3. Particular examples belonging to this class are the round sphere $\mathbb{S}^3$ and the lens spaces $L(p,s)$.

(II) If the orbits of v do not close, supersymmetry requires M_3 to have isometry group $U(1)\times U(1)$. In this case M_3 is a Seifert manifold but v does not necessarily point along the $U(1)$-fibre. A particular example belonging to this class is the ellipsoid $\mathbb{S}^3_b$ of [281].

We will henceforth refer to the manifolds belonging to the setting (I) as *regular*, and to those of setting (II) as *irregular*.[27] Irregular geometries may or may not admit a free $U(1)$ action of the Reeb vector field; see the recent review [421] for a thorough description of the geometric approach to $\mathcal{N}=2$ supersymmetry on three-dimensional Seifert manifolds.

In the five-dimensional case, we will focus on product manifolds $M_3\times\Sigma_h$, where Σ_h is a closed Riemann surface of genus h. Killing spinor solutions in these geometries are built from solutions on M_3, and thus an analogous discussion applies; see the discussion at the end of [420] for further details about the difference in the approach we follow here and that of [231, 422].

[27] We will use the nomenclature "regular geometry" or "regular fibration", meaning that the integral curves of the Killing vector field have regular flow, and similarly for the "irregular fibration". We refrain from distinguish between irregular and quasi-regular cases.

Examples: Regular vs irregular fibrations

We write down some explicit examples of regular and irregular five-dimensional manifolds. The round sphere $\mathbb{S}^5$ and the product $\mathbb{S}^3\times\mathbb{S}^2$ are regular Sasaki-Einstein manifolds. The Sasaki-Einstein manifolds $Y^{p,s}$ studied in [398, 399] are irregular (or quasi-regular). The product manifolds $M_3\times\Sigma_h$ where M_3 is either $\mathbb{S}^3$, $L(p,1)$ or the three-dimensional torus $\mathbb{T}^3$ are regular, while if $M_3=\mathbb{S}^3_b$ is the ellipsoid of [281] it is irregular. Among the irregular manifolds, $\mathbb{S}^3_b\times\Sigma_h$ (as well as replacing $\mathbb{S}^3_b$ with other squashed Seifert three-manifolds) admits a free $U(1)$ action, while $Y^{p,s}$ do not admit any free $U(1)$ action and are described as $U(1)$ fibrations over a warped product $\mathbb{S}^2\rtimes\mathbb{S}^2$. See Appendix 11.B for a classification and discussion of the different types of squashed three-spheres, and Appendix 11.C for a discussion about cohomological localization on $Y^{p,s}$.

One-loop determinants

For regular fibrations, it was shown in [230, 231] that the one-loop contribution of the $\mathcal{N}=1$ vector multiplet to the perturbative partition function on M_{2n+1} is

$$Z_{\rm vec}(M_{2n+1})=\prod_{\alpha\in\triangle}\big(\mathrm{i}\,(\alpha,\phi)\big)^d\prod_{m\neq0}\Big(\frac{m}{r}+\mathrm{i}\,(\alpha,\phi)\Big)^{\mathrm{index}\,\bar\partial^{(m)}}\,, \tag{11.2.10}$$

where as previously $\triangle$ is the root system of the Lie algebra $\mathfrak{g}$ and $(\,\cdot\,,\,\cdot\,)$ is an invariant non-degenerate bilinear form on $\mathfrak{g}$; the power d of the first zero mode factor is given by

$$d=\mathrm{index}\,\bar\partial-\dim H^0(M_{2n+1},\mathbb{R})\,,$$

the difference between the index of the ordinary (untwisted) Dolbeault complex of K_{2n} and the dimension of the space of harmonic functions on M_{2n+1}. The one-loop contribution from an $\mathcal{N}=1$ hypermultiplet in a representation R of the gauge group G is given by

$$Z_{\rm hyp}(M_{2n+1})=\prod_{\rho\in\Lambda_R}\prod_{m\in\mathbb{Z}}\Big(\frac{m}{r}+\mathrm{i}\,(\rho,\phi)+\frac{\Delta}{r}\Big)^{-\mathrm{index}\,\bar\partial^{(m)}}\,, \tag{11.2.11}$$

where Λ_R is the lattice of weights of R and Δ is a constant determined by the conformal scalar field coupling to the curvature in the gauge theory action. In [231] these formulas are applied to $\mathcal{N}=1$ gauge theory on the five-sphere $M_5=\mathbb{S}^5$, viewed as a circle bundle over the projective plane $K_4=\mathbb{P}^2$, with $\Delta=\frac{3}{2}$.

The extension of these formulas to the case of non-constant scalar fields ϕ on K_{2n} can be deduced from the prescription explained in [423, App. B], at least in the case when the kinetic operator L_ϕ is elliptic. In these instances one can apply the index formula "locally" by moving the logarithms of the arguments of the products into the integral and integrating against the index density. For example, for the vector multiplet contribution this prescription gives

$$\begin{aligned}Z_{\rm vec}(M_{2n+1})=\exp\Bigg(\int_{K_{2n}}\sum_{m\in\mathbb{Z}}\mathrm{ch}\big(\mathscr{L}^{\otimes m}\big)\wedge\mathrm{Td}\big(T^{1,0}K_{2n}\big)\sum_{\alpha\in\triangle}\log\Big(\frac{m}{r}+\mathrm{i}(\alpha,\phi)\Big)\\-\dim H^0(M_{2n+1},\mathbb{R})\int_{K_{2n}}\frac{\omega^{\wedge n}}{n!}\sum_{\alpha\in\triangle}\log\big(\mathrm{i}(\alpha,\phi)\big)\Bigg)\,.\end{aligned}$$

For the cohomological localization we shall employ, the further localization to constant ϕ in two dimensions will be immediate (in contrast to the approach of [394]).

The expressions (11.2.10) and (11.2.11) are proven in [230, 231] in the case of five-dimensional K-contact manifolds, with Killing vector field v pointing along the Seifert fibre. In following

sections we will review the main steps in the proof, both in three and five dimensions, and derive the corresponding expressions for the cases in which the Killing vector field v does not point in the direction of the $U(1)$ fibre. For this, we will have to introduce vector multiplets and hypermultiplets, and then topologically twist the field content. For the three-dimensional case we will follow the conventions of [137], while in five dimensions we follow [231, 424].

11.3 $\mathcal{N}=2$ cohomological gauge theories in three dimensions

In this section we study $\mathcal{N}=2$ supersymmetric gauge theories on three-dimensional manifolds. We will first present the theory and its topological twist. Then we will reproduce the formula for the one-loop determinants in the case of a Seifert fibration $M_3 \to K_2$ corresponding to closed orbits of the Killing vector field v. We shall subsequently extend the formulas to the ellipsoid and ellipsoidal lens spaces, corresponding to non-compact orbits of the Killing vector field, by extending the application of the index theorem used in [395], for the ellipsoid, to any squashed Seifert manifold. For the geometric setting we will follow [417, 418], while for the topological twist and derivation of the one-loop determinants, as well as for the normalization of the fields and supersymmetry variations, we will continue to follow [137]. Our conventions are summarized in Appendix 11.A.

11.3.1 Supersymmetric Yang–Mills theory and its cohomological formulation

As usual, we start by placing the gauge theory on flat Euclidean space $\mathbb{R}^3$ and then couple it to background supergravity, following [416, 417]. Let ε and $\widetilde{\varepsilon}$ be two Killing spinors with opposite R-charge, and define the Killing vector field

$$v^\mu = \widetilde{\varepsilon}^{\,\dagger}\,\gamma^\mu \varepsilon \ .$$

Vector multiplet

The $\mathcal{N}=2$ vector multiplet in three dimensions consists of a gauge connection A, a real scalar σ, a complex spinor λ (the gaugino), and an auxiliary real scalar D. The 3d $\mathcal{N}=2$ superalgebra admits a $U(1)_R$ R-symmetry. However, by analogy with the five-dimensional setting, we assume it is enhanced to $SU(2)_R$. For a generic $\mathcal{N}=2$ theory with $U(1)_R$ R-symmetry, one simply neglects the $SU(2)_R$ indices I, J. With this caveat in mind, the spinor $\lambda = (\lambda^I)$ and the real scalar $D = (D^I{}_J)$ carry $SU(2)_R$ indices. The supersymmetry transformations are standard [58, 137]. We denote by $\mathcal{Q}$ the equivariant differential (supersymmetry generator) which is the sum of the two independent supercharges $\frac{1}{2}\left(\widetilde{\mathcal{Q}}_{\widetilde{\varepsilon}} + \mathcal{Q}^\dagger_{\varepsilon}\right)$, and write

$$\begin{aligned}
\mathcal{Q}A_\mu &= \tfrac{\mathrm{i}}{2}\left(\widetilde{\varepsilon}_I^{\,\dagger}\,\gamma_\mu \lambda^I - \lambda_I^\dagger\,\gamma_\mu \varepsilon^I\right) \ , \\
\mathcal{Q}\sigma &= -\tfrac{1}{2}\left(\widetilde{\varepsilon}_I^{\,\dagger}\,\lambda^I + \lambda_I^\dagger\,\varepsilon^I\right) \ , \\
\mathcal{Q}\lambda^I &= -\tfrac{1}{2}\,\gamma^{\mu\nu}\varepsilon^I\,F_{\mu\nu} - D^I{}_J\,\varepsilon^J + \mathrm{i}\,\gamma^\mu \varepsilon^I\,(\mathsf{D}_\mu \sigma) \ , \\
\mathcal{Q}\lambda_I^\dagger &= \tfrac{1}{2}\,\widetilde{\varepsilon}_I^{\,\dagger}\gamma^{\mu\nu}\,F_{\mu\nu} - \widetilde{\varepsilon}_J^{\,\dagger}\,D^J{}_I - \mathrm{i}\,\gamma^\mu \widetilde{\varepsilon}_I^{\,\dagger}\,(\mathsf{D}_\mu \sigma) \ , \\
\mathcal{Q}D_I{}^J &= \tfrac{\mathrm{i}}{2}\left(\widetilde{\varepsilon}_I^{\,\dagger}\,\gamma^\mu(\mathsf{D}_\mu \lambda^J) - (\mathsf{D}_\mu \lambda_I^\dagger)\,\gamma^\mu \varepsilon^J\right) - \tfrac{\mathrm{i}}{2}\left(\widetilde{\varepsilon}_I^{\,\dagger}\,[\sigma, \lambda^J] - [\sigma, \lambda_I^\dagger]\varepsilon^J\right) + (\ I \leftrightarrow J\ ) \ ,
\end{aligned}$$

where F is the curvature of the gauge connection A, and D_μ is the covariant derivative which involves the gauge connection A and also the spin connection when acting on the dynamical spinor fields λ and $\lambda^\dagger$. When the theory is placed on a curved background M_3, one has to add

curvature terms proportional to $\frac{1}{r}$ to the supersymmetry variations $\mathcal{Q}\lambda$, $\mathcal{Q}\lambda^\dagger$ and $\mathcal{Q}D$. These terms will also involve the spinor covariant derivative acting on the Killing spinors from the supergravity background. This procedure is standard and we do not review it here.

Following [137, App. A], we use the Killing vector field v for the topological twist. We set $\widetilde{\varepsilon}=\varepsilon$, and rewrite the spinor fields λ and $\lambda^\dagger$ in the vector multiplet in terms of an odd $\mathfrak{g}$-valued one-form Ψ and an odd $\mathfrak{g}$-valued zero-form χ defined as

$$\Psi_\mu = \tfrac{1}{2}\left(\varepsilon^\dagger\,\gamma_\mu\lambda - \lambda^\dagger\,\gamma_\mu\varepsilon\right) \qquad \text{and} \qquad \chi = \varepsilon^\dagger\,\lambda - \lambda^\dagger\,\varepsilon\ ,$$

which depend on the solution ε of the Killing spinor equation, and hence on the choice of contact structure, but not explicitly on the metric. The field content of the vector multiplet is now written in a cohomological form as

$$A\in\Omega^1(M_3,\mathfrak{g})\ ,\quad \sigma\in\Omega^0(M_3,\mathfrak{g})\ ,\quad \Psi\in\Omega^1(M_3,\mathfrak{g}) \qquad \text{and} \qquad \chi\in\Omega^0(M_3,\mathfrak{g})\ ,$$

with (A,χ) treated as coordinates and (σ,Ψ) as conjugate momenta on field space. We do not include details about the gauge fixing here, and again refer to [137] for the technical details. It suffices to say that the bosonic ghost coordinates are a pair of harmonic zero-forms and the fermionic ghost coordinates are a pair of zero-forms. We use the localizing term

$$\mathcal{Q}V \qquad \text{with} \quad V=\int_{M_3}\left((\mathcal{Q}\lambda)^\dagger\,\lambda+\lambda^\dagger\,(\mathcal{Q}\lambda^\dagger)^\dagger\right)\mathrm{d}\Omega_{M_3}$$

in the path integral which brings the quantum field theory to the fixed point locus

$$F=0 \qquad \text{and} \qquad \sigma=-D=\text{constant}\ .$$

The kinetic operator

Once the fields are in cohomological form, the supersymmetry transformation squares to

$$\mathcal{Q}^2=\mathrm{i}\,\mathsf{L}_\phi \qquad \text{with} \quad \mathsf{L}_\phi=\mathcal{L}_v+\mathcal{G}_\phi\ . \tag{11.3.1}$$

Here L_ϕ is the sum of a Lie derivative along v and a gauge transformation $\mathcal{G}_\phi$ with parameter

$$\phi=\mathrm{i}\,\sigma-v\llcorner A\ .$$

At the end, we shall rotate $\sigma\mapsto\mathrm{i}\,\sigma_0$ and integrate over real $\sigma_0\in\mathfrak{g}$. The localization locus consists of flat connections, and therefore

$$\phi=-\left(\sigma_0+v^\mu\,A^{(0)}_\mu\right)\ \in\ \mathfrak{g}\ ,$$

where $A^{(0)}$ is the point of the moduli space of flat G-connections on M_3 around which we are expanding. If M_3 is simply connected, the only point of the moduli space is the trivial connection. Otherwise, expanding around $A^{(0)}=0$ gives the perturbative part of the partition function. In general, the full answer is given by integrating the partition function over the moduli space of flat G-connections $\mathscr{M}^0_G(M_3)$ supported on M_3, which is given by

$$\mathscr{M}^0_G(M_3)=\mathrm{Hom}\big(\pi_1(M_3),G\big)/G\ , \tag{11.3.2}$$

where the quotient is taken by the conjugation action of G on the holonomy of a connection over representatives of elements in $\pi_1(M_3)$. When M_3 is a circle bundle of degree p over a compact oriented Riemann surface C_g of genus g, there is an explicit presentation of the fundamental group $\pi_1(M_3)$ with generators a_i,b_i,ζ, $i=1,\dots,g$ and the relation

$$\prod_{i=1}^g[a_i,b_i]=\zeta^p\ ,$$

with all other pairwise combinations of generators commuting. An explicit parametrization of the moduli space (11.3.2) in the case $G=U(N)$ can be found in [425, Sec. 6.2].

Hypermultiplets

The field content of an $\mathcal{N}=2$ hypermultiplet in three dimensions consists of the complex scalars $\mathsf{q}=(\mathsf{q}_I)$ with $SU(2)_R$ indices and a complex spinor ψ. These fields are obtained by combining chiral and anti-chiral complex scalars and Weyl spinors. One also needs an auxiliary complex scalar. The supersymmetry transformations are

$$\begin{aligned}\mathcal{Q}\mathsf{q}_I &= -\mathrm{i}\,\widetilde{\varepsilon}_I^{\,\dagger}\,\psi\ ,\\ \mathcal{Q}\psi &= \tfrac{1}{2}\,\gamma^\mu\varepsilon^I\,(\mathsf{D}_\mu\mathsf{q}_I)+\tfrac{\mathrm{i}}{2}\,\sigma\,\mathsf{q}_I\,\varepsilon^I\ ,\end{aligned}$$

plus curvature corrections to be added when the theory is put on M_3. The transformations of the conjugate fields $\mathsf{q}^\dagger,\psi^\dagger$ are the obvious ones, with exchange $\varepsilon\leftrightarrow\widetilde{\varepsilon}$. The topological twist of a hypermultiplet was first performed in [415, App. B].

The Killing spinors are used to introduce a new spinor field

$$\mathsf{q}'=\mathsf{q}_I\,\varepsilon^I\ ,$$

so that the physical fields are all reformulated in terms of spinors. These fields are singlets under the action of $SU(2)_R$; this is instrumental to have the kinetic operator in the desired form. One finds

$$\mathcal{Q}^2=\mathrm{i}\,\mathsf{L}_\phi\qquad\text{with}\quad\mathsf{L}_\phi=\mathcal{L}_v^{\text{spin}}+\mathcal{G}_\phi\ ,$$

as in the vector multiplet. We used the notation $\mathcal{L}_v^{\text{spin}}$ to stress that the Lie derivative is twisted by the spin covariant derivative when acting on spinors on curved manifolds.

To mimic the procedure of [137], a further step is needed: we rearrange the fields again in a cohomological form, combining spinors into differential forms. For this, we need to define a spin^c structure on M_3, and use it to decompose the spinors $\psi_\pm$ and q' into elements of

$$\Omega^0_H(M_3,\mathfrak{g})\oplus\Omega^{(0,1)}_H(M_3,\mathfrak{g})$$

where we decomposed ψ according to the chirality operator as

$$\psi=\psi_++\psi_-\qquad\text{with}\quad\gamma_5\psi_\pm=\pm\,\psi_\pm\qquad\text{and}\qquad\gamma_5=v^\mu\,\gamma_\mu\ .$$

11.3.2 One-loop determinant of the vector multiplet in a regular background

Gaussian integration of the vector multiplet around the fixed point ϕ gives the ratio of fluctuation determinants

$$h(\phi)=\sqrt{\frac{\det\mathrm{i}\,\mathsf{L}_\phi|_{\mathrm{f}}}{\det\mathrm{i}\,\mathsf{L}_\phi|_{\mathrm{b}}}}=\sqrt{\frac{(\det\mathrm{i}\,\mathsf{L}_\phi|_{\Omega^0(M_3,\mathfrak{g})})^3}{\det\mathrm{i}\,\mathsf{L}_\phi|_{\Omega^1(M_3,\mathfrak{g})}\,(\det\mathrm{i}\,\mathsf{L}_\phi|_{H^0(M_3,\mathfrak{g})})^2}}\ ,$$

where the subscripts on the left-hand side refer to the operator acting on fermionic or bosonic fields. Here $H^0(M_3,\mathfrak{g})$ is the space of $\mathfrak{g}$-valued harmonic zero-forms, and L_ϕ is given in (11.3.1). The numerator of $h(\phi)$ includes the contributions from the fermionic field χ and the two fermionic ghost fields, while the denominator includes the contributions from the bosonic field A and the two bosonic ghost fields.

We use the Seifert structure of M_3 to decompose one-forms into horizontal and vertical parts as

$$\Omega^1(M_3,\mathfrak{g})=\Omega^1_V(M_3,\mathfrak{g})\oplus\Omega^1_H(M_3,\mathfrak{g})\cong\Omega^0(M_3,\mathfrak{g})\oplus\Omega^1_H(M_3,\mathfrak{g})\ .$$

We may also identify

$$\Omega^2_H(M_3,\mathfrak{g})\cong\Omega^0(M_3,\mathfrak{g})$$

in three dimensions. The circle bundle structure

$$U(1) \hookrightarrow M_3 \to K_2$$

allows us to further decompose the spaces of zero-forms and horizontal one-forms as

$$\begin{aligned}\Omega^0(M_3,\mathfrak{g}) &= \Omega^0(K_2,\mathfrak{g}) \oplus \bigoplus_{m\neq 0} \Omega^0\big(K_2,\mathscr{L}^{\otimes m}\otimes\mathfrak{g}\big)\ ,\\ \Omega^1_H(M_3,\mathfrak{g}) &= \Omega^1(K_2,\mathfrak{g}) \oplus \bigoplus_{m\neq 0} \Omega^1\big(K_2,\mathscr{L}^{\otimes m}\otimes\mathfrak{g}\big)\ ,\end{aligned} \tag{11.3.3}$$

where we recall that $\mathscr{L}$ is the line bundle associated to the $U(1)$ fibration of M_3. Using the short-hand notation

$$\mathcal{D}^\bullet_m(\phi) = \det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^\bullet(K_2,\mathscr{L}^{\otimes m}\otimes\mathfrak{g})}\ ,$$

we obtain

$$h(\phi) = \frac{1}{\big|\det \mathrm{i}\,\mathsf{L}_\phi|_{H^0(M_3,\mathfrak{g})}\big|}\sqrt{\frac{\mathcal{D}^0_0(\phi)\,\mathcal{D}^2_0(\phi)}{\mathcal{D}^1_0(\phi)}}\prod_{m\neq 0}\sqrt{\frac{\mathcal{D}^0_m(\phi)\,\mathcal{D}^2_m(\phi)}{\mathcal{D}^1_m(\phi)}}\ .$$

The crucial observation at this point is that when the Killing vector field v points along the fibre direction, the decomposition (11.3.3) corresponds to a decomposition in eigenmodes of the Lie derivative operator $\mathcal{L}_v$. The degeneracy of the action of the gauge transformations $\mathcal{G}_\phi$ is resolved in the standard way [58, 137], by decomposing the Lie algebra $\mathfrak{g}$ into its root system as

$$\mathfrak{g} = \bigoplus_{\alpha\in\triangle}\mathfrak{g}_\alpha\ .$$

We finally obtain

$$Z_{\rm vec}(M_3) = \prod_{\alpha\in\triangle}\big(\mathrm{i}\,(\alpha,\phi)\big)^{\frac{1}{2}\chi(K_2)-\dim H^0(M_3,\mathbb{R})}\prod_{m\neq 0}\Big(\frac{m}{r}+\mathrm{i}\,(\alpha,\phi)\Big)^{\mathrm{index}\,\bar\partial^{(m)}} \tag{11.3.4}$$

where we used the fact that the number of remaining modes, after cancellation, is given by the index of the twisted Dolbeault complex

$$\Omega^0(K_2,\mathscr{L}^{\otimes m})\xrightarrow{\bar\partial^{(m)}}\Omega^1(K_2,\mathscr{L}^{\otimes m})\xrightarrow{\bar\partial^{(m)}}\Omega^2(K_2,\mathscr{L}^{\otimes m})\ ,$$

after identification of the complexified de Rham differential with the anti-holomorphic Dolbeault differential. This result agrees with [426]. The first multiplicative term in (11.3.4) is trivial for Seifert homology spheres. The proof of this cohomological localization formula used the properties that M_3 is a Seifert manifold and that v is parallel to the Reeb vector field ξ, and hence that M_3 is a K-contact manifold.

11.3.3 One-loop determinant of the vector multiplet in an irregular background

We now consider the case in which the orbits of the Killing vector field v are not closed. The background geometry is required to have $U(1)\times U(1)$ isometry group in order to preserve supersymmetry [417, 418]. In this instance M_3 is still a Seifert manifold but now v does not point along the fibre. In the spirit of Section 11.3.2, we calculate the ratio of fluctuation determinants through the index theorem. A similar calculation was performed in [395], but there the supersymmetry transformation squares to the sum of a Lie derivative along v, a gauge transformation

and a third transformation which is a sum of R-symmetry and flavour symmetry transformations (determining a new R-symmetry); this is not of the form described in [137, 230], and so our index theory calculations cannot be applied directly in this framework.

We can express v as a linear combination

$$v = a_1\,\xi + a_2\,\widetilde{\xi}$$

of the Reeb vector field ξ and a vector field $\widetilde{\xi}$ which generates a residual $U(1)$ isometry. They are mutually orthogonal and are linear combinations of the generators of the torus isometry. Most of the calculation follows that of the round case from Section 11.3.2, particularly the decomposition (11.3.3) of $\Omega^\bullet(M_3,\mathfrak{g})$ according to the Seifert fibration. However, we now have to face the problem that the eigenmodes of $\mathcal{L}_\xi$ are no longer eigenmodes of $\mathcal{L}_v$.

In [395] it was shown how one can exploit the fact that if the $U(1)$ action generated by the Reeb vector field ξ is free, then the problem can be reduced to the quotient space $M_3/U(1)\cong K_2$. Let us elaborate a bit more on this point. The crucial observation is that the restriction of the operator $\mathrm{i}\mathsf{L}_\phi$ is no longer elliptic, but it is transversally elliptic with respect to the isometry generated by $\widetilde{\xi}$. Following [427], given a first order transversally elliptic differential operator and a subgroup which acts freely, the index can be computed on the quotient space and the Atiyah–Bott localization formula localizes the contributions to the fixed points of the action of the subgroup generated by $\widetilde{\xi}$; this works even when K_2 possesses orbifold points. We then decompose into eigenmodes corresponding to $\widetilde{\xi}$, and the remaining modes after cancellation come from the fixed points of the $U(1)$ action generated by $\widetilde{\xi}$. Therefore for irregular Seifert manifolds we obtain

$$Z_{\rm vec}(M_3) = \prod_{\alpha\in\triangle}\big(\mathrm{i}\,(\alpha,\phi)\big)^{\frac{1}{2}\,\chi(K_2)-\dim H^0(M_3,\mathbb{R})}\prod_f\prod_{m\neq 0}\Big(\frac{m}{\epsilon_f}+\mathrm{i}\,(\alpha,\phi)\Big)^{\frac{1}{2}\,\mathrm{index}\,\bar\partial^{(m)}} \tag{11.3.5}$$

where the second product runs over the fixed points of the $U(1)$ action generated by $\widetilde{\xi}$ on K_2, and ϵ_f is the radius of the circle fibre over the fixed point labelled by f.[28] This formula gives the correct answer for ellipsoids [281, 418], and generalizes the result of [395] to any Seifert manifold which is not K-contact. Notice that an ellipsoid Seifert manifold is necessarily fibered topologically over a sphere $\mathbb{S}^2$ with at most two punctures [421, Sec. 3.5].

11.3.4 One-loop determinant of a hypermultiplet

For the Gaussian integration of a hypermultiplet around a fixed point ϕ, we do not give all the details here since the computation is essentially the same as for the vector multiplet. One finds

$$\sqrt{\frac{\det\mathrm{i}\,\mathsf{L}_\phi|_{\rm f}}{\det\mathrm{i}\,\mathsf{L}_\phi|_{\rm b}}}=\sqrt{\frac{\det\mathrm{i}\,\mathsf{L}_\phi|_{\Omega_H^{(0,1)}(M_3,\mathfrak{g})}}{(\det\mathrm{i}\,\mathsf{L}_\phi|_{\Omega_H^0(M_3,\mathfrak{g})})^2}}\ , \tag{11.3.6}$$

where we used the topological twist described in Section 11.3.1. Recall that this formula only holds if we are allowed to recombine the $SU(2)_R$ singlet spinors $\mathfrak{q}'$ into anti-holomorphic differential forms. This fails for the ellipsoid $\mathbb{S}^3_b$ of [281], for instance, because a spin^c structure is not guaranteed to exist when the dual one-form to the Killing vector v is not a gauge connection for the $U(1)$ fibration.

For round Seifert manifolds, the eigenvalues of $\mathcal{L}_v^{\rm spin}$ are

$$-\frac{\mathrm{i}\,m}{r}-\frac{\mathrm{i}\,\Delta}{r}\ ,$$

[28] The square roots of each fixed point contribution come from the square root of the original ratio of fluctuation determinants. In the regular case, the contributions are equal.

where $m \in \mathbb{Z}$ and Δ is the R-charge the hypermultiplet. The number of remaining modes after cancellations in (11.3.6) is given by the index of the twisted Dolbeault differential, and we obtain

$$Z_{\rm hyp}(M_3) = \prod_{\rho\in\Lambda_R} \prod_{m\in\mathbb{Z}} \Big(\frac{m+\Delta}{r} + {\rm i}\,(\rho,\phi) \Big)^{-{\rm index}\,\bar\partial^{(m)}}$$

where a shift $m \to m+\Delta$ should be included when considering twisted boundary conditions along the fibre, but the contribution of this shift cancels in the computations. This result agrees with [426].

11.3.5 Applications of the cohomological localization formulas

We shall now provide some simple examples illustrating how the localization formulas obtained in this section work to give the correct known results in three dimensions.

Localization on $\mathbb{S}^3$

As a first check, let us examine how to reproduce the three-dimensional localization calculations of [58, Sec. 3] in this framework. We consider the three-sphere $M_3 = \mathbb{S}^3$ of radius r, viewed as a circle bundle of degree one over the projective line $K_2 = \mathbb{P}^1$ via the Hopf fibration, with Euler characteristic $\chi(\mathbb{P}^1) = 2$ and $H^0(\mathbb{S}^3,\mathbb{R}) = \mathbb{R}$. The total Chern class of the holomorphic tangent bundle of $\mathbb{P}^1$ is

$$c(T^{1,0}\mathbb{P}^1) = (1+\omega)^{\wedge 2} = 1 + 2\,\omega = 1 + c_1(T^{1,0}\mathbb{P}^1)\ ,$$

and so the corresponding Todd class is given by

$${\rm Td}(T^{1,0}\mathbb{P}^1) = 1 + \tfrac{1}{2}\,c_1(T^{1,0}\mathbb{P}^1) = 1+\omega\ ,$$

while the Chern characters of the line bundles $\mathscr{L}^{\otimes m} \to \mathbb{P}^1$ are given by

$${\rm ch}(\mathscr{L}^{\otimes m}) = 1 + c_1(\mathscr{L}^{\otimes m}) = 1 + m\,c_1(\mathscr{L}) = 1 + m\,\omega\ .$$

The index of the corresponding twisted Dolbeault complex is thus given by

$${\rm index}\,\bar\partial^{(m)} = \int_{\mathbb{P}^1} (1+m\,\omega)\wedge(1+\omega) = \int_{\mathbb{P}^1} (m+1)\,\omega = m+1\ .$$

The one-loop vector multiplet contribution is then computed to be

$$\begin{aligned} Z_{\rm vec}(\mathbb{S}^3) &= \prod_{\alpha\in\triangle} \prod_{m\neq 0} \Big(\frac{m}{r} + {\rm i}(\alpha,\phi) \Big)^{m+1} \\ &= \prod_{\alpha\in\triangle} \prod_{m=1}^{\infty} \frac{\big(\frac{m}{r} - {\rm i}(\alpha,\sigma_0)\big)^{m+1}}{\big(-\frac{m}{r} - {\rm i}(\alpha,\sigma_0)\big)^{m-1}} \\ &= \prod_{\alpha\in\triangle_+} \prod_{m=1}^{\infty} \frac{\big(\frac{m^2}{r^2} + (\alpha,\sigma_0)^2\big)^{m+1}}{\big(\frac{m^2}{r^2} + (\alpha,\sigma_0)^2\big)^{m-1}} \\ &= \prod_{\alpha\in\triangle_+} \prod_{m=1}^{\infty} \frac{m^4}{r^4} \Big(1 + \frac{r^2\,(\alpha,\sigma_0)^2}{m^2}\Big)^2 \\ &= \prod_{\alpha\in\triangle_+} \Big(\frac{2\sinh \pi\, r\,(\alpha,\sigma_0)}{\pi\, r\,(\alpha,\sigma_0)} \Big)^2\ , \end{aligned}$$

where as previously $\triangle_+$ is the system of positive roots of the Lie algebra $\mathfrak{g}$, and we used the fact that the roots come in positive-negative pairs. In the second line we used the fact that the only flat connection on $\mathbb{S}^3$ is trivial and substituted $\phi = -\sigma_0$, and in the last line we evaluated the infinite product using zeta-function regularization:

$$\prod_{m=1}^{\infty}\left(1+\frac{x^2}{m^2}\right) = \frac{\sinh(\pi\,x)}{\pi\,x} \qquad \text{and} \qquad \prod_{m=1}^{\infty}\frac{m^2}{r^2} = 2\pi\,r\ .$$

The same calculation for a one-loop hypermultiplet determinant gives

$$\begin{aligned} Z_{\rm hyp}(\mathbb{S}^3) &= \prod_{\rho\in\Lambda_R}\prod_{m\in\mathbb{Z}}\left(\frac{m+\Delta}{r}+{\rm i}(\rho,\phi)\right)^{-1-m} \\ &= \prod_{\rho\in\Lambda_R}\frac{\prod_{m=0}^{\infty}\left(\frac{-m-1+\Delta}{r}-{\rm i}(\rho,\sigma_0)\right)^m}{\prod_{m=1}^{\infty}\left(\frac{m-1+\Delta}{r}+{\rm i}(\rho,\sigma_0)\right)^m} \\ &= \prod_{\rho\in\Lambda_R}\prod_{m=1}^{\infty}\left(\frac{\frac{m+1-\Delta}{r}+{\rm i}(\rho,\sigma_0)}{\frac{m-1+\Delta}{r}-{\rm i}(\rho,\sigma_0)}\right)^m \\ &= \prod_{\rho\in\Lambda_R}\mathtt{s}_1\big({\rm i}(1-\Delta)-r\,(\rho,\sigma_0)\big)\ , \end{aligned}$$

where in the last line we inserted the definition of the double-sine function which is the meromorphic function defined by the zeta-function regularized infinite products [428]

$$\mathtt{s}_b(x) = \prod_{m,n=0}^{\infty}\frac{m\,b+n\,b^{-1}+\frac{1}{2}\,(b+b^{-1})-{\rm i}x}{m\,b+n\,b^{-1}+\frac{1}{2}\,(b+b^{-1})+{\rm i}x} \tag{11.3.7}$$

evaluated at $b=1$. These results all agree with the computations of [58, Sec. 3.2] (see also [395, Sec .3.2]).

Localization on ellipsoid $\mathbb{S}^3_b$ and $L(p,1)_b$

We now consider the ellipsoid $\mathbb{S}^3_b$ of [281], with squashing parameter $b>0$ and metric

$${\rm d}s^2_{\mathbb{S}^3_b} = r^2\left(f(\vartheta)^2\ {\rm d}\vartheta\otimes{\rm d}\vartheta + b^2\,\cos^2\vartheta\ {\rm d}\varphi_1\otimes{\rm d}\varphi_1 + b^{-2}\,\sin^2\vartheta\ {\rm d}\varphi_2\otimes{\rm d}\varphi_2\right)$$

induced from the standard metric on $\mathbb{C}^2$ restricted to the locus

$$b^2\,|z_1|^2+b^{-2}\,|z_2|^2 = r^2\ ,$$

where $f(\vartheta)=\sqrt{b^2\,\cos^2\vartheta+b^{-2}\,\sin^2\vartheta}$ and $(\vartheta,\varphi_1,\varphi_2)$ are Hopf coordinates on the usual round sphere $\mathbb{S}^3=\mathbb{S}^3_{b=1}$ of radius r. This defines an irregular fibration with isometry group $U(1)\times U(1)$: The Killing vector field v takes the form

$$v = \frac{b+b^{-1}}{2\,r}\,\xi+\frac{b-b^{-1}}{2\,r}\,\widetilde{\xi}\ ,$$

where $\xi=\frac{\partial}{\partial\theta}$ is the Reeb vector field of the Seifert fibration $\mathbb{S}^3_b\to\mathbb{P}^1$, with $\theta=\frac{1}{2}\,(\varphi_1+\varphi_2)$, and $\widetilde{\xi}=\frac{\partial}{\partial\widetilde{\theta}}$ is the generator of the residual $U(1)$ isometry, with $\widetilde{\theta}=\frac{1}{2}\,(\varphi_1-\varphi_2)$. The fixed points of $\widetilde{\xi}$

correspond to the north and south poles of the base $\mathbb{S}^2 \cong \mathbb{P}^1$, with respective coordinates $\vartheta = 0$, at which the fibre has radius $\epsilon_1 = r\,b$, and $\vartheta = \frac{\pi}{2}$, at which the fibre has radius $\epsilon_2 = r\,b^{-1}$. The index of the twisted Dolbeault complex is a topological invariant, and hence is the same as for round $\mathbb{S}^3$.

Altogether, the localization formula gives

$$\begin{aligned} Z_{\rm vec}(\mathbb{S}^3_b) &= \prod_{\alpha\in\triangle}\prod_{m\neq 0}\left(\left(\frac{m\,b^{-1}}{r}-{\rm i}(\alpha,\sigma_0)\right)^{\frac12}\left(\frac{m\,b}{r}-{\rm i}(\alpha,\sigma_0)\right)^{\frac12}\right)^{m+1} \\ &= \prod_{\alpha\in\triangle}\prod_{m=1}^{\infty}\frac{\left(\frac{m\,b^{-1}}{r}-{\rm i}(\alpha,\sigma_0)\right)^{\frac{m+1}{2}}\left(\frac{m\,b}{r}-{\rm i}(\alpha,\sigma_0)\right)^{\frac{m+1}{2}}}{\left(\frac{m\,b^{-1}}{r}+{\rm i}(\alpha,\sigma_0)\right)^{\frac{m-1}{2}}\left(\frac{m\,b}{r}+{\rm i}(\alpha,\sigma_0)\right)^{\frac{m-1}{2}}} \\ &= \prod_{\alpha\in\triangle_+}\prod_{m=1}^{\infty}\left(\frac{m^2\,b^{-2}}{r^2}+(\alpha,\sigma_0)^2\right)\left(\frac{m^2\,b^2}{r^2}+(\alpha,\sigma_0)^2\right) \\ &= \prod_{\alpha\in\triangle_+}\frac{\sinh\left(\pi\,b\,r\,(\alpha,\sigma_0)\right)\,\sinh\left(\pi\,b^{-1}\,r\,(\alpha,\sigma_0)\right)}{\pi^2\,r^2\,(\alpha,\sigma_0)^2}\ , \end{aligned} \tag{11.3.8}$$

which agrees with [281, Eq. (5.33)] and [395, Eq. (4.24)]. This result is independent of the particular form of the smooth squashing function $f(\vartheta)$ and depends only on its values at the fixed points $\vartheta = 0, \frac{\pi}{2}$; it can therefore be extended to a larger class of backgrounds with the same topology [429].

This result straightforwardly extends to the ellipsoid lens spaces $L(p,1)_b$, with the induced metric on the quotient $\mathbb{S}^3_b/\mathbb{Z}_p$ and associated line bundle $\mathscr{L} \to \mathbb{P}^1$ of degree p, so that now $c_1(\mathscr{L}) = p\,\omega$; see the review [59] for a description of localization on $L(p,1)_b$. The modifications are the same as for the round case, and amount to a shift $\sigma_0 \mapsto \sigma_0 + v^\mu\,A^{(0)}_\mu$, where $A^{(0)}$ is an isolated point of the moduli space of flat G-connections on $L(p,1)$. Such flat connections are classified by conjugacy classes of embeddings of the fundamental group $\pi_1(L(p,1)_b) = \mathbb{Z}_p$ in the gauge group G, and are in one-to-one correspondence with arrays $\vec{m} \in (\mathbb{Z}_p)^{{\rm rank}(G)}$ modulo Weyl symmetry. For the vector multiplet, the localization formula then gives

$$Z_{\rm vec}\big(L(p,1)_b\big) = \prod_{\alpha\in\triangle_+}\frac{\sinh\left(\frac{\pi\,b\,r}{p}\,(\alpha,\sigma_0+{\rm i}\vec{m})\right)\,\sinh\left(\frac{\pi\,b^{-1}\,r}{p}\,(\alpha,\sigma_0+{\rm i}\vec{m})\right)}{\frac{\pi^2\,r^2}{p^2}\,(\alpha,\sigma_0+{\rm i}\vec{m})^2}\ .$$

With similar modifications, one can extend these calculations to any ellipsoid Seifert manifold.

There is no spin^c structure on $\mathbb{S}^3_b$ that can be used to apply our index theory formalism to the hypermultiplet contributions. Nevertheless, the one-loop fluctuation determinant can still be calculated in this case. For example, one could split the twisted Lie derivative as

$$\mathcal{L}^{\rm spin}_v = \frac{2\,b}{r}\,\mathcal{L}^{\rm spin}_{\frac{\partial}{\partial\varphi_1}} + \frac{2\,b^{-1}}{r}\,\mathcal{L}^{\rm spin}_{\frac{\partial}{\partial\varphi_2}}$$

and decompose the fields into eigenmodes of the Lie derivatives in the two orthogonal toroidal directions $\frac{\partial}{\partial\varphi_1}$ and $\frac{\partial}{\partial\varphi_2}$, whose corresponding eigenvalues are then of the form $\frac{m+\Delta}{\epsilon_1} + \frac{n+\Delta}{\epsilon_2}$ where $m,n \in \mathbb{Z}$. Then the one-loop contribution of a hypermultiplet in a representation R of the gauge group G is given by [395, Sec. 4]

$$\begin{aligned} Z_{\rm hyp}(\mathbb{S}^3_b) &= \prod_{\rho\in\Lambda_R}\prod_{m,n=0}^{\infty}\frac{m\,b+n\,b^{-1}+\left(b+b^{-1}\right)\left(1-\frac{\Delta}{2}\right)+{\rm i}r\,(\rho,\sigma_0)}{m\,b+n\,b^{-1}+\left(b+b^{-1}\right)\frac{\Delta}{2}-{\rm i}r\,(\rho,\sigma_0)} \\ &= \prod_{\rho\in\Lambda_R}\mathtt{s}_b\big(\tfrac{\rm i}{2}\,(b+b^{-1})\,(1-\Delta)-r\,(\rho,\sigma_0)\big)\,. \end{aligned} \tag{11.3.9}$$

Note that here the squashing parameter b serves as a zeta-function regulator in the infinite product formula for the double-sine function (11.3.7) in the hypermultiplet contribution. For the $\mathcal{N}=1$ adjoint hypermultiplet with $\Delta=1$, the product over $\Lambda_R=\triangle$ can be split into contributions from positive-negative pairs of roots and the one-loop contribution is trivial: $Z_{\rm hyp}(\mathbb{S}^3_b)=1$.

11.4 $\mathcal{N}=1$ cohomological gauge theories in five dimensions

In this section we derive one-loop fluctuation determinants in various five-dimensional geometries using the Atiyah–Singer index theorem. We first present the supersymmetric gauge theory and its topologically twisted version, and then derive expressions for the one-loop determinants in the regular and irregular cases separately. We subsequently apply the general formalism to some explicit examples, mainly focusing on five-manifolds of the form $M_5=M_3\times\Sigma_h$, with M_3 one of the three-dimensional geometries studied in Section 11.3 and Σ_h a closed Riemann surface of genus h. In this section we adopt the conventions and normalization of [231]. This differs from the rest of the literature on the topic, and in particular the expressions here will only involve anti-holomorphic Dolbeault differentials.[29]

Topologically twisted gauge theories on the five-sphere were studied in [230, 231], which ignited the stream of activity in this area. The next examples considered were the Sasaki-Einstein manifolds $Y^{p,s}$ in [398, 399, 430]. Sasaki-Einstein manifolds are backgrounds which admit $\mathcal{N}=2$ supersymmetry; see Appendix 11.C for a brief review. In [419], further five-dimensional geometries preserving $\mathcal{N}=1$ and $\mathcal{N}=2$ supersymmetry were obtained, following the idea of [416, 417] and adapting it to five dimensions. In particular, one can put an $\mathcal{N}=2$ supersymmetric gauge theory on $\mathbb{S}^3\times\Sigma_h$ and $\mathbb{T}^3\times\Sigma_h$, where $\mathbb{T}^3$ is a three-dimensional torus. On the other hand, the manifolds $L(p,1)\times\Sigma_h$ only admit $\mathcal{N}=1$ supersymmetry. Further geometries admitting Killing spinor solutions, and hence admitting supersymmetric field theories, were obtained in [420] starting from a holographic setting and taking the rigid limit of supergravity. Sasaki-Einstein manifolds, products $M_3\times\Sigma_h$ with M_3 a Seifert three-manifold, and more general $U(1)$ fibrations over products $C_g\times\Sigma_h$ with C_g and Σ_h Riemann surfaces, possibly with orbifold points, are all examples of manifolds studied in [420].

11.4.1 Supersymmetric Yang–Mills theory and its cohomological formulation

Consider five-dimensional $\mathcal{N}=1$ supersymmetric Yang–Mills theory. We define the theory on flat Euclidean spacetime $\mathbb{R}^5$ and then, by coupling it to background supergravity fields, the theory is put on curved manifolds M_5. For manifolds admitting two Killing spinors, the $\mathcal{N}=2$ vector multiplet is described in the $\mathcal{N}=1$ superspace formalism by an $\mathcal{N}=1$ vector multiplet and an $\mathcal{N}=1$ adjoint hypermultiplet. The required modifications to the supersymmetry transformations are described in detail in [422]. Let ε be the five-dimensional Killing spinor on M_5 (see Appendix 11.A for our notation), and define the vector field v through[30]

$$v^\mu=\varepsilon^\dagger\,\Gamma^\mu\varepsilon\ .$$

[29] The normalization in [231] uses the opposite sign for the contact structure, compared to other literature. After the topological twist, some fields will come with additional minus signs, and in particular a two-form in five dimensions, which is usually taken to be self-dual, becomes anti-self-dual here; with our convention, self-dual 2-forms descend to anti-instantons in four dimensions, and vice versa. In practice, the cohomological complex in five dimensions that we will work with only involves the anti-holomorphic Dolbeault differential, while in previous works (see [424] for a review) the contributions from both holomorphic and anti-holomorphic forms are included. The final results will of course be the same in either convention, but the intermediate steps will slightly differ. The only motivation for our choice is to achieve a unified treatment in three and five dimensions. Furthermore, the topological twist of the hypermultiplets involves only anti-holomorphic forms, so this choice also puts the vector multiplet and the hypermultiplets on the same footing.

[30] This differs by a sign from other definitions in the literature, see Footnote 29.

It is a nowhere vanishing Killing vector on M_5.

Vector multiplet

The five-dimensional $\mathcal{N}=1$ vector multiplet consists of a gauge connection A, a scalar σ, a symplectic Majorana spinor λ and an auxiliary real scalar D, where $\lambda=(\lambda^I)$ is a $SU(2)_R$ doublet and $D=(D^I{}_J)$ is a $SU(2)_R$ triplet. The supersymmetry transformations in flat space are

$$\begin{aligned}
\mathcal{Q}_\varepsilon A_\mu &= \mathrm{i}\,\varepsilon_I^\dagger\,\Gamma_\mu\lambda^I\ ,\\
\mathcal{Q}_\varepsilon\sigma &= \varepsilon_I^\dagger\,\lambda^I\ ,\\
\mathcal{Q}_\varepsilon\lambda^I &= -\tfrac{1}{2}\,\Gamma^{\mu\nu}\varepsilon^I\,F_{\mu\nu} - D^I{}_J\,\varepsilon^J + \mathrm{i}\,\Gamma^\mu\varepsilon^I\,(\mathsf{D}_\mu\sigma)\ ,\\
\mathcal{Q}_\varepsilon D_I{}^J &= \mathrm{i}\,\varepsilon_I^\dagger\,\Gamma^\mu(\mathsf{D}_\mu\lambda^J) - \mathrm{i}\,\big[\sigma,\varepsilon_I^\dagger\,\lambda^J\big] + (\ I\leftrightarrow J\ )\ ,
\end{aligned}$$

where F is the curvature of the gauge connection A and D_μ is the covariant derivative, which includes the gauge connection A and also the spin connection when acting on dynamical spinors λ. Curvature corrections proportional to $\frac{1}{r}$ must be added to these flat space transformations when the field theory is put on M_5.

At this point we perform the topological twist. We introduce the one-form Ψ and the horizontal anti-self-dual two-form χ according to[31]

$$\Psi_\mu=\varepsilon_I^\dagger\,\Gamma_\mu\lambda^I \qquad\text{and}\qquad \chi_{\mu\nu}=\varepsilon_I^\dagger\,\Gamma_{\mu\nu}\lambda^I-\eta_\mu\,\varepsilon_I^\dagger\,\Gamma_\nu\lambda^I+\eta_\nu\,\varepsilon_I^\dagger\,\Gamma_\mu\lambda^I\ ,$$

where η is the one-form dual to the Killing vector v. We regard M_5 as a $U(1)$ fibration over a compact Kähler manifold K_4, and when v coincides with the Reeb vector field ξ of the Seifert fibration, then η coincides with the K-contact structure κ of M_5. For squashed geometries, however, $\eta\neq\kappa$.

Contact structure and localization locus

The localizing term we add to the action is the standard one:

$$\mathcal{Q}_\varepsilon V \qquad\text{with}\quad V=\int_{M_5}(\mathcal{Q}_\varepsilon\lambda)^\dagger\,\lambda\,\mathrm{d}\Omega_{M_5}\ ,$$

which in the path integral brings the quantum field theory to the fixed point locus

$$v\llcorner *F=F\ ,\quad \mathsf{D}\sigma=0 \qquad\text{and}\qquad D=-\sigma\otimes\left(\begin{smallmatrix}1&0\\0&-1\end{smallmatrix}\right)\ , \tag{11.4.1}$$

where $*$ is the Hodge duality operator constructed from the metric of M_5.

It is important at this point to stress a major distinction in our setting from that of [230] and subsequent work. When we work with a product of a three-dimensional contact manifold and a Riemann surface, $M_5=M_3\times\Sigma_h$, there is a crucial difference: the contact structure κ lives on M_3, and $\kappa\wedge\mathrm{d}\kappa$ is a volume form on M_3, as is clear from (11.2.6), but it need not be a contact structure on $M_3\times\Sigma_h$. This is important for a choice of compatible metric. For the supersymmetry transformations to be those of a cohomological field theory, one requires the Lie derivative $\mathcal{L}_v$ to commute with the Hodge duality operator. Equivalently, we need v to generate an isometry. The Hodge duality operator on $\Omega^\bullet(M_3\times\Sigma_h)$ takes the form $*_{M_3\times\Sigma_h}=(-1)^\bullet\,*_{M_3}\wedge *_{\Sigma_h}$. This is an important simplification in studying the localization locus on product manifolds.

[31] We are using the same Greek letter χ for a two-form here and for a zero-form in Section 11.3. There should not be any confusion.

The kinetic operator

The supersymmetry transformation squares to $\mathcal{Q}_\varepsilon^2 = \mathrm{i}\,\mathsf{L}_\phi$, with

$$\mathsf{L}_\phi = \mathcal{L}_v + \mathcal{G}_\phi \tag{11.4.2}$$

the sum of the Lie derivative along v and a gauge transformation with parameter $\phi = \mathrm{i}\,\sigma - v \llcorner A$. Since at the end the integration contour for σ must be rotated to the imaginary axis, $\sigma \mapsto \mathrm{i}\,\sigma_0$, we are eventually led to

$$\phi = -\big(\sigma_0 + v^\mu\, A^{(0)}_\mu\big) \ \in \ \mathfrak{g}\ ,$$

with $A^{(0)}$ a connection whose curvature is a solution to the first fixed point equation in (11.4.1). Setting $A^{(0)} = 0$ retains the perturbative partition function, while contributions from non-trivial solutions are related to instantons on the horizontal submanifold.

Hypermultiplets

The field content of a five-dimensional $\mathcal{N} = 1$ hypermultiplet consists of a complex scalar $\mathsf{q} = (\mathsf{q}_I)$, which forms an $SU(2)_R$ doublet, and a complex spinor ψ. These fields are obtained by combining chiral and anti-chiral complex scalars and Dirac spinors. The supersymmetry transformations are

$$\begin{aligned}\mathcal{Q}_\varepsilon \mathsf{q}_I &= -2\,\mathrm{i}\,\varepsilon_I^\dagger\,\psi\ ,\\ \mathcal{Q}_\varepsilon \psi &= \Gamma^\mu \varepsilon^I\,(\mathsf{D}_\mu \mathsf{q}_I) - \sigma\,\mathsf{q}_I\,\varepsilon^I\ .\end{aligned}$$

When coupled to background supergravity fields, additional terms proportional to $\frac{1}{r}$ are to be included.

The topological twist in [231] is then achieved in two steps. First, contract all the $SU(2)_R$ indices, and therefore define the $SU(2)_R$ singlet spinor q' from the scalar q_I as

$$\mathsf{q}' = \mathsf{q}_I\,\varepsilon^I\ .$$

The square of the supersymmetry transformation, which equals the kinetic operator in the action, is

$$\mathcal{Q}_\varepsilon^2 = \mathrm{i}\,\mathsf{L}_\phi \qquad \text{with} \quad \mathsf{L}_\phi = \mathcal{L}_v^{\text{spin}} + \mathcal{G}_\phi\ ,$$

where we indicated explicitly that the Lie derivative is twisted by the spin connection on M_5.

The second step consists in defining a spin^c structure on M_5. For this, in [231] (see also [424, Sec. 3]) the following assumption is made. Let η be the dual one-form to the Killing vector field v. Then, according to [420], the most general metric on the Seifert fibration $M_5 \to K_4$ admitting supersymmetry is of the form

$$\mathrm{d}s^2_{M_5} = \eta \otimes \eta + \mathrm{d}s^2_{K_4}\ ,$$

with transverse Hermitian metric on the Kähler surface K_4. If η is proportional to the contact structure defined by the Seifert fibration, then one can define a canonical spin^c structure on M_5. This condition is equivalent to requiring the orbits of v to be all closed. Manifolds supporting $\mathcal{N} = 2$ supersymmetry belong to this class [419], and the index theorem can be applied in that case.

The spin^c structure identifies, through the action of a representation of the Clifford algebra, spinors with elements of

$$\Omega_H^{(0,\bullet)}(M_5, \mathfrak{g})\ ,$$

so the hypermultiplet is put in cohomological form. See [231, 424] for further details.

After the standard localizing term is added to the action, one has to compute the localization locus. If one considers the trivial solution $A^{(0)}=0$ in the vector multiplet, then the localization locus consists in setting all hypermultiplet fields to zero. It was proven in [398] that this holds for any solution $A^{(0)}$ in the localization locus of the vector multiplet, as long as $\mathrm{d}s^2_{M_5}$ is a Sasaki-Einstein metric.

Supersymmetric Yang–Mills action at the localization locus

Evaluating the full gauge theory action at the fixed point locus on $M_5 = M_3 \times \Sigma_h$ gives

$$S_{\mathrm{cl}}(F,\sigma_0) = \frac{1}{2\,g^2_{\mathrm{YM}}} \int_{M_5} \Big(\big(F \mathbin{\overset{\wedge}{,}} *F\big) + \frac{1}{r}\,(\sigma_0,F)\wedge\kappa\wedge\mathrm{d}\kappa + \frac{1}{r^2}\,(\sigma_0,\sigma_0)\,\kappa\wedge\mathrm{d}\kappa\wedge\omega_{\Sigma_h}\Big)\ , \qquad (11.4.3)$$

where g_{YM} is the Yang–Mills coupling constant. Here κ is the contact structure on the Seifert three-manifold M_3, ω_{Σ_h} is the symplectic structure on the Riemann surface Σ_h and $\frac{1}{2}\,\kappa\wedge\kappa\wedge\omega_{\Sigma_h}$ is the volume form (11.2.8) on $M_3\times\Sigma_h$.

11.4.2 One-loop determinant of the vector multiplet in a regular background

After the topological twist performed in Section 11.4.1, all fields of the vector multiplet are in a cohomological form

$$A\in\Omega^1(M_5,\mathfrak{g})\ ,\quad \sigma\in\Omega^0(M_5,\mathfrak{g})\ ,\quad \Psi\in\Omega^1(M_5,\mathfrak{g}) \qquad \text{and} \qquad \chi\in\Omega^2_{H,-}(M_5,\mathfrak{g})\ ,$$

where by $\Omega^2_{H,\pm}(M_5,\mathfrak{g})$ we denote the spaces of self-dual and anti-self-dual horizontal two-forms with values in the Lie algebra $\mathfrak{g}$. Here we assume that M_5 is a regular background, so that contraction by v separates the horizontal and vertical parts of forms. The gauge connection A is our even coordinate and χ is the odd coordinate on the space of fields. We also have to introduce ghosts, and we refer to [230, 231] for the procedure. For our purposes, it suffices to say that these give two even harmonic scalars and two odd scalars.

Gaussian integration of the vector multiplet around the fixed point gives the ratio of fluctuation determinants

$$h(\phi) = \sqrt{\frac{\det \mathrm{i}\,\mathsf{L}_\phi|_{\mathrm{f}}}{\det \mathrm{i}\,\mathsf{L}_\phi|_{\mathrm{b}}}} = \sqrt{\frac{\det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^2_{H,-}(M_5,\mathfrak{g})}\,(\det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^0(M_5,\mathfrak{g})})^2}{\det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^1(M_5,\mathfrak{g})}\,(\det \mathrm{i}\,\mathsf{L}_\phi|_{H^0(M_5,\mathfrak{g})})^2}}\ ,$$

where $|_{\mathrm{f}}$ (respectively $|_{\mathrm{b}}$) refers to the operator acting on fermionic (respectively bosonic) fields. Here $H^0(M_5,\mathfrak{g})$ is the space of $\mathfrak{g}$-valued harmonic zero-forms on M_5, and the differential operator L_ϕ is given in (11.4.2). The numerator of $h(\phi)$ involves the contributions from the fermionic coordinate χ and the two fermionic ghost coordinates, while the denominator involves the contributions from the bosonic coordinate A and the two bosonic ghost coordinates.

We split

$$\begin{aligned}\Omega^2_H(M_5,\mathfrak{g}) &= \Omega^2_{H,+}(M_5,\mathfrak{g})\oplus\Omega^2_{H,-}(M_5,\mathfrak{g})\ ,\\ \Omega^2_{H,+}(M_5,\mathfrak{g}) &= \Omega^{(2,0)}_H(M_5,\mathfrak{g})\oplus\Omega^{(0,2)}_H(M_5,\mathfrak{g})\oplus\Omega^{(1,1)}_{\mathrm{sympl}}(M_5,\mathfrak{g})\ ,\end{aligned}$$

where $\Omega^{(1,1)}_{\mathrm{sympl}}(M_5,\mathfrak{g})$ are the $\mathfrak{g}$-valued horizontal two-forms proportional to the symplectic structure on the base Kähler manifold K_4. Then

$$\Omega^{(1,1)}_H(M_5,\mathfrak{g}) = \Omega^{(1,1)}_{\mathrm{sympl}}(M_5,\mathfrak{g})\oplus\Omega^2_{H,-}(M_5,\mathfrak{g}) \cong \Omega^0(M_5,\mathfrak{g})\oplus\Omega^2_{H,-}(M_5,\mathfrak{g})\ .$$

Here we consider the regular case, in which v is parallel to the Reeb vector field ξ. With our choices, $v = -\xi = -\frac{\partial}{\partial\theta}$, where $\theta \in [0, 2\pi\, r)$ is the coordinate along the circle fibre. We can decompose horizontal forms according to the fibration structure of $M_5 \to K_4$ as

$$\Omega_H^{(\bullet,\bullet)}(M_5,\mathfrak{g}) = \Omega^{(\bullet,\bullet)}(K_4,\mathfrak{g}) \oplus \bigoplus_{m\neq 0} \Omega^{(\bullet,\bullet)}(K_4,\mathscr{L}^{\otimes m}\otimes\mathfrak{g}) \ .$$

The crucial step now is to recognise that, according to this splitting, the Lie derivative along the Killing vector field v acts on a form $\alpha_m \in \Omega^{(\bullet,\bullet)}(K_4,\mathscr{L}^{\otimes m}\otimes\mathfrak{g})$ as

$$\mathcal{L}_v\alpha_m = -\mathcal{L}_\xi\alpha_m = -\frac{\mathrm{i}\,m}{r}\,\alpha_m \ .$$

The fact that the orbits of v coincide with the orbits of the Reeb vector field ξ is essential here. The action of $\mathrm{i}\,\mathsf{L}_\phi$ on each Kaluza-Klein mode labelled by $m \in \mathbb{Z}$ also includes the gauge transformation $\mathcal{G}_\phi$, whose eigenmodes are found decomposing the Lie algebra $\mathfrak{g}$ into its root system

$$\mathfrak{g} = \bigoplus_{\alpha\in\triangle} \mathfrak{g}_\alpha \ .$$

We are therefore ready to evaluate

$$\begin{aligned} h(\phi) &= \frac{1}{\left|\det \mathrm{i}\,\mathsf{L}_\phi|_{H^0(M_5,\mathfrak{g})}\right|}\,\sqrt{\frac{\det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^2_{H,-}(M_5,\mathfrak{g})}\ \det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^0(M_5,\mathfrak{g})}}{\det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^1_H(M_5,\mathfrak{g})}}} \\ &= \frac{1}{\left|\det \mathrm{i}\,\mathsf{L}_\phi|_{H^0(M_5,\mathfrak{g})}\right|}\,\sqrt{\frac{\mathcal{D}_0^{(1,1)}(\phi)}{\mathcal{D}_0^{(1,0)}(\phi)\,\mathcal{D}_0^{(0,1)}(\phi)}}\ \prod_{m\neq 0}\sqrt{\frac{\mathcal{D}_m^{(1,1)}(\phi)}{\mathcal{D}_m^{(1,0)}(\phi)\,\mathcal{D}_m^{(0,1)}(\phi)}} \ , \end{aligned}$$

where in the second line we denoted

$$\mathcal{D}_m^{(\bullet,\bullet)}(\phi) = \det \mathrm{i}\,\mathsf{L}_\phi|_{\Omega^{(\bullet,\bullet)}(K_4,\mathscr{L}^{\otimes m}\otimes\mathfrak{g})} \ .$$

Standard manipulations at this point [231] (see also [230, App. C]) finally lead to the cohomological localization formula

$$Z_{\mathrm{vec}}(M_5) = \prod_{\alpha\in\triangle}\big(\mathrm{i}\,(\alpha,\phi)\big)^{\frac{1}{12}(c_2(K_4)+c_1(K_4)^2)-\dim H^0(M_5,\mathbb{R})}\ \prod_{m\neq 0}\Big(\frac{m}{r}+\mathrm{i}\,(\alpha,\phi)\Big)^{\mathrm{index}\,\bar\partial^{(m)}} \tag{11.4.4}$$

For a $U(1)$ bundle $M_5 \to C_g \times \Sigma_h$ over the product of two Riemann surfaces of genera g and h, the power of the first multiplicative factor is $(1-g)\,(1-h)-1$.

11.4.3 One-loop determinant of the vector multiplet in an irregular background

We now consider the alternative case of an irregular fibration, whereby v does not point along the $U(1)$ fibre of M_5. Let η be the dual one-form to the Killing vector field v. It is an almost contact structure on M_5; if it is a contact structure, then we are in the situation of Section 11.4.2 above. For the present discussion, we assume that $M_5^{g,h} \to C_g \times \Sigma_h$ is a $U(1)$ fibration over a direct product of two Riemann surfaces, both compact and closed. Rotations along the circle fibre are assumed to act freely on $M_5^{g,h}$.[32] This means that, although the gauge theory could be

[32] The Sasaki-Einstein manifolds $Y^{p,s}$ do not belong to this class, see Appendix 11.C and Section 11.2.4.

put on $M_5^{g,h}$ preserving $\mathcal{N}=1$ supersymmetry when $M_5^{g,h}/U(1)$ admits orbifold points [420], the procedure we describe below does not apply to that case. Since we are in the irregular setting, we need an additional $U(1)$ isometry on $C_g \times \Sigma_h$. In practice, this restricts our considerations to $C_0 = \mathbb{S}^2$ or $C_1 = \mathbb{T}^2$.

Most of the procedure is exactly the same as in Section 11.4.2, particularly the decomposition of differential forms in terms of the Reeb vector field ξ. Nonetheless, we have to face two problems. First, as for the irregular three-dimensional case, we have to bear in mind that forms $\alpha_m \in \Omega^{(\bullet,\bullet)}(C_g \times \Sigma_h, \mathscr{L}^{\otimes m} \otimes \mathfrak{g})$ are no longer eigenmodes of $\mathcal{L}_v$. The other important issue is that now the conditions following from the definition of the two-form χ,

$$v \llcorner \chi = 0 \qquad \text{and} \qquad v \llcorner *\chi = -\chi \ ,$$

cannot be interpreted as saying that χ is a horizontal anti-self-dual two-form. If we express v as a linear combination

$$v = a_1\, \xi + a_2\, \widetilde{\xi} \ ,$$

where ξ is the Reeb vector field and $\widetilde{\xi}$ is a vector field orthogonal to ξ generating a $U(1)$ action, we find that the vertical part of χ may not vanish, but it lies in the subspace orthogonal to $\widetilde{\xi}$. However, more is true: we can decompose χ in terms of a two-form χ_T and a one-form χ_P. Explicitly

$$\chi = \Big(\kappa - \frac{a_1}{a_2}\, \widetilde{\kappa}\Big) \wedge \chi_P + \chi_T \ ,$$

where κ is the contact structure and $\widetilde{\kappa}$ is the one-form dual to $\widetilde{\xi}$, and with further anti-self-duality relations imposed on χ_P and χ_T. In the end, we are left with the same number of degrees of freedom as for a horizontal anti-self-dual two-form. From the more geometric perspective of transverse holomorphic foliations, the most natural point of view is to consider now the index of a new Dolbeault-like operator $\widetilde{\partial}^{\,(m)}$, whose cohomological complex is a deformation of the cohomological complex of the regular case according to the deformation of the transverse holomorphic foliation of $M_5^{g,h}$ as described in [429, Sec. 5] (see also [429, Sec. 7] and [421, Sec. 5] for a discussion about the particular case of ellipsoids).

At this point, we can again follow the approach of [395] and extend it to five dimensions. The action of the Reeb vector field is free, and we can reduce the computations to the quotient space $C_g \times \Sigma_h$. In this way we arrive at the cohomological localization formula

$$Z_{\text{vec}}\big(M_5^{g,h}\big) = \prod_{\alpha \in \triangle} \big(\mathrm{i}\,(\alpha,\phi)\big)^{g\,h-g-h} \prod_f \prod_{m \neq 0} \Big(\frac{m}{\epsilon_f} + \mathrm{i}\,(\alpha,\phi)\Big)^{\frac{1}{2}\,\mathrm{index}\,\bar{\partial}^{(m)}}$$

with an extra product over the fixed points of the additional $U(1)$ action on $C_g \times \Sigma_h$. The length parameter ϵ_f is the radius of the circle fibre over the fixed point labelled by f.

One-loop determinant of the vector multiplet on $M_3 \times \Sigma_h$

We will now specialise the present discussion to product manifolds $M_5^{g,h} = M_3 \times \Sigma_h$, where the classification reduces to the discussion of [417] about the orbits of the three-dimensional Killing vector field on M_3. We shall explicitly compute the Atiyah–Singer index in this case. The Kähler surface $K_4 = C_g \times \Sigma_h$ is endowed with the product Kähler structure $\omega = \omega_{C_g} + \omega_{\Sigma_h}$ and the $U(1)$-bundle projection π is the product of the Seifert fibration $M_3 \to C_g$ and the identity map on Σ_h. The integer cohomology of K_4 has generators $[\omega_{C_g}] \in H^2(C_g, \mathbb{Z})$ and $[\omega_{\Sigma_h}] \in H^2(\Sigma_h, \mathbb{Z})$ in this case, and the first Chern class of the line bundle $\mathscr{L} \to K_4$ associated to the circle fibration is given by $c_1(\mathscr{L}) = \deg(\mathscr{L})\, \omega_{C_g}$. One has $c(T^{1,0}\Sigma_h) = 1 + \chi(\Sigma_h)\, \omega_{\Sigma_h} = 1 + c_1(T^{1,0}\Sigma_h)$ with

$\chi(\Sigma_h) = 2-2h$ the Euler characteristic of the Riemann surface Σ_h, and similarly for C_g. The total Chern class is thus

$$\begin{aligned} c(T^{1,0}K_4) &= c(T^{1,0}C_g) \wedge c(T^{1,0}\Sigma_h) \\ &= \big(1+\chi(C_g)\,\omega_{C_g}\big) \wedge \big(1+\chi(\Sigma_h)\,\omega_{\Sigma_h}\big) \\ &= 1+2\,\big((1-g)\,\omega_{C_g}+(1-h)\,\omega_{\Sigma_h}\big)+4\,(1-g)\,(1-h)\,\omega_{C_g}\wedge\omega_{\Sigma_h} \\ &= 1+c_1(T^{1,0}K_4)+c_2(T^{1,0}K_4)\ , \end{aligned}$$

and the corresponding Todd class is

$$\begin{aligned} \mathrm{Td}(T^{1,0}K_4) &= 1+\tfrac{1}{2}\,c_1(T^{1,0}K_4)+\tfrac{1}{12}\,\big(c_1(T^{1,0}K_4)\wedge c_1(T^{1,0}K_4)+c_2(T^{1,0}K_4)\big) \\ &= 1+(1-g)\,\omega_{C_g}+(1-h)\,\omega_{\Sigma_h}+(1-g)\,(1-h)\,\omega_{C_g}\wedge\omega_{\Sigma_h}\ . \end{aligned}$$

Using $c_1(\mathscr{L}^{\otimes m}) = m\,\deg(\mathscr{L})\,\omega_{C_g}$, the corresponding Chern character is found to be

$$\mathrm{ch}(\mathscr{L}^{\otimes m}) = 1+c_1(\mathscr{L}^{\otimes m})+\tfrac{1}{2}\,c_1(\mathscr{L}^{\otimes m})\wedge c_1(\mathscr{L}^{\otimes m}) = 1+m\,\deg(\mathscr{L})\,\omega_{C_g}\ .$$

The index of the Dolbeault complex in this case is thus given by

$$\begin{aligned} \mathrm{index}\,\bar\partial^{(m)} &= \int_{K_4}\big(1+m\,\deg(\mathscr{L})\,\omega_{C_g}\big)\wedge\big(1+(1-g)\,\omega_{C_g}+(1-h)\,\omega_{\Sigma_h} \\ &\qquad\qquad +(1-g)\,(1-h)\,\omega_{C_g}\wedge\omega_{\Sigma_h}\big) \\ &= \int_{K_4}\big((1-g)\,(1-h)+m\,(1-h)\,\deg(\mathscr{L})\big)\,\omega_{C_g}\wedge\omega_{\Sigma_h} \\ &= (1-h)\int_{C_g}\big(m\,\deg(\mathscr{L})+1-g\big)\,\omega_{C_g} \\ &= (1-h)\,\big[m\,\deg(\mathscr{L})+1-g\big]\ . \end{aligned} \tag{11.4.5}$$

The term in square brackets is the index of the twisted Dolbeault complex associated to the circle bundle $M_3\to C_g$, and we finally get the localization formula

$$Z_{\mathrm{vec}}(M_3\times\Sigma_h) = Z_{\mathrm{vec}}(M_3)^{1-h}\prod_{\alpha\in\triangle_+}(\alpha,\mathrm{i}\,\sigma)^{-2h} \tag{11.4.6}$$

Again, the multiplicative factor in the general localization formula, which includes also contributions from the ghosts, is essential for cancelling the denominator. This formula can be extended to the case in which C_g has orbifold points, as reviewed in [421, Sec. 3.5].

Notice that the localization formula for the one-loop determinants on $M_3\times\Sigma_h$ lifts the perturbative three-dimensional partition function to the perturbative five-dimensional partition function. However, while the full (non-perturbative) partition function on M_3 receives contributions from flat connections on M_3, the full partition function on $M_3\times\Sigma_h$ includes connections $A^{(0)}$ that descend to instantons on $C_g\times\Sigma_h$. The moduli spaces over which we integrate are different. In fact, the pullback to M_5 of flat connections on M_3 are not generally solutions to the fixed point equation (11.4.1). This is a major difference from the partially twisted theory, in which the BPS configurations decompose into flat connections on M_3 and unconstrained connections on Σ_h.

11.4.4 One-loop determinant of a hypermultiplet

For the contribution of a hypermultiplet, we will only consider a regular background here, due to the issues encountered with the vector multiplet discussed in Section 11.4.3. Furthermore, we consider only the one-loop determinant in the perturbative partition function, that is we set $A^{(0)}=0$, hence $\phi=-\sigma_0$. We want to calculate the ratio of fluctuation determinants

$$\sqrt{\frac{\det \mathrm{i}\,\mathsf{L}_{-\sigma_0}|_{\mathrm{f}}}{\det \mathrm{i}\,\mathsf{L}_{-\sigma_0}|_{\mathrm{b}}}}=\sqrt{\frac{\det \mathrm{i}\,\mathsf{L}_{-\sigma_0}|_{\Omega_H^{(0,1)}(M_5,\mathfrak{g})}}{\det \mathrm{i}\,\mathsf{L}_{-\sigma_0}|_{\Omega_H^{0}(M_5,\mathfrak{g})}\ \det \mathrm{i}\,\mathsf{L}_{-\sigma_0}|_{\Omega_H^{(0,2)}(M_5,\mathfrak{g})}}}\ .$$

Applying the same strategy as with the vector multiplet, that is, decomposing the horizontal forms according to the tensor powers of the line bundle $\mathscr{L}$ associated to the $U(1)$ fibration, one arrives at the cohomological localization formula

$$Z^{\mathrm{pert}}_{\mathrm{hyp}}(M_5)=\prod_{\rho\in\Lambda_R}\prod_{m\in\mathbb{Z}}\Big(\frac{m+\Delta}{r}-\mathrm{i}\,(\rho,\sigma_0)\Big)^{-\mathrm{index}\,\bar\partial^{(m)}} \tag{11.4.7}$$

We briefly comment on the cohomological formulation of the hypermultiplet in a squashed or ellipsoid background. The topological twist depends, in general, on the geometric data. However, as discussed in [422] and also in [424, Sec. 3.3], we can turn on the squashing and continuously deform the contact structure, so that the Reeb vector field associated to the new contact structure stays parallel to the Killing vector field v. Then, the cohomological localization applies to the vector multiplet, although using a notion of 'horizontal' which differs from that on the round manifold we started with. For the topological twist of the hypermultiplet, however, a choice of spin^c structure is needed, and hence additional assumptions on the geometry of the base of the $U(1)$ fibration are required, usually that it is Kähler-Einstein. The pragmatic solution of [422] was to take the cohomological form of the hypermultiplet as a definition in a squashed Sasaki-Einstein geometry. This is not, however, a continuous deformation of the hypermultiplet in the round geometry, and we do not follow this strategy here.

11.4.5 Perturbative partition functions

We now come to the first applications of our cohomological localization formulas in five dimensions. Pan proved in [419] that the product manifolds $M_3\times\Sigma_h$ admit $\mathcal{N}=2$ supersymmetry if $M_3=\mathbb{S}^3$ or $M_3=\mathbb{T}^3$, the sphere or the torus. We will work out the full perturbative $\mathcal{N}=2$ partition functions on $\mathbb{S}^3\times\Sigma_h$. We shall then write down the perturbative $\mathcal{N}=1$ partition functions on more general product five-manifolds $M_3\times\Sigma_h$.

$\mathcal{N}=2$ perturbative partition functions on $\mathbb{S}^3\times\Sigma_h$

Regard the three-sphere $\mathbb{S}^3$ as the Hopf fibration of degree one over $C_0=\mathbb{S}^2$. We use the localization formula (11.4.6) with $A^{(0)}=0$ together with (11.4.7) which gives

$$Z^{\mathrm{pert}}_{\mathrm{vec}}(\mathbb{S}^3\times\Sigma_h)=Z_{\mathrm{vec}}(\mathbb{S}^3)^{1-h}\qquad\text{and}\qquad Z^{\mathrm{pert}}_{\mathrm{hyp}}(\mathbb{S}^3\times\Sigma_h)=Z_{\mathrm{hyp}}(\mathbb{S}^3)^{1-h}\ .$$

The full perturbative partition function is given by taking the product of the one-loop vector multiplet determinant with products of the one-loop hypermultiplet determinants over all $\mathcal{N}=2$ hypermultiplets a of conformal dimensions Δ_a in representations R_a of the gauge group G. We then multiply by the Boltzmann weight of the classical action (11.4.3) evaluated at the trivial solution $A^{(0)}=0$, and integrate over the remaining scalar moduli $\sigma_0\in\mathfrak{g}$ using the localization

formulas of Section 11.3.5. We can conjugate σ_0 into a Cartan subalgebra $\mathfrak{t} \subset \mathfrak{g}$ and use the Weyl integral formula to perform the resulting integral with the measure

$$\mathrm{d}\mu(\sigma_0) = \mathrm{d}\sigma_0 \prod_{\alpha\in\triangle_+} (\alpha,\sigma_0)^2 \ , \tag{11.4.8}$$

where $\mathrm{d}\sigma_0$ is the Lebesgue measure on $\mathfrak{t} = \mathbb{R}^{\mathrm{rank}(G)}$. The applicability of the Weyl integral formula is restricted to elements $\sigma_0 \in \mathfrak{t}$ for which the determinant in (11.4.8) is non-vanishing; these are called regular elements, and they form an open dense subset $\mathfrak{t}_{\rm reg} \subset \mathfrak{t}$. After cancelling the Jacobian in the integration measure with the denominator of the vector multiplet one-loop determinant, we obtain the perturbative partition function in the background of [419]:

$$Z^{\rm pert}_{\mathcal{N}=2}(\mathbb{S}^3\times\Sigma_h) = \int_{\mathfrak{t}_{\rm reg}} \mathrm{d}\tilde\sigma\, e^{-\frac{4\pi^2\,\mathrm{vol}(\Sigma_h)}{r g_{\rm YM}^2}(\tilde\sigma,\tilde\sigma)} \prod_{\alpha\in\triangle_+} \sinh\big(\pi\,(\alpha,\tilde\sigma)\big)^{2-2h} \times \prod_a \prod_{\rho_a\in\Lambda_{R_a}} \mathfrak{s}_1\big(\mathrm{i}\,(1-\Delta_a)-(\rho_a,\tilde\sigma)\big)^{1-h}, \tag{11.4.9}$$

where we defined the variable $\tilde\sigma = r\,\sigma_0$ in the Cartan subalgebra $\mathfrak{t}_{\rm reg} \subset \mathfrak{g}$. The $\mathcal{N}=1$ partition function on $\mathbb{S}^3\times\Sigma_h$ without matter and including instanton contributions will be analysed in Section 11.5.1.

$\mathcal{N}=1$ perturbative partition functions on $M_3\times\Sigma_h$

We shall now consider the $\mathcal{N}=1$ perturbative partition functions on more general $U(1)$ fibrations over $C_g\times\Sigma_h$, where C_g is a Riemann surface of genus g, focusing on the case $M_5 = M_3\times\Sigma_h$ where only $M_3\to C_g$ is fibered over C_g. Proceeding as above using the index formula (11.4.5), we arrive at

$$Z^{\rm pert}_{\mathcal{N}=1}(M_3\times\Sigma_h) = \int_{\mathfrak{t}_{\rm reg}} \mathrm{d}\tilde\sigma\, e^{-\frac{\pi\,\mathrm{vol}(C_g)\,\mathrm{vol}(\Sigma_h)}{r^3 g_{\rm YM}^2}(\tilde\sigma,\tilde\sigma)} \prod_{\alpha\in\Delta_+} \sinh\big(\pi\,(\alpha,\tilde\sigma)\big)^{2\,(1-h)\,(1-g)} \times \prod_a \prod_{\rho_a\in\Lambda_{R_a}} \mathfrak{s}_1\big(\mathrm{i}\,(1-\Delta_a)-(\rho_a,\tilde\sigma)\big)^{(1-h)\deg(\mathscr{L})} \ .$$

The products $L(p,1)\times\Sigma_h$ are particular examples [419] with $C_0=\mathbb{S}^2$, for which $\deg(\mathscr{L})=p$. The formalism should also apply when M_3 is a more general Seifert homology sphere which admits a contact structure, and in particular for the lens spaces $M_3 = L(p,s)$. It would also be interesting to extend the formalism to the case in which Σ_h has punctures.

11.4.6 Contact instantons and their pushdown to four dimensions

We shall now work out solutions to the fixed point equation (11.4.1) on $M_5 = M_3\times\Sigma_h$, and then study their pushdown to four dimensions.

Regular fibrations

We first focus on regular Seifert manifolds. We want to solve the equation

$$v \lrcorner *F = F \tag{11.4.10}$$

on $M_3 \times \Sigma_h$, where $v = -\xi$ is the Killing vector field with ξ the Reeb vector field of the Seifert fibration $M_3 \to C_g$. On K-contact five-manifolds, the solutions to this equation are refered to as contact instantons [230], and their moduli spaces are studied in [431]. We can rewrite (11.4.10) as

$$*F = \kappa \wedge F$$

or equivalently

$$v \llcorner F = 0 \qquad \text{and} \qquad F_{H,-} = 0 \ , \tag{11.4.11}$$

where $F_{H,\pm}$ denote the self-dual and anti-self-dual horizontal parts of the curvature two-form F. Let us decompose the gauge connection as

$$A = A_\theta\, \kappa + A_H \ ,$$

where κ is (minus) the contact structure on M_3 dual to v, so that

$$F = \big(A_\theta\, \mathrm{d}\kappa + \mathrm{d}A_\theta \wedge \kappa + \mathrm{d}A_H\big) - \mathrm{i}\,\big(A_H \wedge A_H + [A_\theta, A_H] \wedge \kappa\big) \ .$$

We partly follow the treatment of [231, Sec. 3.2]. The first equation $v \llcorner F = 0$ reads

$$\mathsf{D}_H A_\theta = \mathrm{d}A_\theta + \mathrm{i}[A_H, A_\theta] = 0 \ , \tag{11.4.12}$$

so A_θ is covariantly constant along $C_g \times \Sigma_h$. At this point, it is useful to prove that both the $\mathfrak{g}$-valued function A_θ and one-form A_H are invariant under translations along the fibre, generated by v. For this, we choose the gauge[33]

$$\mathcal{L}_v A = 0 \ . \tag{11.4.13}$$

Then

$$\mathcal{L}_v A_\theta = v \llcorner \mathrm{d}A_\theta = -\mathrm{i} v \llcorner [A_H, A_\theta] = 0 \qquad \text{and} \qquad \mathcal{L}_v A_H = v \llcorner \mathrm{d}A_H + \mathrm{d}(v \llcorner A_H) = 0 \ ,$$

where $v \llcorner \mathrm{d}A_H = 0$ follows from the gauge fixing condition (11.4.13).

From (11.4.12) it follows that the curvature F has only a horizontal part given by

$$F = F_H = A_\theta\, \mathrm{d}\kappa + \mathrm{d}A_H - \mathrm{i}\, A_H \wedge A_H \ .$$

At this point we use the second equation $F_{H,-} = 0$. The surviving self-dual part belongs to the vector space

$$F_{H,+} \ \in \ \Omega^{(1,1)}_{\text{sympl}}(C_g \times \Sigma_h, \mathfrak{g}) \oplus \Omega^{(2,0)}(C_g \times \Sigma_h, \mathfrak{g}) \oplus \Omega^{(0,2)}(C_g \times \Sigma_h, \mathfrak{g}) \ .$$

This implies that $\mathrm{d}A_H$ is proportional to the Kähler two-form $\omega_{C_g} + \omega_{\Sigma_h}$ on the base, and recalling the relation (11.2.7) between $\mathrm{d}\kappa$ and the Kähler form on $C_g \times \Sigma_h$, altogether we arrive at a curvature which is of the form

$$F = f_H\, \big(\omega_{C_g} + \omega_{\Sigma_h}\big) + F^{(2,0)} + F^{(0,2)} \ ,$$

where

$$F^{(2,0)} = -\mathrm{i}\, \big[(A_H)_y, (A_H)_z\big]\, \mathrm{d}y \wedge \mathrm{d}z \qquad \text{and} \qquad F^{(0,2)} = -\mathrm{i}\, \big[(A_H)_{\bar y}, (A_H)_{\bar z}\big]\, \mathrm{d}\bar y \wedge \mathrm{d}\bar z \ ,$$

and we have chosen local complex coordinates $(y, \bar y) \in C_g$ and $(z, \bar z) \in \Sigma_h$. The function $f_H \in \Omega^0(C_g \times \Sigma_h, \mathfrak{g})$ is a purely four-dimensional quantity.

To summarise, we arrive at a solution $A = A_\theta\, \kappa + A_H$ on $M_3 \times \Sigma_h$, where A_H is a connection on $C_g \times \Sigma_h$, and A_θ is a $\mathfrak{g}$-valued function which is constant along the fibre and covariantly constant

[33] We avoid formal considerations involved in the gauge fixing procedure. The details are exactly as in [230, 231].

on $C_g \times \Sigma_h$ with respect to A_H. The curvature F of A lives on $C_g \times \Sigma_h$ and is self-dual (from the five-dimensional point of view). The Yang–Mills action evaluated at these connections gives

$$\begin{aligned} S_{\rm YM}(F) &= \frac{1}{2\,g_{\rm YM}^2} \int_{M_5} \big(F \wedge\!\!\!\!\wedge *F\big) \\ &= \frac{1}{2\,g_{\rm YM}^2} \int_{M_5} \kappa \wedge \big(F_{H,+} \wedge\!\!\!\!\wedge F_{H,+}\big) \\ &= -\frac{\pi\, r}{g_{\rm YM}^2} \int_{C_g \times \Sigma_h} \big(F_{H,+} \wedge\!\!\!\!\wedge F_{H,+}\big) \\ &= \frac{8\pi^3\, r\,(\vec m, \vec n)}{g_{\rm YM}^2} \ , \end{aligned}$$

where we integrated over the circular fibre, of radius r, and used the fact that $F_{H,+}$ is independent of the fibre direction. The integer vectors $\vec m, \vec n \in \mathbb{Z}^{{\rm rank}(G)}$ are the gauge fluxes through C_g and Σ_h, respectively, which can be identified with weights of the Lie algebra $\mathfrak{g}$. Then, from Künneth's theorem, $(\vec m, \vec n) \in \mathbb{Z}$ is proportional to the second Chern character ${\rm ch}_2(P) \in H^4(C_g \times \Sigma_h, \mathbb{Q}) = H^2(C_g, \mathbb{Q}) \otimes H^2(\Sigma_h, \mathbb{Q})$ of the principal G-bundle $P \to C_g \times \Sigma_h$ on which A_H is a connection.

Five-dimensional gauge theories also have a topological global $U(1)_{\rm inst}$ symmetry [215], with conserved current

$$J_{\rm inst} = *\big(F \wedge\!\!\!\!\wedge F\big) \ .$$

The $U(1)_{\rm inst}$ charge of the current $J_{\rm inst}$ is the instanton number computed from F. The derivation given above makes it clear that this topological symmetry is related to the $U(1)$ invariance under rotation of the Seifert fibre.

Pushdown

Let us now describe the pushdown of these solutions. The localization of the supersymmetric gauge theory onto connections constant along the fibre, whose curvature descends to four dimensions, is reminiscent of the framework of [432], where the topologically twisted theory on $\mathbb{S}^2 \times \mathbb{S}^2$ was studied. In particular, the vertical component A_θ of the gauge field A is covariantly constant on the four-dimensional base manifold $C_g \times \Sigma_h$, thus A_θ is a scalar field on $C_g \times \Sigma_h$ constrained in exactly the same way as the scalar σ. Following [433], we can redefine our vector multiplet and hypermultiplets in terms of four-dimensional supersymmetry multiplets. The real scalar σ is combined with the vertical component A_θ to give a complex scalar $\phi = {\rm i}\,\sigma - v \llcorner A$, together with a purely four-dimensional gauge connection A_H. This reduction brings the $\mathcal{N} = 1$ five-dimensional vector multiplet down to the $\mathcal{N} = 2$ four-dimensional vector multiplet. Similar manipulations can be done for the hypermultiplets. The five-dimensional Majorana spinor ε breaks down into one left and one right chiral four-dimensional Killing spinor.

In [432], it is explained how to topologically twist the $\mathcal{N} = 2$ gauge theory on any four-dimensional manifold admitting a $U(1)$ isometry. We can then simply borrow their results: The fixed point equations in four dimensions are

$$[F, \phi] = [F, \phi^\dagger] = [\phi, \phi^\dagger] = 0 \ , \quad w \llcorner \mathsf{D}_H \phi^\dagger = 0 \qquad \text{and} \qquad w \llcorner F - {\rm i}\,{\rm d}\phi = 0 \ ,$$

where w is the vector field used for the twist. The vanishing Lie brackets imply that we can conjugate all fields A_H, σ and $v \llcorner A$ into the same Cartan subalgebra of $\mathfrak{g}$ simultaneously. This means that covariantly constant scalars σ can be taken to be constant. As pointed out in [432, Sec. 4.3], to obtain the full partition function one should include not only the sum over gauge fluxes $\vec m$ and $\vec n$ through the two surfaces C_g and Σ_h (which we have equivalently obtained from

direct computations in five dimensions), but also the Nekrasov partition functions which sum over point-like instantons corresponding to the fixed points of the action of the maximal torus of the symmetry group given by the direct product of the gauge group with the isometry group.

To explicitly compute the instanton contributions, we could take $C_0 = \Sigma_0 = \mathbb{S}^2$, where our result coincides with that of [244, Sec. 2]; they proceed in the other direction, starting from the theory on $\mathbb{S}^2 \times \mathbb{S}^2$ with Ω-background and then lifting it to $\mathbb{S}^3_b \times \mathbb{S}^2$.

While contributions from four-dimensional point-like instantons on $C_g \times \Sigma_h$ are hard to compute in more general geometries, we may hope to recover the full answer from a resurgent analysis, as explained in [434] for $\mathbb{S}^4$.

Irregular fibrations

The irregular case is more subtle. In this case, we may attempt to proceed as in the regular case, but now with the Killing vector v no longer pointing along the fibre direction. In other words, the equation

$$v \llcorner *F = F \tag{11.4.14}$$

cannot be interpreted in terms of horizontal and anti-self-dual components. Instead, we can decompose the gauge field as

$$A = A_\eta\, \eta + A_T \ ,$$

where η is the one-form dual to v, A_η is the component of the gauge connection along the direction of the isometry generated by v, and A_T is the transverse gauge connection. That is, we replace the notion of horizontal with that of transverse, which is natural in the present context [420]. Then the condition $v \llcorner F = 0$, necessary to fulfill (11.4.14), can be solved in an analogous way as for the regular case, leading to

$$\mathsf{D}_T A_\eta = \mathrm{d}A_\eta + \mathrm{i}[A_T, A_\eta] = 0 \qquad \text{and} \qquad \mathcal{L}_v A_\eta = 0 = \mathcal{L}_v A_T \ .$$

From the analysis of the regular case above, it is clear that we can again pushdown the theory to four dimensions. However, this time we do not reduce to the base $C_g \times \Sigma_h$ of the Seifert fibration, but instead to the submanifold transverse to η (equivalently, the submanifold orthogonal to v). This is what one expects by the construction of [420].

11.5 q-deformed Yang–Mills theories from cohomological localization

This final section is devoted to the study of five-dimensional supersymmetric Yang–Mills theory on $\mathbb{S}^3_b \times \Sigma_h$, where $\mathbb{S}^3_b$ is either the squashed sphere of [405] or the ellipsoid of [281] (see Appendix 11.B for details); recall from Section 11.2.1 that these are the five-dimensional theories that naturally descend from six-dimensional superconformal field theories on squashed geometries. We show how q-deformations of Yang–Mills theory on a Riemann surface Σ_h arise from our localization procedure. Our formula for the partition function of the standard q-deformed Yang–Mills theory in Section 11.5.1 improves the result of [394, 396], wherein the Gaussian term was not retained; as we discuss, this Boltzmann factor is important for applications to holography. Both treatments of [394] and [396] focus on the zero area limit where $\mathrm{vol}(\Sigma_h) \to 0$, which hides the fact that the resulting q-deformed Yang–Mills theory has $p = 1$. We elucidate the geometric significance of this new q-deformation through an analysis of the resulting matrix model on the sphere $\Sigma_0 = \mathbb{S}^2$, by adapting the procedure of [409] to the present case.

11.5.1 Localization on $\mathbb{S}^3 \times \Sigma_h$

Consider the $\mathcal{N} = 1$ partition function without matter on $\mathbb{S}^3 \times \Sigma_h$, beyond the perturbative calculation of Section 11.4.5. More precisely, we include the full set of solutions obtained in Section 11.4.6, but discard the four-dimensional point-like instantons.[34] The resulting partition function is given by

$$Z_{\mathcal{N}=1}(\mathbb{S}^3 \times \Sigma_h) = \sum_{A^{(0)}} \int_{\mathfrak{t}_{\rm reg}} \mathrm{d}\sigma_0 \, e^{-S_{\rm cl}(F,\sigma_0)} \prod_{\alpha \in \triangle_+} \sinh\left(\pi \, r \, (\alpha, \phi)\right)^{2-2h} ,$$

with $S_{\rm cl}$ the action evaluated at the localization locus:

$$S_{\rm cl}(F, \sigma_0) = S_{\rm YM}(F) + \frac{1}{2\, g_{\rm YM}^2} \int_{\mathbb{S}^3 \times \Sigma_h} \left(\frac{1}{r^2} \, (\sigma_0, \sigma_0) \, \kappa \wedge \mathrm{d}\kappa \wedge \omega_{\Sigma_h} + \frac{1}{r} \, (\sigma_0, F) \, \wedge \kappa \wedge \mathrm{d}\kappa \right) .$$

The first summand in the action, $S_{\rm YM}(F)$, is the five-dimensional Yang–Mills action evaluated in Section 11.4.6:

$$\begin{aligned} S_{\rm YM}(F) &= \frac{1}{2\, g_{\rm YM}^2} \int_{\mathbb{S}^3 \times \Sigma_h} \kappa \wedge \left(F_{H,+} \owedge F_{H,+} \right) \\ &= -\frac{\pi \, r}{g_{\rm YM}^2} \int_{\mathbb{S}^2 \times \Sigma_h} \left(F|_{\mathbb{S}^2 \times \Sigma_h} \owedge F|_{\mathbb{S}^2 \times \Sigma_h} \right) \\ &= \frac{8\pi^3 \, r}{g_{\rm YM}^2} \, (\vec{m}, \vec{n}) \ , \end{aligned} \tag{11.5.1}$$

where in the second equality we used the defining equations, and in the last equality $(\vec{m}, \vec{n}) \in \mathbb{Z}$ is proportional to the second Chern character ch_2 associated to the pertinent principal G-bundle over $\mathbb{S}^2 \times \Sigma_h$.

The sum in the partition function runs over the G-connections $A^{(0)}$ whose curvature F satisfies the fixed point equation

$$v \llcorner *F = F \ .$$

These solutions were studied in Section 11.4.6. It is straightforward to check that connections whose curvature lives on Σ_h, that is $F = F|_{\Sigma_h}$, do not satisfy this equation. In other words, connections that are flat when projected onto $\mathbb{S}^3$, and which thus belong to the localization locus of a purely three-dimensional theory, do not belong to the localization locus of the fully twisted five-dimensional theory. The flat connection on $\mathbb{S}^3$ does belong to the localization locus of the five-dimensional theory which is partially twisted along Σ_h (with arbitrary $A|_{\Sigma_h}$), but not to the localization locus of the fully twisted theory; the partition function of the field theory topologically twisted along Σ_h is important from the perspective of the six-dimensional theory and its reduction to four-dimensional theories of class $\mathcal{S}$ [394, 396, 244]. The same argument can be made for connections which are flat along Σ_h, which then only belong to the localization locus of the five-dimensional theory which is partially twisted along $\mathbb{S}^3$ (with arbitrary $A|_{\mathbb{S}^3}$) [411, 412, 414].

We can write

$$Z_{\mathcal{N}=1}(\mathbb{S}^3 \times \Sigma_h) = \sum_{\vec{m}, \vec{n} \in \mathbb{Z}^{\mathrm{rank}(G)}} \int_{\mathfrak{t}_{\rm reg}} \mathrm{d}\phi \, e^{-S_{\rm cl}(\vec{m}, \vec{n}; \phi)} \prod_{\alpha \in \triangle_+} \sinh\left(\pi \, r \, (\alpha, \phi)\right)^{2-2h} . \tag{11.5.2}$$

[34] This sector of the partition function was named "perturbative" in [244]. However, throughout the present chapter the term perturbative has been used to refer to the expansion around the trivial connection, while the subsector of the full partition function we are considering now includes a much wider class of solutions. A proper definition hinted at in [244, Sec. 2.4] might be "partition function neglecting codimension four field configurations". These are non-perturbative with respect to an expansion in the geometric area parameter $\mathrm{vol}(\Sigma_h)$.

We shifted the integration variable to $\phi = -\sigma_0 - v \llcorner A^{(0)}$. This is the natural variable to use when descending to a four-dimensional description, as the sum in (11.5.2) is taken over the gauge fluxes $\vec{m}$ and $\vec{n}$ through $\mathbb{S}^2$ and Σ_h, respectively; the first and second Chern characters associated to the principal G-bundle over $\mathbb{S}^2 \times \Sigma_h$ are obtained by using the pairing $(\,\cdot\,,\,\cdot\,)$. The action evaluated at the fixed points consists of three terms:

$$S_{\rm cl}(\vec{m},\vec{n};\phi) = S_{\rm YM}(\vec{m},\vec{n}) + \frac{1}{2\,g_{\rm YM}^2}\int_{\mathbb{S}^3\times\Sigma_h}\Big(\frac{1}{r^2}\,(\phi,\phi)\,\kappa\wedge{\rm d}\kappa\wedge\omega_{\Sigma_h} - \frac{1}{r}\,(\phi,F|_{\mathbb{S}^2\times\Sigma_h})\wedge\kappa\wedge{\rm d}\kappa\Big)\ ,$$

where the first summand is written in (11.5.1). Passing to the scaled variable $\tilde{\phi} = r\,\phi$, we get

$$\begin{aligned} S_{\rm cl}(\vec{m},\vec{n};\tilde{\phi}) &= S_{\rm YM}(\vec{m},\vec{n}) \\ &\quad + \frac{1}{g_{\rm YM}^2}\Big(\frac{8\pi^2\,{\rm vol}(\Sigma_h)}{r}\,(\tilde{\phi},\tilde{\phi}) + \frac{2\pi{\rm i}}{r^2}\int_{\mathbb{S}^3\times\Sigma_h}\frac{\rm i}{4\pi}\,(\tilde{\phi},F|_{\mathbb{S}^2\times\Sigma_h})\wedge\kappa\wedge{\rm d}\kappa\Big)\ . \end{aligned} \tag{11.5.3}$$

Zero flux sector

We will now restrict ourselves to those connections for which $S_{\rm YM}(\vec{m},\vec{n}) = 0$. Contributions with non-vanishing second Chern character are exponentially suppressed. We turn off the gauge fluxes through $\mathbb{S}^2$, $\vec{m} = \vec{0}$, and follow standard techniques from two-dimensional Yang–Mills theory, see in particular [423]. Using

$$\frac{1}{2}\int_{\mathbb{S}^3\times\Sigma_h}\frac{\rm i}{2\pi}\,(\tilde{\phi},F|_{\mathbb{S}^2\times\Sigma_h})\wedge\kappa\wedge{\rm d}\kappa = 4\pi^2\,r^3\,(\tilde{\phi},\vec{n})\ ,$$

and summing over $\vec{n}\in\mathbb{Z}^{{\rm rank}(G)}$, the third summand in the action (11.5.3) produces the delta-function constraint

$$\frac{4\pi^2\,r}{g_{\rm YM}^2}\,\tilde{\phi} = \vec{k}$$

for some integer vector $\vec{k}\in\mathbb{Z}^{{\rm rank}(G)}$ which can be identified with a regular weight of $\mathfrak{g}$. Plugging this into the remaining Gaussian part of the action gives

$$\frac{g_{\rm YM}^2}{4\pi\,r}\,\frac{{\rm vol}(\Sigma_h)}{\pi\,r^2}\,(\vec{k},\vec{k}\,)\ .$$

In the one-loop determinant, we obtain

$$\sinh\big(\pi\,r\,(\alpha,\phi)\big) = \sinh\big(\pi\,(\alpha,\tilde{\phi})\big) = \sinh\Big(\frac{g_{\rm YM}^2}{4\pi\,r}\,(\alpha,\vec{k}\,)\Big) = -\big[(\alpha,\vec{k}\,)\big]_q$$

where

$$q = e^{-g_{\rm str}} \qquad \text{with} \quad g_{\rm str} := \frac{g_{\rm YM}^2}{2\pi\,r}\ . \tag{11.5.4}$$

The final form of the partition function is then

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3\times\Sigma_h) = \sum_{\vec{k}\in(\mathbb{Z}^{{\rm rank}(G)})_{\rm reg}}\ \prod_{\alpha\in\triangle_+}\big[(\alpha,\vec{k}\,)\big]_q^{2-2h}\ q^{\frac{1}{2}\,(\vec{k},\vec{k}\,)\,\frac{{\rm vol}(\Sigma_h)}{\pi\,r^2}}\ . \tag{11.5.5}$$

From (11.5.4) we identify the six-dimensional radius $\beta = \frac{g_{\rm YM}^2}{2\pi}$, and it is instructive to rewrite

$$q^{\frac{1}{2}\,(\vec{k},\vec{k}\,)\,\frac{{\rm vol}(\Sigma_h)}{\pi\,r^2}} = \exp\Big(-(\vec{k},\vec{k}\,)\,\frac{2\pi\,\beta\ {\rm vol}(\Sigma_h)}{{\rm vol}_\kappa(\mathbb{S}^3)}\Big)\ , \tag{11.5.6}$$

where $\mathrm{vol}_\kappa(\mathbb{S}^3)$ is the volume of $\mathbb{S}^3$ taken with respect to the contact structure κ, in our normalization.

The expression (11.5.5) is the partition function of the standard q-deformed two-dimensional Yang–Mills theory on Σ_h (cf. Section 11.2.2), at $p = 1$ and with Gaussian weight corrected by the ratio of volumes between the fibre Riemann surface Σ_h and the total space sphere $\mathbb{S}^3$. This ratio of volumes matches exactly with [407]. For unitary gauge group $G = U(N)$, this is reminiscent of the large N free energy of gauge theories with a holographic dual. In particular, for $\Sigma_h = \Sigma_0 = \mathbb{S}^2$, the N^3 behaviour of the free energy at large N in the present theory follows immediately from the large N limit of q-deformed Yang–Mills theory, which in turn is given by the free energy of Chern–Simons theory on $\mathbb{S}^3$ as we are in the case $p = 1$. This indeed gives the right answer for a gauge theory on a Sasaki-Einstein five-manifold which has a holographic dual [231]. See [241] for a discussion and a similar example where the N^3 behaviour of the free energy in five dimensions is extracted from Chern–Simons gauge theory in three dimensions. See also [235] for the localization of five-dimensional maximally supersymmetric Yang–Mills theory to a three-dimensional subsector, and the relation with Chern–Simons theory.

We notice that the string coupling g_{str} depends only on the ratio $\frac{g^2_{\mathrm{YM}}}{r}$. This is consistent with dimensional reduction from the $(2,0)$ theory on $\mathbb{S}^1 \times \mathbb{S}^3 \times \Sigma_h$ discussed in Section 11.2.1. In the six-dimensional setting, g^2_{YM} plays the role of the circumference of the circle $\mathbb{S}^1$ on which we have reduced. On the other hand, the small area limit $\mathrm{vol}(\Sigma_h) \to 0$ gives the superconformal index of the four-dimensional gauge theory on $\mathbb{S}^1 \times \mathbb{S}^3$, consistently with the conjecture of [406]. From (11.5.6) it is evident that the small $\mathrm{vol}(\Sigma_h)$ limit is the same as the large $\mathrm{vol}_\kappa(\mathbb{S}^3)$ limit which decompactifies the three-sphere $\mathbb{S}^3$ to $\mathbb{R}^3$.

Reinstating gauge fluxes

We have so far restricted ourselves to the sector where $S_{\mathrm{YM}}(\vec{m},\vec{n}) = 0$. In general we have to consider additional solutions which are given by connections whose curvature lives on $\mathbb{S}^2 \times \Sigma_h$ with non-trivial second Chern character $\mathrm{ch}_2 \neq 0$, and

$$S_{\mathrm{YM}}(\vec{m},\vec{n}) = \frac{8\pi^3\, r}{g^2_{\mathrm{YM}}}\,(\vec{m},\vec{n}) = 2\pi\,\mathrm{i}\Big(-\,\mathrm{i}\,\frac{4\pi^2\, r}{g^2_{\mathrm{YM}}}\,(\vec{m},\vec{n})\Big)\ ,$$

where as above $\vec{n}$ is the gauge flux through Σ_h and $\vec{m}$ is the gauge flux through $\mathbb{S}^2$. The full partition function includes this term in the action, along with a sum over $\vec{m} \in \mathbb{Z}^{\mathrm{rank}(G)}$. The bracketed term has exactly the same coefficient as a BF-type term in the action. This means that the procedure we used for the $\vec{m} = \vec{0}$ sector should be modified by a shift $\tilde{\phi} \mapsto \tilde{\phi} - \mathrm{i}\,\vec{m}$ after the sum over all $\vec{n} \in \mathbb{Z}^{\mathrm{rank}(G)}$. We finally arrive at

$$Z_{\mathcal{N}=1}(\mathbb{S}^3 \times \Sigma_h) = \sum_{\vec{m}\in\mathbb{Z}^{\mathrm{rank}(G)}}\ \sum_{\vec{k}\in(\mathbb{Z}^{\mathrm{rank}(G)})_{\mathrm{reg}}}\ \prod_{\alpha\in\triangle_+}\Big(\big[(\alpha,\vec{k}-\mathrm{i}\,\vec{m})\big]_q\,\big[(\alpha,\vec{k}+\mathrm{i}\,\vec{m})\big]_q\Big)^{1-h}$$
$$\times\ q^{\frac{1}{2}\,(\vec{k}-\mathrm{i}\,\vec{m},\vec{k}-\mathrm{i}\vec{m})\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\, r^2}}\ .$$

The q-deformed measure has in fact precisely the right form to support non-trivial fluxes $\vec{m}$.

11.5.2 Localization on squashed $\mathbb{S}^3_b \times \Sigma_h$

There exist two types of squashings of $\mathbb{S}^3$ that can be lifted to five dimensions with $\mathcal{N} = 2$ supersymmetry. The first type is the familiar case of [281]. From the point of view of cohomological localization, the squashing simply corresponds to a rescaling of the fibre radius, and the results

for the one-loop determinants on $\mathbb{S}^3_b \times \Sigma_h$ are exactly the same as for the round sphere $\mathbb{S}^3$, but with fibre radius $\epsilon \neq r$, where r is the radius of the base $\mathbb{S}^2$. See Appendix 11.B for further details.

The other squashed sphere, which is a regular fibration, is that of [405]. We did not give a formal derivation of the one-loop determinants in the cohomological gauge theory for this background. However, we know that the Killing vector field v has closed orbits and is parallel to the Reeb vector field ξ. We can therefore lift the results from $\mathbb{S}^3_b$ to $\mathbb{S}^3_b \times \Sigma_h$, and the one-loop determinants are given by

$$Z_{\rm vec}(\mathbb{S}^3_b \times \Sigma_h) = \Bigg(\prod_{\alpha \in \triangle_+} \frac{\sinh\big(\pi\,\epsilon_1\,(\alpha,\sigma_0)\big)\,\sinh\big(\pi\,\epsilon_2\,(\alpha,\sigma_0)\big)}{\pi^2\,\epsilon_1\,\epsilon_2\,(\alpha,\sigma_0)^2} \Bigg)^{1-h}$$

$$Z_{\rm hyp}(\mathbb{S}^3_b \times \Sigma_h) = \Bigg(\prod_{\rho \in \Lambda_R} \mathtt{s}_b\big(\tfrac{\rm i}{2}\,(b+b^{-1})\,(1-\Delta) - r\,(\rho,\sigma_0)\big) \Bigg)^{1-h} ,$$

which agrees with [244, Sec. 2]. We adopted the notation

$$\epsilon_1 = r\,b \qquad \text{and} \qquad \epsilon_2 = r\,b^{-1} \qquad \text{with} \quad b = \frac{1-{\rm i}\,u}{\sqrt{1+u^2}}$$

from Appendix 11.B.

The computation of the full $\mathcal{N}=1$ partition function without matter on $\mathbb{S}^3_b \times \Sigma_h$ proceeds exactly as in Section 11.5.1, with only a modification in the one-loop determinant. This results in a bi-orthogonalization of the q-deformed measure, and the partition function in the sector of vanishing second Chern character $\mathrm{ch}_2 = 0$ is given by

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \Sigma_h) = \sum_{\vec{k} \in (\mathbb{Z}^{\mathrm{rank}(G)})_{\rm reg}} \prod_{\alpha \in \triangle_+} \Big(\big[b\,(\alpha,\vec{k}\,)\big]_q \,\big[b^{-1}\,(\alpha,\vec{k}\,)\big]_q \Big)^{1-h}\; q^{\frac{1}{2}\,(\vec{k},\vec{k}\,)\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\,r^2}} .$$

The extension to the full partition function including non-trivial gauge fluxes $\vec{m}$ through the base $\mathbb{S}^2$ is exactly as described in Section 11.5.1, as we are presently working in the regular case. After inclusion of gauge fluxes $\vec{m} \neq \vec{0}$, our result appears to be only in partial agreement with [244, 410], where the field theory is first defined on $\mathbb{S}^3_b \times \mathbb{R}^2$, then dimensionally reduced, and finally put on the Riemann surface Σ_h with a twist. That procedure allows for additional background fluxes for the flavour symmetry, which do not appear in our framework nor in [394].

11.5.3 Localization on ellipsoid $\mathbb{S}^3_b \times \Sigma_h$

We now consider the geometry $\mathbb{S}^3_b \times \Sigma_h$ for the ellipsoid $\mathbb{S}^3_b$ of [281]. Most of the steps are the same as in the round $\mathbb{S}^3$ case of Section 11.5.1. The action at the localization locus is again given by

$$S_{\rm cl}(F,\sigma_0) = S_{\rm YM}(F) + \frac{1}{2\,g^2_{\rm YM}} \int_{\mathbb{S}^3 \times \Sigma_h} \Big(\frac{1}{r^2}\,(\sigma_0,\sigma_0)\,\kappa \wedge {\rm d}\kappa \wedge \omega_{\Sigma_h} + \frac{1}{r}\,(\sigma_0,F) \wedge \kappa \wedge {\rm d}\kappa \Big) ,$$

with

$$S_{\rm YM}(F) = \frac{1}{2\,g^2_{\rm YM}} \int_{\mathbb{S}^3 \times \Sigma_h} \big(F \wedge *F\big) .$$

The zero flux sector of the partition function is

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \Sigma_h) = \sum_{\vec{n} \in \mathbb{Z}^{\mathrm{rank}(G)}} \int_{\mathfrak{t}_{\rm reg}} {\rm d}\sigma_0\, e^{-S_{\rm cl}(\vec{0},\vec{n};\sigma_0)} \times \prod_{\alpha \in \triangle_+} \Big(\sinh\big(\pi\,r\,b\,(\alpha,\sigma_0)\big)\,\sinh\big(\pi\,r\,b^{-1}\,(\alpha,\sigma_0)\big) \Big)^{1-h} . \tag{11.5.7}$$

Changing variable $\tilde{\sigma} = r\,\sigma_0$ and repeating the same steps used in Section 11.5.2, we arrive at

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \Sigma_h) = \sum_{\vec{k}\in(\mathbb{Z}^{\mathrm{rank}(G)})_{\mathrm{reg}}} e^{-\frac{g_{\mathrm{str}}}{2}\,(\vec{k},\vec{k})\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\,r^2}} \times \prod_{\alpha\in\triangle_+}\left(\sinh\left(b\,\frac{g_{\mathrm{str}}\,(\alpha,\vec{k})}{2}\right)\sinh\left(b^{-1}\,\frac{g_{\mathrm{str}}\,(\alpha,\vec{k})}{2}\right)\right)^{1-h} ,$$

where as before we defined the string coupling $g_{\mathrm{str}} = \frac{g^2_{\mathrm{YM}}}{2\pi\,r} = \frac{\beta}{r}$. The final expression can be rewritten in terms of $q = e^{-g_{\mathrm{str}}}$ as

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \Sigma_h) = \sum_{\vec{k}\in(\mathbb{Z}^{\mathrm{rank}(G)})_{\mathrm{reg}}} q^{\frac{1}{2}\,(\vec{k},\vec{k})\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\,r^2}} \prod_{\alpha\in\triangle_+}\Big(\big[b\,(\alpha,\vec{k})\big]_q\,\big[b^{-1}\,(\alpha,\vec{k})\big]_q\Big)^{1-h} .$$

From (11.5.7) we see that the perturbative partition function on $\mathbb{S}^3_b \times \mathbb{S}^2$, retaining only the contribution from the trivial flat connection $A^{(0)} = 0$, coincides with the perturbative partition function of Chern–Simons gauge theory on the lens space $L(p,1)$, continued to arbitrary values $p = b^2 \in \mathbb{R}$:

$$Z^{\mathrm{pert}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \mathbb{S}^2) = \int_{\mathfrak{t}_{\mathrm{reg}}} \mathrm{d}\sigma\, e^{-\frac{1}{2\,g_{\mathrm{str}}}\,(\sigma,\sigma)\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\,r^2}} \prod_{\alpha\in\triangle_+}\sinh\left(b\,\frac{(\alpha,\sigma)}{2}\right)\sinh\left(b^{-1}\,\frac{(\alpha,\sigma)}{2}\right) ,$$

where we rescaled $\sigma = g_{\mathrm{str}}\,\sigma_0$. This becomes more evident if we use instead an asymmetric length scaling to define the variable $\tilde{\sigma} = \epsilon_1\,\sigma_0 = r\,b\,\sigma_0$. For general genus h we then arrive at the discrete matrix model

$$Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \Sigma_h) = \frac{1}{b^{\mathrm{rank}(G)}} \sum_{\vec{k}\in(\mathbb{Z}^{\mathrm{rank}(G)})_{\mathrm{reg}}} q^{\frac{1}{2}\,(\vec{k},\vec{k})\,\frac{\mathrm{vol}(\Sigma_h)}{\pi\,r^2}} \prod_{\alpha\in\triangle_+}\Big(\big[(\alpha,\vec{k})\big]_q\,\big[b^{-2}\,(\alpha,\vec{k})\big]_q\Big)^{1-h} .$$

Further details and analysis of this matrix model for genus $h = 0$ and gauge group $G = U(N)$, along the lines of [409], are provided in Section 11.5.4 below.

Reinstating the additional contributions of contact instantons, with non-vanishing fluxes through the base $\mathbb{S}^2$ of the $U(1)$ fibration of $\mathbb{S}^3_b$, is a much more subtle issue. This is because the $U(1)$-action now involves the Killing vector field v which differs from the Reeb vector field ξ on $\mathbb{S}^3_b$, so that contractions with v do not separate the horizontal and vertical parts of the differential forms involved.

11.5.4 The matrix model

We focus now on the partition function $Z^{\vec{m}=\vec{0}}_{\mathcal{N}=1}(\mathbb{S}^3_b \times \mathbb{S}^2)$ for gauge group $G = U(N)$. The partition function is formally the same for either the squashed sphere or the ellipsoid $\mathbb{S}^3_b$. Only the geometric meaning of the squashing parameter b is different in the two cases, in particular b is a complex number of unit modulus $|b| = 1$ for the squashed sphere and $b > 0$ is real for the ellipsoid. However, as the partition function can be analytically continued in both cases, there is no difference in practice.

Our goal is then to study the discrete random matrix ensemble with partition function

$$\mathcal{Z}_N(b) = \frac{1}{b^N\,N!} \sum_{\vec{\ell}\in\mathbb{Z}^N} e^{-\frac{g_{\mathrm{str}}}{2}\,\sum_{j=1}^N \ell_j^2} \prod_{1\leq j<k\leq N} 4\sinh\left(\frac{g_{\mathrm{str}}}{2}\,(\ell_j-\ell_k)\right)\sinh\left(\frac{g_{\mathrm{str}}}{2\,b^2}\,(\ell_j-\ell_k)\right) , \quad (11.5.8)$$

which can be identified with the discrete version of the bi-orthogonal Stieltjes-Wigert ensemble studied in [83]. If $p := b^2 \in \mathbb{Z}$, the continuous version of the bi-orthogonal Stieltjes-Wigert

ensemble provides the partition function of Chern–Simons theory on the lens space $L(p,1)$. In the limit $b \to 1$ we recover the partition function of q-deformed Yang–Mills theory constructed from the monopole bundle over $\mathbb{S}^2$ with $p = 1$, whose continuous counterpart is Chern-Simon theory on $\mathbb{S}^3$ (with analytically continued level). In the present setting, b^2 may be any positive real number, not necessarily integer, and indeed the analysis of the bi-orthogonal Stieltjes-Wigert ensemble in [83] does not rely on p being integer. We shall now clarify the geometric significance of the dependence on the squashing parameter b of this q-deformed Yang–Mills theory.

Semi-classical expansion

Following [409], we will begin by performing a modular inversion of the series (11.5.8) to obtain the dual description of the q-deformed Yang–Mills matrix model in terms of instanton degrees of freedom. For this, we consider the function

$$F_b(x_1,\dots,x_N) = e^{-\frac{g_{\rm str}}{4}\,\sum_{j=1}^N x_j^2} \prod_{1\leq j<k\leq N} 2\sinh\Big(\frac{g_{\rm str}}{2\,b^2}\,(x_j-x_k)\Big)$$

of continuous variables $(x_1,\dots,x_N)\in\mathbb{R}^N$. Its Fourier transform is given by

$$\begin{aligned}\widehat{F}_b(y_1,\dots,y_N) &:= \int_{\mathbb{R}^N} {\rm d}x\, e^{\sum_{j=1}^N (2\pi{\rm i}x_j\,y_j - \frac{g_{\rm str}}{4}\,x_j^2)} \prod_{1\leq j<k\leq N}\Big(e^{\frac{(x_j-x_k)\,g_{\rm str}}{2\,b^2}} - e^{-\frac{(x_j-x_k)\,g_{\rm str}}{2\,b^2}}\Big)\\ &= e^{-\frac{4\pi^2}{g_{\rm str}}\,\sum_{j=1}^N \left(y_j+\frac{{\rm i}(N-1)\,g_{\rm str}}{4\pi\,b^2}\right)^2+\frac{N\,(N-1)^2\,g_{\rm str}}{2\,b^4}}\\ &\quad\times\int_{\mathbb{R}^N}{\rm d}u\,e^{-\frac{g_{\rm str}}{4}\,\sum_{j=1}^N u_j^2}\prod_{1\leq j<k\leq N}\Big(e^{\frac{g_{\rm str}\,u_j}{b^2}+\frac{4\pi\,{\rm i}\,y_j}{b^2}}-e^{\frac{g_{\rm str}\,u_k}{b^2}+\frac{4\pi\,{\rm i}\,y_k}{b^2}}\Big)\ ,\end{aligned}$$

where we completed squares and changed integration variables to $u_j = x_j - \frac{4\pi{\rm i}}{g_{\rm str}}\,y_j + \frac{N-1}{b^2}$. We now change integration variables again, in the usual way for matrix models with hyperbolic interactions, by defining

$$z_j = e^{\frac{g_{\rm str}\,u_j}{b^2}+\frac{2\,g_{\rm str}}{b^4}}$$

to get

$$\begin{aligned}\widehat{F}_b(y_1,\dots,y_N) &= \Big(\frac{b^2}{g_{\rm str}}\Big)^N e^{-\frac{g_{\rm str}}{2\,b^4}\,N\,(N^2+1)}\,e^{-\frac{4\pi^2}{g_{\rm str}}\,\sum_{j=1}^N\left(y_j+\frac{{\rm i}(N-1)\,g_{\rm str}}{4\pi\,b^2}\right)^2}\\ &\quad\times\int_{(0,\infty)^N}{\rm d}z\,e^{-\frac{b^4}{4\,g_{\rm str}}\,\sum_{j=1}^N(\log z_j)^2}\prod_{1\leq j<k\leq N}\Big(z_j\,e^{\frac{4\pi\,{\rm i}\,y_j}{b^2}}-z_k\,e^{\frac{4\pi\,{\rm i}\,y_k}{b^2}}\Big)\ .\end{aligned}$$

The integral expression we have arrived at is exactly the same as in [409, Eq. (3.14)] under the identification of the string coupling constant $\tilde g_{\rm str}$ there as $\tilde g_{\rm str} = \frac{g_{\rm str}}{b^2}$, which as we have seen is the coupling that reproduces the standard q-deformed Yang–Mills theory. We can therefore evaluate the integral using Stieltjes-Wigert polynomials to get

$$\begin{aligned}\widehat{F}_b(y_1,\dots,y_N) &= \Big(\frac{4\pi}{g_{\rm str}}\Big)^{\frac{N}{2}}\,e^{\frac{g_{\rm str}}{6\,b^4}\,N\,(N-1)\,(N-2)}\,e^{-\frac{4\pi^2}{g_{\rm str}}\,\sum_{j=1}^N\left(y_j+\frac{{\rm i}(N-1)\,g_{\rm str}}{4\pi\,b^2}\right)^2}\\ &\quad\times\prod_{1\leq j<k\leq N}\Big(e^{\frac{4\pi\,{\rm i}\,y_j}{b^2}}-e^{\frac{4\pi\,{\rm i}\,y_k}{b^2}}\Big)\ .\end{aligned}$$

At this point, we apply the convolution theorem for Fourier transformations to get

$$\begin{aligned}Z_b(y_1,\dots,y_N) &:= \int_{\mathbb{R}^N}{\rm d}x\,e^{2\pi\,{\rm i}\,\sum_{j=1}^N x_j\,y_j}\,F_b(x_1,\dots,x_N)\,F_1(x_1,\dots,x_N)\\ &= \int_{\mathbb{R}^N}{\rm d}t\ \widehat{F}_b\Big(\frac{y_1-t_1}{2},\dots,\frac{y_N-t_N}{2}\Big)\,\widehat{F}_1\Big(\frac{y_1+t_1}{2},\dots,\frac{y_N+t_N}{2}\Big)\ .\end{aligned}$$

After some calculation, we arrive finally at

$$Z_b(y_1,\dots,y_N) = e^{-\frac{2\pi^2}{g_{\rm str}}\,\sum_{j=1}^N\,y_j^2}\;\mathcal{W}_b(y_1,\dots,y_N)\ ,$$

with weight given by

$$\begin{aligned}\mathcal{W}_b(y_1,\dots,y_N) = \Big(\frac{4\pi}{g_{\rm str}}\Big)^N\,e^{\frac{g_{\rm str}}{12}\,N\,(N^2-1)\,(1+b^{-4})} \\ \times\int_{\mathbb{R}^N}{\rm d}t\,e^{-\frac{2\pi^2}{g_{\rm str}}\,\sum_{j=1}^N\,t_j^2}\prod_{1\leq j<k\leq N}2\Big(\cos\pi\left(y_{jk}\,(1+b^{-2})+t_{jk}\,(1-b^{-2})\right) \\ -\,\cos\pi\left(y_{jk}\,(1-b^{-2})+t_{jk}\,(1+b^{-2})\right)\Big)\ ,\end{aligned} \tag{11.5.9}$$

where we adopted the shorthand notation $y_{jk} := y_j - y_k$ and $t_{jk} := t_j - t_k$. This correctly reproduces the weight of [409, Eq. (3.20)] in the limit $b = 1$.

The final step in developing the semi-classical expansion of the partition function (11.5.8) is Poisson resummation, and we finally arrive at

$$\mathcal{Z}_N(b) = \frac{1}{b^N\,N!}\sum_{\vec{\ell}\in\mathbb{Z}^N}Z_b(\vec{\ell}\,) = \frac{1}{b^N\,N!}\sum_{\vec{\ell}\in\mathbb{Z}^N}e^{-\frac{2\pi^2}{g_{\rm str}}\,(\vec{\ell},\vec{\ell}\,)}\;\mathcal{W}_b(\vec{\ell}\,)\ . \tag{11.5.10}$$

This expression admits the standard interpretation as a sum over instanton solutions of the two-dimensional gauge theory: Since the q-deformation arises here through one-loop determinants in the initial five-dimensional field theory, at the classical level this theory is just ordinary Yang–Mills theory on the sphere $\mathbb{S}^2$. The exponential prefactors in the series (11.5.10) are then the classical contributions to the gauge theory path integral from the Yang–Mills action evaluated on instantons of topological charge $\ell_j \in \mathbb{Z}$ corresponding to a Dirac monopole of the j-th factor of the maximal torus $U(1)^N \subset U(N)$, while the integrals (11.5.9) are the fluctuation determinants around each instanton.

Rational limit and Chern–Simons theory on $L(p,s)$

Up to now the derivation of (11.5.10) worked for every positive real value of the squashing parameter b. Let us now specialise the squashing parameter to the rational values

$$b^2 = \frac{p}{s}\ \in\ \mathbb{Q}\ ,$$

where p and s are coprime positive integers with $1 \leq s \leq p$. From the five-dimensional perspective that we started with, the ellipsoid Seifert manifold $\mathbb{S}^3_b$ then has the topology of a lens space $L(p,s)$, viewed as a circle bundle over $\mathbb{S}^2$ with two marked points [421]; the exceptional fibres over the marked points respectively makes them $\mathbb{Z}_p$ and $\mathbb{Z}_s$ orbifold points. The first Chern class of the line V-bundle $\mathscr{L}(p,s)$ over the $\mathbb{P}^1$ orbifold associated to $L(p,s)$ is

$$c_1\big(\mathscr{L}(p,s)\big) = \frac{s}{p}\,\omega_{\mathbb{P}^1}\ ,$$

which cancels the local curvatures at the marked points of $\mathbb{P}^1$ to ensure that the total degree of the Seifert fibration is zero. This is also homeomorphic to the 'fake' lens space which is the quotient $\mathbb{S}^3/\mathbb{Z}_p$ by the free $\mathbb{Z}_p$-action

$$\big(z_1,z_2\big)\longmapsto\big(e^{2\pi{\rm i}s/p}\,z_1,e^{2\pi\,{\rm i}/p}\,z_2\big)\ , \tag{11.5.11}$$

where $\mathbb{S}^3$ is regarded as the unit sphere in $\mathbb{C}^2$. From the two-dimensional perspective, we will now show that the instanton expansion (11.5.10) retains topological information reflecting its five-dimensional origin, by rewriting it in terms of flat connection contributions to $U(N)$ Chern–Simons gauge theory on the lens spaces $L(p,s)$.

Since

$$\pi_1\big(L(p,s)\big) = \mathbb{Z}_p \ ,$$

gauge inequivalent flat $U(N)$ connections are labelled by N-tuples $\vec{m} \in (\mathbb{Z}_p)^N$, which are torsion magnetic charges coming from the pullback of a Yang–Mills instanton on the sphere $\mathbb{S}^2$ to a flat connection on $L(p,s)$ [397]. Let us then rewrite the series variables $\vec{\ell} \in \mathbb{Z}^N$ in (11.5.10) as

$$\vec{\ell} = \vec{m} + p\,\vec{l}\,,$$

with $\vec{l} \in \mathbb{Z}^N$. We can rewrite the interactions among eigenvalues from (11.5.9) as

$$4\,\sin\frac{\pi\,s}{p}\,\big(m_{jk} - t_{jk} + p\,(l_j - l_k)\big)\,\sin\pi\,\big(m_{jk} + t_{jk} + p\,(l_j - l_k)\big)\ ,$$

and thus it depends only on the values of $\vec{\ell} \in \mathbb{Z}^N$ modulo p, that is, on $\vec{m} \in (\mathbb{Z}_p)^N$.[35] They are also invariant under the Weyl symmetry group S_N of the gauge group $U(N)$, so that we can reduce the sum over all N-tuples $\vec{m} \in (\mathbb{Z}_p)^N$ to the ordered ones with

$$m_N \geq m_{N-1} \geq \cdots \geq m_1 \ .$$

Thus the partition function (11.5.10) depends only on how many times the integers $k \in \{0,1,\dots,p-1\}$ appear in the string $(m_1,\dots,m_N)$. We denote these multiplicities as $\mathbf{N} = (N_0, N_1, \dots, N_{p-1})$, which by construction are p-component partitions of the rank N:

$$N_k \geq 0 \qquad \text{and} \qquad \sum_{k=0}^{p-1} N_k = N \ .$$

Under this reordering the Weyl symmetry breaks according to

$$S_N \longrightarrow S_{N_0} \times S_{N_1} \times \cdots \times S_{N_{p-1}} \ .$$

The partition function (11.5.10) is then rewritten as

$$\mathcal{Z}_N(p,s) = \Big(\frac{s}{p}\Big)^{N/2} \sum_{\mathbf{N}\vdash N} \frac{1}{\prod\limits_{k=0}^{p-1} N_k!}\ \mathcal{W}_{p,s}\big(0^{N_0}, 1^{N_1}, \dots, (p-1)^{N_{p-1}}\big)$$
$$\times \sum_{\vec{l}\in\mathbb{Z}^N} \prod_{k=0}^{p-1} \exp\Big(-\frac{2\pi^2}{g_{\rm str}} \sum_{j=N_0+N_1\cdots+N_{k-1}+1}^{N_k} (k + p\,l_j)^2\Big)\ ,$$

where here $k^{N_k} := (k,\dots,k)$ is the N_k-vector whose entries are all equal to k. As in [409, Sec. 3.3], we identify in the second line a product of elliptic theta-functions

$$\vartheta_3(\tau|z) = \sum_{l\in\mathbb{Z}} e^{\pi{\rm i}\tau\,l^2 + 2\pi{\rm i}l\,z}$$

[35] Strictly speaking, this is only true if the integers p and s have the same even/odd parity, which we tacitly assume.

which enables us to write

$$\mathcal{Z}_N(p,s) = \sum_{\mathbf{N}\vdash N} e^{-\frac{2\pi^2}{g_{\rm str}}\sum_{k=0}^{p-1} N_k\,k^2}\ \mathcal{W}_{p,s}\big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)\prod_{k=0}^{p-1}\frac{\vartheta_3\big(\frac{2\pi{\rm i}p^2}{g_{\rm str}}\big|\frac{2\pi{\rm i}p\,k}{g_{\rm str}}\big)^{N_k}}{N_k!}\ . \tag{11.5.12}$$

We can write the fluctuation weight $\mathcal{W}_{p,s}\big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)$ explicitly in its integral form (11.5.9) and reorganize the integration variables t_j into subsets t_j^J with $J\in\{0,1,\dots,p-1\}$ and $j\in\{1,\dots,N_J\}$. We then shift integration variables as $u_j^J := t_j^J - j$ to get

$$\begin{aligned}
&\mathcal{W}_{p,s}\big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)\\
&\quad=\Big(\frac{4\pi}{g_{\rm str}}\Big)^N e^{\frac{g_{\rm str}}{12}N(N^2-1)\left(1+\frac{s^2}{p^2}\right)}\\
&\qquad\times\prod_{J=0}^{p-1}\int_{\mathbb{R}^{N_J}}{\rm d}u^J\,e^{-\frac{2\pi^2}{g_{\rm str}}\sum_{j=1}^{N_J}(u_j^J+j)^2}\prod_{1\le j<k\le N_J}4\,\sin\frac{\pi\,s}{p}\left(u_j^J-u_k^J\right)\sin\pi\left(u_j^J-u_k^J\right)\\
&\qquad\quad\times\prod_{0\le J<K\le p-1}\prod_{j=1}^{N_J}\prod_{k=1}^{N_K}4\,\sin\frac{\pi\,s}{p}\left(u_j^J-u_k^K+J-K\right)\sin\pi\left(u_j^J-u_k^K+J-K\right)\ .
\end{aligned}\tag{11.5.13}$$

The products here which are independent of (p,s) combine to give a standard Weyl determinant, while the (p,s)-dependent products carry the information about the surgery data of the Seifert homology sphere $X(s/p)$.

In fact, if we drop the product of theta-functions from the sum (11.5.12) and rescale the string coupling as before to $\tilde g_{\rm str} = s\,g_{\rm str}/p$, we can recognise the analytically continued partition function of $U(N)$ Chern–Simons gauge theory at level $k\in\mathbb{Z}$ on the lens space $L(p,s)$: the exponential prefactor is recognized as the classical contribution to the path integral from the Chern–Simons action evaluated on the flat $U(N)$ connection labelled by $\mathbf{N}$ [133, 397], with the analytic continuation

$$\tilde g_{\rm str} = \frac{s\,g_{\rm str}}{p} = \frac{2\pi{\rm i}}{k+N}\ .$$

Moreover, after a straightforward change of integration variables (and subsequent analytic continuation), the integral expression (11.5.13) is easily seen to agree with the multi-eigenvalue integral formula from [435, Thm 7] for the contribution to the one-loop fluctuation determinant from the flat connection $\mathbf{N}$. Thus the full partition function (11.5.12) can be written as

$$\mathcal{Z}_N(p,s) = \sum_{\mathbf{N}\vdash N} Z^{\rm CS}_{p,s}(\mathbf{N})\prod_{k=0}^{p-1}\frac{\vartheta_3\big(\frac{2\pi{\rm i}p^2}{g_{\rm str}}\big|\frac{2\pi{\rm i}p\,k}{g_{\rm str}}\big)^{N_k}}{N_k!}\ ,$$

where

$$Z^{\rm CS}_{p,s}(\mathbf{N}) := \exp\left(\frac{2\pi^2\,s}{\tilde g_{\rm str}\,p}\sum_{k=0}^{p-1}N_k\,k^2\right)\mathcal{W}_{p,s}\big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)$$

is the contribution to the Chern–Simons partition function from the point of the moduli space of flat connections on the lens space $L(p,s)$ labelled by

$$\vec m = \big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)\ .$$

The connection between Chern–Simons theory on lens spaces $L(p,s)$ with $s>1$ and q-deformed Yang–Mills theory was also obtained in [397], but in a much different and more complicated

fashion. There the two-dimensional gauge theory is defined on the collection of exceptional divisors of the four-dimensional Hirzebruch–Jung space $X_4(p,s)$, which is the minimal resolution of the $A_{p,s}$ singularity defined by the same orbifold action (11.5.11) on $\mathbb{C}^2$. The corresponding partition function depends explicitly on the intersection moduli e_i of the exceptional divisors, which in the three-dimensional case translate into framing integers that enter the surgery construction of the Seifert space $L(p,s)$; after stripping away the Chern–Simons fluctuation determinants, the resulting partition function computes the contribution of fractional instantons to the partition function of topologically twisted $\mathcal{N}=4$ Yang–Mills theory on $X_4(p,s)$ [397]. This is not the case here. Like the topological Chern–Simons theory, our two-dimensional gauge theory partition function (11.5.12) is independent of the framing integers e_i and depends only on the pair of integers (p,s) which uniquely determine $L(p,s)$ up to homeomorphism. In particular, stripping away the Chern–Simons fluctuation determinants $\mathcal{W}_{p,s}\big(0^{N_0},1^{N_1},\dots,(p-1)^{N_{p-1}}\big)$ would leave a (p,s)-independent partition function proportional to $\vartheta_3\big(\frac{2\pi\mathrm{i}}{g_{\rm str}}\big|0\big)^N$ [397], which is the contribution of fractional instantons to the partition function of $\mathcal{N}=4$ gauge theory on $X_4(1,1)\cong\mathcal{O}_{\mathbb{P}^1}(-1)$. This suggests that our squashing of the two-dimensional q-deformed gauge theory on $\mathbb{S}^2$ is, like the standard theory at $p=1$, also related to the Calabi-Yau geometry of the resolved conifold $\mathcal{O}_{\mathbb{P}^1}(-1)\oplus\mathcal{O}_{\mathbb{P}^1}(-1)$; the topological string interpretation of this theory is certainly worthy of further investigation.

Large N limit

For any finite value of the rank N, the partition function $\mathcal{Z}_N(b)$ is a continuous function of the squashing parameter $b>0$. In this sense our squashed q-deformations of two-dimensional Yang–Mills theory are continuations of the lens space theories analysed above: Since the set of rational b^2 is dense in the space of all squashing parameters $b>0$, any partition function can be expressed as a limit of the two-dimensional gauge theories whose geometric meanings were explained above. It would be interesting to understand more precisely what the underlying geometry means for generic real values $b>0$.

However, we do expect the partition function (11.5.8) to experience a phase transition in the large N regime, triggered by the discreteness of the matrix model, at least for large enough values of the squashing parameter b. The standard q-deformed Yang–Mills theory on $\mathbb{S}^2$ undergoes a phase transition for $p>2$ [409], and we can extrapolate this to our more general setting. The eigenvalue distribution $\rho(\lambda)$ of a discrete random matrix ensemble is subject to the constraint

$$\rho(\lambda)\leq 1\ ,$$

which in the present case is always fulfilled at large N when $b\leq\sqrt{2}$. It would be interesting to see how the phase transition appears at $b>\sqrt{2}$ in terms of the bi-orthogonal Stieltjes-Wigert polynomials of [83, Sec. 4.1]. It was argued in [83], and later proved in [436], that around the trivial flat connection the discrete and continuous versions of the Stieltjes-Wigert ensemble are essentially the same, thus the zero-instanton sector of our squashing of q-deformed Yang–Mills theory can be obtained exactly via bi-orthogonal polynomials. For the case $b=1$ this gives the full partition function of q-deformed Yang–Mills theory, since the only flat connection on $L(1,1)\cong\mathbb{S}^3$ is trivial.

Introduce the 't Hooft coupling

$$t:=g_{\rm str}\,N$$

and take the 't Hooft limit $N\to\infty, g_{\rm str}\to 0$ with t fixed. In this limit, the partition function (11.5.8) is proportional to the Chern–Simons matrix model on $L(p,1)$ around the trivial connection, continued to $p=b^2\in\mathbb{R}$. Equivalently, from the instanton expansion (11.5.10) we infer that, as long as the fluctuations $\mathcal{W}_b(\vec{\ell}\,)$ give sub-leading contributions, all instanton contributions are

suppressed except for the trivial one. Taking the large N limit of [83, Eq. (2.26)], in the large N regime we obtain

$$\mathcal{Z}_N^{(\infty)}(b) = 2^{N\,(N+1)} \left(\frac{2\pi}{g_{\rm str}}\right)^{\frac{N}{2}} \exp\left(- \frac{N^2\,t}{12\,b^4}\left(3\,b^8 + 6\,b^4 - 13\right) - \frac{N^2\,b^8}{t^2}\,F_{\rm CS}^{(0)}\left(\frac{t}{b^4}\right)\right) ,$$

where

$$F_{\rm CS}^{(0)}(t) = \frac{t^3}{12} - \frac{\pi^2 t}{6} - {\rm Li}_3\big(e^{-t}\big) + \zeta(3)$$

is the planar free energy of Chern–Simons theory on $\mathbb{S}^3$ with (analytically continued) 't Hooft coupling t. For $b \leq \sqrt{2}$ this solution is exact, and indeed the free energy of the supersymmetric gauge theory on $\mathbb{S}^3_b \times \mathbb{S}^2$ exhibits the N^3 behaviour in the strong coupling region for $t \to \infty$, as in the case of the five-sphere $\mathbb{S}^5$ [231]. However, for higher values of b, this solution ceases to be valid for large t and we expect the strong coupling region to have a different solution.

11.A Spinor conventions

For field theories in three dimensions we follow the normalization and conventions of [137]. We work in the $\mathcal{N} = 2$ formalism, and for theories with $\mathcal{N} = 4$ supersymmetry only $SU(2)_R \subset SU(2)_C \times SU(2)_H$ is manifest. In five dimensions, we follow [231]. We work in the $\mathcal{N} = 1$ formalism, with $SU(2)_R$ R-symmetry, using the letters $I, J, \dots$ for the indices. In theories admitting $\mathcal{N} = 2$ supersymmetry, only $SU(2)_R \subset SU(2)_R \times U(1)_r \subset SO(5)_R$ is manifest, where $SU(2)_R \times U(1)_r$ is the maximal R-symmetry group preserved by the product manifolds considered in the main text and $SO(5)_R$ the R-symmetry in five-dimensional flat space. In both three and five dimensions, $SU(2)_R$ indices are raised and lowered with the Levi-Civita symbol ϵ_{IJ} or ϵ^{IJ}, with the convention $\epsilon_{12} = -1 = -\epsilon^{12}$.

We do not write Lorentz spin indices explicitly in the spinors, and understand that they are contracted using the charge conjugation matrix C, a real antisymmetric matrix satisfying

$$C\,\Gamma^\mu = (\Gamma^\mu)^\top\,C\ .$$

With this choice, spinor components are taken to be Grassmann-even, and anticommutation is a consequence of $C^\top = -C$. This is also in agreement with the conventions in [419, 420], where Killing spinors from rigid supergravity are taken to be Grassmann-even symplectic Majorana spinors. Our notation for Killing spinors is then as follows: ε^I satisfies

$$\big(\epsilon_{IJ}\,C\varepsilon^J\big)^* = \varepsilon^I\ ,$$

with C the charge conjugation matrix. This implies

$$\varepsilon^\dagger_I = \big(\epsilon_{IJ}\,\varepsilon^J\big)^\top\ .$$

We impose the following reality conditions on the fields in the five-dimensional theories. The scalars $(\mathfrak{q}^\dagger)^I$ in a hypermultiplet are related to $\mathfrak{q}_I$ by complex conjugation and transposition. As we are working in Euclidean space, there is no reality condition on the spinor fields, and we have to choose a half-dimensional integration cycle in the configuration space of fields. The gauginos λ^I are symplectic Majorana spinors,

$$\big(\epsilon_{IJ}\,C\lambda^J\big)^* = \lambda^I\ ,$$

and we take as a definition of the fields $\lambda^\dagger_I$ the equation

$$\lambda^\dagger_I = \big(\epsilon_{IJ}\,\lambda^J\big)^\top\ .$$

The reasoning for the spinor ψ in a hypermultiplet is analogous. See also [424, Sec. 2.1] for discussion about the treatment of the reality condition for spinors. The sole difference between our conventions and those of [231, 424] is a factor $\sqrt{-1}$ in the definition of the scalar σ in the vector multiplet, as in those references the rotated field (which we denote σ_0) is taken from the very beginning.

In three dimensions, the gauginos λ^I are not subject to additional constraints, and we impose $\lambda_I^\dagger$ to be related to λ^I as in Minkowski signature, following [58, 137], and similarly for $\psi^\dagger$ and ψ.

11.B Squashed three-spheres

Three types of squashed sphere $\mathbb{S}^3_b$ that preserve at least $\mathcal{N}=2$ supersymmetry (four supercharges) exist in the literature: the squashed sphere called the "familiar case" in [281], the squashed sphere of [405], and the ellipsoid of [281] which was originally called the "less familiar case" of squashed sphere. We have ordered them according to their increasing deviation from the standard round sphere $\mathbb{S}^3$. In the following we briefly describe and discuss them within the cohomological field theory formalism, see also [396, Sec. 7] for related discussion.

The simplest case of squashed sphere is the "familiar case" of [281], for which the one-loop determinants are the same as in the round case up to rescaling of variables. This squashed sphere is obtained by simply changing the radius of the Hopf fibre with respect to the radius of the base $\mathbb{S}^2$, so the background has isometry group $SU(2)\times U(1)$. The Killing spinor is covariantly constant, as on the round $\mathbb{S}^3$. The Killing vector field v has compact orbits and coincides with the generator of rotations along the Hopf fibre, hence it is parallel to the Reeb vector field ξ. The computation of the one-loop determinants on this geometry is very simple in our setting: it is clear from construction (see (11.3.4)–(11.3.5)) that only the radius of the circle fibre enters the one-loop determinant, and the result of [281] follows immediately from Källén's formula. More generally, the one-loop determinants on the squashed lens spaces $\mathbb{S}^3_b/\mathbb{Z}_p$, with $p\in\mathbb{Z}_{>0}$ and $\mathbb{S}^3_b$ the "familiar" squashed sphere of [281], are given by the same formula as for the round lens space, with the proper scaling of the size of the fibre.

Another squashing that preserves $\mathcal{N}=2$ supersymmetry is the non-trivial squashed sphere of [405]. This is obtained by twisted dimensional reduction from $\mathbb{S}^3\times\mathbb{R}$ with round $\mathbb{S}^3$. One employs a Scherk–Schwarz compactification to put the theory on $\mathbb{S}^3\times\mathbb{S}^1$ by identifying

$$\exp\left(2\pi\,\beta\,\frac{\partial}{\partial t}+\pi\,\beta\,\mathsf{R}\right)X \;\sim\; X$$

for any field X, where β is the radius of $\mathbb{S}^1$, t is the coordinate along $\mathbb{R}$ and R is the generator of an R-symmetry transformation. This R-symmetry action twists the compactification by including a finite rotation on $\mathbb{S}^3$ in the periodic identification of fields along $\mathbb{S}^1$, and hence also twists the dimensional reduction when sending $\beta\to 0$. With this procedure, one is able to preserve an $SU(2)\times U(1)$ isometry group and to obtain a metric on $\mathbb{S}^3$ as for the "familiar case" of [281], hence it does not modify the metric on the base $\mathbb{S}^2$ of the Hopf fibration. However, the Killing spinor now becomes a non-constant field. In [417, Sec. 5.2] it was shown that the supergravity background of [405] admits $\mathcal{N}=4$ supersymmetry. The Killing vector field v has closed orbits, and therefore Källén's localization directly applies. In this squashed sphere, the Killing spinor depends on a parameter u related to the radius of the circle fibre through

$$\frac{r^2}{\epsilon^2}=1+u^2\ ,$$

where r is the radius of the round base $\mathbb{S}^2$ and ϵ is the radius of the $U(1)$ fibre. This relation arises from the twisted dimensional reduction. Since the Killing vector field is non-constant, although

it points along the squashed Hopf fibre one has to consider the complex dependence on $1 \pm \mathrm{i}\, u$: at the two fixed points the fibre has radius, respectively, given by

$$\epsilon_1 = r\, b \qquad \text{and} \qquad \epsilon_2 = r\, b^{-1} \qquad \text{with} \quad b = \sqrt{\frac{1-\mathrm{i}\, u}{1+\mathrm{i}\, u}}\ .$$

This justifies the applicability of the localization formula (11.3.5), and explains why the one-loop determinants with this non-trivial squashing are formally the same as for the ellipsoid of [281].

Finally we consider the ellipsoid of [281], which is the squashed sphere considered mostly in the main text. It has only $U(1) \times U(1)$ isometry group as the squashing also deforms the metric on the base $\mathbb{S}^2$ of the fibration. It is defined as the locus in $\mathbb{C}^2$ satisfying

$$b^2\, |z_1|^2 + b^{-2}\, |z_2|^2 = r^2\ , \tag{11.B.1}$$

with $b = \sqrt{\epsilon_1/\epsilon_2}$ and $r = \sqrt{\epsilon_1\, \epsilon_2}$. The supergravity background only preserves four supercharges. As explained in [417, Sec. 5.1] and highlighted in the main text, the orbits of the Killing vector field v are not closed, and v does not point along the fibre direction. It splits into two vector fields, generating two $U(1)$ isometries with closed orbits. The localized partition function on this geometry is discussed in Section 11.3.5.

Notice the different geometric meaning of the squashing parameter b in the squashed sphere [405] and in the ellipsoid [281]. For the ellipsoid $b > 0$ is real, while for the squashed sphere $b \in \mathbb{C}$ with $|b| = 1$. In both cases the partition functions can be analytically continued to arbitrary complex values $b \in \mathbb{C} \setminus \{0\}$, and the expressions are given by (11.3.8) and (11.3.9). Thus, in practice, we can compute the one-loop determinants on the regular background $\mathbb{S}^3_b$ of [405] and then continue the result to $b > 0$, or vice versa. The limit $b \to 0^+$ in the squashed sphere and ellipsoid also has different geometric meaning. For the squashed sphere, $b \to 0^+$ means $u \to -\mathrm{i}$, and corresponds to blowing up the twisted Hopf fibre. In the ellipsoid, $b \to 0^+$ reduces the geometric locus (11.B.1) to $|z_2| = 0$, hence the $\mathbb{S}^3_b$ geometry degenerates to $\mathbb{C}$. In both cases the geometry becomes non-compact, and we do not expect the cohomological localization to work in this limit.

11.C Sasaki-Einstein five-manifolds

There exists a family of Sasaki-Einstein metrics on five-manifolds M_5 which topologically are $U(1)$-fibrations over the product $\mathbb{S}^2 \times \mathbb{S}^2$ [437]. The simplest case is the familiar conifold $T^{1,1}$, which is homeomorphic to $\mathbb{S}^3 \times \mathbb{S}^2$ and so its associated line bundle $\mathscr{L}_{1,1}$ has Chern class $c_1(\mathscr{L}_{1,1}) = \omega_1$ given by the generator $[\omega_1] \in H^2(\mathbb{S}^2, \mathbb{Z})$ of the first base factor. More generally, there is an infinite family of irregular backgrounds labeled by a pair of integers (p, s), and denoted $Y^{p,s}$ [438]. The first Chern class associated to the circle bundle $Y^{p,s} \to \mathbb{S}^2 \times \mathbb{S}^2$ is

$$c_1(\mathscr{L}_{p,s}) = p\, \omega_1 + s\, \omega_2$$

where $[\omega_1]$ and $[\omega_2]$ generate the second cohomology $H^2(\mathbb{S}^2, \mathbb{Z})$ of the respective factors of the base. When p and s are coprime, $Y^{p,s}$ is again topologically $\mathbb{S}^3 \times \mathbb{S}^2$.

Field theories with $\mathcal{N} = 1$ supersymmetry on these manifolds have been studied in [398, 399] via application of the index theorem. Except for the simplest case $\mathbb{S}^3 \times \mathbb{S}^2$, the Killing vector field v does not have closed orbits. Moreover, the Reeb vector field ξ does not act freely on the base space of the fibration, which is a warped product $\mathbb{S}^2 \rtimes \mathbb{S}^2$. These manifolds have a toric action, and in fact admit a free $U(1)$ action, which however is generated by a different vector field from the Reeb vector field. Therefore our formalism does not apply. In [398], the way around this problem was to use the same idea that we did, but in the opposite direction. The manifolds $Y^{p,s}$ can be obtained as a quotient $\mathbb{S}^3 \times \mathbb{S}^3/U(1)$, with $U(1)$ acting freely on the six-dimensional

manifold $\mathbb{S}^3 \times \mathbb{S}^3$. One can compute the one-loop determinants using the index of the twisted Dolbeault complex in six dimensions, and then use the fact that there is a free $U(1)$ action to push it down to $Y^{p,s}$. In the case of the conifold $T^{1,1}$, the vacuum moduli spaces of instantons have been described in this way by [439].

In fact, the most natural way to look at these geometries is as follows. Consider $\mathbb{C}^4$ as the direct product $\mathbb{C}^2 \times \mathbb{C}^2$, with a sphere $\mathbb{S}^3 \subset \mathbb{C}^2$ embedded in each factor in the standard way. There is a $U(1)^4$ action on $\mathbb{C}^4$, where each $U(1)_i$ acts on the corresponding factor of $\mathbb{C}$ in $\mathbb{C}^4$, with associated equivariant parameter ϵ_i^{-1}, $i = 1, \ldots, 4$. There is a further $U(1)_T$ action on $\mathbb{C}^4$ with charges which are functions of two integer parameters p and s; explicitly, $U(1)_T$ acts with charges $(p+s, p-s, -p, -s)$. Then the action of $U(1)_T$ is free on $\mathbb{S}^3 \times \mathbb{S}^3 \subset \mathbb{C}^4$, and taking the quotient $\mathbb{S}^3 \times \mathbb{S}^3 / U(1)_T$ gives $Y^{p,s}$ with a residual toric action by $U(1)^3$, but none of the remaining $U(1)$ actions is free. The one-loop determinants then appear as products over four integers m_i, $i = 1, \ldots, 4$, each one corresponding to an eigenvalue $\frac{m_i}{\epsilon_i}$ where ϵ_i^{-1} is the equivariant parameter for rotation in the i-th plane $\mathbb{C}$ in $\mathbb{C}^4$. These four integers are constrained by one linear relation, corresponding to the quotient by the freely acting $U(1)_T$, which reads

$$(p+s)\, m_1 + (p-s)\, m_2 - p\, m_3 - s\, m_4 = 0 \ .$$

Chapter 12

$T\overline{T}$-deformation of q-Yang–Mills theory

12.1 Introduction to the chapter

Two-dimensional quantum field theories provide a playground for the study of exactly solvable models, and for testing the relationships and dualities with other areas such as integrable systems, statistical mechanics and string theory. Recent broad interest has been attracted to the solvable irrelevant deformation by the $T\overline{T}$ operator [45, 46], which is present in any local relativistic two-dimensional quantum field theory (see [440] for a review). The novelty of this deformation is marked by two key properties. First, it does not alter the integrability of a system. Second, when the field theory is compactified on a circle, the evolution of the energy levels with the parameter τ, encoding the strength of the deformation, is described by a first order inhomogeneous differential equation of Burgers type. The deformation induced by the operator $T\overline{T}$, henceforth called the $T\overline{T}$-deformation, was interpreted in [441] as a random fluctuation of the background geometry. A further step in this direction was made in [442, 443], where a path integral formulation of the $T\overline{T}$-deformed theory was put forward: it was proven in [442, 443] that the deformation of a given quantum field theory by the $T\overline{T}$ operator is equivalent to coupling the undeformed theory to flat space Jackiw–Teitelboim (JT) gravity. These ideas are similar in spirit to the interpretation of the $T\overline{T}$-deformation as a field-dependent spacetime coordinate transformation [444].

After the original formalism for $T\overline{T}$-deformed field theories considered by [45, 46], a wide range of other aspects of the $T\overline{T}$-deformation have been investigated. The original flat space deformation was extended to two-dimensional quantum field theories on AdS_2 in [445]. Their role in $\mathrm{AdS}_3/\mathrm{CFT}_2$ holography was investigated in [446–454]. The $T\overline{T}$-deformation of Wess–Zumino–Witten (WZW) models was studied from the string theory perspective in the target space theory [455] and also in the gauged worldsheet sigma-model [456]. Supersymmetric extensions were considered by [457–462]. Generalized $T\overline{T}$-deformations were discussed in [463–465], while the extensions to higher-dimensional field theories is considered by [466] using holography, and more recently in [467] from direct analysis of the renormalization group flow equation. Other facets of the perturbation by the $T\overline{T}$ operator considered recently include the study of the modular properties of the partition functions of deformed theories [468, 469], the extension of the $T\overline{T}$-deformation to non-relativistic systems [470], and correlation functions in conformal field theories on curved manifolds [471, 472]. A bridge between the Polyakov loop and the $T\overline{T}$-deformation of a bosonic field theory has been established in [473].

In this chapter we are concerned with the $T\overline{T}$-deformations of two-dimensional gauge theories. A simple proposal for the $T\overline{T}$-deformation of Yang–Mills theory on a Riemann surface was advocated by [474]: due to the simple form of the evolution of the two-dimensional Yang–Mills

Hamiltonian with the deformation parameter τ, the $T\overline{T}$-deformed version of the theory simply amounts to replacing the quadratic Casimir that appears in the usual heat kernel expansion of the partition function according to

$$C_2 \longmapsto \frac{C_2}{1-\tau\,\frac{g_{\rm YM}^2}{2}\,C_2}\ , \tag{12.1.1}$$

where $g_{\rm YM}$ is the Yang–Mills coupling constant. Following this proposal, the phase structure of the $U(N)$ theory on the sphere was studied at large N in [9], using standard field theory techniques. The prescription (12.1.1) was derived by [475] by directly coupling the heat kernel expansion, which depends on the area of the Riemann surface, to JT gravity and performing the gravitational path integral.

The aim of the present chapter is to study the analogous features of the q-deformation of two-dimensional Yang–Mills theory. To justify the effect of the $T\overline{T}$-deformation given by (12.1.1), we follow a different route than [475]. We consider the first order formalism for two-dimensional Yang–Mills theory, which rewrites it as a deformation of BF theory, and hence as an example of an 'almost' topological gauge theory, in the sense which we make precise in Section 12.2. In this latter general class of theories we can study the effect of the $T\overline{T}$-deformation precisely through its coupling to JT gravity, and we reproduce the prescription of [474] as a corollary of a more general result, using standard Abelianization techniques to evaluate the path integral. At the same time, thanks to the almost topological nature of these two-dimensional theories, we can employ cutting and gluing techniques of topological quantum field theory to rigorously justify the extension of the $T\overline{T}$-deformation, which is only well-defined on flat space, to curved Riemann surfaces such as the sphere, which was not addressed by [474, 475]. In a certain sense, the two deformations are compatible: the q-deformation results from a modification of the path integral measure, leaving the quadratic Casimir unchanged, while the $T\overline{T}$-deformation modifies only the Hamiltonian, i.e. the Casimir, but nothing else.

With our techniques, we are able to explore how various facets of two-dimensional Yang–Mills theory are affected by the $T\overline{T}$-deformation. For example, we obtain closed expressions for Wilson loop observables as well as the partition functions on Riemann surfaces with marked points. However, a number of noteworthy features are lost under the deformation. For example, the $T\overline{T}$-deformation of q-deformed Yang–Mills theory is no longer related to Chern–Simons theory (or a deformation thereof) on a circle bundle over the Riemann surface. Moreover, the large N factorization property, which splits the $U(N)$ gauge theory into chiral and anti-chiral sectors, no longer holds after deformation. This splitting is a crucial ingredient in the derivation of the large N string theory dual of two-dimensional Yang–Mills theory, thus casting doubt on the existence of such a string theory description of the $T\overline{T}$-deformed theory. This can be physically understood by mapping the $T\overline{T}$-deformed gauge theory onto a system of N non-relativistic fermions on a circle, which are now subjected to non-local interactions leading to long-range correlations between the fermions.

A central point of the current chapter is the analysis of the large N limit of the $T\overline{T}$-deformation of q-deformed $U(N)$ Yang–Mills theory on the sphere. We show that the main features of the undeformed theory are preserved, namely there is a third order phase transition induced by instantons. Furthermore, the $T\overline{T}$-deformation lowers the critical line as the strength τ of the deformation is increased. On the other hand, it extends the class of line bundles for which the phase transition occurs. We also show that these results continue to hold in the refinement of the theory, known as (q,t)-deformed Yang–Mills theory, whereby the region of the small coupling phase is further reduced by the refinement.

The rest of the chapter is organized as follows. In Section 12.2 we present our formalism for generic almost topological gauge theories. In Section 12.3 we then focus on Yang–Mills theory in

two dimensions together with its q-deformation and subsequent refinement which depend, among other continuous moduli, on a discrete parameter $p \in \mathbb{Z}$; we present their $T\overline{T}$-deformation and study how their well-known properties are changed by the deformation. Sections 12.4-12.5 are dedicated to the study of the phase structure at large N, where we find that the expected phase transition extends to $p < 2$ as a consequence of the $T\overline{T}$-deformation. Two appendices at the end of the chapter contain some technical details that supplement the analyses of the main text.

Conventions. To avoid excessive repetition of the word 'deformation', we will only explicitly state it when using the terminology '$T\overline{T}$-deformation'. The q-deformed Yang–Mills theory and its refinement, (q,t)-deformed Yang–Mills theory, will be henceforth simply called 'q-Yang–Mills theory' and '(q,t)-Yang–Mills theory', respectively.

12.2 $T\overline{T}$-deformation of almost topological gauge theories

Consider the partition function of a gauge theory $\mathscr{T}$ with compact connected gauge group G on a Riemann surface Σ which is described by the insertion of a non-local operator $\mathcal{O}(\Phi)$ in the path integral of a two-dimensional topological quantum field theory:

$$\mathcal{Z}_{\mathscr{T}}[\Sigma] = \int \mathscr{D}\Phi \; e^{-S_{\text{TQFT}}(\Phi)} \, \mathcal{O}(\Phi) =: \big\langle \mathcal{O}(\Phi) \big\rangle_{\text{TQFT}} \, .$$

Here Φ collectively denotes the fields of the theory and $\mathscr{D}\Phi$ is a gauge-invariant measure on the space of fields, while the action $S_{\text{TQFT}}(\Phi)$ defines a topological field theory. Notable examples of such theories are two-dimensional Yang–Mills theory and its relatives, which arise from a BF-type topological gauge theory through a deformation that is precisely of this type, as we will review in Section 12.3. These theories will be the focus of subsequent sections. Nevertheless, one may also consider deformations of the topologically twisted sigma-models of [476, 477] or of other classes of two-dimensional topological field theories [478] by some non-local operator, and our considerations in this section also pertain to these more general gauge theories.

The spacetime, on which our field theory is defined, is a Riemann surface Σ, possibly with s marked points decorated with representations $R_1, \dots, R_s$ of the gauge group G, in which case the partition function is denoted by

$$\mathcal{Z}_{\mathscr{T}}[\Sigma; R_1, \dots, R_s] \, .$$

The surface Σ is allowed to have boundaries, and the partition function will be understood as a function of suitable boundary conditions, which in particular include the holonomies of the gauge connection along the one-dimensional boundaries. Field theories without gauge symmetries can be considered as well in this framework as a special case with trivial gauge group.

Theories of this class are amenable to the $T\overline{T}$-deformation, albeit defined on a curved spacetime Σ, thanks to their "almost" topological nature. According to [442, 443] (see also [479]), $T\overline{T}$-deformation is equivalent to coupling the field theory to two-dimensional topological gravity. If the theory we start with is topological, the gravitational sector of the path integral can be integrated out with no effect. However, if non-local operators have been inserted, they couple to the gravitational sector and the $T\overline{T}$-deformation is represented symbolically as

$$\mathcal{Z}_{\mathscr{T}}[\Sigma] \quad \xrightarrow{T\overline{T}\text{-deformation}} \quad \mathcal{Z}^{T\overline{T}}_{\mathscr{T}}[\Sigma]$$

with

$$\mathcal{Z}^{T\overline{T}}_{\mathscr{T}}[\Sigma] = \left\langle \frac{1}{\mathcal{Z}_{\text{JT}}} \int \mathscr{D}\mathbf{e} \; \mathcal{O}(\Phi;\mathbf{e}) \; \exp\Big(\tfrac{1}{2\tau} \int_\Sigma (\mathbf{e}-\mathbf{f}) \wedge (\mathbf{e}-\mathbf{f}) \Big) \right\rangle_{\text{TQFT}} \, .$$

This is the path integral of JT gravity, normalized by the pure gravity partition function $\mathcal{Z}_{\rm JT}$. The integration is over the coframe field $\mathbf{e}$ of the target space, with $\mathbf{f}$ the coframe field of the worldsheet Σ, and the path integral measure is induced by the metric

$$\delta s^2 = \int_\Sigma \delta\mathbf{e} \wedge \delta\mathbf{e} \ .$$

The notation $\mathcal{O}(\Phi;\mathbf{e})$ means that the non-local operator has the same form as before, but now lives in the manifold with coframe field $\mathbf{e}$. For derivative-free operators $\mathcal{O}(\Phi)$ this simply means that, in every integral, we have to replace the original volume form ω on Σ, written in terms of the coframe field of Σ as $\mathbf{f}\wedge\mathbf{f}$, by the target space volume form $\mathbf{e}\wedge\mathbf{e}$. This presentation is equivalent to the change of variables described in [444], but for the purposes of the present work we find it convenient to use the explicit path integral presentation.

The proof of equivalence with the $T\overline{T}$-deformation presented in [443] relies on showing that the gravitational path integral is one-loop exact, and reproduces the $T\overline{T}$-deformed partition function. This ties in nicely with the arguments of [442] that the gravitational dressing provided by the $T\overline{T}$-deformation is a semi-classical effect. We shall see this explicitly for the class of non-local operators that we ultimately consider below.

Remark. We always consider the surface Σ to be equipped with a Riemannian metric, and therefore use the Euclidean gravity path integral, following the conventions of [479] (which agree with those of [474, 444]). The deformation parameter τ in this thesis then differs by a sign from the conventions of [443, 480] which work in Lorentzian signature.

Remark. It was noted in [475] (see also [480, App. A] for relevant discussion) that there is a subtlety in the normalization of the path integral measure $\mathscr{D}\mathbf{e}$: as will be manifest below, assuming the naive normalization of the measure and performing the gravity path integral, one does not recover the undeformed theory in the limit $\tau \to 0$. Imposing the latter condition instead leads to a choice of normalization for the path integral measure $\mathscr{D}\mathbf{e}$ which depends on $\mathcal{O}(\Phi)$. In particular, the order of the path integrals do not commute, and the topological gravity degrees of freedom should always be introduced inside the correlator $\langle\,\cdot\,\rangle_{\rm TQFT}$. Below we will provide a more extensive comparison between the present analysis and that of [475], and this technical aspect will play an important role.

With the application to two-dimensional Yang–Mills theory along with its generalizations and deformations in mind, we now specialize our analysis to the case in which the functional dependence of $\mathcal{O}(\Phi)$ is through operators of the form

$$\mathcal{O}(\Phi) = \exp\left(-\frac{\lambda}{2N}\int_\Sigma V(\phi,\psi)\ \omega\right) \ ,$$

where ω is the normalized volume form on Σ and the potential $V(\phi,\psi)$ is a scalar functional of scalar fields ϕ and possibly spinor fields ψ. The coupling is $\frac{\lambda}{N}$, where λ is a 't Hooft parameter and N is the rank of the gauge group G. When coupled to topological gravity, this operator becomes

$$\mathcal{O}(\Phi;\mathbf{e}) = \exp\left(-\frac{\lambda}{2N}\int_\Sigma V(\phi,\psi)\ \mathbf{e}\wedge\mathbf{e}\right) \ ,$$

and the $T\overline{T}$-deformed partition function reads

$$\mathcal{Z}_{\mathscr{T}}^{T\overline{T}}[\Sigma] = \int \mathscr{D}\Phi\ \frac{1}{\mathcal{Z}_{\rm JT}}\int \mathscr{D}\mathbf{e}\ \exp\Big(-S_{\rm TQFT}(\Phi) - \lambda\int_\Sigma\big(\tfrac{1}{2N}V(\phi,\psi)\ \mathbf{e}\wedge\mathbf{e} - \tfrac{1}{2\tau}\,(\mathbf{e}-\mathbf{f})\wedge(\mathbf{e}-\mathbf{f})\big)\Big) \, .$$

We have chosen a non-standard definition of the parameter τ, including an overall factor λ, which will be convenient in the forthcoming discussion. After simple manipulation and integrating over

$\mathbf{e}' = \mathbf{e} - \mathbf{f}$, one obtains

$$\begin{aligned}\mathcal{Z}_{\mathscr{T}}^{T\overline{T}}[\Sigma] &= \int \mathscr{D}\Phi \ \exp\Big(-S_{\text{TQFT}}(\Phi) - \tfrac{\lambda}{2}\int_\Sigma N^{-1}\,V(\phi,\psi)\ \mathbf{f}\wedge\mathbf{f} \\ &\qquad + \tfrac{\lambda}{2}\int_\Sigma \tfrac{\tau\,N^{-1}}{1-\tau\,N^{-1}\,V(\phi,\psi)}\big(N^{-1}\,V(\phi,\psi)\,\mathbf{f}\big)\wedge\big(N^{-1}\,V(\phi,\psi)\,\mathbf{f}\big)\Big) \\ &= \int \mathscr{D}\Phi \ \exp\Big(-S_{\text{TQFT}}(\Phi) - \tfrac{\lambda}{2N}\int_\Sigma \tfrac{V(\phi,\psi)}{1-\frac{\tau}{N}\,V(\phi,\psi)}\,\omega\Big)\ , \end{aligned} \tag{12.2.1}$$

which correctly reproduces the prescription of [474] for the $T\overline{T}$-deformation given by

$$\frac{\lambda}{2N}\,V(\phi,\psi) \longmapsto \frac{\frac{\lambda}{2N}\,V(\phi,\psi)}{1-\frac{\tau}{N}\,V(\phi,\psi)}\ . \tag{12.2.2}$$

In (12.2.1) we use a $V(\phi,\psi)$-dependent normalization of the gravity path integral that cancels a factor from the Gaussian integration. Had we not done so, we would not recover the undeformed partition function in the limit $\tau \to 0$. This is the incarnation of the subtleties in the normalization of the path integral measure discussed in [475, 480].

$T\overline{T}$-deformation in curved spacetime

The $T\overline{T}$-deformation is only well-defined in flat space. However, in the present setting, we argue that, since the underlying theory is topological, for suitable insertions $\mathcal{O}(\Phi)$ we can define the $T\overline{T}$-deformation on flat space, and then put the theory on a curved manifold Σ. We now explain this point more rigorously.

Thanks to the cutting and gluing property of topological quantum field theories [481, 482], one can decompose Σ into disks, cylinders and pairs of pants, obtaining the same theory on each piece (see Figure 12.1). Such components have boundaries, and one should impose suitable boundary conditions on the fields. The disk, the cylinder and the pair of pants are homeomorphic respectively to the complex plane $\mathbb{C}$, the punctured plane $\mathbb{C}^\times$ and the doubly-punctured plane $\mathbb{C}^{\times\times}$. Therefore, we reduce the topological quantum field theory on flat components, which are many copies of the complex plane $\mathbb{C}$, with either zero, one or two holes. At this point, we insert $\mathcal{O}(\Phi)$ on each component, and turn on gravity. Since each component is flat, the $T\overline{T}$-deformation prescription is well-defined on each component.

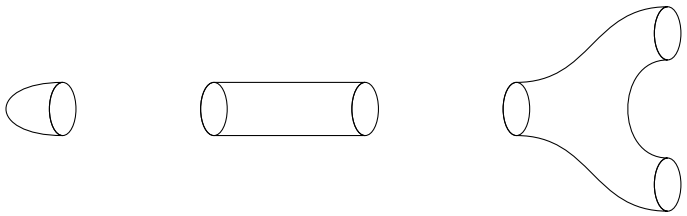


Figure 12.1. The disk (left), the cylinder (center) and the pair of pants (right), homeomorphic to the complex plane with respectively zero, one or two holes.

After performing the JT gravity path integral, we can glue back together the pieces and reassemble Σ (see Figure 12.2). Topological gravity couples to bulk geometry and does not change the boundary data, at least for operator insertions $\mathcal{O}(\Phi)$ of the form in (12.2.1) (we will briefly comment on the most general case shortly). Hence the gluing goes exactly as without $T\overline{T}$-deformation, and we obtain a $T\overline{T}$-deformed theory on Σ.

At this point it is worthwhile mentioning the proposal [480] that $T\overline{T}$-deformation in curved spacetime corresponds to massive gravity. In the present setting, we could equivalently take the

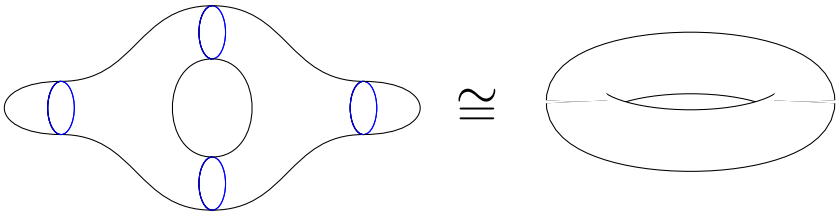


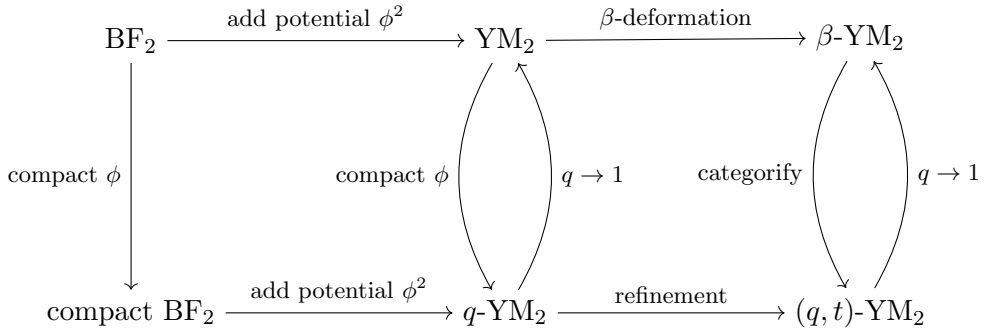
Figure 12.2. Obtaining a torus from elementary pieces. On the left, the gluing is an integration over boundary conditions (in blue).

(Euclidean version of the) proposal of [480] as the definition of $T\overline{T}$-deformation on curved two-dimensional manifolds. A step towards a rigorous definition of generic $T\overline{T}$-deformed theories on curved manifolds has also been taken in [483].

12.3 Two-dimensional Yang–Mills and q-Yang–Mills theories

In this chapter, as in the other ones in Part III, we study two-dimensional Yang–Mills theory, its q-deformation and its subsequent refinement to two-dimensional (q,t)-deformed Yang–Mills theory. These are examples of almost topological gauge theories of the type discussed in Section 12.2, where the underlying topological field theory is two-dimensional BF theory (sometimes also referred to as two-dimensional topological Yang–Mills theory) whose fields are a scalar field ϕ on Σ in the adjoint representation of the gauge algebra $\mathfrak{g}$ and the curvature F^A of a gauge connection A on (a trivial principal G-bundle over) Σ. Ordinary Yang–Mills theory on Σ corresponds to a deformation of this BF theory by a non-local operator $\mathcal{O}(\phi)$ which adds a potential $V(\phi) = \mathrm{Tr}\,\phi^2$ to the BF action. This theory can be β-deformed by modifying the discrete matrix model which arises for $\beta = 2$ to a general β-ensemble. One can further deform the underlying BF theory by making the field ϕ compact, that is, taking it to be valued in the adjoint representation of the gauge group G. Adding the potential $V(\phi)$ deforms this theory to q-Yang–Mills theory which can be subsequently refined to (q,t)-Yang–Mills theory, that is a categorification of the β-ensemble. The initial theories, with non-compact ϕ, can then be regarded as classical limits $q \to 1$ of the theories with compact scalar ϕ.

We depict these relationships between the various incarnations of Yang–Mills theory on Σ through the diagram



In this section we use the formalism developed in Section 12.2 to study the $T\overline{T}$-deformation of the Yang–Mills theories appearing in this diagram.

12.3.1 $T\overline{T}$-deformation of two-dimensional Yang–Mills theory

In [474] the $T\overline{T}$-deformation of two-dimensional Yang–Mills theory on Σ was obtained, through explicit solution of the flow equation

$$\frac{\partial\mathcal{L}(\tau)}{\partial\tau} = \det_{\mu,\nu=1,2}\big[T_{\mu\nu}(\tau)\big] \ ,$$

to all orders in $\tau \in [0,\infty)$. This equation is to be solved with the initial condition on the deformed Lagrangian $\mathcal{L}(\tau)$ that requires $\mathcal{L}(0) = \mathcal{L}_{\rm YM}$ to be the Yang–Mills Lagrangian, and $T_{\mu\nu}(\tau)$ is the Hilbert energy-momentum tensor of the two-dimensional field theory. The same deformation has recently been obtained in [475] through the coupling with JT gravity. We rederive the result by exploiting the equivalent first order formulation as a deformation of BF theory. The argument is as follows: Yang–Mills theory is a pure gauge theory, but it is equivalent to a BF theory with additional Gaussian term for the scalar $\phi \in \Omega^0(\Sigma,\mathfrak{g})$ given by (see [423])

$$S_{\rm YM} = \frac{N}{2\lambda}\int_\Sigma \mathrm{Tr}\ F^A * F^A = \int_\Sigma \mathrm{Tr}\left(\mathrm{i}\,\phi\,F^A + \frac{\lambda}{2N}\,\phi^2\,\omega\right) \ ,$$

where the equality is understood to hold on-shell. Here $F^A \in \Omega^2(\Sigma,\mathfrak{g})$ is the curvature of a gauge connection A on Σ, Tr is an invariant quadratic form on the Lie algebra $\mathfrak{g}$, ω is the symplectic structure on Σ and $*$ is the Hodge operator constructed from the Riemannian metric compatible with ω. The Yang–Mills coupling is

$$g^2_{\rm YM} = \lambda\,N^{-1} \ .$$

The first term is the action of two-dimensional BF theory which is topological, thus the $T\overline{T}$-deformation of the first order formulation only changes the potential $V(\phi) = \mathrm{Tr}\ \phi^2$ as in (12.2.2). From this point, the derivation of the heat kernel expansion using Abelianization of the path integrals goes exactly as in [423]: one conjugates the scalar field ϕ into a Cartan subalgebra of $\mathfrak{g}$ using gauge invariance and the Weyl integral formula, and then integrates over the root components A_α of the gauge connections with respect to the root space decomposition of the Lie algebra $\mathfrak{g}$. Then two-dimensional Yang–Mills theory can be $T\overline{T}$-deformed by replacing the quadratic Casimir of representations R of G according to[36]

$$C_2(R) \longmapsto C_2^{T\overline{T}}(R,\tau) := \frac{C_2(R)}{1-\frac{\tau}{N^3}\,C_2(R)} \ . \tag{12.3.1}$$

This derivation immediately extends to the generalized Yang–Mills theory of [484], where higher order Casimir operators are included by adding higher degree terms to the potential $V(\phi)$. These can include multi-trace terms, since the derivation does not rely on the explicit form of $V(\phi)$. The $T\overline{T}$-deformation of the generalized two-dimensional Yang–Mills theory is then directly obtained from (12.2.2). The final answer for the partition function of generalized Yang–Mills theory is then

$$\mathcal{Z}^{T\overline{T}}_{\text{gen-YM}}[\Sigma] = \sum_R \dim(R)^{\chi(\Sigma)}\,\exp\Big(-\frac{\lambda}{2N}\,\frac{C_{\rm gen}(R)}{1-\frac{\tau}{N^3}\,C_{\rm gen}(R)}\Big) \ , \tag{12.3.2}$$

where $C_{\rm gen}(R)$ includes the quadratic and higher Casimir operators. The sum runs over isomorphism classes of irreducible representations R of G [384] with dimension $\dim(R)$, the coupling λ is identified with the area A of the surface Σ, and $\chi(\Sigma)$ is the Euler characteristic of Σ.

[36] We are slightly changing the normalization of the topological gravity action, $\frac{1}{\tau} \mapsto \frac{N^2}{\tau}$, to make the right-hand side well-defined at all τ for every N. That this is the correct normalization in general follows from the flow equations in [474].

Comparison with the literature

Since the partition function of $T\overline{T}$-deformed two-dimensional Yang–Mills theory has been derived in different ways in the literature [474, 475, 485], it is appropriate to now pause and discuss our result.

Formula (12.3.2), or more precisely its original version with $C_{\rm gen}(R) = C_2(R)$, was first proposed in [474], although it was not rigorously justified for curved surfaces Σ. The proposal of [474] was also the starting point of previous work [9] studying the phase structure of the $T\overline{T}$-deformed theory. Here we have provided the derivation, following an argument similar to that of [475] but with a few important differences.

In [475] topological gravity is introduced after integrating out the gauge fields. In particular, as carefully explained there, the JT gravity path integral is representation-dependent and is inserted inside the sum over irreducible representations. Schematically

$$\sum_R Z_R(\omega) \xrightarrow{T\overline{T}\text{-deformation of [475]}} \sum_R \int \mathscr{D}_R\mathbf{e}\; Z_R(\mathbf{e}\wedge\mathbf{e}) \ ,$$

where $Z_R(\omega)$ is the summand in (12.3.2), and we have stressed its dependence on the volume form ω of Σ. The normalization of the measure $\mathscr{D}_R\mathbf{e}$ is taken to be R-dependent.

Therefore, the procedure of [475] does not deform the original path integral, but deforms each summand in the expression obtained after Abelianization [423]. The technique we adopted, instead, describes a deformation of the full path integral, and proves that the Abelianization takes place also in the $T\overline{T}$-deformed theory. The two results coincide, as expected. Indeed the gauge fields do not enter in the definition of the operator $\mathcal{O}(\phi)$, which couples to gravity. For this reason, the integration over the coframe field is expected to commute with the integration over the gauge fields.

12.3.2 $T\overline{T}$-deformed q-Yang–Mills theory

We can extend the argument above to q-deformed Yang–Mills theory: this deformation modifies the domain of integration, making the scalar field ϕ compact, i.e. taking $\phi \in \Omega^0(\Sigma, G)$ to be valued in the Lie group G instead of its Lie algebra $\mathfrak{g}$, without altering the action [408, 136]. In this case Abelianization proceeds by conjugating ϕ into the maximal torus of G. The $T\overline{T}$-deformation thus changes the potential for the (now compact) scalar, exactly as in the case of ordinary two-dimensional Yang–Mills theory. The final answer for the $T\overline{T}$-deformed partition function of q-Yang–Mills theory on a surface Σ of genus g_Σ with s boundaries is

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}[\Sigma; g_1, \dots, g_s] = \sum_R \dim_q(R)^{\chi(\Sigma)}\, q^{\frac{p}{2}\, C_2^{T\overline{T}}(R,\tau)}\ \chi_R(g_1)\cdots\chi_R(g_s) \ , \tag{12.3.3}$$

with the identification of the q-parameter $q = e^{-\lambda}$. Here $p \in \mathbb{Z}$ is a discrete parameter, the $T\overline{T}$-deformed Casimir is defined in (12.3.1), and

$$\chi(\Sigma) = 2 - 2g_\Sigma - s$$

is the Euler characteristic of Σ. The boundary conditions $g_1, \dots, g_s \in G$ are the holonomies of the gauge connection around the boundaries, with characters χ_R in the representation R, and $\dim_q(R)$ is the quantum dimension of R. For closed surfaces Σ, the formula (12.3.3) is simply

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}[\Sigma] = \sum_R \dim_q(R)^{2-2g_\Sigma}\, q^{\frac{p}{2}\, C_2^{T\overline{T}}(R,\tau)} \ .$$

Again, the argument straightforwardly extends to generalized q-deformed Yang–Mills theory, with additional higher degree terms added to the potential $V(\phi)$.

The partition function of the ordinary $T\overline{T}$-deformed Yang–Mills theory from Section 12.3.1 above is recovered by taking the limit

$$p \to \infty \qquad \text{and} \qquad \lambda \to 0 \qquad \text{with} \quad \lambda\, p = A \text{ fixed} \ , \tag{12.3.4}$$

where A is the area of Σ.

q-Yang–Mills theory on the disk and on the cylinder

As we have shown, the procedure of $T\overline{T}$-deformation works for every Riemann surface Σ, possibly with boundary. Special roles are played by the disk and cylinder partition functions. On the disk we have

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\Big[\ \bigcirc\!\!\!\!\!\;\big) \ ;g\Big] = \sum_R \dim_q(R)\ q^{\frac{p}{2}\,C_2^{T\overline{T}}(R,\tau)}\ \chi_R(g)\ .$$

Gluing two disks whose boundaries have opposite orientations and using the orthogonality of the characters we get the $T\overline{T}$-deformed partition function on the sphere $\mathbb{S}^2$:

$$\int_G \mathrm{d}g\ \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\Big[\ \bigcirc\!\!\!\!\!\;\big) \ ;g\Big]\ \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\Big[\big(\!\!\!\!\!\;\bigcirc \ ;g^{-1}\Big] = \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\Big[\ \bigcirc\!\!\!\!\!\!\;\big(\big) \ \Big] = \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\big[\mathbb{S}^2\big]\ ,$$

where $\mathrm{d}g$ is the invariant Haar measure on G.

The cylinder partition function is

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\Big[\big(\!\!\!\!\;\boxed{}\!\!\!\;\big) \ ;g_{\mathrm{in}},g_{\mathrm{out}}\Big] = \sum_R q^{\frac{p}{2}\,C_2^{T\overline{T}}(R,\tau)}\ \chi_R(g_{\mathrm{in}}^{-1})\,\chi_R(g_{\mathrm{out}})\ ,$$

where we have already taken into account the orientation in the definition of the boundary condition g_{in}. In the topological limit $\lambda = 0$, it serves as a propagator: attaching it to any surface Σ replaces the holonomy g_{in} by g_{out}. At non-zero area though, attaching a cylinder has a notable effect which effectively increases the coupling. With our choice of normalization for τ, the effect of gluing a cylinder to Σ is precisely the same as in the theory without $T\overline{T}$-deformation.

Supersymmetry

An additional consistency check for our formulas comes from the minimal supersymmetric extension of Yang–Mills theory. Two-dimensional Yang–Mills theory and its q-deformation are equivalent to their supersymmetric counterparts. The BRST multiplet is (A,ϕ,ψ), with ψ a Grassmann-odd one-form on Σ with values in the Lie algebra $\mathfrak{g}$, and the action is schematically modified as

$$S_{\mathrm{YM}} \longmapsto S_{\mathrm{YM}} + \int_\Sigma \mathrm{Tr}(\psi\wedge\psi)\ .$$

The equivalence is straightforwardly checked by integrating out ψ. On the other hand, the new term is topological and hence is insensitive to the $T\overline{T}$-deformation. We can thus first $T\overline{T}$-deform and integrate out ψ afterwards, obtaining again the result (12.3.3). So regardless of the route followed, the $T\overline{T}$-deformation of two-dimensional Yang–Mills theory and its generalizations always provides the same answer with or without supersymmetry.

Refinement

Let us now consider the refinement of q-deformed Yang–Mills theory [486], also known as (q,t)-deformed Yang–Mills theory. The refinement leaves the action unchanged but modifies the path integral measure [486]. Therefore we can $T\overline{T}$-deform the theory and the Abelianization technique continues to work, hence the $T\overline{T}$-deformation modifies the partition function of (q,t)-Yang–Mills theory only in the Gaussian potential, according to (12.2.2). We will give more details later on in Section 12.5.5.

θ-angle

In q-Yang–Mills theory, the θ-angle term is introduced in the path integral as a linear term in the potential $V(\phi)$, and it descends from a chemical potential for D2-branes in the construction of [487, 408]. Therefore it will also couple to the gravitational path integral after $T\overline{T}$-deformation, and will enter in the denominator of the deformed potential through

$$-\frac{\lambda}{2N}\,C_2(R)+\mathrm{i}\,\theta\,C_1(R)\longmapsto\frac{-\frac{\lambda}{2N}\,C_2(R)+\mathrm{i}\,\theta\,C_1(R)}{1-\frac{\tau}{N^3}\left[C_2(R)-\frac{2\,\mathrm{i}\,\theta\,N}{\lambda}\,C_1(R)\right]}\ ,$$

where C_1 is the first Casimir of G, which is non-zero only for non-simply connected gauge groups.

12.3.3 Wilson loops, marked points and q a root of unity

One may also include Wilson loop operators in an irreducible representation R of G along a closed curve $\mathcal{C}$ on Σ:

$$W_R(\mathcal{C})=\mathrm{Tr}_R\,\mathcal{P}\exp\oint_{\mathcal{C}}A\ .$$

We assume for simplicity that $\mathcal{C}$ does not wind around any handle of Σ.

The expectation value of a collection of s Wilson loops in $T\overline{T}$-deformed two-dimensional Yang–Mills theory, both the ordinary and q-deformed versions, is computed as follows. Cut Σ along the s cycles $\mathcal{C}_1,\dots,\mathcal{C}_s$, obtaining $s+1$ components: s of them have disk topology and the last is the remainder of Euler characteristic $\chi(\Sigma)-s$. The next step is to compute the $T\overline{T}$-deformed partition function on each component, which is a wavefunction of the holonomies along the boundaries $\mathcal{C}_1,\dots,\mathcal{C}_s$. Then glue the components pairwise back together by integrating over G. The example of $\mathbb{S}^2$ with three loops is depicted in Figure 12.3.

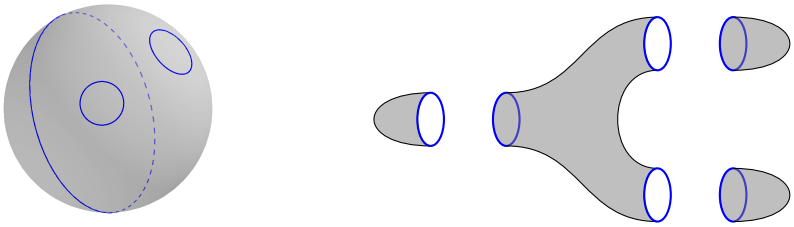


Figure 12.3. A sphere with three Wilson loops is cut into three disks plus a remaining pair of pants.

In this way we find the normalized expectation value

$$\begin{aligned}\big\langle W_{R_1}(\mathcal{C}_1)\cdots W_{R_s}(\mathcal{C}_s)\big\rangle=\frac{1}{\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}[\Sigma]}\sum_R\dim_q(R)^{\chi(\Sigma)-s}\,q^{\frac{p-A_W}{2}\,C_2^{T\overline{T}}(R,\tau)}\\ \times\prod_{i=1}^{s}\sum_{\widetilde{R}_i}\dim_q(\widetilde{R}_i)\,q^{\frac{a_i}{2}\,C_2^{T\overline{T}}(\widetilde{R}_i,\tau)}\,N^{\widetilde{R}_i}{}_{R_iR}\ ,\end{aligned}$$

where a_i is the area enclosed by the loop $\mathcal{C}_i$, R_i is the representation label of the i^{th} loop, and $\widetilde{R}_i$ is a summation variable denoting an irreducible representation associated to the quantization on the i^{th} component. Geometrically, R is associated to the remainder, R_i to the i^{th} cut and $\widetilde{R}_i$ to the i^{th} disk. We have also denoted

$$A_W = \sum_{i=1}^{s} a_i \ ,$$

and we assume $A_W < p$ to ensure convergence of the first series. The quantities $N^{R_1}{}_{R_2 R_3}$ are fusion coefficients obtained from the integration over the holonomies; for unitary gauge group they are the Littlewood-Richardson coefficients. This formula differs from the original theory simply in the replacement of the Casimir as in (12.3.1).

In the limit in which the loops shrink to a point, $a_i \to 0$, we obtain (after dropping the normalization $\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}[\Sigma]^{-1}$) the partition function of $T\overline{T}$-deformed q-Yang–Mills theory on a surface Σ with s marked points decorated with irreducible representations $R_1, \dots, R_s$:

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}[\Sigma; R_1, \dots, R_s] = \sum_{R} \dim_q(R)^{\chi(\Sigma)}\, q^{\frac{p}{2} C_2^{T\overline{T}}(R,\tau)} \prod_{i=1}^{s} \sum_{\widetilde{R}_i} \frac{\dim_q(\widetilde{R}_i)}{\dim_q(R)}\, N^{\widetilde{R}_i}{}_{R_i R} \ .$$

Analogous formulas hold for the $T\overline{T}$-deformation of ordinary Yang–Mills theory.

Lost connection with Chern–Simons theory

For $\tau = 0$ and $0 < |q| < 1$ (with possibly $q \in \mathbb{C}$), when $|q| \to 1$ at roots of unity, the sum over representations terminates for gauge group $G = U(N)$ [89]. After $T\overline{T}$-deformation, the quadratic Casimir part is modified and the cancellations that truncate the series no longer take place.

The usual connection between q-Yang–Mills theory and Chern–Simons theory thus no longer holds. For the same reason, even when q is a root of unity, one cannot understand Wilson loops in the $T\overline{T}$-deformed version in terms of observables in Chern–Simons theory, or some deformation thereof, living in the total space of a degree p circle bundle over Σ.

Furthermore, when looking for the modular matrices S and T of $PSL(2,\mathbb{Z})$ in expressions such as (12.3.3), we recall that matrix elements like

$$\mathsf{S}_{R\widetilde{R}} \qquad \text{and} \qquad \mathsf{T}_{R\widetilde{R}}$$

are defined for integrable representations R and $\widetilde{R}$, while in our case the sum runs over all the irreducible representations. In this sense the $T\overline{T}$-deformation spoils the modular properties of the theory at q a root of unity, as could have been foreseen from the explicit form of (12.3.1).[37]

12.3.4 Breakdown of factorization

We have seen that the usual connection with Chern–Simons theory is lost as soon as the $T\overline{T}$-deformation is turned on. In the following we discuss another well-known central feature of two-dimensional Yang–Mills theory that does not hold after $T\overline{T}$-deformation: the factorization of the partition function $\mathcal{Z}_{q\text{-YM}}$ into chiral and anti-chiral sectors [487, 408, 409, 489]. This strongly suggests that the usual large N string theory picture of two-dimensional Yang–Mills theory breaks down after $T\overline{T}$-deformation.

[37] Gauging the $T\overline{T}$-deformed WZW model of [456] does not yield a connection with Chern–Simons theory, or some deformation thereof. The relation with Chern–Simons theory is indeed a special property of the conformal fixed point [488].

Let $\mathscr{H}$ be the Hilbert space of states of the theory, and endow it with the basis $\{|R\rangle\}$ in one-to-one correspondence with isomorphism classes of irreducible unitary representations of G [384]. Adopting a common shorthand, we call it the representation basis. The normalization is

$$\langle R|\widetilde{R}\rangle = \dim(R)^{\chi(\Sigma)}\ \delta_{R\widetilde{R}}\ .$$

The factorization property relies on being able to make a replacement

$$q^{\frac{p}{2}\,C_2(R)} \longrightarrow q^{\frac{p}{2}\,C_2(R_+)}\,q^{\frac{p}{2}\,C_2(R_-)}\ ,$$

where R_+ and R_- are known as "chiral" and "anti-chiral" representations, which correspond to states

$$|R_\pm\rangle \in \mathscr{H}_\pm\ ,$$

in the factorized Hilbert space $\mathscr{H}_+ \otimes \mathscr{H}_-$. It is clear from (12.3.1)–(12.3.3) that the factorization breaks down at $\tau \neq 0$.

Quantization of the $T\overline{T}$-deformed theory

Consider the unitary gauge group $G = U(N)$ and the surface Σ as a fibration over $\mathbb{S}^1$, with the circle interpreted as the Euclidean time direction. By definition, the partition function of q-Yang–Mills theory on Σ is given by

$$\mathcal{Z}_{q\text{-YM}}(\lambda) = \mathrm{Tr}_{\mathscr{H}}\, e^{-\frac{\lambda\,p}{2N}\,\widehat{H}_{q\text{-YM}}} = \sum_R \langle R|e^{-\frac{\lambda\,p}{2N}\,\widehat{H}_{q\text{-YM}}}|R\rangle\ ,$$

where $\widehat{H}_{q\text{-YM}}$ is the Hamiltonian, and we have taken the trace over the Hilbert space $\mathscr{H}$ in the representation basis, which diagonalizes $\widehat{H}_{q\text{-YM}}$ with eigenvalues $C_2(R)$. A generic deformation controlled by a parameter τ which triggers an RG flow would produce

$$\mathcal{Z}^{\rm def}_{q\text{-YM}}(\lambda,\tau) = \mathrm{Tr}_{\mathscr{H}(\tau)}\, e^{-\frac{\lambda\,p}{2N}\,\widehat{H}(\tau)} = \sum_{R(\tau)} \langle R(\tau)|e^{-\frac{\lambda\,p}{2N}\,\widehat{H}(\tau)}|R(\tau)\rangle\ ,$$

deforming both the Hamiltonian to $\widehat{H}(\tau)$ and the Hilbert space to $\mathscr{H}(\tau)$. The basis $\{|R(\tau)\rangle\}$ would reduce to the representation basis when sending $\tau \to 0$. Note that, in this general framework, since the Hilbert space changes, one may need to include additional states. However, when the deformation is by the composite operator $T\,\overline{T}$, the explicit form of the deformed eigenvalues of $\widehat{H}(\tau)$ is known, and in particular no new eigenvalues arise. For q-Yang–Mills theory we obtain explicitly

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}(\lambda,\tau) = \sum_{R(0)} \langle R(0)|e^{-\frac{\lambda\,p}{2N}\,\widehat{H}(\tau)}|R(0)\rangle\ ,$$

with the deformed Hamiltonian

$$\widehat{H}(\tau) = \frac{\widehat{H}_{q\text{-YM}}}{1-\frac{\tau}{N^3}\,\widehat{H}_{q\text{-YM}}}$$

diagonalized by the representation basis $\{|R\rangle\} = \{|R(0)\rangle\}$ for all $\tau \geq 0$. Only the eigenvalues $C_2^{T\overline{T}}(R,\tau)$ are different. Therefore, although in general from the knowledge of the eigenvalues one cannot exclude that additional degenerate states arise in the $T\overline{T}$-deformed theory, we see that this is not the case for two-dimensional Yang–Mills theory and its relatives. Indeed, having found explicitly the $T\overline{T}$-deformed partition function, the presence of additional states at $\tau > 0$ should have a null net contribution, but this is not possible from the explicit, strictly positive form of the eigenvalues.

In conclusion, the partition function of $T\overline{T}$-deformed q-Yang–Mills theory, and hence also ordinary Yang–Mills theory through the limit (12.3.4), is the trace of the exponential of the $T\overline{T}$-deformed Hamiltonian found in [474] taken in an undeformed Hilbert space of states.

Free fermion formulation

Let us now focus on ordinary (without q-deformation) two-dimensional Yang–Mills theory for definiteness. While the factorization structure of q-Yang–Mills theory is richer, in the sense that already at finite N one sees a factorization into chiral and anti-chiral building blocks, the breakdown of these properties happens at a fundamental level, which is more clearly seen by looking directly at $T\overline{T}$-deformed ordinary Yang–Mills theory.

It is well-known that the Hilbert space $\mathscr{H}$ factorizes at large N as [490]

$$\mathscr{H} \xrightarrow{N\to\infty} \mathscr{H}_+ \otimes \mathscr{H}_- \ , \tag{12.3.5}$$

with the representation basis factorizing accordingly as

$$|R\rangle \xrightarrow{N\to\infty} |R_+\rangle \otimes |R_-\rangle \ ,$$

with $|R_\pm\rangle \in \mathscr{H}_\pm$. The Hilbert spaces $\mathscr{H}_+$ and $\mathscr{H}_-$ are known as the "chiral" and "anti-chiral" sectors, respectively. From the factorization (12.3.5) one has [490]

$$\begin{aligned}\lim_{N\to\infty} \mathcal{Z}_{\rm YM}(A) &= \left(\sum_{R_+} \langle R_+|e^{-\frac{A}{2N}\,\widehat{H}_{\rm YM}}|R_+\rangle\right)\left(\sum_{R_-} \langle R_-|e^{-\frac{A}{2N}\,\widehat{H}_{\rm YM}}|R_-\rangle\right)\\ &= \left(\mathrm{Tr}_{\mathscr{H}_+}\, e^{-\frac{A}{2N}\,\widehat{H}_{\rm YM}}\right)\left(\mathrm{Tr}_{\mathscr{H}_-}\, e^{-\frac{A}{2N}\,\widehat{H}_{\rm YM}}\right)\end{aligned}$$

where we dropped overall constants. In the $T\overline{T}$-deformed theory, the large N factorization of the Hilbert space (12.3.5) continues to hold according to the discussion above, but the trace can no longer be factorized into a product of traces.

We will now further elucidate this point through the equivalence with a system of N non-relativistic fermions [491, 492] (non-perturbative corrections were studied in [493]). In the mapping of two-dimensional $U(N)$ Yang–Mills theory to a system of N free fermions on $\mathbb{S}^1$ [491, 492], the ground state corresponds to the state where the fermions occupy the N lowest energy levels, as in the left panel of Figure 12.4. In the representation basis, the ground state is described by the trivial representation, while higher-dimensional representations are mapped to excited states, in which fermions have jumped to higher energy levels.

At finite N, excitations above the positive Fermi surface, or below the negative Fermi surface, may arise from any fermion, as depicted in the central and right panels of Figure 12.4. In the large N limit the two Fermi surfaces decouple, and excitations above the positive Fermi level (respectively below the negative Fermi level) correspond to fermions close to that surface, thus with positive (respectively negative) energy, jumping to higher (respectively lower) unoccupied levels.

Jumps of order N sites are exponentially suppressed with N, and can only be seen from a non-perturbative analysis [493]. Therefore the factorization is interpreted as a disentanglement of the Fermi surfaces.

In the $T\overline{T}$-deformed theory the two Fermi surfaces remain entangled even in the large N limit. The crucial difference between the picture of [491] and its $T\overline{T}$-deformed version lays in the interpretation of the Casimir in terms of free fermions. While at $\tau = 0$ it is a confining quadratic potential, this interpretation is lost at $\tau > 0$. Indeed, by expanding the $T\overline{T}$-deformed potential (12.2.2) in a geometric series, we do not obtain a confining potential for fermions, but instead infinitely many non-local interaction terms which introduce long-distance correlations.

A consequence of these additional interactions is that, even at $N \to \infty$, the energy required for a fermion to jump to another level does not only depend on the energy separation between the initial and final state, but it is also a function of the levels occupied by all of the other fermions. From a conformal field theory perspective, this casts doubt on the existence of a string theory dual to $T\overline{T}$-deformed two-dimensional Yang–Mills theory, unless it is a highly exotic one.

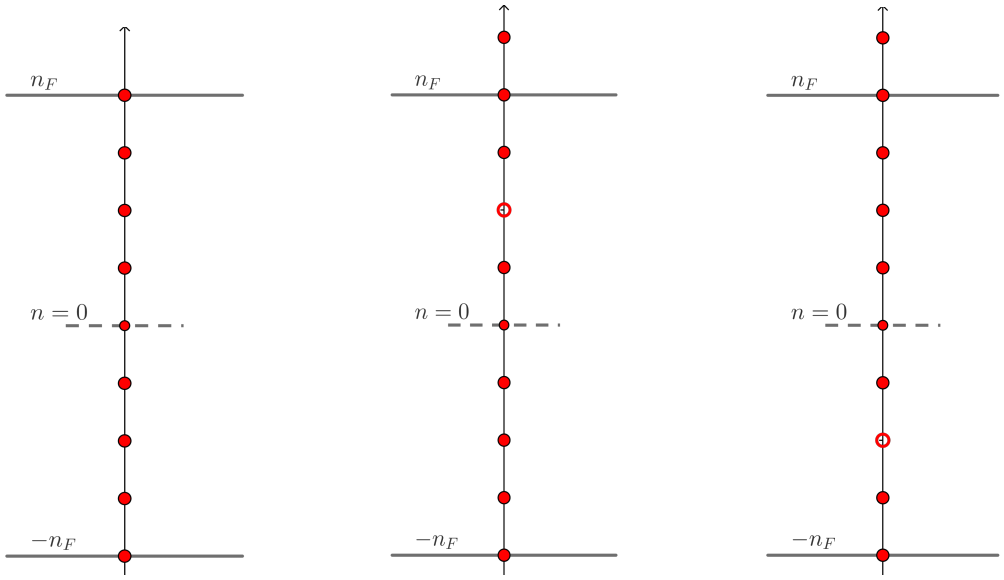


Figure 12.4. Non-relativistic fermions: ground state (left) and two excited states (center, right). The two excited states are excitations over the positive Fermi surface. In the center, a fermion occupying a positive energy level jumps above the positive Fermi surface: this will correspond to a chiral state at large N. On the right, a fermion occupying a negative energy level jumps above the positive Fermi surface: this will be exponentially suppressed at large N.

12.4 Phase transitions in $T\overline{T}$-deformed Yang–Mills theory

In this section, we study the partition function of $T\overline{T}$-deformed $U(N)$ Yang–Mills theory on $\Sigma = \mathbb{S}^2$:

$$\mathcal{Z}^{T\overline{T}}_{\mathbb{S}^2}(A,\tau) = \sum_R (\dim R)^2 \exp\left(-\frac{A}{2N}\left(\frac{C_2(R)}{1-\frac{\tau}{N^3}C_2(R)}\right)\right) \tag{12.4.1}$$

where, as customary, A is the product of the area and the 't Hooft gauge coupling. The sum runs over isomorphism classes of irreducible $U(N)$ representations R, which are in one-to-one correspondence with N-tuples R such that

$$+\infty > R_1 \geq R_2 \geq \cdots \geq R_N > -\infty \ . \tag{12.4.2}$$

It is useful to change variables

$$h_i = -R_i + i - \frac{N+1}{2} \qquad \forall i = 1, \ldots, N \ . \tag{12.4.3}$$

In these variables the partition function of $T\overline{T}$-deformed $U(N)$ Yang–Mills theory reads

$$\mathcal{Z}^{T\overline{T}}_{\mathbb{S}^2}(A,\tau) = \frac{1}{N!\,G(N+1)^\chi} \sum_{\vec{h}\in\mathbb{Z}^N} \Delta(\vec{h})^\chi \ \exp\left\{-\frac{\frac{A}{2N}\left[\sum_{j=1}^N h_j^2 - \frac{N(N^2+1)}{12}\right]}{1-\frac{\tau}{N^3}\left[\sum_{j=1}^N h_j^2 - \frac{N(N^2+1)}{12}\right]}\right\} \ , \tag{12.4.4}$$

where we used the symmetry of the sum to lift the restriction (12.4.2) to the principal Weyl chamber, letting the sum run over unordered $\vec{h} = (h_1, \ldots, h_N) \in \mathbb{Z}^N$. The shift proportional to

$-\frac{1}{12}$ in the Casimir would simply give an overall factor at $\tau = 0$, but it becomes relevant at $\tau > 0$. Here G is the Barnes G-function, which for integer argument can be written as

$$G(N+1) = \prod_{j=1}^{N-1} j! \tag{12.4.5}$$

and $\Delta(\vec{h})$ is the Vandermonde determinant

$$\Delta(\vec{h}) = \prod_{1 \le i < j \le N} (h_i - h_j) \ . \tag{12.4.6}$$

Following the standard procedure for the large N analysis [193], we introduce the variables

$$x := \frac{i}{N}, \quad r(x) := \frac{r_i}{N}, \quad h(x) := -r(x) + x - \frac{1}{2},$$

and, sending $N \to \infty$, replace $N^{-1}\sum_{i=1}^{N}$ with $\int_0^1 dx$. Sending $N \to \infty$ the leading contribution to the partition function (12.4.4) comes from the saddle point configuration that minimizes the effective action

$$S_{\text{eff}}[h] = -\int_0^1 dx \int_0^1 dy \log|h(y) - h(x)| + \frac{A}{2}\sum_{j=0}^{\infty} \tau^j \left[\int_0^1 dx h(x)^2 - \frac{1}{12}\right]^{j+1}.$$

where we have expanded the $T\overline{T}$-deformed Casimir in a geometric series, which is allowed as long as the $T\overline{T}$-deformed theory is well-defined.

At this point, we introduce the eigenvalue density ρ, as usual, according to:

$$\rho(h)dh = dx,$$

which is normalized:

$$\int dh \rho(h) = 1. \tag{12.4.7}$$

Moreover, the discreteness of the eigenvalues imposes the additional constraint

$$\rho(h) \le 1. \tag{12.4.8}$$

The action functional (12.4) becomes:

$$S[\rho] = -\int du \rho(u) \int dv \rho(v) \log|u - v| + \frac{A}{2}\sum_{j=0}^{\infty} \tau^j \left[\int du \rho(u) u^2 - \frac{1}{12}\right]^{j+1}.$$

We are then led to look for a distribution $\rho(h)$ that satisfies the singular integral equation

$$\mathrm{P}\int du \frac{\rho(u)}{h-u} = \frac{A}{2} h \sum_{j=0}^{\infty} (j+1)\tau^j \left[\int du \rho(u) u^2 - \frac{1}{12}\right]^j = 0, \tag{12.4.9}$$

where the symbol $\mathrm{P}\int$ means the principal value of the integral.

It is hard to face the saddle point equation (12.4.9) analytically, due to the appearance of ρ on the right-hand side. Nevertheless, the $T\overline{T}$-deformation only introduced powers of the second moment of the eigenvalue distribution, and the dependence on h remains factorized. This allows for a perturbative solution in τ, and we will solve equation (12.4.9) to all orders.

12.4.1 Perturbative solution

At zeroth order in τ, the theory obviously reduces to pure Yang–Mills and equation (12.4.9) describes the Douglas–Kazakov distribution [193]. Indeed, for $j = 0$ the equation reduces to the saddle point equation of a Gaussian matrix model:

$$\mathrm{P}\int du \frac{\rho(u)}{h-u} = \frac{A}{2}h, \tag{12.4.10}$$

which is solved by the celebrated Wigner semicircle distribution

$$\rho(h) = \frac{A}{2\pi}\sqrt{\frac{4}{A} - h^2}, \qquad \mathrm{supp}\rho = \left[-\frac{2}{\sqrt{A}}, \frac{2}{\sqrt{A}}\right]. \tag{12.4.11}$$

However, the solution must satisfy the constraint (12.4.8) $\rho \leq 1$, meaning that the present one-cut solution only holds up to $A_{cr}^{(0)} = \pi^2$. For the moment, we focus on the perturbative analysis in the small coupling phase $A < A_{cr}$, and discuss the strong coupling phase $A > A_{cr}$ in the next subsection.

The second moment of the Wigner semicircle distribution (12.4.11) is:

$$\frac{A}{2\pi}\int_{-2/\sqrt{A}}^{2/\sqrt{A}} dh \left(\sqrt{\frac{4}{A} - h^2}\right) h^2 = \frac{1}{A}. \tag{12.4.12}$$

A check of the consistency condition for the geometric expansion at this order:

$$\tau\left(\frac{1}{A} - \frac{1}{12}\right) < 1,$$

leads to

$$A > A_{lb}^{(0)} = \frac{12\tau}{12+\tau}. \tag{12.4.13}$$

In particular, this restriction is removed when $\tau \to 0$, as it should for the undeformed limit. We remark, however, that this is an $\mathcal{O}(1)$ estimation of A_{lb}, and not a true constraint, which must be imposed on the full (all-orders) result.

We now proceed to the next order in perturbation theory, corresponding to $j = 0, 1$ in the geometric expansion. The saddle point equation at order τ is:

$$\mathrm{P}\int du \frac{\rho(u)}{h-u} = \frac{Ab_1}{2}h, \tag{12.4.14}$$

where we have denoted

$$b_1 \equiv b_1(A,\tau) = 1 + \tau\left(\frac{1}{A} - \frac{1}{12}\right).$$

As we are in the small coupling phase, $A < A_{cr}^{(0)} = \pi^2$, we have that $b_1(A,\tau) > 1$. The saddle point equation (12.4.14) is again satisfied by the Wigner semicircular distribution, now with parameter Ab_1, that is:

$$\rho(h) = \frac{Ab_1}{2\pi}\sqrt{\frac{4}{Ab_1} - h^2}, \qquad \mathrm{supp}\rho = \left[-\frac{2}{\sqrt{Ab_1}}, \frac{2}{\sqrt{Ab_1}}\right]. \tag{12.4.15}$$

From this it stems that the second moment at first order in τ is $1/Ab_1$. The constraint (12.4.8) implies $Ab_1 < \pi^2$ and, as $b_1 > 1$, in particular we get

$$\frac{\pi^2 - 2\tau}{1 - \frac{\tau}{6}} = A_{cr}^{(1)} < A_{cr}^{(0)} = \pi^2.$$

We now consider a generic order k in the perturbative expansion in the parameter τ. The general procedure is clear from order 1, and can be iterated, giving order by order a Wigner semicircle distribution with different coefficients. The second moment, approximated at previous order, is $1/Ab_{k-1}$, and the saddle point equation reduces to

$$\mathrm{P}\int du \frac{\rho(u)}{h-u} = \frac{Ab_k}{2} h, \tag{12.4.16}$$

with generic multiplicative factor

$$b_k \equiv b_k(A,\tau) = \sum_{j=0}^{k} (j+1)\, t^j \left(\frac{1}{Ab_{k-1}} - \frac{1}{12}\right)^j. \tag{12.4.17}$$

Notice that we have a recursive way to calculate the b_k's, only depending on the previous one, although in a nontrivial way.

The solution is given by

$$\rho(h) = \frac{Ab_k}{2\pi}\sqrt{\frac{4}{Ab_k} - h^2}, \qquad \mathrm{supp}\rho = \left[-\frac{2}{\sqrt{Ab_k}}, \frac{2}{\sqrt{Ab_k}}\right], \tag{12.4.18}$$

as long as the condition $Ab_k < \pi^2$ holds. In particular, as $b_k = 1 + \mathcal{O}(\tau)$, we have that the critical value of the area is lowered from the pure Yang–Mills case, i.e. $A_{cr}^{(k)} < \pi^2$, at least for τ small enough. Consistently, the constraint guarantees order by order that:

$$\tau\left(\frac{1}{Ab_k} - \frac{1}{12}\right) \geq \tau\left(\frac{1}{\pi^2} - \frac{1}{12}\right) \geq 0.$$

We will now obtain the full solution to (12.4.9) by including all orders in τ. This formally corresponds to evaluate recursive relation (12.4.17) for all k, and the eigenvalue distribution is then given by the Wigner semicircle expression with parameter Ab_∞. From expression (12.4.17) one recursively infers that

$$b_k \leq b_{k-1} + (k+1)\tau^k \left(\frac{1}{Ab_{k-1}} - \frac{1}{12}\right)^k \quad \Longrightarrow \quad |b_k - b_{k-1}| \to 0,$$

and therefore $b_\infty(A,\tau)$ is given by the solution of the equation:

$$b_\infty = \lim_{k\to\infty} \sum_{j=0}^{k} (j+1)\, t^j \left(\frac{1}{Ab_{k-1}} - \frac{1}{12}\right)^j. \tag{12.4.19}$$

Samples of the convergence of b_k are given in Figure 12.5. Writing the right hand side of (12.4.19) as the derivative of a geometric series,[38] b_∞ is determined by solving:

$$b_\infty = \left[1 - \tau\left(\frac{1}{Ab_\infty} - \frac{1}{12}\right)\right]^{-2}. \tag{12.4.20}$$

This leads to a cubic equation in b_∞, but only one of the three solutions satisfies

$$b_\infty(A,\tau) \xrightarrow{\tau\to 0} 1,$$

[38] We can do that for $A > A_{lb}$, as we assumed at the beginning.

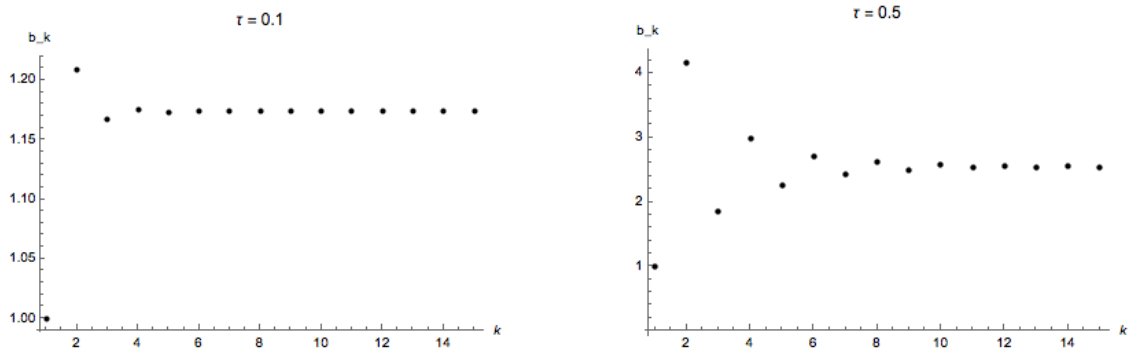


Figure 12.5. Convergence of the sequence $\{b_k\}_k$, for $\tau = 0.1$ (left) and $\tau = 0.5$ (right).

hence we uniquely identify our solution for $b_\infty(A,\tau)$. Explicitly:

$$b_\infty(A,\tau) = \frac{1+\frac{2\tau}{A}\left(1+\frac{\tau}{12}\right)+\sqrt{1+\frac{4\tau}{A}\left(1+\frac{\tau}{12}\right)}}{2\left(1+\frac{\tau}{12}\right)^2}. \tag{12.4.21}$$

The small coupling region is defined by the condition $Ab_\infty < \pi^2$, whence the critical value for the coupling is

$$A_{cr}(\tau) = \pi^2\left(1-\tau\left(\frac{1}{\pi^2}-\frac{1}{12}\right)\right)^2, \tag{12.4.22}$$

as long as $\tau < \frac{12\pi^2}{12-\pi^2}$, and no positive solution for τ bigger than the mentioned value. At this point, we ought to check the consistency of our initial assumption: we developed a perturbative expansion in τ, and then solved it to all orders, assuming the existence of a region $A > A_{lb}$ for which

$$\tau\left(\frac{1}{Ab_\infty}-\frac{1}{12}\right) < 1,$$

corresponding to:

$$Ab_\infty(A,\tau) > \frac{12\tau}{12+\tau}. \tag{12.4.23}$$

As we have the explicit expression for $b_\infty(A,\tau)$, we can see that the infimum of the left hand side of (12.4.23), as a function of A is exactly the right hand side, that is, b_∞ takes exactly the expression for which the lower bound is pushed to $A_{lb} = 0$. This means that our procedure holds for any $A > 0, \tau \geq 0$, or, in other words, the assumption we made to start with the perturbative procedure is always verified in the region of validity of the heat kernel expansion.

To summarize, we have proved that, after the $T\overline{T}$-deformation, we still have a small coupling phase $0 < A < A_{cr}(\tau)$ analogous to the undeformed case, with eigenvalue distribution given by a Wigner semicircle. Nevertheless, the effect of the deformation is to modify the parameter of the distribution, as well as moving the original critical value [193]. In particular, for small values of τ, the value $A_{cr}(\tau)$ of the critical area is a decreasing function, hence $A_{cr} \leq \pi^2$, whilst when $\tau \geq \frac{12\pi^2}{12-\pi^2}$ we do not have any small coupling phase, due to the constraint $A > 0$, and only the strong coupling phase exists. The eigenvalue density in the small coupling phase $0 < A < A_{cr}(\tau)$ is:

$$\rho(h) = \frac{Ab_\infty(A,\tau)}{2\pi}\sqrt{\frac{4}{Ab_\infty(A,\tau)}-h^2}, \qquad \text{supp}\rho = \left[-\frac{2}{\sqrt{Ab_\infty(A,\tau)}},\frac{2}{\sqrt{Ab_\infty(A,\tau)}}\right], \tag{12.4.24}$$

where $b_\infty(A,\tau)$ is given in (12.4.21).

12.4.2 Strong coupling phase

Throughout the solution showed above, we had to impose an upper bound to the coupling A in order not to violate the constraint (12.4.8). When $A > A_{cr}(\tau)$ the Wigner semicircle distribution is not allowed anymore, so that we have to look for a two-cut solution of the saddle point equation (12.4.9). We follow again a perturbative approach, reproducing the procedure of [193] order by order. According to what we have seen in the small coupling phase, the $T\overline{T}$-deformation introduces a nontrivial dependence on the parameters A and τ, but preserves the form of the Douglas–Kazakov solution. Therefore the two-cut solution, if any, must be of the form:

$$\rho(h) = \begin{cases} \varphi(h), & h \in [-\alpha, -\beta] \cup [\beta, \alpha]\,, \\ 1, & h \in [-\beta, \beta]\,, \end{cases}$$

where $0 \leq \beta \leq \alpha$ depend, in general, on A and τ. Plugging this expression into (12.4.9) we get:

$$\mathrm{P}\int du \frac{\varphi(u)}{h-u} = \log\left(\frac{h-b}{h+b}\right) + \frac{A}{2} h \sum_{j=0}^{\infty} (j+1)\,\tau^j \left[\int du \rho(u) u^2 - \frac{1}{12}\right]^j . \tag{12.4.25}$$

The idea is again to proceed perturbatively in τ, evaluating the second moment on the right hand side, using the approximation at previous order. Again, this will only account for a modification $A \mapsto A d_k$, with

$$d_k = \sum_{j=0}^{k} (j+1)\,\tau^j \left[\int_{-\alpha}^{\alpha} du \rho(u) u^2 - \frac{1}{12}\right]^j , \tag{12.4.26}$$

where the second moment of the distribution is evaluated at order $k-1$.

We directly treat the problem at a generic order k, knowing that $d_0 = 1$, and hence the initial step of our procedure corresponds to the result of [193]. We define a complex function

$$\Phi(z) = \int \frac{\varphi(u)}{u-z} du, \tag{12.4.27}$$

for $z \notin [-\alpha, \alpha]$. On one hand, when z approaches a real number $h \in [-\alpha, \alpha]$, we can write

$$\Phi_+(h) - \Phi_-(h) = 2\pi \mathrm{i} \varphi(h) \mathbb{1}_U(h), \tag{12.4.28}$$

where $\Phi_\pm(h) := \lim_{\varepsilon \to 0} \Phi(h \pm \mathrm{i}\varepsilon)$ and $\mathbb{1}_U$ is the characteristic function of the set

$$U := [-\alpha, -\beta] \cup [\beta, \alpha] \equiv [-\alpha, \alpha] \setminus (-\beta, \beta)\,.$$

Once we obtain a complex solution to the saddle point equation (12.4.25), we can recover the φ by evaluating the left hand side of (12.4.28) as the discontinuity at the branch cut of the complex solution for $h \in [-\alpha, \alpha]$. Such a complex solution is:

$$\Phi(z) = -\frac{1}{2\pi \mathrm{i}} \sqrt{(\alpha^2 - z^2)(\beta^2 - z^2)} \oint_{\gamma_U} du \frac{\frac{A d_k}{2} u + \log\left(\frac{u-\beta}{u+\beta}\right)}{(u-z)\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}, \tag{12.4.29}$$

for some path γ_U in the complex plane around the cut U. After an adequate deformation of the contour integral, one gets:

$$\Phi(z) = -\frac{A d_k}{2} z - \log\left(\frac{z-\beta}{z+\beta}\right) - \sqrt{(\alpha^2 - z^2)(\beta^2 - z^2)} \int_{-\beta}^{\beta} \frac{du}{(u-z)\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}. \tag{12.4.30}$$

The first two terms are obtained from the residue theorem, and the last one accounts for the branch cut of the logarithm along $[-\beta,\beta]$. When z approaches the real axis, the logarithm has a discontinuity of $2\pi\mathrm{i}$ if $z\to h\in[-\beta,\beta]$, and has no discontinuity out of that interval, while the third term is discontinuous in all $z\in[-\alpha,\alpha]$. Thus we get:

$$\Phi_+(h)-\Phi_-(h)=-2\pi\mathrm{i}\mathbb{1}_{[-\beta,\beta]}-2\mathrm{i}\,\mathrm{sign}(h)\sqrt{(\alpha^2-h^2)(h^2-\beta^2)}\int_{-\beta}^{\beta}\frac{du}{(u-h)\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}. \tag{12.4.31}$$

The sign function appears because, for $h\in[\beta,\alpha]$, one approaches the branch cut of the square root from the proper direction, i.e. $\Phi_+-\Phi_-$ corresponds to "counter-clockwise minus clockwise". For $h\in[-\alpha,-\beta]$, instead, the branch cut is approached from the converse direction.

Therefore, comparing with (12.4.28), one arrives at:

$$\begin{aligned}\rho(h)&=\varphi(h)\mathbb{1}_U+\mathbb{1}_{[-\beta,\beta]}\\&=\frac{1}{\pi}\mathrm{sign}(h)\sqrt{(\alpha^2-h^2)(h^2-\beta^2)}\int_{-\beta}^{\beta}\frac{du}{(h-u)\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}.\end{aligned} \tag{12.4.32}$$

It is easy to check that eigenvalue distribution in a positive function of h in all $[-\alpha,\alpha]$ and is identically 1 in the interval $[-\beta,\beta]$.

Until this point we simply reproduced the procedure of [193], which applies also for our generalized case. Notice that the eigenvalue distribution ρ apparently does not yield an explicit dependence on Ad_k; nevertheless, the parameters α,β will depend on it, and so will do ρ. The boundaries α,β can be fixed by the asymptotic expansion of $\Phi(z)$, for instance by comparison between (12.4.30) and the definition (12.4.27). From this latter we have:

$$\begin{aligned}\Phi(z)&=-\frac{1}{z}\left(\int_U du\varphi(u)+\frac{1}{z^2}\int_U du\varphi(u)u^2+\dots\right)\\&=-\frac{1}{z}\left[(1-2\beta)+\frac{1}{z^2}\left(\int_{-\alpha}^{\alpha}du\rho(u)u^2-\frac{2}{3}\beta^3\right)+\dots\right].\end{aligned} \tag{12.4.33}$$

On the other hand, the explicit expression (12.4.30) implies:

$$\begin{aligned}\Phi(z)&=-\frac{Ad_k}{2}z-\log\left(1-\frac{2\beta}{z}+\dots\right)+z\left(1-\frac{\alpha^2+\beta^2}{2z^2}+\dots\right)\int_{-\beta}^{\beta}du\frac{\left[1+\frac{u^2}{z^2}+\frac{u^4}{z^4}+\dots\right]}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}\\&=z\left[-\frac{Ad_k}{2}z+\int_{-\beta}^{\beta}\frac{du}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}\right]+\frac{1}{z}\left[2\beta+\int_{-\beta}^{\beta}du\frac{u^2-\frac{\alpha^2+\beta^2}{2}}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}\right]+\mathcal{O}\left(\frac{1}{z^3}\right)\end{aligned} \tag{12.4.34}$$

The comparison at $\mathcal{O}(z)$ imposes the constraint

$$\int_{-\beta}^{\beta}\frac{du}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}=\frac{Ad_k}{2}\quad\Longrightarrow\quad\alpha=\frac{4}{Ad_k}K\left(\frac{\beta}{\alpha}\right), \tag{12.4.35}$$

where $K(\cdot)$ is the complete elliptic integral of first kind. Analogously, from comparison at $\mathcal{O}(z^{-1})$ one gets the constraint:

$$\begin{aligned}&2\beta-\frac{\alpha^2+\beta^2}{2}\int_{-\beta}^{\beta}\frac{du}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}+\int_{-\beta}^{\beta}\frac{u^2du}{\sqrt{(\alpha^2-u^2)(\beta^2-u^2)}}=-1+2\beta\\&\Longrightarrow\quad K\left(\frac{\beta}{\alpha}\right)\left[2E\left(\frac{\beta}{\alpha}\right)-\left(1-\frac{\beta^2}{\alpha^2}\right)K\left(\frac{\beta}{\alpha}\right)\right]=\frac{Ad_k}{4},\end{aligned} \tag{12.4.36}$$

where $E(\cdot)$ is the complete elliptic integral of second kind, and we plugged in (12.4.35) to simplify the expression. The result is clearly the same as [193], up to a rescaling $A \mapsto Ad_k$.

As we are interested in knowing the second moment of the distribution ρ, we may use $\mathcal{O}(z^{-3})$ of the expansion above to obtain the dependence of the integral expression on the other parameters. It leads to:

$$\begin{aligned}\int_{-\alpha}^{\alpha} du\rho(u)u^2 &= \int_{-\beta}^{\beta} du \frac{-u^4 + \frac{\alpha^2+\beta^2}{2}u^2 + \frac{\left(\alpha^2-\beta^2\right)^2}{8}}{\sqrt{\left(\alpha^2-u^2\right)\left(\beta^2-u^2\right)}} \\ &= \frac{\left(\alpha^2-\beta^2\right)^2}{16} Ad_k + \alpha\left(\alpha^2+\beta^2\right)\left[K\left(\frac{\beta}{\alpha}\right) - E\left(\frac{\beta}{\alpha}\right)\right] \\ &\quad - \frac{2}{3}\alpha^3\left[\left(2+\frac{\beta^2}{\alpha^2}\right)K\left(\frac{\beta}{\alpha}\right) - 2\left(1+\frac{\beta^2}{\alpha^2}\right)E\left(\frac{\beta}{\alpha}\right)\right].\end{aligned} \tag{12.4.37}$$

We can use the properties of the elliptic integrals to extract information about the dependence on Ad_k. In particular, from the first two conditions (12.4.35)-(12.4.36) we get that, for $Ad_k \to \pi^2$, one recovers the same parameters as approaching the critical point from below, that is, $(\alpha = 2/\pi, \beta = 0)$. More specifically, for Ad_k close to π^2, we may approximate the elliptic integrals, and obtain the first terms of the expansion of α and β around $Ad_k = \pi^2$:

$$\begin{aligned}\alpha &= \frac{1}{\pi}\left[2 - \frac{Ad_k-\pi^2}{\pi^2} + \frac{5}{4}\left(\frac{Ad_k-\pi^2}{\pi^2}\right)^2 + \dots\right], \\ \beta &= \frac{1}{\pi}\left[2\sqrt{2}\left(\frac{\left(Ad_k-\pi^2\right)}{\pi^2}\right)^{\frac{1}{2}} - \frac{15}{4\sqrt{2}}\left(\frac{Ad_k-\pi^2}{\pi^2}\right)^{\frac{3}{2}} + \dots\right].\end{aligned} \tag{12.4.38}$$

Moreover, concerning the second moment, approximated close to the critical point, we get:

$$\int_{-\alpha}^{\alpha} du\rho(u)u^2 = \frac{1}{\pi^2}\left[1 - \frac{Ad_k-\pi^2}{\pi^2} + 3\left(\frac{Ad_k-\pi^2}{\pi^2}\right)^2 + \dots\right]. \tag{12.4.39}$$

At this point, we are able to determine the full nonperturbative expression for the eigenvalue density $\rho(h)$, close enough to the critical point. That is: on one hand, we have a formal recursive expression for the coefficients d_k of the perturbative expansion in τ at strong coupling, while on the other hand, if we want to determine the order of the phase transition, we need to know an explicit expression for the dependence of the eigenvalue density on the parameters A and τ. Notice however, that local information close to the critical point is enough to characterize the phase transition. For these reasons, we look for a full (nonperturbative) solution, approximating close to the critical point. The solution we will find will be only valid up to order $(Ad_\infty - \pi^2)^2$.

The formal limit $k \to \infty$ of this expression (12.4.26) leads to the equation:

$$d_\infty = \left[1 - \tau\left(\int_{-\alpha}^{\alpha} du\rho(u)u^2 - \frac{1}{12}\right)\right]^{-2}, \tag{12.4.40}$$

and the approximated solution close to the critical point is found plugging expression (12.4.39), obtaining:

$$d_\infty \approx \left[1 - \tau\left[\frac{1}{\pi^2}\left(1 - \frac{Ad_\infty-\pi^2}{\pi^2}\right) - \frac{1}{12}\right]\right]^{-2}, \tag{12.4.41}$$

which again admits only one solution compatible with $\lim_{\tau\to 0} d_\infty(A,\tau) = 1$. As a side remark, we highlight that the defining equation for d_∞ starts to differ from the one for b_∞ only at order $(Ad_\infty - \pi^2)^2$, implying that b_∞ and d_∞ will coincide up to the first derivative when evaluated at the critical point A_{cr} (same 1-jet at A_{cr}).

12.4.3 Third order phase transition

In this subsection, we study the free energy from the point of view of small and large area, that is $A < A_{cr}(\tau)$ and $A > A_{cr}(\tau)$ respectively, with the aim to determine the order of the phase transition. The free energy of the system is defined as:

$$\mathcal{F}_N(A,\tau) = -\frac{1}{N^2}\log \mathcal{Z}_N(A,\tau)\,. \tag{12.4.42}$$

In the large N limit the derivative with respect to the control parameter A is given by:

$$\frac{\partial \mathcal{F}}{\partial A} = \frac{1}{2}\sum_{j=0}^{\infty}\tau^j\left[\int du\rho(u)u^2 - \frac{1}{12}\right]^{j+1}. \tag{12.4.43}$$

Before passing to the direct evaluation, we notice that:

$$\frac{\partial \mathcal{F}}{\partial A} = F'_{\mathrm{DK}}(Ac_\infty)\sum_{j=0}^{\infty}\tau^j\left[\int du\rho(u)u^2 - \frac{1}{12}\right]^{j}, \tag{12.4.44}$$

where by $F'_{\mathrm{DK}}(A)$ we mean the first derivative of the free energy obtained by Douglas and Kazakov [193], and c_∞ is a shorthand:

$$c_\infty = \begin{cases} b_\infty, & A < A_{cr}(\tau); \\ d_\infty, & A > A_{cr}(\tau). \end{cases}$$

Therefore

$$\frac{\partial \mathcal{F}}{\partial A} = F'_{\mathrm{DK}}(Ac_\infty)\left[1-\tau\left(\int du\rho(u)u^2 - \frac{1}{12}\right)\right]^{-1}. \tag{12.4.45}$$

Taking advantage of the defining equation (12.4.20) and (12.4.40) for b_∞ and d_∞ respectively, we can rewrite:

$$\frac{\partial \mathcal{F}}{\partial A} = F'_{\mathrm{DK}}(Ac_\infty)\sqrt{c_\infty}. \tag{12.4.46}$$

When $A < A_{cr}$, the latter expression is calculated using the distribution at small coupling:

$$\left.\frac{\partial \mathcal{F}}{\partial A}\right|_{A<A_{cr}} = \frac{\sqrt{b_\infty}}{2}\left(\frac{1}{Ab_\infty}-\frac{1}{12}\right) = \frac{\sqrt{b_\infty}}{2}\left[\frac{1}{\pi^2}\left(1-\frac{Ab_\infty-\pi^2}{\pi^2}+\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)^2+\dots\right)-\frac{1}{12}\right]. \tag{12.4.47}$$

Analogously, in the strong coupling phase $A > A_{cr}(\tau)$ we ought to use the eigenvalue distribution at strong coupling, which, approximating close to the critical point, provides the expression:

$$\left.\frac{\partial \mathcal{F}}{\partial A}\right|_{A>A_{cr}} = \frac{\sqrt{d_\infty}}{2}\left[\frac{1}{\pi^2}\left(1-\frac{Ad_\infty-\pi^2}{\pi^2}+3\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)^2+\dots\right)-\frac{1}{12}\right].$$

By construction of the two-cut solution, we know that:

$$Ab_\infty \xrightarrow{A\to A_{cr}^-} \pi^2 \xleftarrow{A\to A_{cr}^+} Ad_\infty \tag{12.4.48}$$

which guarantees $\frac{\partial \mathcal{F}}{\partial A}$ is continuous at the critical point, thus the transition is at least of second order. We in fact have that:

$$\begin{aligned}\left.\frac{\partial \mathcal{F}}{\partial A}\right|_{A>A_{cr}} - \left.\frac{\partial \mathcal{F}}{\partial A}\right|_{A<A_{cr}} &= \frac{1}{2}\left(\frac{1}{\pi^2}-\frac{1}{12}\right)\left(\sqrt{d_\infty}-\sqrt{b_\infty}\right) \\ &\quad -\frac{1}{2\pi^2}\left[\sqrt{d_\infty}\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)-\sqrt{b_\infty}\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)\right] \\ &\quad +\frac{1}{2\pi^2}\left[3\sqrt{d_\infty}\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)^2-\sqrt{b_\infty}\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)^2\right]+\dots.\end{aligned} \tag{12.4.49}$$

Taking a further derivative with respect to A we get:

$$\begin{aligned}\left.\frac{\partial^2\mathcal{F}}{\partial A^2}\right|_{A>A_{cr}} - \left.\frac{\partial^2\mathcal{F}}{\partial A^2}\right|_{A<A_{cr}} &= \frac{1}{4}\left(\frac{1}{\pi^2}-\frac{1}{12}\right)\left[\frac{d'_\infty}{\sqrt{d_\infty}}-\frac{b'_\infty}{\sqrt{b_\infty}}\right]\\ &-\frac{1}{4\pi^2}\left[\frac{d'_\infty}{\sqrt{d_\infty}}\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)-\frac{b'_\infty}{\sqrt{b_\infty}}\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)\right]\\ &-\frac{1}{2\pi^4}\left[\sqrt{d_\infty}\left(d_\infty+Ad'_\infty\right)-\sqrt{b_\infty}\left(b_\infty+Ab'_\infty\right)\right]\\ &+\frac{1}{4\pi^2}\left[3\frac{d'_\infty}{\sqrt{d_\infty}}\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)^2-\frac{b'_\infty}{\sqrt{b_\infty}}\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)^2\right]\\ &+\frac{1}{\pi^4}\left[\sqrt{d_\infty}\left(d_\infty+Ad'_\infty\right)\left(\frac{Ad_\infty-\pi^2}{\pi^2}\right)-\sqrt{b_\infty}\left(b_\infty+Ab'_\infty\right)\left(\frac{Ab_\infty-\pi^2}{\pi^2}\right)\right]+\dots\end{aligned} \tag{12.4.50}$$

Is is straightforward to see that the second, fourth and fifth term vanish at the critical point. The first and third term are more subtle, because the involve derivatives of the coefficients b_∞, d_∞. However, we have that both coefficients are defined by formally the same expression, but using the eigenvalue distribution at small or large coupling respectively. Regarding b_∞, we can evaluate its derivative using (12.4.20):

$$b'_\infty = -\frac{2\tau}{\pi^2}\frac{b_\infty^{5/2}}{1-Ab_\infty^{3/2}}+\dots,$$

where the dots represent term that vanish at the critical point. The same can be done for d_∞ using (12.4.40), to obtain:

$$d'_\infty = -\frac{2\tau}{\pi^2}\frac{d_\infty^{5/2}}{1-Ad_\infty^{3/2}}+\dots.$$

Hence we infer that the first derivatives b'_∞ and d'_∞ of the coefficients coincide at the critical point (this fails to be true for higher derivatives). Consequently, the phase transition is again of third order.

It is remarkable that the $T\overline{T}$-deformation introduced a nontrivial dependence on the coupling A, but in such a way that it does not affect the order of the phase transition.

12.4.4 Instanton analysis

Gross and Matytsin presented evidence for the phase transition to be triggered by instantons [385]. The existence of a phase transition can be closely related to the discreteness of the matrix model (10.1.2) [385, 494, 436]. In [495], Yang–Mills theory on the sphere is described in terms of N nonrelativistic free fermions on $\mathbb{S}^1$, with instantons corresponding to different winding numbers for fermions at a given position, and the phase transition occurs due to the condensation of fermions in (discrete) momentum space, so again the discreteness turned out to be essential to permit a phase transition. From this general argument, and taking into account expression (12.4.1), the statement is expected to hold also in the $T\overline{T}$-deformed version of two-dimensional Yang–Mills, as the effect of the deformation, at the level of the matrix model, is to replace the discrete Gaussian weight with a discrete weight whose potential has also additional multitrace contributions. Thus, in this section we look at the role played by instantons in the phase transition.

Instantons in the undeformed theory

By instanton, we mean a solution of the classical Yang–Mills equation of motion which is gauge inequivalent to the trivial one. Those solutions are in one to one correspondence with collections

of N monopole charges:

$$\ell := (\ell_1, \dots, \ell_N) \in \mathbb{Z}^N \,.$$

The action for a given classical configuration is:

$$S_{inst}\,(\ell) = \frac{N}{2A} \sum_{j=1}^{N} (2\pi\ell_j)^2 \,, \tag{12.4.51}$$

and the partition function splits into the sum of contributions from instanton sectors:

$$\mathcal{Z}_N^{(\mathrm{YM})} = \sum_{\ell \in \mathbb{Z}^N} w\,(\ell)\, e^{-S_{inst}(\ell)}. \tag{12.4.52}$$

This is the content of Witten's result [496] extending the Duistermaat–Heckman theorem. From the point of view of the Abelianization procedure [497], each ℓ_j is the first Chern class of a $U(1)$-bundle, see [497, 423, 384] for more details. Later on, this viewpoint was the one adopted in [494] to estimate the instanton contributions in the case of q-deformed Yang–Mills theory on $\mathbb{S}^2$.

A practical difficulty is to evaluate the weights $w\,(\ell)$, which was done in [495] (and in [409] for the q-deformed case), through the method of Poisson resummation. In [385], the contribution of the single-monopole sector $\ell = (1, 0, \dots, 0)$ was calculated, showing that this correction to the saddle point approximation at large N is exponentially suppressed,

$$\frac{w\,(1,0,\dots,0)\, e^{-S_{inst}(1,0,\dots,0)}}{w\,(0,\dots,0)\, e^{-S_{inst}(0,\dots,0)}} \propto e^{-\frac{2\pi^2}{A} N \gamma\left(\frac{A}{\pi^2}\right)},$$

where the function $\gamma\,(\cdot)$ is the one introduced by Gross and Matytsin [385] and is given by

$$\gamma(x) = \sqrt{1-x} - \frac{x}{2} \log\left(\frac{1+\sqrt{1-x}}{1-\sqrt{1-x}}\right). \tag{12.4.53}$$

In particular, as $\gamma\,(x) > 0$ for $0 < x < 1$ and $\gamma\,(1) = 0$, in the small coupling phase contributions from instanton sectors are exponentially suppressed at large N, but they become more and more relevant as the critical point is approached. In this sense, the weight $w\,(\ell)$ acts as a counterpart of the Boltzmann factor $e^{-S(\ell)}$, and at the critical point those two contributions are exactly balanced.

Instantons in the $T\overline{T}$-deformed theory

The evaluation of the full instanton expansion for the deformed theory would correspond to find an explicit expression of the form:

$$\mathcal{Z}_N\,(A,\tau) = \sum_{\ell \in \mathbb{Z}^N} w\,(\ell)\, e^{-S_{inst}(\ell)},$$

where now the weights and the action include the effects of the deformation by the $T\overline{T}$-operator. We have found strong evidence that this can be done in an analytic way, but the weights obtained are not very enlightening and unsuitable for the purpose of this section. Instead, we will only look at the first instanton correction, and how it affects the model. A discussion on the Poisson resummation for the full explicit expression is relegated to Appendix 12.A.

As a first step, the partition function can be rewritten as a sum over Fourier transforms of contributions from each representation:

$$\mathcal{Z}_N\,(A,\tau) = \sum_{\ell \in \mathbb{Z}^N} Z_\ell, \tag{12.4.54}$$

where each instanton sector contributes as:

$$Z_\ell = \int_{\mathbb{R}^N} \prod_{i=1}^{N} dh_i e^{-\mathrm{i}2\pi \sum_{i=1}^{N} \ell_i h_i} \prod_{i<j} \left(\frac{h_i - h_j}{j-i}\right)^2 e^{-\frac{A}{2N}\sum_{j=0}^{\infty}\left(\frac{\tau}{N^3}\right)^j \left(\sum_{i=1}^{N} h_i^2 - \frac{N\left(N^2+2\right)}{12}\right)^{j+1}}. \quad (12.4.55)$$

Consider the single-monopole sector corresponding to $\ell = (\ell_1, 0, \ldots, 0)$ (we will eventually set $\ell_1 = \pm 1$). It contributes to the partition function as:

$$Z_{(\ell_1,0,\ldots,0)} = \int_{\mathbb{R}^N} \prod_{i=1}^{N} dh_i e^{-N^2 S_{\ell_1}[h]},$$

where

$$S_{\ell_1}[h] = -\frac{2}{N^2}\sum_{i<j} \log\left(\frac{h_i - h_j}{j-i}\right) + \frac{A}{2N^3}\sum_{j=0}^{\infty}\left(\frac{\tau}{N^3}\right)^j \left(\sum_{i=1}^{N} h_i^2 - \frac{N\left(N^2+2\right)}{12}\right)^{j+1} + \frac{2\pi\mathrm{i}}{N^2}\ell_1 h_1. \quad (12.4.56)$$

This means that the correction to the large N action, with respect to the vacuum sector, is of $\mathcal{O}(N^{-1})$. This implies that, at large N, we can perform $N-1$ integrals using the saddle point approximation for the eigenvalue distribution, and eventually treat the integration over h_1 separately [385, 494]. Notice that the saddle point for h_1 will be, in general, complex, due to the purely imaginary "Fourier interaction" with ℓ_1. Nevertheless, as we are interested in the suppressing factor, we will avoid the technicalities involved in determining the imaginary part of the instanton contribution.

Taking the large N limit, we have:

$$Z_{(\ell_1,0,\ldots,0)} = \mathcal{C}^{N-1}\int_{-\infty}^{\infty} dh_1 e^{-NS_{\ell_1}[h_1]}, \quad (12.4.57)$$

where $\mathcal{C}$ is the integral over any variable $h_2, \ldots, h_N$, calculated with the saddle point approximation, and the effective action for the scaled variable $h = \frac{h_1}{N}$ is:

$$S_{\ell_1}[h] = -2\int du\rho(u)\log(h-u) - 2 + \frac{A}{2}\sum_{j=0}^{\infty}\tau^j\left(h^2 - \frac{1}{12}\right)^{j+1} + 2\pi\mathrm{i}\ell_1 h, \quad (12.4.58)$$

and the density ρ is the one obtained in (12.4.24) for small coupling $A < A_{cr}$. The saddle point for the effective action is given by:

$$2\mathrm{P}\int du\frac{\rho(u)}{h-u} + Ah\sum_{j=0}^{\infty}(j+1)\tau^j\left(h^2 - \frac{1}{12}\right)^j + 2\pi\mathrm{i}\ell_1 = 0$$

which, using the fact that ρ satisfies the saddle point for $\ell_1 = 0$, simplifies into:

$$2\pi\mathrm{i}\left(\rho(h) + \ell_1\right) = 0.$$

This leads to the saddle point:

$$h^2 = \left(\frac{2}{Ab_\infty}\right)^2\left(Ab_\infty - \pi^2\ell_1\right),$$

which, since the analysis is being brought on in the small coupling phase $Ab_\infty < \pi^2$, gives a purely imaginary saddle point:

$$h = \frac{2\pi\mathrm{i}}{Ab_\infty}\mathrm{sign}(\ell_1)\sqrt{\ell_1^2 - \frac{Ab_\infty}{\pi^2}}. \quad (12.4.59)$$

The result is exactly the same obtained in the undeformed case [385], up to a rescaling $A \mapsto Ab_\infty$. In particular:

$$\frac{Z_{(1,0,\dots,0)}}{Z_{(0,\dots,0)}} = \mathcal{C}' e^{-N\frac{2\pi^2}{Ab_\infty}\gamma\left(\frac{Ab_\infty}{\pi^2}\right)}, \tag{12.4.60}$$

where $\mathcal{C}'$ is an overall constant and $\gamma(\cdot)$ is the function (12.4.53). Hence, the same conclusions of the undeformed case [385] hold: from the perspective of the small coupling expansion, the phase transition is triggered by instantons.

A more thorough analysis of the single-monopole instanton correction in the $T\overline{T}$-deformed case, shows that the relevance of instantons increases with τ, for $\tau < \frac{12\pi^2}{12-\pi^2}$, and there is no suppression at all for $\tau \geq \frac{12\pi^2}{12-\pi^2}$. This explains why the phase transition occurs earlier in the deformed case, that is $A_{cr}(\tau) \leq A_{cr}(0)$: indeed, the function $\frac{1}{Ab_\infty}\gamma\left(\frac{Ab_\infty}{\pi^2}\right)$ decreases faster than the function $\frac{1}{A}\gamma\left(\frac{A}{\pi^2}\right)$, hence the first instanton sector becomes relevant at a lower value of A, in comparison to the undeformed case. This is presented in Figures 12.6 and 12.7.

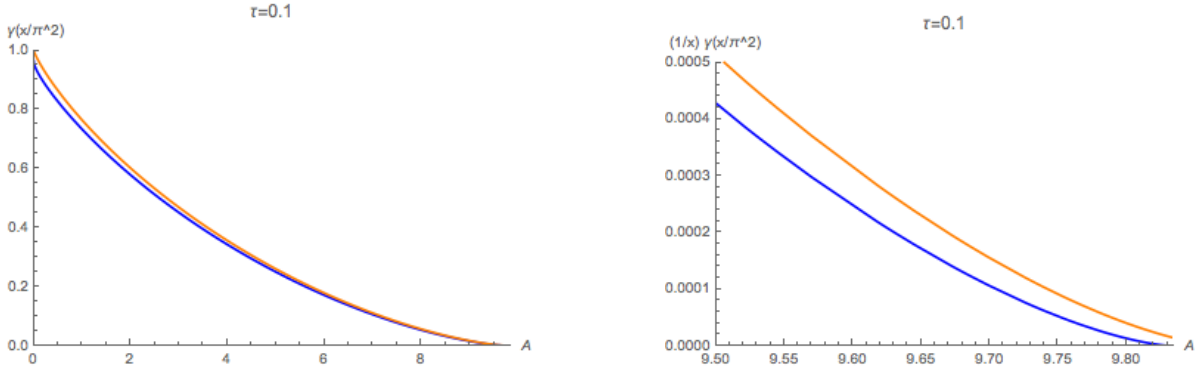


Figure 12.6. On the left: a comparison of the function $\gamma\left(\frac{x}{\pi^2}\right)$ for the deformed (blue) and undeformed (orange) case. On the right: a zoom on the tail of the function $\frac{1}{x}\gamma\left(\frac{x}{\pi^2}\right)$ for the deformed (blue) and undeformed (orange) case. The plots are at $\tau = 0.1$.

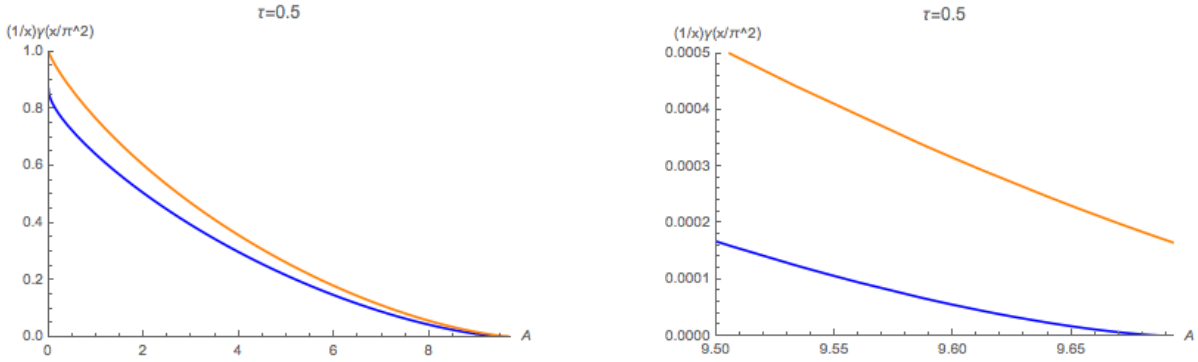


Figure 12.7. On the left: a comparison of the function $\gamma\left(\frac{x}{\pi^2}\right)$ for the deformed (blue) and undeformed (orange) case. On the right: a zoom on the tail of the function $\frac{1}{x}\gamma\left(\frac{x}{\pi^2}\right)$ for the deformed (blue) and undeformed (orange) case. The plots are at $\tau = 0.5$.

12.5 Phase transitions in $T\overline{T}$-deformed q-Yang–Mills theory

Two-dimensional $U(N)$ Yang–Mills theory on $\mathbb{S}^2$ undergoes a third order phase transition [193], henceforth called the Douglas-Kazakov (DK) transition, which is induced by instanton insta-

bilities [385]. The same third order phase transition is experienced by the q-deformed theory [498, 494, 409] when $p > 2$. The critical value of the coupling $\lambda_{\rm cr}$ decreases monotonically with increasing p and one eventually recovers the DK transition in the limit $p \to \infty$ [498, 494, 409], see Figure 12.8. This means that the q-deformation extends the region in parameter space corresponding to a weak coupling phase.

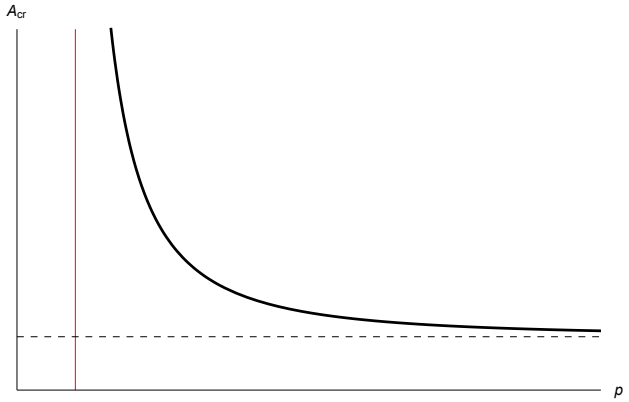


Figure 12.8. The critical curve of q-Yang–Mills theory, in terms of the parameter $A = \lambda/p$ as a function of p. The horizontal asymptote (dashed) is the DK critical point $A_{\rm cr} = \pi^2$. The vertical asymptote (red) is the point $p = 2$. This plot is inspired by [498].

We have shown in Section 12.4 that $T\overline{T}$-deformed (but not q-deformed) Yang–Mills theory also undergoes a DK-type transition for $0 \leq \tau < \tau_{\rm max}$, with

$$\frac{1}{\tau_{\rm max}} = \frac{1}{\pi^2} - \frac{1}{12} \ . \tag{12.5.1}$$

The critical value of the area parameter decreases with increasing τ, and eventually no weak coupling phase exists when τ approaches $\tau_{\rm max}$ [9]. Therefore the $T\overline{T}$-deformation reduces the region of parameter space corresponding to a weak coupling phase.

The goal of this section is to analyze the large N phase diagram when both the q-deformation and the $T\overline{T}$-deformation are turned on. In the following we summarize our large N results. In Section 12.5.1 we present the large N formalism when both deformations are turned on and study the weak coupling regime, while Section 12.5.2 is dedicated to a study of the critical surface. Section 12.5.3 discusses the role of instanton contributions, while Section 12.5.4 is dedicated to a study of the phase transition and the strong coupling regime. Finally, in Section 12.5.5 we comment on the large N limit of the refined theory.

Large N results

Before diving into the detailed analysis of the large N phase structure, we summarize here our main findings:

- $T\overline{T}$-deformed q-Yang–Mills theory undergoes a third order phase transition when $p > p_0$. Remarkably, we find that $p_0 < 2$.

- The slice of parameter space giving a weak coupling phase is extended, relative to the pure $T\overline{T}$-deformation, and is reduced only relative to the q-deformation. This interpolates perfectly between the effects discovered respectively in [498, 494, 409] and [9], recovering the single-deformation scenarios as limiting cases.

- The phase transition is induced by instantons. The shrinking of the weak coupling region is explained by the fact that in the $T\overline{T}$-deformed theory the suppression factor of the instantons is smaller, hence their effect becomes relevant at lower values of the coupling λ.

12.5.1 Large N limit of $T\overline{T}$-deformed q-Yang–Mills theory

At this stage we are ready to apply the formalism of Section 12.4 to the harder problem of q-Yang–Mills theory.

In terms of the shifted weights $\vec{h} \in \mathbb{Z}^N$ introduced in (12.4.3), the partition function of $T\overline{T}$-deformed $U(N)$ q-Yang–Mills theory is

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}(\lambda,\tau) = \frac{1}{N!}\sum_{\vec{h}\in\mathbb{Z}^N}\left(\frac{\Delta_q(\vec{h})}{\Delta_q(\emptyset)}\right)^{\chi}\exp\left(-\frac{\frac{\lambda p}{2N}\left(\sum\limits_{j=1}^{N}h_j^2-\frac{N\,(N^2+1)}{12}\right)}{1-\frac{\tau}{N^3}\left(\sum\limits_{j=1}^{N}h_j^2-\frac{N\,(N^2+1)}{12}\right)}\right), \tag{12.5.2}$$

where $\Delta_q(\vec{h})$ is a q-deformation of the Vandermonde determinant

$$\Delta_q(\vec{h}) = \prod_{1\leq i<j\leq N} 2\sinh\frac{\lambda\,(h_i-h_j)}{2N}\ ,$$

and we used the shorthand notation $\Delta_q(\emptyset) := \Delta_q(h_i = i)$. Here $\Delta_q(\emptyset)$ plays the role of a q-deformation of the Barnes G-function defined in (12.4.5). We have also left χ arbitrary, $0 < \chi$, although we recall that its physical value is $\chi = \chi(\mathbb{S}^2) = 2$.

We now take the large N limit of (12.5.2), which we stress is a 't Hooft limit with 't Hooft coupling λ, while the Yang–Mills coupling is λ/N. In this limit, the contribution of $\Delta_q(\emptyset)^{-\chi}$ is given by [498]

$$\lim_{N\to\infty}\chi\log\Delta_q(\emptyset) = -\frac{\chi}{\lambda^2}\,F_0^{\text{CS}}(\lambda)\ ,$$

where $F_0^{\text{CS}}(\lambda)$ is the planar free energy of $U(N)$ Chern–Simons theory on the three-sphere $\mathbb{S}^3$:

$$F_0^{\text{CS}}(\lambda) = \frac{\lambda^3}{12} - \frac{\pi^2}{6}\,\lambda - \text{Li}_3(e^{-\lambda}) + \zeta(3)\ .$$

The analogue of the saddle point equation (12.4.9) in this q-deformed setting is

$$\text{P}\int \mathrm{d}u\,\rho(u)\coth\frac{\lambda\,(h-u)}{2} = \frac{2p}{\chi}\,h\sum_{k=0}^{\infty}(k+1)\,\tau^k\left(\int_{\text{supp}(\rho)}\mathrm{d}u\,\rho(u)\,u^2-\frac{1}{12}\right)^k\ . \tag{12.5.3}$$

The solution will be a function

$$\rho(h) := \rho\big(h;\lambda,\tfrac{2p}{\chi},\tau\big)$$

depending parametrically on the couplings, which is normalized and satisfies the constraint from (12.4.8): $\rho(h) \leq 1$. We notice also that χ only enters the large N limit in the combination $\frac{2p}{\chi}$, and hence is simply a rescaling of p.

We solve the saddle point equation (12.5.3) perturbatively, as in Section 12.4 above. Assuming a one-cut solution, the zeroth order solution is [498, 494, 409]

$$\rho^{(0)}(h) = \frac{2p}{\pi\,\chi}\tan^{-1}\sqrt{\frac{e^{\chi\,\lambda/2p}}{\cosh^2\frac{\lambda\chi}{4p}\,h}-1}\ ,$$

with support

$$\mathrm{supp}\big(\rho^{(0)}\big) = \big[-\alpha^{(0)}, \alpha^{(0)}\big] \qquad \text{where} \quad \alpha^{(0)} = \frac{2}{\lambda}\cosh^{-1} e^{\chi\,\lambda/4p} \ .$$

We have put the superscript $^{(0)}$ everywhere to remind us that this is the zeroth order solution in a perturbative expansion in τ. The second moment of this distribution is

$$\mu_2^{(0)} = \int_{-\alpha^{(0)}}^{\alpha^{(0)}} \mathrm{d}u\, \rho^{(0)}(u)\, u^2 = \frac{\chi^2}{12p^2} + \frac{1}{3\lambda^2}\Big(\pi^2 + 6\,\mathrm{Li}_2\big(e^{-\chi\,\lambda/2p}\big)\Big) + \frac{8p}{\chi\,\lambda^3}\Big(\mathrm{Li}_3\big(e^{-\chi\,\lambda/2p}\big) - \zeta(3)\Big) \ .$$

The next order approximation of (12.5.3) is

$$\mathrm{P}\int \mathrm{d}u\, \rho(u) \coth\frac{\lambda\,(h-u)}{2} = \frac{2p}{\chi}\, b_{q,1}\, h \ ,$$

where

$$b_{q,1} := b_{q,1}\big(\lambda, \tfrac{2p}{\chi}, \tau\big) = 1 + \tau\left(\mu_2^{(0)} - \frac{1}{12}\right) \ .$$

The solution at this order then will be again as in [498, 494, 409], but with a renormalized value of p given by

$$p \longmapsto \frac{2p}{\chi}\, b_{q,1} \ .$$

Iterating this argument, at a generic order $O(\tau^k)$ the saddle point equation is the same but the renormalization of p at this order is

$$\frac{2p}{\chi}\, b_{q,k} \ .$$

The parameter $b_{q,k}$ is obtained using the approximation $\mu_2^{(k-1)}$, which is itself a function of $b_{q,k-1}$. We find

$$b_{q,k} = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{1-x^{k+1}}{1-x} - 1\right)\Big|_{x=\tau\left(\mu_2^{(k-1)} - \frac{1}{12}\right)} \ ,$$

and the convergence at $k \to \infty$ is guaranteed by the convergence of the geometric series defining the $T\overline{T}$-deformation. Although the study of the limiting value $b_{q,\infty}$ is based on exactly the same arguments as for b_∞ in [9], it is difficult to find explicit formulas due to the q-deformation. We provide more details on $b_{q,\infty}$ and an approximate study in the large p regime in Appendix 12.B.

Even without an explicit expression, we can extract information about $b_{q,\infty}$ from its defining equation

$$b_{q,\infty} = \left(1 - \tau\left(\mu_2 - \frac{1}{12}\right)\right)^{-2} \ , \tag{12.5.4}$$

with μ_2 depending itself on $b_{q,\infty}$. From this equation we already see that $b_{q,\infty} \geq 1$, with equality only at $\tau = 0$.

We also have to check the consistency of the $T\overline{T}$-deformation. Looking back at (12.3.1), we have to find for what values of τ the inequality

$$\tau\left(\mu_2 - \frac{1}{12}\right) < 1$$

is satisfied, so that the deformed Casimir is a well-defined (positive) deformation of the quadratic Casimir of $U(N)$. From (12.5.4), the left-hand side of this inequality is $1 - 1/\sqrt{b_{q,\infty}}$, with $b_{q,\infty} \geq 1$, hence the $T\overline{T}$-deformation is well-posed for all non-negative values of τ. Thus the theory is well-defined at large N all along the RG flow triggered by the $T\overline{T}$-deformation. This was not obvious from (12.3.1) and we regard it as a strong consistency check.

12.5.2 Critical curves

We now set χ equal to its physical value $\chi = \chi(\mathbb{S}^2) = 2$.

The solution we found in Section 12.5.1 above holds as long as ρ satisfies the requirement (12.4.8). From the formulas above, and the property $|\tan^{-1}(x)| \leq \frac{\pi}{2}$, we have

$$\rho(h) < \frac{p\, b_{q,\infty}}{2}$$

and the system undergoes a phase transition only for those values of $p > p_0$, where $p_0 := p_0(\tau)$ is implicitly defined by

$$p_0\, b_{q,\infty}(p_0, \tau) = 2 \ .$$

At $\tau = 0$, this was used in [498, 494, 409] to show that for $p \leq \chi(\mathbb{S}^2) = 2$ there is only one phase, while two phases separated by a critical line in the (λ, p)-plane appear for $p > 2$.

We have seen in Section 12.5.1 above that $b_{q,\infty} \geq 1$, which implies that for all $\tau > 0$ a phase transition takes place whenever $p > p_0$ with $p_0 < 2$. As soon as the $T\overline{T}$-deformation is turned on, the theory with $p = 2$ develops a strong coupling phase, with critical line descending from ∞ to a finite value of λ.

For $p > p_0(\tau)$, the theory presents two phases, separated by a codimension-one critical surface in the octant $(\lambda > 0, p > 0, \tau \geq 0)$, parametrized by $\lambda = \lambda_{\rm cr}(p, \tau)$. This surface should be seen as a one-parameter family of critical curves, parametrized by $\tau \geq 0$, describing the evolution of the critical curve of Figure 12.8 along the RG flow induced by the $T\overline{T}$-deformation, with τ playing the role of the "time". See Figure 12.9 for a schematic picture. (Note that Figure 12.9 represents just a rough illustration of the actual critical surface.)

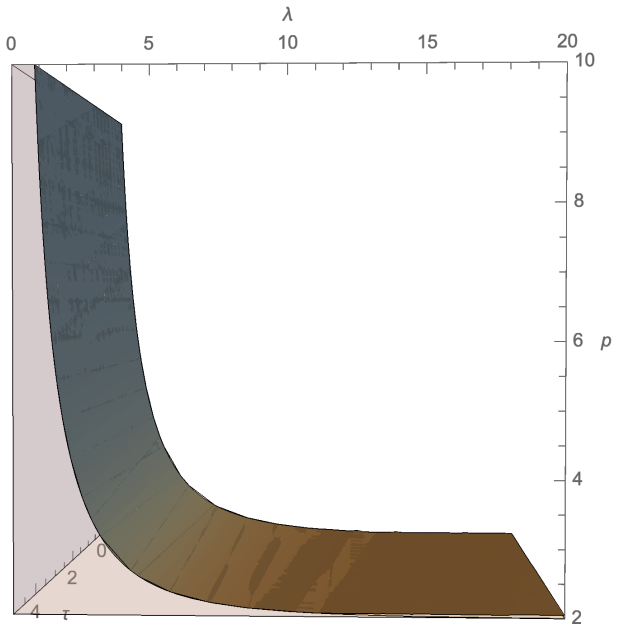


Figure 12.9. Schematic plot of the critical surface. The gray region represents the weak coupling phase.

This critical surface is defined implicitly by the equation

$$\lambda_{\rm cr} = p\, b_{q,\infty} \log\left(1 + \tan^2 \frac{\pi}{p\, b_{q,\infty}}\right) \ . \tag{12.5.5}$$

It is important to bear in mind that $b_{q,\infty}$ depends on λ, and it must be evaluated at the critical value $\lambda_{\rm cr}$ in the right-hand side of (12.5.5).

Without an explicit expression for $b_{q,\infty}$ we cannot provide a formula for $\lambda_{\rm cr} = \lambda_{\rm cr}(p,\tau)$ describing the critical surface. Nevertheless, the lessons learned from the study of Appendix 12.B are that $b_{q,\infty}$ is a monotonically decreasing function of p, eventually approaching b_∞ from above as $p \to \infty$. In conclusion, from Appendix 12.B we find that $b_{q,\infty}$ decreases with p, and from (12.5.4) we see that $b_{q,\infty}$ increases with τ. This matches precisely with the known effects of the two deformations taken separately. This, together with (12.5.5), implies

$$p^{-1}\, A_{\rm cr}(\tau) \leq \lambda_{\rm cr}(p,\tau) \leq \lambda_{\rm cr}(p,0) \ ,$$

which means that the volume in the octant $(\lambda > 0, p > 0, \tau \geq 0)$ of the parameter space describing a weak coupling phase is reduced, relative to only the $T\overline{T}$-deformation, and is enhanced relative to only the q-deformation.

$p = 1$ case

Since $b_{q,\infty}$ is an increasing function of τ, one may expect that, for τ sufficiently large, the phase transition also takes place at $p = 1$. In other words, one may wonder whether eventually $p_0(\tau) < 1$ for sufficiently large τ. Unfortunately, our analysis of $b_{q,\infty}$ is not reliable in the limit $p \to 1$, and we cannot draw any conclusions in this direction.

The parameter p has a geometric meaning as the degree of the holomorphic line bundle

$$\mathcal{O}(-p) \longrightarrow \Sigma \ .$$

For every $p > 1$, the total space is a resolution of the Kleinian (or du Val) singularity $\mathbb{C}^2/\mathbb{Z}_p$, singling out $p = 1$ as a special case. See also [499] for a discussion on the physical relevance of $p > 1$.

On the other hand, there is nothing special about $p = 1$ compared to $p = 2$ in the original construction of [487, 408].

The fate of the theory for $p = 1$ with $T\overline{T}$-deformation has been clarified in [500], showing that the phase transition is extended all the way down to $p = 1$ for appropriate values of τ.

12.5.3 Instanton analysis

A classic result of two-dimensional Yang–Mills theory on $\mathbb{S}^2$ is that the phase transition is triggered by instantons [385]; by 'instanton' here we mean a solution to the classical Yang–Mills equation of motion which is gauge-inequivalent to the trivial connection. In the weak coupling phase, the Boltzmann weight of the saddle point configuration dominates the partition function at large N. However, beyond a critical value of the coupling, non-perturbative contributions cease to be suppressed and compete with the Boltzmann weight, inducing a phase transition.

The same mechanism is at work in the $T\overline{T}$-deformed theory [9]. Here the deformation reduces the suppression factor of the instantons, and therefore the phase transition takes place at lower values of the coupling.

In q-Yang–Mills theory, again the unstable instantons are the cause of the phase transition [498, 409]. Here we show that the same arguments apply to the $T\overline{T}$-deformed theory. We find that, as in the situation without q-deformation, instantons are less suppressed in the $T\overline{T}$-deformed theory.

We start by rewriting the sphere partition function as

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}(\lambda,\tau) = \frac{1}{N!} \sum_{\vec{\ell} \in \mathbb{Z}^N} Z_{\vec{\ell}}(\lambda,\tau) \ , \tag{12.5.6}$$

where $Z_{\vec{\ell}}$ encodes the instanton contributions, and can be obtained from a modular transformation of the partition function written in the representation basis. See Appendix 12.C for further details.

The complete analysis for q-Yang–Mills theory was carried out in [498, 409], and we show how it is adapted to the $T\overline{T}$-deformed theory in Appendix 12.C. Although we cannot get a closed expression, we show that it can be evaluated order by order in τ.

As pointed out already in [385], focusing on the first instanton sector gives clearer insights into understanding how the non-perturbative effects kick in. We therefore consider the one-instanton sector, for which

$$\vec{\ell} = (\ell_1, 0, \dots, 0) \qquad \text{with} \quad \ell_1 = \pm 1 \ .$$

Its contribution to the partition function $\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}$ is

$$Z_{(\ell_1,0,\dots,0)}(\lambda,\tau) = \frac{1}{\Delta_q(\emptyset)^2} \int_{\mathbb{R}^N} \mathrm{d}\vec{h}\, e^{-2\pi\,\mathrm{i}\,\ell_1\,h_1} \prod_{1\le i<j\le N} 4\sinh^2\frac{\lambda\,(h_i-h_j)}{2N} \times \exp\left(-\frac{\lambda\,p}{2N} \frac{\sum\limits_{i=1}^N h_i^2 - \frac{N\,(N^2-1)}{12}}{1-\tau\left(\sum\limits_{i=1}^N h_i^2 - \frac{N\,(N^2-1)}{12}\right)} \right) .$$

In the large N limit, the contribution of ℓ_1 to the eigenvalue density is of order $O(N)$, hence sub-leading against the $O(N^2)$ contributions from the rest of the action. Therefore in the large N 't Hooft limit we can integrate over the eigenvalues $h_2, \dots, h_N$ using the eigenvalue density $\rho(h)$ found in Section 12.5.1 above. Rescaling the integration variable h_1 to $h = h_1/N$, we obtain

$$Z_{(\ell_1,0,\dots,0)}(\lambda,\tau) = \frac{\mathcal{Z}_{N-1}(\lambda,\tau)}{\Delta_q(\emptyset)^2} \int_{\mathbb{R}} \mathrm{d}h\, e^{-N\,S_{\text{eff}}[h]} \ ,$$

where $\mathcal{Z}_{N-1}$ comes from integrating out the remaining $N-1$ eigenvalues, and is equal to the partition function of the $U(N-1)$ theory in the zero-instanton sector.[39] The effective action functional is given by

$$S_{\text{eff}}[h] = -2 \int_{\text{supp}(\rho)} \mathrm{d}u\, \rho(u) \log\left|\sinh\frac{\lambda\,(h-u)}{2}\right| + \frac{\lambda\,p\,b_{q,\infty}}{2}\,h^2 - 2\pi\,\mathrm{i}\,\ell_1\,h \ ,$$

where we used the definition (12.5.4) of $b_{q,\infty}$ to simplify the expression. We obtain the saddle point equation for the first eigenvalue given by

$$p\,b_{q,\infty}\,h - \frac{2\pi\,\mathrm{i}\,\ell_1}{\lambda} = \mathrm{P}\int \mathrm{d}u\,\rho(u)\coth\frac{\lambda\,(h-u)}{2} \ . \tag{12.5.7}$$

This is a saddle point equation for h, with $\rho(u)$ known. At this point we notice that, as expected, (12.5.7) is the same equation found in [498, 494], except for the renormalization $p \mapsto p\,b_{q,\infty}$. We can therefore read off the solution from [498, 494] to get

$$h = \begin{cases} \frac{2\,\mathrm{i}\,\ell_1}{\lambda}\tan^{-1}\sqrt{\frac{e^{-\lambda/p\,b_{q,\infty}}}{\cos^2\frac{\pi}{p\,b_{q,\infty}}} - 1} \ , & p\,b_{q,\infty} > 2 \ , \\ \frac{2\pi\,\mathrm{i}\,\ell_1}{\lambda} \ , & p\,b_{q,\infty} \le 2 \ , \end{cases} \tag{12.5.8}$$

where we used $|\ell_1| = 1$ to simplify

$$\cosh\left(\frac{\lambda}{2p\,b_{q,\infty}}\,\frac{2\pi\,\mathrm{i}\,|\ell_1|}{\lambda}\right) = \cos\frac{\pi}{p\,b_{q,\infty}} \ .$$

[39] The symmetry breaking $U(N) \to U(1)\times U(N-1)$ in the one-instanton sector is explained in Appendix 12.C.

From (12.5.8) it follows that there is no phase transition for p below a critical value $p_0(\tau)$, defined such that

$$p\, b_{q,\infty}(\lambda, p, \tau) \leq 2 \qquad \text{for} \quad \lambda > 0 \qquad \text{when} \quad p \leq p_0(\tau) \ ,$$

because then the instanton contributions are suppressed for all values of λ. On the other hand, when $p > p_0(\tau)$ and $p\, b_{q,\infty} > 2$ we have (dropping an irrelevant overall constant)

$$\frac{Z_{(\ell_1,0,\dots,0)}(\lambda,\tau)}{Z_{(0,0,\dots,0)}(\lambda,\tau)} = \exp\left(-\frac{N}{\lambda p}\,\gamma(\lambda, p\, b_{q,\infty})\right)$$

where γ is the function defined in [498], which in turn is a one-parameter deformation of the suppression function found in [385]. When

$$\frac{e^{-\lambda/p\, b_{q,\infty}}}{\cos^2\frac{\pi}{p\, b_{q,\infty}}} - 1 > 0 \ ,$$

corresponding to $\lambda < \lambda_{\rm cr}$, the function γ is a positive decreasing function of λ, for any fixed p. However, it becomes purely imaginary when $\lambda > \lambda_{\rm cr}$, implying that the one-instanton sector is no longer suppressed and its contribution becomes relevant.

12.5.4 Strong coupling phase

For values of λ such that, for given p and τ, the eigenvalue density $\rho(h)$ found in Section 12.5.1 above does not satisfy the constraint (12.4.8), we have to drop the assumption of a one-cut solution and find another distribution $\rho(h)$ satisfying the bound.

The strategy is the same as that followed in Section 12.5.1 above for the weak coupling phase: expanding the saddle point equation (12.5.3) as a power series in τ, we can solve it iteratively. We do not spell out the technical details here, as they are exactly as in [498, 494, 409], up to the renormalization $p \mapsto \frac{2p}{\chi}\, d_{q,k}$. The coefficient

$$d_{q,k} = \sum_{j=0}^{k} (j+1)\, \tau^j \left(\mu_2^{(k-1)} - \frac{1}{12}\right)^j \tag{12.5.9}$$

is formally the same as $b_{q,k}$, but they differ in that $b_{q,k}$ is computed using the second moment μ_2 of the distribution $\rho(h)$ at weak coupling in (12.5.4), whereas for $d_{q,k}$ the moment μ_2 corresponds to $\rho(h)$ at strong coupling in (12.5.9). We distinguish the weak and strong coupling solutions by $\rho_{\rm weak}$ and $\rho_{\rm strong}$. The complete solution would require finding $d_{q,\infty}$.

Third order phase transition

Even without closed expressions available for $b_{q,\infty}$ and $d_{q,\infty}$, we can extract information from the form of the eigenvalue density and from what is known at $\tau = 0$. In fact, at $\tau = 0$ the phase transition is of third order [498, 494, 409], which means that $\log \mathcal{Z}_{q\text{-YM}}$ is twice continuously differentiable along the critical curve $\lambda = \lambda_{\rm cr}(p)$.

The crucial feature is that one can extract the derivative $\frac{\partial}{\partial\lambda} \log \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}$ from the second moment $\mu_2[\rho]$ in each phase, and a third order phase transition implies that $\mu_2[\rho_{\rm weak}]$ and $\mu_2[\rho_{\rm strong}]$ agree up to their first derivatives at $\tau = 0$. We now exploit this fact to describe the behaviours of $b_{q,\infty}$ and $d_{q,\infty}$ close to the critical curve. Using the defining expressions (12.5.4) and (12.5.9) for $b_{q,\infty}$ and $d_{q,\infty}$, we can expand in λ around $\lambda = \lambda_{\rm cr}$ to get

$$\begin{aligned} b_{q,\infty} &= f_0^{\rm weak}(b_{q,\infty}|_{\lambda=\lambda_{\rm cr}}) + (\lambda-\lambda_{\rm cr})^2\ f_2^{\rm weak}(b_{q,\infty}|_{\lambda=\lambda_{\rm cr}}) + O\big((\lambda-\lambda_{\rm cr})^3\big) \ , \\ d_{q,\infty} &= f_0^{\rm strong}(d_{q,\infty}|_{\lambda=\lambda_{\rm cr}}) + (\lambda-\lambda_{\rm cr})^2\ f_2^{\rm strong}(d_{q,\infty}|_{\lambda=\lambda_{\rm cr}}) + O\big((\lambda-\lambda_{\rm cr})^3\big) \ . \end{aligned}$$

The agreement of μ_2 at weak and strong coupling at $\tau = 0$ can be used to show that $f_0^{\text{weak}} = f_0^{\text{strong}}$ at the critical point for all $\tau \geq 0$, which in turn guarantees that $b_{q,\infty}$ and $d_{q,\infty}$ agree up to the first derivative. Essentially, by direct inspection one finds that the defining equation for $d_{q,\infty}$ is exactly the same as for $b_{q,\infty}$ at order $O(\lambda - \lambda_{\text{cr}})$.

On the other hand, one can check that, after inclusion of the renormalization of p at weak or strong coupling, the first derivative $\frac{\partial}{\partial\lambda} \log \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}$ depends only on $b_{q,\infty}$ (respectively $d_{q,\infty}$) at weak (respectively strong) coupling, and not on their derivatives. This is done again by expanding the integral formula for μ_2 close to λ_{cr}, and is a consequence of the very simple way in which the parameters $b_{q,\infty}$ and $d_{q,\infty}$ enter.

As an immediate consequence, the second derivative $\frac{\partial^2}{\partial\lambda^2} \log \mathcal{Z}^{T\overline{T}}_{q\text{-YM}}\big|_{\lambda=\lambda_{\text{cr}}}$ depends only on the 1-jets of $b_{q,\infty}$ and $d_{q,\infty}$ at λ_{cr}, which we have argued to match. We therefore find that the phase transition is of third order.

We refer to [9] for further discussion, since the details of the argument do not depend on the explicit form of $\rho(h)$ once we zoom in close to the critical curve.

12.5.5 Refinement

Consider the refinement of $U(N)$ q-Yang–Mills theory [486] with refinement parameter $t = q^\beta$, for $\beta \in \mathbb{Z}_{>0}$; the unrefined limit $t = q$ then corresponds to $\beta = 1$. A refined definition of the shifted weight variables $\vec{h} \in \mathbb{Z}^N$ introduced in (12.4.3) is given by

$$h_i = -R_i + \beta\left(i - \frac{N+1}{2}\right) \qquad \text{for} \quad i = 1, \dots, N \ .$$

In this basis we find that the $T\overline{T}$-deformed partition function on $\mathbb{S}^2$ is

$$\mathcal{Z}^{T\overline{T}}_{(q,\,t)\text{-YM}}(\lambda,\tau) = \frac{1}{N!} \sum_{\vec{h}\in\mathbb{Z}^N} \frac{\Delta_{(q,t)}(\vec{h})\,\Delta_{(q,t)}(-\vec{h})}{-\Delta_{(q,t)}(\emptyset)^2} \ \exp\left(-\frac{\frac{\lambda\,p}{2N}\left(\sum\limits_{i=1}^N h_i^2 - \beta^2\,\frac{N\,(N^2-1)}{12}\right)}{1 - \frac{\tau}{N^3}\left(\sum\limits_{i=1}^N h_i^2 - \beta^2\,\frac{N\,(N^2-1)}{12}\right)}\right) , \tag{12.5.10}$$

in which the $T\overline{T}$-deformation only changes the refined (or β-deformed) quadratic Casimir, relative to the case $\tau = 0$. The Macdonald measure $\Delta_{(q,t)}(\vec{h})$ in (12.5.10) is given for $t = q^\beta$ as

$$\Delta_{(q,t)}(\vec{h}) = \prod_{m=0}^{\beta-1} \prod_{1\leq i<j\leq N} 2\sinh\frac{\lambda\,(h_i - h_j + m)}{2N} \ ,$$

and $\Delta_{(q,t)}(\emptyset)$ is obtained by setting $h_i - h_j = \beta\,(i-j)$. This discrete matrix model has been thoroughly studied in [436].

It was proven in [501] that the 't Hooft limit of (q,t)-Yang–Mills theory coincides with that of q-Yang–Mills theory with rescaled coupling $\lambda' = \beta\,\lambda$. The proof of [501] is straightforwardly adapted to the $T\overline{T}$-deformed setting, and we find that q-Yang–Mills theory and its refinement coincide in the 't Hooft limit after rescaling

$$\lambda' = \beta\,\lambda \qquad \text{and} \qquad \tau' = \beta^2\,\tau \ .$$

When $\beta > 1$ the refinement non-trivially modifies the deformation parameter τ, and the weak coupling region in parameter space is drastically reduced relative to the unrefined case $\beta = 1$.

12.A Instantons in $T\overline{T}$-deformed Yang–Mills theory and Poisson resummation

This appendix is dedicated to a more accurate calculation of the instanton contributions in the present model. We will follow the procedure of [495] to evaluate the weights $w(\ell)$ in the $T\overline{T}$-deformed version of the theory (12.4.1). The full instanton expansion is obtained starting from the formula

$$\mathcal{Z}_N(A,\tau) = \sum_{\ell\in\mathbb{Z}^N} Z_\ell,$$

where Z_ℓ is the Fourier transform of the sector corresponding to a given representation, that is,

$$Z_\ell = \int \prod_i dh_i e^{-2\pi \mathrm{i}\sum_i \ell_i h_i} \prod_{i<j}\left(\frac{h_i-h_j}{i-j}\right)^2 e^{-\frac{A}{2N}\sum_{j=0}^{\infty}\left(\frac{\tau}{N^3}\right)^j\left[\sum_i\left(h_i-i+\frac{N}{2}\right)\left(h_i+i-\frac{N}{2}\right)\right]^{j+1}}. \tag{12.A.1}$$

We can expand the contribution arising from the $T\overline{T}$-deformation:

$$\begin{aligned} Z_\ell = \mathcal{C}\int \prod_i dh_i e^{-2\pi\mathrm{i}\sum_i \ell_i h_i}\prod_{i<j}(h_i-h_j)^2 e^{-\frac{A}{2N}\sum_i h_i^2} \\ \times\left\{\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{A}{2N}\right)^n\sum_{j=0}^{\infty}c_j(n)\left(\frac{\tau}{N^3}\right)^j\left[\sum_i h_i^2-\frac{N(N^2-1)}{12}\right]^j\right\}, \end{aligned} \tag{12.A.2}$$

where $\mathcal{C}$ is an irrelevant overall factor and $\{c_j(n)\}_j$ are the coefficients of the series expansion of $\left(\frac{x}{1-x}\right)^n$. Neglecting the shift in the Casimir, and putting the focus on the last term, we obtain its Fourier transform as:

$$\int dh_i e^{-2\pi\mathrm{i}\ell_i h_i-\frac{A}{2N}h_i^2}\left[\frac{\tau}{N^3}h_i^2\right]^j = \left(\frac{\tau}{N^3}\right)^j\Gamma\left(\frac{1}{2}+j\right){}_1F_1\left(\frac{1}{2}+j,\frac{1}{2},-\frac{N}{2A}\ell_i^2\right),$$

where ${}_1F_1$ is the confluent hypergeometric function.

The key observation to go further is that, inside the integral (12.A.2), three ingredients appear: the Gaussian measure, the Vandermonde determinant and a totally symmetric polynomial in the variables h_i^2. Therefore, performing the integration with a single Vandermonde determinant, which is a totally antisymmetric polynomial, one obtains again the Vandermonde multiplying some totally symmetric polynomial (or total symmetrisation of hypergeometric functions), up to overall constant factors. The result is indeed:

$$\begin{aligned} f_1(\ell) &:= \int\prod_i dh_i e^{-2\pi\mathrm{i}\sum_i\ell_i h_i-\frac{A}{2N}\sum_i h_i^2}\prod_{i<j}(h_i-h_j)\left\{\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{A}{2N}\right)^n\sum_{j=0}^{\infty}c_j(n)\left(\frac{\tau}{N^3}\right)^j\left[\sum_i h_i^2\right]^j\right\} \\ &= \mathcal{C}'\prod_{i<j}(\ell_i-\ell_j)e^{-\frac{N}{2A}\sum_i\ell_i^2}\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{A}{2N}\right)^n\sum_{j=0}^{\infty}c_j(n)\left(\frac{\tau}{N^3}\right)^j\mathcal{S}\left({}_1F_1\left(\frac{1}{2}+\nu_i+j,\frac{1}{2}+\tilde{\nu}_i,-\frac{N}{2A}\ell_i^2\right)\right), \end{aligned}$$

where $\mathcal{S}(x)$ is a totally symmetric polynomial of order N in N variables, and $\nu_i,\tilde{\nu}_i\in\mathbb{N}$. Therefore, the arguments presented in [495] hold also in this case. Taking care of the polynomials that appear due to the $T\overline{T}$-deformation, and one could retrieve the exact form of Z_ℓ, by taking the convolution of two expressions as given above

$$Z_\ell = (f_1 * f_1)(\ell).$$

Although the discussion presented in this Appendix is qualitative, it illustrates how the exact instanton contribution should in principle be obtainable following standard methods. In particular,

the shift in the quadratic Casimir can be reintroduced, and, using the binomial expansion, one can apply the calculations sketched here, taking care of the coefficients, to retrieve the exact contribution of each instanton sector.

12.B Approximate solution for $b_{q,\infty}$

In Section 12.5.1 we have studied the weak coupling phase in the large N limit of $T\overline{T}$-deformed q-Yang–Mills theory. Turning on the $T\overline{T}$-deformation amounts to replacing $p \mapsto p\, b_{q,\infty}$, where $b_{q,\infty}$ depends on p, λ and τ, and is implicitly determined by (12.5.4). In this appendix we analyze (12.5.4) in the small q-deformation regime, that is, $p \to \infty$ with $\lambda p = A$ fixed. We set $\chi = 2$; the χ-dependence is eventually reinstated by replacing p with $\frac{2p}{\chi}$.

Expanding (12.5.4) at large p we get

$$b_{q,\infty} = \left(1 - \tau\left(\frac{1}{A\, b_{q,\infty}} + \frac{1}{6p^2\, b_{q,\infty}^2} + \frac{A}{72p^4\, b_{q,\infty}^3} + O\big(p^{-6}\big) - \frac{1}{12}\right)\right)^{-2} \ . \tag{12.B.1}$$

At $O(p^{-2})$ this equation admits only one solution satisfying

$$\lim_{\tau\to 0} b_{q,\infty} = 1 \ ,$$

as required in order to recover the correct behaviour when the $T\overline{T}$-deformation is turned off. This solution is a decreasing function of p, for fixed A and τ, and converges from above to the corresponding quantity b_∞ of $T\overline{T}$-deformed Yang–Mills theory as p becomes large. The explicit expression for this solution is rather lengthy and cumbersome, hence we do not write it explicitly. Instead, we plot the solution as a function of p, for different values of A and τ, in Figures 12.10, 12.11 and 12.12.

From these plots we see the qualitative behaviour of $b_{q,\infty}$, and for all fixed choices of A and τ it converges asymptotically to the value b_∞. This guarantees that $b_{q,\infty}$ indeed arises as a q-deformation of b_∞, and b_∞ is correctly recovered in the limit (12.3.4). In particular, when $\tau > 0$ we have

$$1 < b_\infty < b_{q,\infty} \ ,$$

that is, $b_{q,\infty}$ is a decreasing function of p which converges to b_∞ from above.

We now look at how these features are altered when the $O(p^{-4})$ contribution is taken into account. Going to the next non-trivial order introduces a dependence on $b_{q,\infty}^{-3}$ in the right-hand side of (12.B.1). The values of $b_{q,\infty}$ are thus found by solving a degree seven polynomial equation, but six solutions will be spurious. We do not dive into an analytic approach, and instead numerically illustrate the behaviour of the solution as a function of p, for different values of τ. From what we have learnt at $O(p^{-2})$, it is sufficient to limit ourselves to $A = 1$ and a few values of τ. The solutions are plotted in Figures 12.13 and 12.14 for $\tau = 0.1$ and $\tau = 1$, respectively. We see that the solution is again a decreasing function of p, which changes rapidly for small p and is almost constant at large p. We also check that the solution approaches 1 as $\tau \to 0$, in agreement with our analytic study. The conclusions therefore remain unchanged after the inclusion of $O(p^{-4})$ corrections.

12.C Instantons in $T\overline{T}$-deformed q-Yang–Mills theory

In the analyses of [498, 409], where the Poisson resummation of the heat kernel expansion of the partition function is done explicitly for the q-deformed theory, it was found that the instantons are responsible for the phase transition also in the q-deformed case. An analogous result was

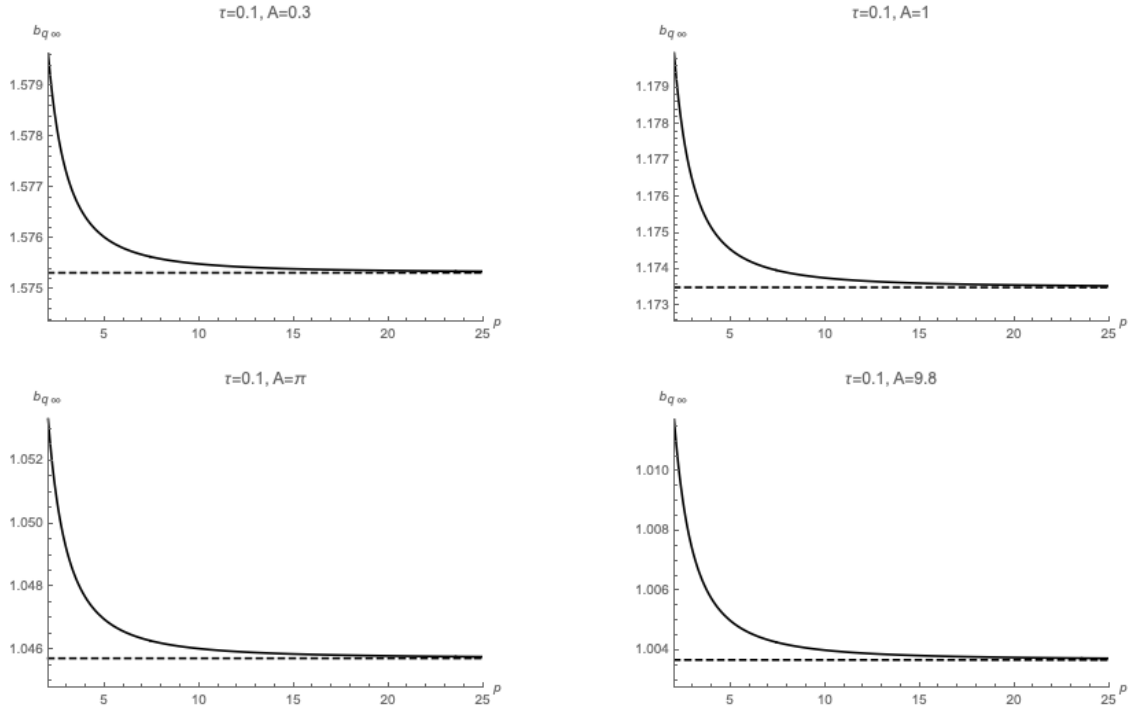


Figure 12.10. Plot of $b_{q,\infty}$ as a function of p at $\tau = 0.1$. The four plots correspond to $A = 0.3, 1, \pi, 9.8$. The dashed horizontal line is the corresponding value of b_∞ in $T\overline{T}$-deformed Yang–Mills theory without q-deformation.

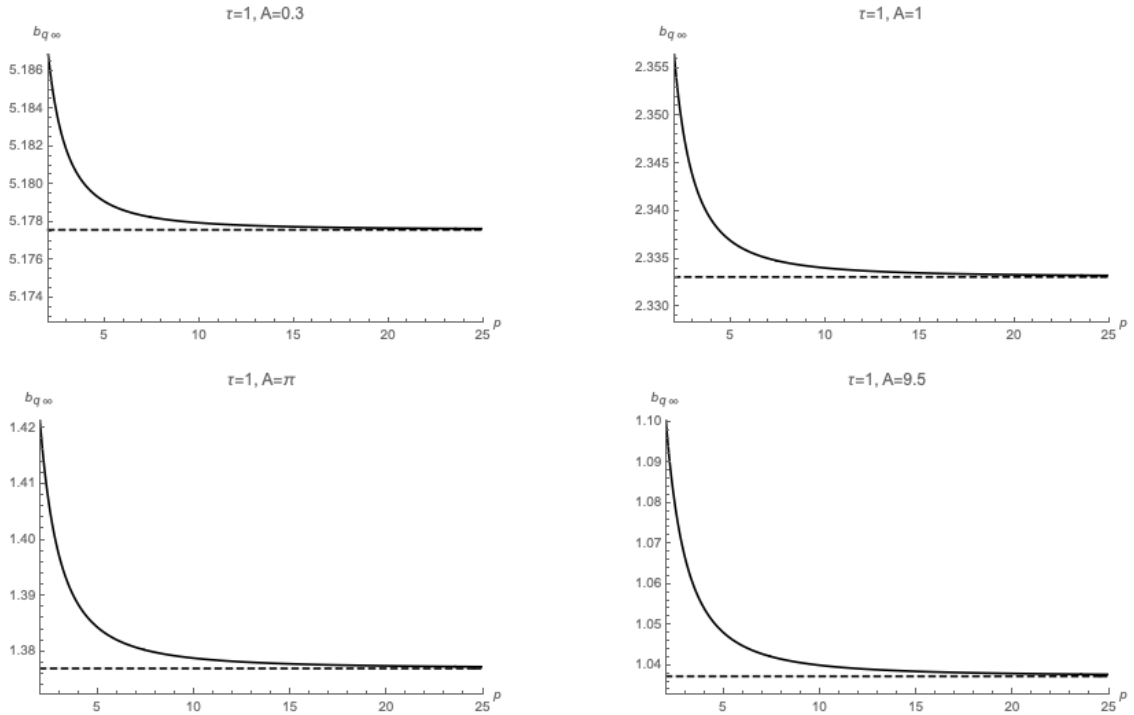


Figure 12.11. Plot of $b_{q,\infty}$ as a function of p at $\tau = 1$. The four plots correspond to $A = 0.3, 1, \pi, 9.5$. The dashed horizontal line is the corresponding value of b_∞ in $T\overline{T}$-deformed Yang–Mills theory without q-deformation.

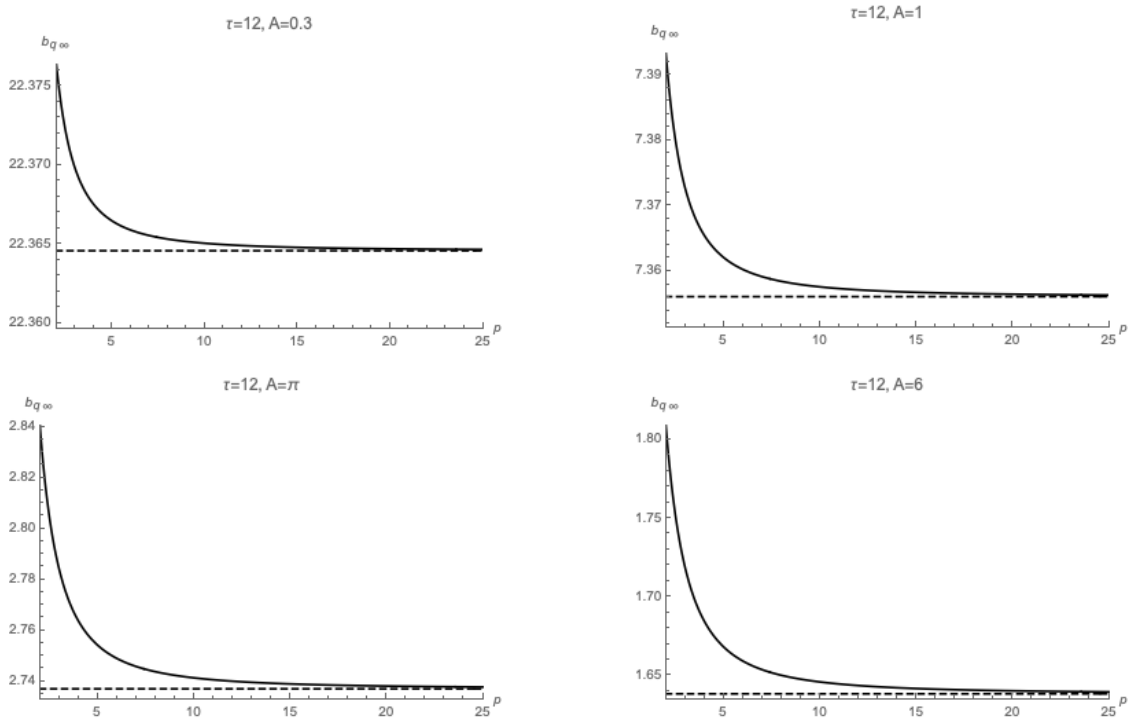


Figure 12.12. Plot of $b_{q,\infty}$ as a function of p at $\tau = 12$. The four plots correspond to $A = 0.3, 1, \pi, 6$. The dashed horizontal line is the corresponding value of b_∞ in $T\overline{T}$-deformed Yang–Mills theory without q-deformation.

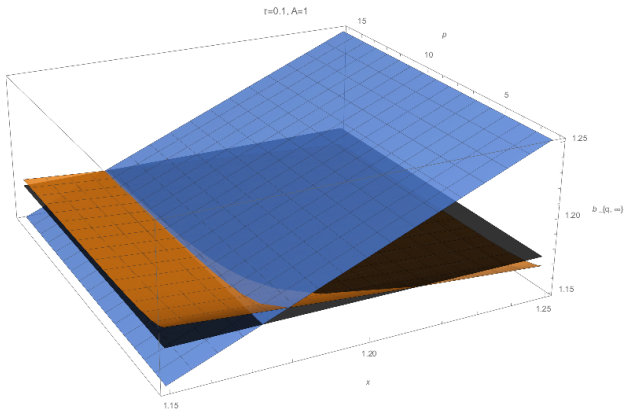


Figure 12.13. Plot of $b_{q,\infty}$ as a function of p, at $A = 1$ and $\tau = 0.1$. The blue surface is the function x on the left-hand side of (12.B.1). The orange surface is the right-hand side of (12.B.1) as a function of p and x. $b_{q,\infty}$ is determined by the intersection of these two surfaces. The black horizontal surface is the asymptotic value b_∞.

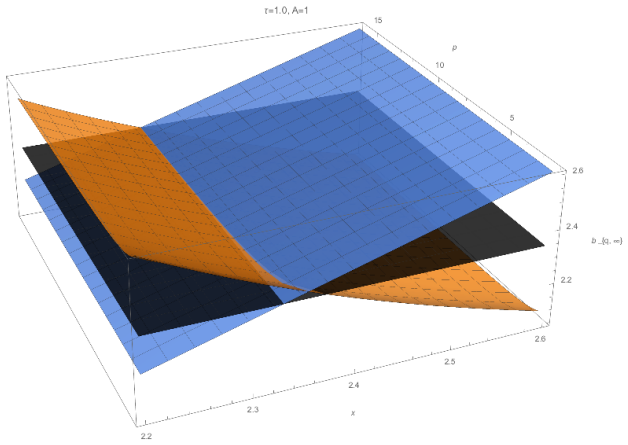


Figure 12.14. Plot of $b_{q,\infty}$ as a function of p, at $A = 1$ and $\tau = 1.0$. The blue surface is the function x on the left-hand side of (12.B.1). The orange surface is the right-hand side of (12.B.1) as a function of p and x. $b_{q,\infty}$ is determined by the intersection of these two surfaces. The black horizontal surface is the asymptotic value b_∞.

presented in [494], where only the first instanton sector was taken into account. From the results of [9], we expect that this property is not affected by the $T\overline{T}$-deformation.

Two-dimensional Yang–Mills instantons

Let us start with some generalities concerning instantons in two-dimensional Yang–Mills theory. We start by rewriting the partition function of $T\overline{T}$-deformed q-Yang–Mills theory on $\mathbb{S}^2$ as in (12.5.6):

$$\mathcal{Z}^{T\overline{T}}_{q\text{-YM}}(\lambda,\tau) = \frac{1}{N!}\sum_{\vec{\ell}\in\mathbb{Z}^N} Z_{\vec{\ell}}(\lambda,\tau)\ , \tag{12.C.1}$$

where $Z_{\vec{\ell}}$ encodes the contribution of an instanton labelled by $\vec{\ell} = (\ell_1,\ell_2,\dots,\ell_N) \in \mathbb{Z}^N$. Two-dimensional $U(N)$ Yang–Mills instantons are given by diagonal $\mathfrak{u}(N)$-valued gauge fields

$$A = \mathrm{diag}\,(A_{\ell_1}, A_{\ell_2}, \dots, A_{\ell_N})$$

where A_{ℓ_i} is a Dirac monopole potential of charge $\ell_i \in \mathbb{Z}$. Each entry is a gauge connection in the monopole bundle over $\mathbb{S}^2$ of magnetic charge ℓ_i,

$$\mathcal{L}^{\otimes\ell_i} \longrightarrow \mathbb{S}^2\ ,$$

where $\mathcal{L}$ is the canonical line bundle of $\mathbb{P}^1$ (we identify $\mathbb{P}^1 \cong \mathbb{S}^2$). Each instanton configuration determines a splitting of the $U(N)$ gauge bundle on $\mathbb{S}^2$, which in turn describes a symmetry breaking

$$U(N) \longrightarrow \prod_{l\in\mathbb{Z}} U(N_l)\ ,$$

where N_l encodes the degeneracy of the magnetic charges: N_l counts how many times the integer $l \in \mathbb{Z}$ appears in the string $\vec{\ell} = (\ell_1,\ell_2,\dots,\ell_N) \in \mathbb{Z}^N$ (we omit factors in the product with $N_l = 0$). A generic configuration breaks the gauge group $U(N)$ to its maximal torus $U(1)^N$, while the one-instanton sector describes a soft breaking $U(N) \to U(1)\times U(N-1)$. The trivial connection $A = 0$ (the zero-instanton sector) is the only gauge field preserving the full $U(N)$ symmetry.

The restriction to gauge-inequivalent configurations reduces the coweights $\vec{\ell} \in \mathbb{Z}^N$ to the Weyl chamber

$$\ell_1 \geq \ell_2 \geq \cdots \geq \ell_N\ ,$$

but from the symmetry of the partition function we can drop this restriction at the cost of an overall factor $(N!)^{-1}$.

Complete instanton partition function

We will now focus on the instanton expansion (12.C.1) and give the first steps towards understanding the complete instanton partition function, including all instanton contributions. However, due to the sophistication of the $T\overline{T}$-deformation, we cannot provide a complete answer and so we only present a partial analysis here.

Each summand $Z_{\vec{\ell}}$ in (12.C.1) can be computed as the Fourier transform of the contribution Z_R of an irreducible $U(N)$ representation R:

$$Z_{\vec{\ell}}(\lambda,\tau) = \frac{1}{\Delta_q(\emptyset)^2}\int_{\mathbb{R}^N}\left(\prod_{i=1}^N \mathrm{d}h_i\, e^{-2\pi\,\mathrm{i}\,\ell_i\,h_i}\right)\prod_{1\leq i<j\leq N} 4\sinh^2\frac{\lambda\,(h_i-h_j)}{2N} \times\ \exp\left(-\frac{\frac{\lambda p}{2N}\left(\sum\limits_{j=1}^N h_j^2-\frac{N\,(N^2+1)}{12}\right)}{1-\frac{\tau}{N^3}\left(\sum\limits_{j=1}^N h_j^2-\frac{N\,(N^2+1)}{12}\right)}\right).$$

We can expand the effect of the $T\overline{T}$-deformation in a double power series, finding

$$Z_{\vec{\ell}}(\lambda,\tau) = K_{N,\lambda,p}\int_{\mathbb{R}^N}\left(\prod_{i=1}^N \mathrm{d}h_i\, e^{-\frac{\lambda p}{2N}\,(h_i^2+\frac{4\pi\,\mathrm{i}\,N}{\lambda p}\,\ell_i\,h_i)}\right)\prod_{1\leq i<j\leq N} 4\sinh^2\frac{\lambda\,(h_i-h_j)}{2N} \times\ \sum_{n=0}^\infty\frac{1}{n!}\left(-\frac{\lambda p}{2N}\right)^n\sum_{k=0}^\infty c_k(n)\left(\frac{\tau}{N^3}\right)^k\left(\sum_{i=1}^N h_i^2-\frac{N\,(N^2-1)}{12}\right)^k,$$

where $\{c_k(n)\}$ are the coefficients of the Taylor expansion of the function $\left(\frac{x}{1-x}\right)^n$ around $x=0$, and

$$K_{N,\lambda,p} = e^{\frac{\lambda p\,(N^2-1)}{24N}}\,\Delta_q(\emptyset)^{-2}$$

is an overall factor. A direct computation of $Z_{\vec{\ell}}$ is difficult already for $\tau=0$. However, by exploiting the Weyl denominator formula, we can compute a different function $\tilde{Z}_{\vec{\ell}}$, which we define in the same way as $Z_{\vec{\ell}}$ but with only a single power of the q-deformed Vandermonde determinant $\Delta_q(\vec{h})$, instead of the square $\Delta_q(\vec{h})^2$ which enters the expression for $Z_{\vec{\ell}}$. Explicitly,

$$\tilde{Z}_{\vec{\ell}}(\lambda,\tau) = K_{N,\lambda,p}\int_{\mathbb{R}^N}\left(\prod_{i=1}^N \mathrm{d}h_i\, e^{-\frac{\lambda p}{2N}\,(h_i^2+\frac{4\pi\,\mathrm{i}\,N}{\lambda p}\,\ell_i\,h_i)}\right)\prod_{1\leq i<j\leq N} 2\sinh\frac{\lambda\,(h_i-h_j)}{2N} \times\ \sum_{n=0}^\infty\frac{1}{n!}\left(-\frac{\lambda p}{2N}\right)^n\sum_{k=0}^\infty c_k(n)\left(\frac{\tau}{N^3}\right)^k\left(\sum_{i=1}^N h_i^2-\frac{N\,(N^2-1)}{12}\right)^k. \tag{12.C.2}$$

At $\tau=0$, the Weyl denominator formula gives [498, 409]

$$\tilde{Z}_{\vec{\ell}}(\lambda,0) = K'_{N,\lambda,p}\,e^{-\frac{2\pi^2 N}{\lambda p}\sum_{i=1}^N \ell_i^2}\sum_{1\leq i<j\leq N}\sigma_{ij}\sin\left(\frac{\ell_i-\ell_j}{2p}\right),$$

where $\sigma_{ij}=+1$ if the permutation of the first N integers which sends i and j to the first and second positions respectively is even, and $\sigma_{ij}=-1$ if the permutation is odd. Here $K'_{N,\lambda,p}$ is another overall constant that we will not keep track of.

Turning on the $T\overline{T}$-deformation corresponds to introducing powers of the original quadratic Casimir, which in the expansion in (12.C.2) corresponds to introducing the terms with $k > 0$. The Fourier transform of each summand in (12.C.2) then gives

$$\begin{aligned}\tilde{Z}_{\vec{\ell}}(\lambda,\tau) = \tilde{K}_{N,\lambda,p}\, e^{-\frac{2\pi^2 N}{\lambda p}\sum_{i=1}^{N}\ell_i^2} \sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{\lambda p}{2N}\right)^n \sum_{k=0}^{\infty} c_k(n)\left(\frac{\tau}{N^3}\right)^k \\ \times \sum_{1\le i<j\le N}\sigma_{ij}\left(P^{\mathrm{s}}_{2k}(\vec{\ell})\sin\left(\frac{\ell_i-\ell_j}{2p}\right) - (\ell_i-\ell_k)\,P^{\mathrm{c}}_{2k}(\vec{\ell})\cos\left(\frac{\ell_i-\ell_j}{2p}\right)\right)\,,\end{aligned} \tag{12.C.3}$$

where P^{s}_{2k} and P^{c}_{2k} are totally symmetric polynomials of degree $2k$ in N variables, which can be explicitly computed order by order in τ.

At this point, the expression for $Z_{\vec{\ell}}$ could be obtained by the Fourier convolution of two functions of the form (12.C.3). The explicit calculation is rather cumbersome and should be performed order by order in τ; we do not attempt it here. However, we stress that, for each instanton sector $\vec{\ell}$, every order in the perturbative expansion can in principle be evaluated with the generalization of the strategy of [498, 409] that we have just sketched.

Notation and conventions

For completeness, we collect here definitions and conventions used through the main text. We closely follow the standard conventions in the random matrix literature.

Basic mathematical notation

$\mathbb{R}$ and $\mathbb{C}$ are the fields of real and complex numbers, respectively, $\mathbb{N}$ indicates the natural numbers and $\mathbb{Z}$ the integers. To avoid possible confusions, we sometimes adopt notations such as $\mathbb{Z}_{\geq 0}$, $\mathbb{Z}_{>0}$ and so on. Besides, $\mathbb{S}^d$ is the d-dimensional sphere, $\mathbb{T}^d \cong \left(\mathbb{S}^1\right)^d$ the d-dimensional torus, and $\mathbb{P}^d$ is the d-complex dimensional projective space.

We alternatively use the symbols $x \to x_0^-$ and $x \uparrow x_0$ to represent the limit of x to x_0 taken from below, and likewise for $x \to x_0^+$ and $x \downarrow x_0$ for the limit taken from above.

The notation $\mathbb{1}_{\mathscr{S}}$ represents the indicator function on the domain $\mathscr{S}$, defined as $\mathbb{1}_{\mathscr{S}}(x) = 1$ if $x \in \mathscr{S}$ and $\mathbb{1}_{\mathscr{S}}(x) = 0$ otherwise. This should not be confused with the identity endomorphism, denoted Id in Chapter 3.

Geometric conventions

We extensively use the existence of a canonical Kähler structure on every closed Riemann surface and implicitly endow the Riemann surface with it, when needed. Expressions such as $\mathbb{S}^2 \cong \mathbb{P}^1$ are to be interpreted in this vein.

Random matrix theory conventions

In the context of matrix models, the partition function is defined to be the N-fold eigenvalue integral over the whole support of the weight function. In this dissertation it is always denoted by $\mathcal{Z}$, possibly decorated with subscripts and/or superscripts indicating specifics of the theory under consideration. Explicitly, for a weight function $w(u)$ and integration cycle $\mathcal{C}$ we define

$$\mathcal{Z} = \frac{1}{N!} \int_{\mathcal{C}^N} \prod_{1 \leq a < b \leq N} |u_a - u_b|^2 \prod_{a=1}^{N} w(u_a) \, \mathrm{d}u_a .$$

The free energy is $\mathcal{F} \propto \frac{1}{N^2} \log \mathcal{Z}$, with N-independent proportionality factor. The symbol log means the principal branch of the complex logarithm, even if we typically deal with real-valued $\mathcal{Z}$, so most of the times log could be replaced by ln for practical purposes.[40] Whenever the sign matters (i.e. when the integration domain is non-compact) $\mathcal{F}$ is taken with opposite sign with respect to $\log \mathcal{Z}$. This choice is customary in most instances, but differs from some of the literature on 5d gauge theories, as discussed and motivated in Chapter 6. The reason for the $1/N^2$ factor is

[40] The imaginary part of $\mathcal{F}$ is nevertheless of physical interest, because it encodes mixed background Chern–Simons couplings.

to obtain a finite result at large N. For this precise reason, in some chapters additional prefactors have been included in the definition of $\mathcal{F}$, depending on the specific limit considered.

Whenever a large N limit is involved, ρ denotes the density of eigenvalues, which is compactly supported. All the limits are to be understood in a distributional sense: we will not claim ρ to be a function, but rather we determine a probability measure $\rho(u)\mathrm{d}u$ on some compact domain determined by the problem. In this way, $\mathcal{F}$ is always a function, differentiable at least once if $\rho(u)$ depends continuously on the parameters. In most cases, stronger results for ρ can be obtained by mild adaptations of existing rigorous results to the cases of interest to us, but this is not true in general. Since ρ is not a physical observable, this subtlety does not alter the final results, hence we do not seek sharper statements than convergence in distribution.

Amount of supersymmetry

The standard way to specify the amount of supersymmetry in a QFT in d dimensions is by indicating the number $\mathcal{N}$ of Killing spinors, whose component are contracted with the supercharges. While counting supercharges is unambiguous, the number of components of a spinor vary according to the dimension, and also according to the possible realizations of supersymmetry algebras in the given dimension. Here we write down the dictionary explicitly for completeness.

- In 6d, the $\mathcal{N} = (2,0)$ theory has sixteen supercharges and is maximally supersymmetric. The $\mathcal{N} = (1,0)$ theories have eight supercharges and hence are less constrained.
- In 5d, $\mathcal{N} = 1$ theories have eight supercharges. The only realization of 5d $\mathcal{N} = 2$ theories comes from the 6d $\mathcal{N} = (2,0)$ compactified on a circle.
- In 4d, $\mathcal{N} = 1$ means four supercharges, $\mathcal{N} = 2$ is eight supercharges and $\mathcal{N} = 4$ is the maximally supersymmetric theory with sixteen supercharges.
- In 3d, $\mathcal{N} = 2$ means four supercharges. Likewise, $\mathcal{N} = 4$ means eight supercharges. ABJ(M) theories have enhanced $\mathcal{N} = 6$ at the superconformal point, thus twelve supercharges. The maximally supersymmetric theory with sixteen supercharges has $\mathcal{N} = 8$, and admits various UV realizations, including the dimensional reduction of 4d $\mathcal{N} = 4$ or from ABJM theory with Chern–Simons level $k = 1$.

Bibliography

[1] L. Santilli and M. Tierz, *Exact results and Schur expansions in quiver Chern-Simons-matter theories*, *JHEP* **10** (2020) 022 [2008.00465].

[2] L. Santilli and M. Tierz, *Complex (super)-matrix models with external sources and q-ensembles of Chern-Simons and ABJ(M) type*, *J. Phys. A* **53** (2020) 425201 [1805.10543].

[3] L. Santilli and M. Tierz, *Phase transitions and Wilson loops in antisymmetric representations in Chern-Simons-matter theory*, *J. Phys. A* **52** (2019) 385401 [1808.02855].

[4] L. Santilli, *Phases of five-dimensional supersymmetric gauge theories*, *JHEP* **07** (2021) 088 [2103.14049].

[5] L. Santilli and M. Tierz, SQED_3 *and* SQCD_3*: Phase transitions and integrability*, *Phys. Rev. D* **100** (2019) 061702 [1906.09917].

[6] L. Santilli and M. Tierz, *Exact equivalences and phase discrepancies between random matrix ensembles*, *J. Stat. Mech.* **2008** (2020) 083107 [2003.10475].

[7] L. Santilli and M. Tierz, *Multiple phases and meromorphic deformations of unitary matrix models*, *Nucl. Phys. B* **976** (2022) 115694 [2102.11305].

[8] L. Santilli, R. J. Szabo and M. Tierz, *Five-dimensional cohomological localization and squashed q-deformations of two-dimensional Yang-Mills theory*, *JHEP* **06** (2020) 036 [2003.09411].

[9] L. Santilli and M. Tierz, *Large N phase transition in* $T\overline{T}$ *-deformed 2d Yang-Mills theory on the sphere*, *JHEP* **01** (2019) 054 [1810.05404].

[10] L. Santilli, R. J. Szabo and M. Tierz, $T\overline{T}$*-deformation of q-Yang-Mills theory*, *JHEP* **11** (2020) 086 [2009.00657].

[11] L. Santilli and M. Tierz, *Phase transition in complex-time Loschmidt echo of short and long range spin chain*, *J. Stat. Mech.* **2006** (2020) 063102 [1902.06649].

[12] L. Santilli and M. Tierz, *Riemannian Gaussian distributions, random matrix ensembles and diffusion kernels*, *Nucl. Phys. B* **973** (2021) 115582 [2011.13680].

[13] L. Santilli and M. Tierz, *Schur expansion of random-matrix reproducing kernels*, *J. Phys. A* **54** (2021) 435202 [2106.04168].

[14] L. Santilli and M. Tierz, *Crystal bases and three-dimensional* $\mathcal{N} = 4$ *Coulomb branches*, *JHEP* **03** (2022) 073 [2111.05206].

[15] P. Deligne, P. Etingof, D. S. Freed, L. C. Jeffrey, D. Kazhdan, J. W. Morgan, D. R. Morrison and E. Witten, eds., *Quantum fields and strings: a course for mathematicians. Vol. 1, 2*. AMS, Providence, RI, 1999.

[16] T. Eguchi, P. B. Gilkey and A. J. Hanson, *Gravitation, Gauge Theories and Differential Geometry*, *Phys. Rept.* **66** (1980) 213.

[17] G. 't Hooft, *A Planar Diagram Theory for Strong Interactions*, *Nucl. Phys. B* **72** (1974) 461.

[18] E. Brezin, C. Itzykson, G. Parisi and J. B. Zuber, *Planar Diagrams*, *Commun. Math. Phys.* **59** (1978) 35.

[19] S. Coleman, *Aspects of Symmetry: Selected Erice Lectures*. Cambridge University Press, Cambridge, U.K., 1985, 10.1017/CBO9780511565045.

[20] E. Witten, *Baryons in the 1/n Expansion*, *Nucl. Phys. B* **160** (1979) 57.

[21] Y. Y. Goldschmidt, *1/N Expansion in Two-dimensional Lattice Gauge Theory*, *J. Math. Phys.* **21** (1980) 1842.

[22] M. Billò, M. Caselle, A. D'Adda and S. Panzeri, *Finite temperature lattice QCD in the large N limit*, *Int. J. Mod. Phys. A* **12** (1997) 1783 [hep-th/9610144].

[23] T. DeGrand, *Lattice baryons in the 1/N expansion*, *Phys. Rev. D* **86** (2012) 034508 [1205.0235].

[24] B. Lucini and M. Panero, *SU(N) gauge theories at large N*, *Phys. Rept.* **526** (2013) 93 [1210.4997].

[25] M. García Pérez, *Prospects for large N gauge theories on the lattice*, *PoS* **LATTICE2019** (2020) 276 [2001.10859].

[26] I. R. Klebanov, *String theory in two-dimensions*, in *Spring School on String Theory and Quantum Gravity*, 7, 1991, hep-th/9108019.

[27] P. Di Francesco, P. H. Ginsparg and J. Zinn-Justin, *2-D Gravity and random matrices*, *Phys. Rept.* **254** (1995) 1 [hep-th/9306153].

[28] J. M. Maldacena, *The Large N limit of superconformal field theories and supergravity*, *Adv. Theor. Math. Phys.* **2** (1998) 231 [hep-th/9711200].

[29] S. S. Gubser, I. R. Klebanov and A. M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett. B* **428** (1998) 105 [hep-th/9802109].

[30] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150].

[31] J. M. Maldacena, *Wilson loops in large N field theories*, *Phys. Rev. Lett.* **80** (1998) 4859 [hep-th/9803002].

[32] D. Bessis, C. Itzykson and J. B. Zuber, *Quantum field theory techniques in graphical enumeration*, *Adv. Appl. Math.* **1** (1980) 109.

[33] J. Zinn-Justin, *Quantum field theory and critical phenomena*, vol. 77 of *Int.l Series of Monographs on Physics*. Oxford University Press, Oxford, UK, 1989.

[34] J. Zinn-Justin, *Phase transitions and renormalization group*, Oxford graduate texts. Oxford University Press, Oxford, UK, 2007.

[35] E. Witten, *Introduction to cohomological field theories*, *Int. J. Mod. Phys. A* **6** (1991) 2775.

[36] E. Witten, *Topological Quantum Field Theory*, *Commun. Math. Phys.* **117** (1988) 353.

[37] R. J. Szabo, *Equivariant localization of path integrals*. Springer, 2000, 10.1007/3-540-46550-2, [hep-th/9608068].

[38] V. Pestun and M. Zabzine, *Introduction to localization in quantum field theory*, *J. Phys. A* **50** (2017) 443001 [1608.02953].

[39] M. L. Mehta, *Random Matrices*, Pure and Applied Mathematics. Elsevier Science, 2004.

[40] P. J. Forrester, *Log-gases and random matrices*, vol. 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010, 10.1515/9781400835416.

[41] G. Akemann, J. Baik and P. Di Francesco, *The Oxford Handbook of Random Matrix Theory*, Oxford Handbooks in Mathematics. Oxford University Press, 2011.

[42] E. Witten, *Quantum Field Theory and the Jones Polynomial*, *Commun. Math. Phys.* **121** (1989) 351.

[43] O. Aharony, O. Bergman, D. L. Jafferis and J. Maldacena, *N=6 superconformal Chern-Simons-matter theories, M2-branes and their gravity duals*, *JHEP* **10** (2008) 091 [0806.1218].

[44] O. Aharony, O. Bergman and D. L. Jafferis, *Fractional M2-branes*, *JHEP* **11** (2008) 043 [0807.4924].

[45] A. Cavaglià, S. Negro, I. M. Szécsényi and R. Tateo, *$T\bar{T}$-deformed 2D Quantum Field Theories*, *JHEP* **10** (2016) 112 [1608.05534].

[46] F. A. Smirnov and A. B. Zamolodchikov, *On space of integrable quantum field theories*, *Nucl. Phys.* **B915** (2017) 363 [1608.05499].

[47] A. Berele and A. Regev, *Hook Young-diagrams with applications to combinatorics and to representations of Lie-superalgebras*, *Adv. Math.* **64** (1987) 118.

[48] E. Langmann, R. J. Szabo and K. Zarembo, *Exact solution of noncommutative field theory in background magnetic fields*, *Phys. Lett.* **B569** (2003) 95 [hep-th/0303082].

[49] E. Langmann, R. J. Szabo and K. Zarembo, *Exact solution of quantum field theory on noncommutative phase spaces*, *JHEP* **01** (2004) 017 [hep-th/0308043].

[50] N. Seiberg and E. Witten, *Gauge dynamics and compactification to three-dimensions*, in *Conference on the Mathematical Beauty of Physics (In Memory of C. Itzykson)*, 6, 1996, hep-th/9607163.

[51] M. Bullimore, T. Dimofte and D. Gaiotto, *The Coulomb Branch of 3d $\mathcal{N} = 4$ Theories*, *Commun. Math. Phys.* **354** (2017) 671 [1503.04817].

[52] A. Hanany and E. Witten, *Type IIB superstrings, BPS monopoles, and three-dimensional gauge dynamics*, *Nucl. Phys. B* **492** (1997) 152 [hep-th/9611230].

[53] A. Braverman, M. Finkelberg and H. Nakajima, *Coulomb branches of* 3*d* $\mathcal{N} = 4$ *quiver gauge theories and slices in the affine Grassmannian*, *Adv. Theor. Math. Phys.* **23** (2019) 75 [1604.03625].

[54] M. Kashiwara, *On crystal bases*, in *Canadian Math. Conf. Proc.*, vol. 16, (Providence, RI), p. 155–197, AMS, 1995.

[55] D. Bump and A. Schilling, *Crystal Bases*. World Scientific, Singapore, 2017, 10.1142/9876.

[56] V. Pestun, *Localization of gauge theory on a four-sphere and supersymmetric Wilson loops*, *Commun. Math. Phys.* **313** (2012) 71 [0712.2824].

[57] V. Pestun et al., *Localization techniques in quantum field theories*, *J. Phys.* **A50** (2017) 440301 [1608.02952].

[58] A. Kapustin, B. Willett and I. Yaakov, *Exact Results for Wilson Loops in Superconformal Chern-Simons Theories with Matter*, *JHEP* **03** (2010) 089 [0909.4559].

[59] B. Willett, *Localization on three-dimensional manifolds*, *J. Phys.* **A50** (2017) 443006 [1608.02958].

[60] M. Mariño, *Lectures on localization and matrix models in supersymmetric Chern-Simons-matter theories*, *J. Phys.* **A44** (2011) 463001 [1104.0783].

[61] L. J. Mordell, *The definite integral* $\int_{-\infty}^{\infty} \frac{e^{ax^2+bx}}{e^{cx}+d} dx$ *and the analytic theory of numbers*, *Acta Math.* **61** (1933) 323.

[62] J. G. Russo, G. A. Silva and M. Tierz, *Supersymmetric U(N) Chern–Simons-Matter Theory and Phase Transitions*, *Commun. Math. Phys.* **338** (2015) 1411 [1407.4794].

[63] G. Giasemidis and M. Tierz, *Mordell integrals and Giveon-Kutasov duality*, *JHEP* **01** (2016) 068 [1511.00203].

[64] J. G. Russo and F. A. Schaposnik, $\mathcal{N} = 2$ *Chern-Simons-matter theories without vortices*, *JHEP* **07** (2017) 062 [1704.03266].

[65] D. L. Jafferis and X. Yin, *Chern-Simons-Matter Theory and Mirror Symmetry*, 0810.1243.

[66] K. Okuyama, *A Note on the Partition Function of ABJM theory on* S^3, *Prog. Theor. Phys.* **127** (2012) 229 [1110.3555].

[67] J. G. Russo and G. A. Silva, *Exact partition function in* $U(2) \times U(2)$ *ABJM theory deformed by mass and Fayet-Iliopoulos terms*, *JHEP* **12** (2015) 092 [1510.02957].

[68] H. Awata, S. Hirano and M. Shigemori, *The Partition Function of ABJ Theory*, *PTEP* **2013** (2013) 053B04 [1212.2966].

[69] M. Mariño and P. Putrov, *Exact Results in ABJM Theory from Topological Strings*, *JHEP* **06** (2010) 011 [0912.3074].

[70] M. Honda, *Direct derivation of "mirror" ABJ partition function*, *JHEP* **12** (2013) 046 [1310.3126].

[71] D. Gaiotto and A. Tomasiello, *The gauge dual of Romans mass*, *JHEP* **01** (2010) 015 [0901.0969].

[72] N. Drukker et al., *Roadmap on Wilson loops in 3d Chern–Simons-matter theories*, *J. Phys.* **A53** (2020) 173001 [1910.00588].

[73] N. Drukker and D. Trancanelli, *A Supermatrix model for N=6 super Chern-Simons-matter theory*, *JHEP* **02** (2010) 058 [0912.3006].

[74] D. Martelli and J. Sparks, *Moduli spaces of Chern-Simons quiver gauge theories and AdS(4)/CFT(3)*, *Phys. Rev.* **D78** (2008) 126005 [0808.0912].

[75] Y. Imamura and K. Kimura, *On the moduli space of elliptic Maxwell-Chern-Simons theories*, *Prog. Theor. Phys.* **120** (2008) 509 [0806.3727].

[76] D. L. Jafferis and A. Tomasiello, *A Simple class of N=3 gauge/gravity duals*, *JHEP* **10** (2008) 101 [0808.0864].

[77] F. Benini, C. Closset and S. Cremonesi, *Quantum moduli space of Chern-Simons quivers, wrapped D6-branes and AdS4/CFT3*, *JHEP* **09** (2011) 005 [1105.2299].

[78] T. Kitao, K. Ohta and N. Ohta, *Three-dimensional gauge dynamics from brane configurations with (p,q)-fivebrane*, *Nucl. Phys.* **B539** (1999) 79 [hep-th/9808111].

[79] O. Bergman, A. Hanany, A. Karch and B. Kol, *Branes and supersymmetry breaking in three-dimensional gauge theories*, *JHEP* **10** (1999) 036 [hep-th/9908075].

[80] A. Guarino, D. L. Jafferis and O. Varela, *String Theory Origin of Dyonic N=8 Supergravity and Its Chern-Simons Duals*, *Phys. Rev. Lett.* **115** (2015) 091601 [1504.08009].

[81] A. Guarino, J. Tarrio and O. Varela, *Romans-mass-driven flows on the D2-brane*, *JHEP* **08** (2016) 168 [1605.09254].

[82] K.-M. Lee and S. Lee, *1/2-BPS Wilson Loops and Vortices in ABJM Model*, *JHEP* **09** (2010) 004 [1006.5589].

[83] Y. Dolivet and M. Tierz, *Chern-Simons matrix models and Stieltjes-Wigert polynomials*, *J. Math. Phys.* **48** (2007) 023507 [hep-th/0609167].

[84] M. Tierz, *Soft matrix models and Chern-Simons partition functions*, *Mod. Phys. Lett. A* **19** (2004) 1365 [hep-th/0212128].

[85] A. Mironov and A. Morozov, *On the complete perturbative solution of one-matrix models*, *Phys. Lett.* **B771** (2017) 503 [1705.00976].

[86] A. Mironov and A. Morozov, *Sum rules for characters from character-preservation property of matrix models*, *JHEP* **08** (2018) 163 [1807.02409].

[87] D. García-García and M. Tierz, *Matrix models for classical groups and Toeplitz±Hankel minors with applications to Chern-Simons theory and fermionic models*, *J. Phys. A* **53** (2020) 345201 [1901.08922].

[88] S. Ramanujan, *Some definite integrals connected with Gauss's sums*, *Messenger of Math.* **XLIV** (1915) 75.

[89] S. G. Naculich and H. J. Schnitzer, *Level-rank duality of the U(N) WZW model, Chern-Simons theory, and 2-D qYM theory*, *JHEP* **06** (2007) 023 [hep-th/0703089].

[90] I. G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford Mathematical Monographs. Oxford University Press, Walton Street, Oxford, UK, 1995.

[91] D. Bump, *Lie Groups*, vol. 225 of *Graduate texts in mathematics*. Springer-Verlag, New York, US, 2013, 10.1007/978-1-4614-8024-2.

[92] D. Bump and A. Gamburd, *On the averages of characteristic polynomials from classical groups*, *Commun. Math. Phys.* **265** (2006) 227 [math-ph/0502043].

[93] A. Morozov, *Cauchy formula and the character ring*, *Eur. Phys. J.* **C79** (2019) 76 [1812.03853].

[94] A. Kapustin and M. J. Strassler, *On mirror symmetry in three-dimensional Abelian gauge theories*, *JHEP* **04** (1999) 021 [hep-th/9902033].

[95] M. Tierz, *Exact solution of Chern-Simons-matter matrix models with characteristic/orthogonal polynomials*, *JHEP* **04** (2016) 168 [1601.06277].

[96] D. L. Jafferis, *The Exact Superconformal R-Symmetry Extremizes Z*, *JHEP* **05** (2012) 159 [1012.3210].

[97] C. Closset, T. T. Dumitrescu, G. Festuccia, Z. Komargodski and N. Seiberg, *Contact Terms, Unitarity, and F-Maximization in Three-Dimensional Superconformal Theories*, *JHEP* **10** (2012) 053 [1205.4142].

[98] C. Closset, T. T. Dumitrescu, G. Festuccia, Z. Komargodski and N. Seiberg, *Comments on Chern-Simons Contact Terms in Three Dimensions*, *JHEP* **09** (2012) 091 [1206.5218].

[99] E. A. Coutsias and N. D. Kazarinoff, *Disorder, renormalizability, theta functions and cornu spirals*, *Physica D: Nonlinear Phenomena* **26** (1987) 295.

[100] M. Berry and J. Goldberg, *Renormalisation of curlicues*, *Nonlinearity* **1** (1988) 1.

[101] S. Benvenuti and S. Pasquetti, *3D-partition functions on the sphere: exact evaluation and mirror symmetry*, *JHEP* **05** (2012) 099 [1105.2551].

[102] J. G. Russo and M. Tierz, *Quantum phase transition in many-flavor supersymmetric QED_3*, *Phys. Rev. D* **95** (2017) 031901 [1610.08527].

[103] S. Lee, *Superconformal field theories from crystal lattices*, *Phys. Rev.* **D75** (2007) 101901 [hep-th/0610204].

[104] S. Kim, S. Lee, S. Lee and J. Park, *Abelian Gauge Theory on M2-brane and Toric Duality*, *Nucl. Phys.* **B797** (2008) 340 [0705.3540].

[105] K. Hosomichi, K.-M. Lee, S. Lee, S. Lee and J. Park, *N=4 Superconformal Chern-Simons Theories with Hyper and Twisted Hyper Multiplets*, *JHEP* **07** (2008) 091 [0805.3662].

[106] D. Ž. Djoković and G. Hochschild, *Semisimplicity of 2-graded Lie algebras, II*, *Illinois Journal of Mathematics* **20** (1976) 134.

[107] T. Kimura, *Linking loops in ABJM and refined theory*, *JHEP* **07** (2015) 030 [1503.01462].

[108] N. Kubo and S. Moriyama, *Two-Point Functions in ABJM Matrix Model*, *JHEP* **05** (2018) 181 [1803.07161].

[109] S. Gukov, A. S. Schwarz and C. Vafa, *Khovanov-Rozansky homology and topological strings*, *Lett. Math. Phys.* **74** (2005) 53 [hep-th/0412243].

[110] S. Gukov, A. Iqbal, C. Kozcaz and C. Vafa, *Link Homologies and the Refined Topological Vertex*, *Commun. Math. Phys.* **298** (2010) 757 [0705.1368].

[111] N. M. Temme, *Error functions, Dawson's and Fresnel integrals*, in *NIST Handbook of Mathematical Functions* (F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clark, eds.), ch. 7. Cambridge University Press, New York, NY, USA, 2010.

[112] S. Hyun and S.-H. Yi, *Non-compact Topological Branes on Conifold*, *JHEP* **11** (2006) 075 [hep-th/0609037].

[113] N. Drukker, D. Gaiotto and J. Gomis, *The Virtue of Defects in 4D Gauge Theories and 2D CFTs*, *JHEP* **06** (2011) 025 [1003.1112].

[114] B. Robinson and C. F. Uhlemann, *Supersymmetric D3/D5 for massive defects on curved space*, *JHEP* **12** (2017) 143 [1709.08650].

[115] Y. Hatsuda, M. Honda, S. Moriyama and K. Okuyama, *ABJM Wilson Loops in Arbitrary Representations*, *JHEP* **10** (2013) 168 [1306.4297].

[116] S. Hirano, K. Nii and M. Shigemori, *ABJ Wilson loops and Seiberg duality*, *PTEP* **2014** (2014) 113B04 [1406.4141].

[117] K. Okuyama, *D-Brane Amplitudes in Topological String on Conifold*, *Phys. Lett.* **B645** (2007) 275 [hep-th/0606048].

[118] D. Anninos and G. A. Silva, *Solvable Quantum Grassmann Matrices*, *J. Stat. Mech.* **1704** (2017) 043102 [1612.03795].

[119] M. Tierz, *Polynomial solution of quantum Grassmann matrices*, *J. Stat. Mech.* **1705** (2017) 053203 [1703.02454].

[120] S. Gukov and M. Stošić, *Homological Algebra of Knots and BPS States*, *Proc. Symp. Pure Math.* **85** (2012) 125 [1112.0030].

[121] E. Gorsky, S. Gukov and M. Stošić, *Quadruply-graded colored homology of knots*, *Fundamenta Mathematicae* **243** (2018) 209 [1304.3481].

[122] P. Di Francesco, *2D quantum gravity, matrix models and graph combinatorics*, in *Marie Curie Training Course: Applications of Random Matrices in Physics* (E. Brezin, V. Kazakov, D. Serban, P. Wiegmann and A. Zabrodin, eds.), vol. 221 of *NATO Advanced Study Institute series II: Mathematics, physics and chemistry*, (Dordrecht, Netherlands), pp. 33–88, Springer, 2006, math-ph/0406013.

[123] H. S. Snyder, *Quantized space-time*, *Phys. Rev.* **71** (1947) 38.

[124] N. Seiberg and E. Witten, *String theory and noncommutative geometry*, *JHEP* **09** (1999) 032 [hep-th/9908142].

[125] F. Lizzi and R. J. Szabo, *Noncommutative geometry and string duality*, *PoS Corfu98* **073** (1999) [hep-th/9904064].

[126] S. Minwalla, M. Van Raamsdonk and N. Seiberg, *Noncommutative perturbative dynamics*, *JHEP* **02** (2000) 020 [hep-th/9912072].

[127] M. R. Douglas and N. A. Nekrasov, *Noncommutative field theory*, *Rev. Mod. Phys.* **73** (2001) 977 [hep-th/0106048].

[128] R. J. Szabo, *Quantum field theory on noncommutative spaces*, *Phys. Rept.* **378** (2003) 207 [hep-th/0109162].

[129] H. Steinacker, *Non-commutative geometry and matrix models*, *PoS* **QGQGS2011** (2011) 004 [1109.5521].

[130] J. de Jong and R. Wulkenhaar, *Nonperturbative evaluation of the partition function for the real scalar quartic QFT on the Moyal plane at weak coupling*, *J. Math. Phys.* **60** (2019) 083504 [1809.09453].

[131] H. Grosse, A. Hock and R. Wulkenhaar, *Solution of all quartic matrix models*, 1906.04600.

[132] H. Grosse, A. Hock and R. Wulkenhaar, *Solution of the self-dual Φ^4 QFT-model on four-dimensional Moyal space*, *JHEP* **01** (2020) 081 [1908.04543].

[133] M. Mariño, *Chern-Simons theory, matrix integrals, and perturbative three manifold invariants*, *Commun. Math. Phys.* **253** (2004) 25 [hep-th/0207096].

[134] M. Mariño, *Chern-Simons theory, matrix models, and topological strings*, *Int. Ser. Monogr. Phys.* **131** (2005) 1.

[135] C. Beasley and E. Witten, *Non-Abelian localization for Chern-Simons theory*, *J. Diff. Geom.* **70** (2005) 183 [hep-th/0503126].

[136] M. Blau and G. Thompson, *Chern-Simons theory on S1-bundles: Abelianisation and q-deformed Yang-Mills theory*, *JHEP* **05** (2006) 003 [hep-th/0601068].

[137] J. Källén, *Cohomological localization of Chern-Simons theory*, *JHEP* **08** (2011) 008 [1104.5353].

[138] M. Romo and M. Tierz, *Unitary Chern-Simons matrix model and the Villain lattice action*, *Phys. Rev. D* **86** (2012) 045027 [1103.2421].

[139] D. S. Dean and S. N. Majumdar, *Extreme value statistics of eigenvalues of Gaussian random matrices*, *Phys. Rev. E* **77** (2008) 041108 [0801.1730].

[140] M. Aganagic, A. Klemm, M. Mariño and C. Vafa, *Matrix model as a mirror of Chern-Simons theory*, *JHEP* **02** (2004) 010 [hep-th/0211098].

[141] N. Halmagyi and V. Yasnov, *The Spectral curve of the lens space matrix model*, *JHEP* **11** (2009) 104 [hep-th/0311117].

[142] D. J. Grabiner, *Brownian motion in a Weyl chamber, non-colliding particles, and random matrices*, *Ann. Inst. H. Poincaré Probab. Statist.* **35** (1999) 177 [math/9708207].

[143] J. Baik and T. M. Suidan, *Random matrix central limit theorems for nonintersecting random walks*, *Ann. Probab.* **35** (2007) 1807 [math/0605212].

[144] S. de Haro and M. Tierz, *Brownian motion, Chern-Simons theory, and 2-D Yang-Mills*, *Phys. Lett. B* **601** (2004) 201 [hep-th/0406093].

[145] Y. Takahashi and M. Katori, *Noncolliding Brownian Motion with Drift and Time-Dependent Stieltjes-Wigert Determinantal Point Process*, *J. Math. Phys.* **53** (2012) 103305 [1207.4351].

[146] E. Langmann and R. J. Szabo, *Duality in scalar field theory on noncommutative phase spaces*, *Phys. Lett.* **B533** (2002) 168 [hep-th/0202039].

[147] T. R. Morris, *Checkered surfaces and complex matrices*, *Nucl. Phys.* **B356** (1991) 703.

[148] Harish-Chandra, *Differential operators on a semisimple Lie algebra*, *Amer. J. Math.* **79** (1957) 87.

[149] C. Itzykson and J. B. Zuber, *The Planar Approximation. 2.*, *J. Math. Phys.* **21** (1980) 411.

[150] N. Yu. Reshetikhin and V. G. Turaev, *Ribbon graphs and their invariants derived from quantum groups*, *Commun. Math. Phys.* **127** (1990) 1.

[151] N. Y. Reshetikhin and V. G. Turaev, *Invariants of three manifolds via link polynomials and quantum groups*, *Invent. Math.* **103** (1991) 547.

[152] A. Yu. Alekseev, A. Recknagel and V. Schomerus, *Brane dynamics in background fluxes and noncommutative geometry*, *JHEP* **05** (2000) 010 [hep-th/0003187].

[153] J. Pawelczyk and H. Steinacker, *A Quantum algebraic description of D branes on group manifolds*, *Nucl. Phys.* **B638** (2002) 433 [hep-th/0203110].

[154] L. Chekhov, *Matrix models with hard walls: Geometry and solutions*, *J. Phys.* **A39** (2006) 8857 [hep-th/0602013].

[155] D. S. Dean and S. N. Majumdar, *Large deviations of extreme eigenvalues of random matrices*, *Phys. Rev. Lett.* **97** (2006) 160201 [cond-mat/0609651].

[156] S. N. Majumdar and G. Schehr, *Top eigenvalue of a random matrix: large deviations and third order phase transition*, *J. Stat. Mech.* **2014** (2014) P01012 [1311.0580].

[157] A. Dhar, A. Kundu, S. N. Majumdar, S. Sabhapandit and G. Schehr, *Exact Extremal Statistics in the Classical 1D Coulomb Gas*, *Phys. Rev. Lett.* **119** (2017) 060601 [1704.08973].

[158] P. Vivo, *Large deviations of the maximum of independent and identically distributed random variables*, *Eur. J. Phys.* **36** (2015) 055037 [1507.05442].

[159] Y. Makeenko, *Loop equations in matrix models and in 2-D quantum gravity*, *Mod. Phys. Lett. A* **6** (1991) 1901.

[160] F. D. Cunden, P. Facchi, M. Ligabò and P. Vivo, *Universality of the third-order phase transition in the constrained Coulomb gas*, *J. Stat. Mech.* **1705** (2017) 053303 [1702.05071].

[161] F. D. Cunden, P. Facchi, M. Ligabò and P. Vivo, *Universality of the weak pushed-to-pulled transition in systems with repulsive interactions*, *J. Phys. A* **51** (2018) [1711.09141].

[162] R. J. Baxter, *Statistical mechanics of a one-dimensional Coulomb system with a uniform charge background*, *Math. Proc. Cambridge Phil. Soc.* **59** (1963) 779–787.

[163] S. A. Yost, *Supermatrix models*, *Int. J. Mod. Phys.* **A7** (1992) 6105 [hep-th/9111033].

[164] L. Alvarez-Gaume and J. L. Mañes, *Supermatrix models*, *Mod. Phys. Lett.* **A6** (1991) 2039.

[165] J. Alfaro, R. Medina and L. F. Urrutia, *The Itzykson-Zuber integral for U(m/n)*, *J. Math. Phys.* **36** (1995) 3085 [hep-th/9412012].

[166] T. Guhr, *Gelfand-Tzetlin coordinates for the unitary supergroup*, *Commun. Math. Phys.* **176** (1996) 555.

[167] E. M. Moens and J. van der Jeugt, *A determinantal formula for supersymmetric Schur polynomials*, *J. Algebraic Combin.* **17** (2003) 283.

[168] S. Matsumoto and S. Moriyama, *ABJ Fractional Brane from ABJM Wilson Loop*, *JHEP* **03** (2014) 079 [1310.8051].

[169] S. Matsuno and S. Moriyama, *Giambelli Identity in Super Chern-Simons Matrix Model*, *J. Math. Phys.* **58** (2017) 032301 [1603.04124].

[170] T. Furukawa and S. Moriyama, *Jacobi-Trudi Identity in Super Chern-Simons Matrix Model*, *SIGMA* **14** (2018) 049 [1711.04893].

[171] G. Szegő, *On certain Hermitian forms associated with the Fourier series of a positive function*, *Comm. Sém. Math. Univ. Lund* **Tome Supplémentaire** (1952) 228–238.

[172] C. P. Herzog, I. R. Klebanov, S. S. Pufu and T. Tesileanu, *Multi-Matrix Models and Tri-Sasaki Einstein Spaces*, *Phys. Rev. D* **83** (2011) 046001 [1011.5487].

[173] D. L. Jafferis, I. R. Klebanov, S. S. Pufu and B. R. Safdi, *Towards the F-Theorem: N=2 Field Theories on the Three-Sphere*, *JHEP* **06** (2011) 102 [1103.1181].

[174] J. Hong and J. T. Liu, *Subleading corrections to the S^3 free energy of necklace quiver theories dual to massive IIA*, *JHEP* **11** (2021) 183 [2103.17033].

[175] I. R. Klebanov, F. Popov and G. Tarnopolsky, *TASI Lectures on Large N Tensor Models*, *PoS* **TASI2017** (2018) 004 [1808.09434].

[176] M. Mariño, *Les Houches lectures on matrix models and topological strings*, hep-th/0410165.

[177] R. C. Santamaria, M. Mariño and P. Putrov, *Unquenched flavor and tropical geometry in strongly coupled Chern-Simons-matter theories*, *JHEP* **10** (2011) 139 [1011.6281].

[178] T. Azeyanagi, M. Fujita and M. Hanada, *From the planar limit to M-theory*, *Phys. Rev. Lett.* **110** (2013) 121601 [1210.3601].

[179] E. Brezin and V. A. Kazakov, *Exactly Solvable Field Theories of Closed Strings*, *Phys. Lett. B* **236** (1990) 144.

[180] D. J. Gross and A. A. Migdal, *Nonperturbative Two-Dimensional Quantum Gravity*, *Phys. Rev. Lett.* **64** (1990) 127.

[181] M. R. Douglas and S. H. Shenker, *Strings in Less Than One-Dimension*, *Nucl. Phys. B* **335** (1990) 635.

[182] V. Periwal and D. Shevitz, *Unitary matrix models as Exactly solvable string theories*, *Phys. Rev. Lett.* **64** (1990) 1326.

[183] J. Baik, P. Deift and K. Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, *J. Amer. Math. Soc.* **12** (1999) 1119 [math/9810105].

[184] D. Aldous and P. Diaconis, *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem*, *Bull. Amer. Math. Soc.* **36** (1999) 413.

[185] M. Moshe and J. Zinn-Justin, *Quantum field theory in the large N limit: A Review*, *Phys. Rept.* **385** (2003) 69 [hep-th/0306133].

[186] G. Arias-Tamargo, D. Rodriguez-Gomez and J. G. Russo, *On the UV completion of the $O(N)$ model in $6-\epsilon$ dimensions: a stable large-charge sector*, *JHEP* **09** (2020) 064 [2003.13772].

[187] S. Giombi and J. Hyman, *On the large charge sector in the critical O(N) model at large N*, *JHEP* **09** (2021) 184 [2011.11622].

[188] D. Orlando, S. Reffert and T. Schmidt, *Following the flow for large N and large charge*, *Phys. Lett. B* **825** (2022) 136881 [2110.07616].

[189] R. Moser, D. Orlando and S. Reffert, *Convexity, large charge and the large-N phase diagram of the φ^4 theory*, *JHEP* **02** (2022) 152 [2110.07617].

[190] D. J. Gross and E. Witten, *Possible Third Order Phase Transition in the Large N Lattice Gauge Theory*, *Phys. Rev. D* **21** (1980) 446.

[191] S. R. Wadia, *N = Infinity Phase Transition in a Class of Exactly Soluble Model Lattice Gauge Theories*, *Phys. Lett. B* **93** (1980) 403.

[192] S. R. Wadia, *A Study of U(N) Lattice Gauge Theory in 2-dimensions*, 1212.2906.

[193] M. R. Douglas and V. A. Kazakov, *Large N phase transition in continuum QCD in two-dimensions*, *Phys. Lett.* **B319** (1993) 219 [hep-th/9305047].

[194] A. Gorsky, A. Milekhin and S. Nechaev, *Two faces of Douglas-Kazakov transition: from Yang-Mills theory to random walks and beyond*, *Nucl. Phys. B* **950** (2020) 114849 [1604.06381].

[195] J. Russo and K. Zarembo, *Large N Limit of N=2 SU(N) Gauge Theories from Localization*, *JHEP* **10** (2012) 082 [1207.3806].

[196] J. G. Russo and K. Zarembo, *Evidence for Large-N Phase Transitions in N=2* Theory*, *JHEP* **04** (2013) 065 [1302.6968].

[197] J. Russo and K. Zarembo, *Massive N=2 Gauge Theories at Large N*, *JHEP* **11** (2013) 130 [1309.1004].

[198] J. G. Russo and K. Zarembo, *Localization at Large N*, in *Proceedings, 100th anniversary of the birth of I.Ya. Pomeranchuk (Pomeranchuk 100): Moscow, Russia, June 5-6, 2013*, pp. 287–311, 2014, 1312.1214, DOI.

[199] J. G. Russo, *$\mathcal{N}$ = 2 gauge theories and quantum phases*, *JHEP* **12** (2014) 169 [1411.2602].

[200] K. Zarembo, *Strong-Coupling Phases of Planar N=2* Super-Yang-Mills Theory*, *Theor. Math. Phys.* **181** (2014) 1522 [1410.6114].

[201] X. Chen-Lin, J. Gordon and K. Zarembo, *$\mathcal{N} = 2^*$ super-Yang-Mills theory at strong coupling*, *JHEP* **11** (2014) 057 [1408.6040].

[202] A. Barranco and J. G. Russo, *Large N phase transitions in supersymmetric Chern-Simons theory with massive matter*, *JHEP* **03** (2014) 012 [1401.3672].

[203] L. Anderson and K. Zarembo, *Quantum Phase Transitions in Mass-Deformed ABJM Matrix Model*, *JHEP* **09** (2014) 021 [1406.3366].

[204] L. Anderson and J. G. Russo, *ABJM Theory with mass and FI deformations and Quantum Phase Transitions*, *JHEP* **05** (2015) 064 [1502.06828].

[205] L. Anderson and N. Drukker, *More Large N limits of 3d gauge theories*, *J. Phys. A* **50** (2017) 345401 [1701.04409].

[206] T. Nosaka, K. Shimizu and S. Terashima, *Large N behavior of mass deformed ABJM theory*, *JHEP* **03** (2016) 063 [1512.00249].

[207] T. Nosaka, K. Shimizu and S. Terashima, *Mass Deformed ABJM Theory on Three Sphere in Large N limit*, *JHEP* **03** (2017) 121 [1608.02654].

[208] A. Nedelin, *Phase transitions in 5D super Yang-Mills theory*, *JHEP* **07** (2015) 004 [1502.07275].

[209] S. A. Hartnoll and S. Kumar, *Higher rank Wilson loops from a matrix model*, *JHEP* **08** (2006) 026 [hep-th/0605027].

[210] J. G. Russo and K. Zarembo, *Wilson loops in antisymmetric representations from localization in supersymmetric gauge theories*, *Rev. Math. Phys.* **30** (2018) 1840014 [1712.07186].

[211] X. Chen-Lin and K. Zarembo, *Higher Rank Wilson Loops in N = 2* Super-Yang-Mills Theory*, *JHEP* **03** (2015) 147 [1502.01942].

[212] M. Mezei and S. S. Pufu, *Three-sphere free energy for classical gauge groups*, *JHEP* **02** (2014) 037 [1312.0920].

[213] I. R. Klebanov, S. S. Pufu and B. R. Safdi, *F-Theorem without Supersymmetry*, *JHEP* **10** (2011) 038 [1105.4598].

[214] M. Tierz, *Wilson loops and free energies in* 3*d* $\mathcal{N} = 4$ *SYM: exact results, exponential asymptotics and duality*, *PTEP* **2019** (2019) 053B01 [1804.10845].

[215] N. Seiberg, *Five-dimensional SUSY field theories, nontrivial fixed points and string dynamics*, *Phys. Lett. B* **388** (1996) 753 [hep-th/9608111].

[216] C. Cordova, T. T. Dumitrescu and K. Intriligator, *Deformations of Superconformal Theories*, *JHEP* **11** (2016) 135 [1602.01217].

[217] C.-M. Chang, *5d and 6d SCFTs Have No Weak Coupling Limit*, *JHEP* **09** (2019) 016 [1810.04169].

[218] E. Witten, *Phase transitions in M theory and F theory*, *Nucl. Phys. B* **471** (1996) 195 [hep-th/9603150].

[219] D. R. Morrison and N. Seiberg, *Extremal transitions and five-dimensional supersymmetric field theories*, *Nucl. Phys. B* **483** (1997) 229 [hep-th/9609070].

[220] K. A. Intriligator, D. R. Morrison and N. Seiberg, *Five-dimensional supersymmetric gauge theories and degenerations of Calabi-Yau spaces*, *Nucl. Phys. B* **497** (1997) 56 [hep-th/9702198].

[221] H. Hayashi, C. Lawrie, D. R. Morrison and S. Schafer-Nameki, *Box Graphs and Singular Fibers*, *JHEP* **05** (2014) 048 [1402.2653].

[222] F. Apruzzi, C. Lawrie, L. Lin, S. Schäfer-Nameki and Y.-N. Wang, *5d Superconformal Field Theories and Graphs*, *Phys. Lett. B* **800** (2020) 135077 [1906.11820].

[223] C. Closset, M. Del Zotto and V. Saxena, *Five-dimensional SCFTs and gauge theory phases: an M-theory/type IIA perspective*, *SciPost Phys.* **6** (2019) 052 [1812.10451].

[224] F. Apruzzi, C. Lawrie, L. Lin, S. Schäfer-Nameki and Y.-N. Wang, *Fibers add Flavor, Part I: Classification of 5d SCFTs, Flavor Symmetries and BPS States*, *JHEP* **11** (2019) 068 [1907.05404].

[225] F. Apruzzi, C. Lawrie, L. Lin, S. Schäfer-Nameki and Y.-N. Wang, *Fibers add Flavor, Part II: 5d SCFTs, Gauge Theories, and Dualities*, *JHEP* **03** (2020) 052 [1909.09128].

[226] L. Bhardwaj and G. Zafrir, *Classification of 5d N=1 gauge theories*, *JHEP* **12** (2020) 099 [2003.04333].

[227] J. Eckhard, S. Schäfer-Nameki and Y.-N. Wang, *Trifectas for T_N in 5d*, *JHEP* **07** (2020) 199 [2004.15007].

[228] A. Collinucci and R. Valandro, *The role of U(1)'s in 5d theories, Higgs branches, and geometry*, *JHEP* **10** (2020) 178 [2006.15464].

[229] L. Bhardwaj, *Flavor symmetry of 5d SCFTs. Part II. Applications*, *JHEP* **04** (2021) 221 [2010.13235].

[230] J. Källén and M. Zabzine, *Twisted supersymmetric 5D Yang-Mills theory and contact geometry*, *JHEP* **05** (2012) 125 [1202.1956].

[231] J. Källén, J. Qiu and M. Zabzine, *The perturbative partition function of supersymmetric 5D Yang-Mills theory with matter on the five-sphere*, *JHEP* **08** (2012) 157 [1206.6008].

[232] K. Hosomichi, R.-K. Seong and S. Terashima, *Supersymmetric Gauge Theories on the Five-Sphere*, *Nucl. Phys. B* **865** (2012) 376 [1203.0371].

[233] H.-C. Kim and S. Kim, *M5-branes from gauge theories on the 5-sphere*, *JHEP* **05** (2013) 144 [1206.6339].

[234] G. Lockhart and C. Vafa, *Superconformal Partition Functions and Non-perturbative Topological Strings*, *JHEP* **10** (2018) 051 [1210.5909].

[235] M. Mezei, S. S. Pufu and Y. Wang, *Chern-Simons theory from M5-branes and calibrated M2-branes*, *JHEP* **08** (2019) 165 [1812.07572].

[236] J. A. Minahan, *Matrix models for 5d super Yang–Mills*, *J. Phys. A* **50** (2017) 443015 [1608.02967].

[237] D. L. Jafferis and S. S. Pufu, *Exact results for five-dimensional superconformal field theories with gravity duals*, *JHEP* **05** (2014) 032 [1207.4359].

[238] J. Källén, J. A. Minahan, A. Nedelin and M. Zabzine, *N^3-behavior from 5D Yang-Mills theory*, *JHEP* **10** (2012) 184 [1207.3763].

[239] J. A. Minahan, A. Nedelin and M. Zabzine, *5D super Yang-Mills theory and the correspondence to AdS_7/CFT_6*, *J. Phys. A* **46** (2013) 355401 [1304.1016].

[240] B. Assel, J. Estes and M. Yamazaki, *Wilson Loops in 5d N=1 SCFTs and AdS/CFT*, *Annales Henri Poincaré* **15** (2014) 589 [1212.1202].

[241] G. Giasemidis, R. J. Szabo and M. Tierz, *Supersymmetric gauge theories, Coulomb gases and Chern-Simons matrix models*, *Phys. Rev. D* **89** (2014) 025016 [1310.3122].

[242] C.-M. Chang, M. Fluder, Y.-H. Lin and Y. Wang, *Romans Supergravity from Five-Dimensional Holograms*, *JHEP* **05** (2018) 039 [1712.10313].

[243] M. Fluder and C. F. Uhlemann, *Precision Test of AdS_6/CFT_5 in Type IIB String Theory*, *Phys. Rev. Lett.* **121** (2018) 171603 [1806.08374].

[244] P. M. Crichigno, D. Jain and B. Willett, *5d Partition Functions with A Twist*, *JHEP* **11** (2018) 058 [1808.06744].

[245] C. F. Uhlemann, *Exact results for 5d SCFTs of long quiver type*, *JHEP* **11** (2019) 072 [1909.01369].

[246] C. F. Uhlemann, *Wilson loops in 5d long quiver gauge theories*, *JHEP* **09** (2020) 145 [2006.01142].

[247] J. A. Minahan and A. Nedelin, *Phases of planar 5-dimensional supersymmetric Chern-Simons theory*, *JHEP* **12** (2014) 049 [1408.2767].

[248] J. A. Minahan and A. Nedelin, *Five-dimensional gauge theories on spheres with negative couplings*, *JHEP* **02** (2021) 102 [2007.13760].

[249] T. J. Hollowood and S. P. Kumar, *Partition function of $\mathcal{N} = 2^*$ SYM on a large four-sphere*, *JHEP* **12** (2015) 016 [1509.00716].

[250] J. G. Russo, *Properties of the partition function of $\mathcal{N} = 2$ supersymmetric QCD with massive matter*, *JHEP* **07** (2019) 125 [1905.05267].

[251] P. Jefferson, H.-C. Kim, C. Vafa and G. Zafrir, *Towards Classification of 5d SCFTs: Single Gauge Node*, 1705.05836.

[252] C.-M. Chang, M. Fluder, Y.-H. Lin and Y. Wang, *Spheres, Charges, Instantons, and Bootstrap: A Five-Dimensional Odyssey*, *JHEP* **03** (2018) 123 [1710.08418].

[253] M. Fluder and C. F. Uhlemann, *Evidence for a 5d F-theorem*, *JHEP* **02** (2021) 192 [2011.00006].

[254] H. Neuberger, *Scaling Regime at the Large N Phase Transition of Two-dimensional Pure Gauge Theories*, *Nucl. Phys. B* **340** (1990) 703.

[255] D. R. Morrison, S. Schafer-Nameki and B. Willett, *Higher-Form Symmetries in 5d*, *JHEP* **09** (2020) 024 [2005.12296].

[256] F. Albertini, M. Del Zotto, I. García Etxebarria and S. S. Hosseini, *Higher Form Symmetries and M-theory*, *JHEP* **12** (2020) 203 [2005.12831].

[257] P. Benetti Genolini and L. Tizzano, *Instantons, symmetries and anomalies in five dimensions*, *JHEP* **04** (2021) 188 [2009.07873].

[258] O. Bergman, D. Rodríguez-Gómez and G. Zafrir, *5d superconformal indices at large N and holography*, *JHEP* **08** (2013) 081 [1305.6870].

[259] C. Hwang, J. Kim, S. Kim and J. Park, *General instanton counting and 5d SCFT*, *JHEP* **07** (2015) 063 [1406.6793].

[260] A. Bourget, J. F. Grimminger, A. Hanany, M. Sperling and Z. Zhong, *Magnetic Quivers from Brane Webs with O5 Planes*, *JHEP* **07** (2020) 204 [2004.04082].

[261] O. Bergman and D. Rodríguez-Gómez, *The Cat's Cradle: deforming the higher rank E_1 and $\tilde{E}_1$ theories*, *JHEP* **02** (2021) 122 [2011.05125].

[262] X. Li and F. Yagi, *Thermodynamic limit of Nekrasov partition function for 5-brane web with O5-plane*, *JHEP* **06** (2021) 004 [2102.09482].

[263] M. Esole and S.-H. Shao, *M-theory on Elliptic Calabi-Yau Threefolds and 6d Anomalies*, 1504.01387.

[264] A. Collinucci and R. Valandro, *A string theory realization of special unitary quivers in 3 dimensions*, *JHEP* **11** (2020) 157 [2008.10689].

[265] J. Teschner, ed., *New Dualities of Supersymmetric Gauge Theories*, Mathematical Physics Studies. Springer, Cham, Switzerland, 2016, 10.1007/978-3-319-18769-3.

[266] N. Hama, K. Hosomichi and S. Lee, *Notes on SUSY Gauge Theories on Three-Sphere*, *JHEP* **03** (2011) 127 [1012.3512].

[267] D. Gaiotto and E. Witten, *S-Duality of Boundary Conditions In N=4 Super Yang-Mills Theory*, *Adv. Theor. Math. Phys.* **13** (2009) 721 [0807.3720].

[268] S. Schindler, *Some transplantation theorems for the generalized Mehler transform and related asymptotic expansions*, *Trans. Amer. Math. Soc.* **155** (1971) 257.

[269] O. A. Daahluis, *Hypergeometric function*, in *NIST Handbook of Mathematical Functions* (F. W. J. Olver, D. W. Lozier, R. F. Boisvert and C. W. Clark, eds.), ch. 15. Cambridge University Press, New York, NY, USA, 2010.

[270] A. Gil, J. Segura and N. M. Temme, *Computing the conical function $P^{\mu}_{-1/2+i\tau}(x)$*, *SIAM J. Sci. Comput.* **31** (2009) 1716.

[271] T. M. Dunster, *Conical functions with one or both parameters large*, *Proceedings of the Royal Society of Edinburgh: Section A Mathematics* **119** (1991) 311–327.

[272] S. Flügge, *Practical Quantum Mechanics*, Classics in Mathematics. Springer, 1998, 10.1007/978-3-642-61995-3.

[273] M. A. Olshanetsky and A. M. Perelomov, *Classical integrable finite dimensional systems related to Lie algebras*, *Phys. Rept.* **71** (1981) 313.

[274] M. Hallnäs and S. Ruijsenaars, *A Recursive Construction of Joint Eigenfunctions for the Hyperbolic Nonrelativistic Calogero-Moser Hamiltonians*, *International Mathematics Research Notices* **2015** (2015) 10278 [1305.4759].

[275] T. Nishioka, Y. Tachikawa and M. Yamazaki, *3d Partition Function as Overlap of Wavefunctions*, *JHEP* **08** (2011) 003 [1105.4390].

[276] D. R. Gulotta, C. P. Herzog and S. S. Pufu, *From Necklace Quivers to the F-theorem, Operator Counting, and T(U(N))*, *JHEP* **12** (2011) 077 [1105.2817].

[277] G. J. Heckman and E. M. Opdam, *Root systems and hypergeometric functions. I*, *Compositio Math.* **64** (1987) 329.

[278] M. Isachenkov and V. Schomerus, *Superintegrability of d-dimensional Conformal Blocks*, *Phys. Rev. Lett.* **117** (2016) 071602 [1602.01858].

[279] M. Isachenkov and V. Schomerus, *Integrability of conformal blocks. Part I. Calogero-Sutherland scattering theory*, *JHEP* **07** (2018) 180 [1711.06609].

[280] M. Dedushenko, S. S. Pufu and R. Yacoby, *A one-dimensional theory for Higgs branch operators*, *JHEP* **03** (2018) 138 [1610.00740].

[281] N. Hama, K. Hosomichi and S. Lee, *SUSY Gauge Theories on Squashed Three-Spheres*, *JHEP* **05** (2011) 014 [1102.4716].

[282] T. Tao, *Topics in random matrix theory*, vol. 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012, 10.1090/gsm/132.

[283] J. Baik, P. Deift and T. Suidan, *Combinatorics and random matrix theory*, vol. 172 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2016.

[284] F. Colomo and A. G. Pronko, *Third-order phase transition in random tilings*, *Phys. Rev.* **E88** (2013) 042125 [1306.6207].

[285] F. Colomo and A. G. Pronko, *Thermodynamics of the Six-Vertex Model in an L-Shaped Domain*, *Commun. Math. Phys.* **339** (2015) 699 [1501.03135].

[286] A. Okounkov, *Infinite wedge and random partitions*, *Selecta Math. (N.S.)* **7** (2001) 57 [math/9907127].

[287] A. Okounkov, *The uses of random partitions*, in *XIVth International Congress on Mathematical Physics*, pp. 379–403. World Sci. Publ., Hackensack, NJ, 2005. math-ph/0309015.

[288] A. Borodin and G. Olshanski, *Distributions on partitions, point processes, and the hypergeometric kernel*, *Comm. Math. Phys.* **211** (2000) 335 [math/9904010].

[289] A. Okounkov, SL(2) *and z-measures*, in *Random matrix models and their applications*, vol. 40 of *Math. Sci. Res. Inst. Publ.*, pp. 407–420. Cambridge Univ. Press, Cambridge, 2001. math/0002135.

[290] A. Borodin and G. Olshanski, *z-measures on partitions, Robinson-Schensted-Knuth correspondence, and $\beta = 2$ random matrix ensembles*, in *Random matrix models and their applications*, vol. 40 of *Math. Sci. Res. Inst. Publ.*, pp. 71–94. Cambridge Univ. Press, Cambridge, 2001. math/9905189.

[291] E. Barnes, *The theory of the G-function*, *Quart. J. Math.* **31** (1899) 264.

[292] J.-M. Stéphan, *Extreme boundary conditions and random tilings*, *SciPost Phys. Lect. Notes* **26** (2021) 1 [2003.06339].

[293] K. Johansson, *A multi-dimensional markov chain and the meixner ensemble*, *Ark. Mat.* **48** (2010) 79 [0707.0098].

[294] K. Hikami and T. Imamura, *Vicious walkers and hook Young tableaux*, *J. Phys. A* **36** (2003) 3033 [cond-mat/0209512].

[295] D. García-García and M. Tierz, *Toeplitz minors and specializations of skew Schur polynomials*, *J. Combin. Theory* **172** (2020) 105201 [1706.02574].

[296] C. Andréief, *Note sur une relation entre les intégrales définies des produits des fonctions*, *Mém . Soc. Sci. Phys. Nat. Bordeaux* **2** (1886) 1.

[297] P. J. Forrester, *Meet Andréief, Bordeaux 1886, and Andreev, Kharkov 1882—1883*, *Random Matrices: Theory and Applications* **08** (2019) 1930001 [1806.10411].

[298] K. Johansson, *Shape fluctuations and random matrices*, *Comm. Math. Phys.* **209** (2000) 437 [math/9903134].

[299] A. Böttcher and B. Silbermann, *Toeplitz matrices and determinants with Fisher-Hartwig symbols*, *J. Funct. Anal.* **63** (1985) 178.

[300] A. Böttcher and H. Widom, *Two elementary derivations of the pure Fisher-Hartwig determinant*, *Integral Equations Operator Theory* **53** (2005) 593 [math/0312198].

[301] A. Böttcher and S. M. Grudsky, *Spectral properties of banded Toeplitz matrices*, vol. 96. Siam, 2005.

[302] J. Baik, *Random vicious walks and random matrices*, *Comm. Pure Appl. Math.* **53** (2000) 1385 [math/0001022].

[303] C. A. Tracy and H. Widom, *Level spacing distributions and the Airy kernel*, *Commun. Math. Phys.* **159** (1994) 151 [hep-th/9211141].

[304] M. Hisakado, *Unitary matrix models and Painleve III*, *Mod. Phys. Lett.* **A11** (1996) 3001 [hep-th/9609214].

[305] M. Hisakado, *Unitary matrix models with a topological term and discrete time Toda equation*, *Phys. Lett.* **B395** (1997) 208 [hep-th/9611177].

[306] A. Borodin, A. Okounkov and G. Olshanski, *Asymptotics of Plancherel measures for symmetric groups*, *J. Amer. Math. Soc.* **13** (2000) 481 [math/9905032].

[307] K. Johansson, *The longest increasing subsequence in a random permutation and a unitary random matrix model*, *Math. Res. Lett.* **5** (1998) 63.

[308] K. Johansson, *Discrete orthogonal polynomial ensembles and the Plancherel measure*, *Annals Math.* **153** (2001) 259 [math/9906120].

[309] E. Witten, *Analytic Continuation Of Chern-Simons Theory*, *AMS/IP Stud. Adv. Math.* **50** (2011) 347 [1001.2933].

[310] S. D. Cutkosky, *Introduction to Algebraic Geometry*, vol. 188 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2018.

[311] F. D. Cunden, P. Facchi, M. Ligabò and P. Vivo, *Third-order phase transition: random matrices and screened Coulomb gas with hard walls*, *J. Stat. Phys.* **175** (2019) 1262 [1810.12593].

[312] Y. Chen and N. Mekareeya, *The Hilbert series of U/SU SQCD and Toeplitz Determinants*, *Nucl. Phys.* **B850** (2011) 553 [1104.2045].

[313] E. Basor, Y. Chen and N. Mekareeya, *The Hilbert Series of $\mathcal{N} = 1$ $SO(N_c)$ and $Sp(N_c)$ SQCD, Painlevé VI and Integrable Systems*, *Nucl. Phys.* **B860** (2012) 421 [1112.3848].

[314] Y. Chen, N. Jokela, M. Jarvinen and N. Mekareeya, *Moduli space of supersymmetric QCD in the Veneziano limit*, *JHEP* **09** (2013) 131 [1303.6289].

[315] N. S. Witte and P. J. Forrester, *Gap probabilities in the finite and scaled Cauchy random matrix ensembles*, *Nonlinearity* **13** (2000) 1965 [math-ph/0009022].

[316] P. W. Brouwer, *Generalized circular ensemble of scattering matrices for a chaotic cavity with non-ideal leads*, *Phys. Rev. B* **51** (1995) 16878 [cond-mat/9501025].

[317] W. Koepf and M. Masjed-Jamei, *A generic polynomial solution for the differential equation of hypergeometric type and six sequences of orthogonal polynomials related to it*, *Integral Transforms Spec. Funct.* **17** (2006) 559.

[318] M. Masjed-Jamei, F. Marcellán and E. J. Huertas, *A finite class of orthogonal functions generated by Routh-Romanovski polynomials*, *Complex Var. Elliptic Equ.* **59** (2014) 162.

[319] P. A. Lesky, *Endliche und unendliche Systeme von kontinuierlichen klassischen Orthogonalpolynomen*, *Z. Angew. Math. Mech.* **76** (1996) 181.

[320] R. Koekoek, P. A. Lesky and R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their q-analogues*, Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2010, 10.1007/978-3-642-05014-5.

[321] K. Jordaan and F. Toókos, *Orthogonality and asymptotics of pseudo-Jacobi polynomials for non-classical parameters*, *J. Approx. Theory* **178** (2014) 1.

[322] Z. Song and R. Wong, *Asymptotics of pseudo-Jacobi polynomials with varying parameters*, *Stud. Appl. Math.* **139** (2017) 179.

[323] R. Wong, *Asymptotics of orthogonal polynomials*, *Int. J. Numer. Anal. Model.* **15** (2018) 193.

[324] V. Aldaya, J. Bisquert and J. Navarro-Salas, *The Quantum relativistic harmonic oscillator: Generalized Hermite polynomials*, *Phys. Lett.* **A156** (1991) 381.

[325] V. Romanovski, *Sur quelques classes nouvelles de polynomes orthogonaux*, *C. R. Acad. Sci.* **188** (1929) 1023.

[326] A. P. Raposo, H. J. Weber, D. E. Alvarez-Castillo and M. Kirchbach, *Romanovski polynomials in selected physics problems*, *Central Eur. J. Phys.* **5** (2007) 253 [0706.3897].

[327] R. Askey, *Beta integrals and the associated orthogonal polynomials*, in *Number theory, Madras 1987*, vol. 1395 of *Lecture Notes in Math.*, pp. 84–121. Springer, Berlin, 1989. DOI.

[328] S.-E. Ekström, C. Garoni and S. Serra-Capizzano, *Are the eigenvalues of banded symmetric toeplitz matrices known in almost closed form?*, *Experime. Math.* **27** (2018) 478.

[329] M. Barrera, A. Böttcher, S. M. Grudsky and E. A. Maximenko, *Eigenvalues of even very nice Toeplitz matrices can be unexpectedly erratic*, in *The diversity and beauty of applied operator theory*, vol. 268 of *Oper. Theory Adv. Appl.*, pp. 51–77. Birkhäuser/Springer, Cham, 2018. 1710.05243. DOI.

[330] S. Jain, S. Minwalla, T. Sharma, T. Takimi, S. R. Wadia and S. Yokoyama, *Phases of large N vector Chern-Simons theories on $S^2 \times S^1$*, *JHEP* **09** (2013) 009 [1301.6169].

[331] J. A. Minahan, *Matrix models with boundary terms and the generalized Painleve II equation*, *Phys. Lett. B* **268** (1991) 29.

[332] J. A. Minahan, *Flows and solitary waves in unitary matrix models with logarithmic potentials*, *Nucl. Phys. B* **378** (1992) 501 [hep-th/9111012].

[333] B. Sundborg, *The Hagedorn transition, deconfinement and N=4 SYM theory*, *Nucl. Phys. B* **573** (2000) 349 [hep-th/9908001].

[334] O. Aharony, J. Marsano, S. Minwalla, K. Papadodimas and M. Van Raamsdonk, *The Hagedorn - deconfinement phase transition in weakly coupled large N gauge theories*, *Adv. Theor. Math. Phys.* **8** (2004) 603 [hep-th/0310285].

[335] L. Alvarez-Gaume, C. Gomez, H. Liu and S. Wadia, *Finite temperature effective action, AdS(5) black holes, and 1/N expansion*, *Phys. Rev. D* **71** (2005) 124023 [hep-th/0502227].

[336] L. Alvarez-Gaume, P. Basu, M. Mariño and S. R. Wadia, *Blackhole/String Transition for the Small Schwarzschild Blackhole of AdS(5)x S**5 and Critical Unitary Matrix Models*, *Eur. Phys. J. C* **48** (2006) 647 [hep-th/0605041].

[337] T. Azuma, P. Basu and S. R. Wadia, *Monte Carlo Studies of the GWW Phase Transition in Large-N Gauge Theories*, *Phys. Lett. B* **659** (2008) 676 [0710.5873].

[338] M. Hanada and J. Maltz, *A proposal of the gauge theory description of the small Schwarzschild black hole in $AdS_5 \times S^5$*, *JHEP* **02** (2017) 012 [1608.03276].

[339] D. Berenstein, *Submatrix deconfinement and small black holes in AdS*, *JHEP* **09** (2018) 054 [1806.05729].

[340] M. Hanada, G. Ishiki and H. Watanabe, *Partial Deconfinement*, *JHEP* **03** (2019) 145 [1812.05494].

[341] M. Hanada, A. Jevicki, C. Peng and N. Wintergerst, *Anatomy of Deconfinement*, *JHEP* **12** (2019) 167 [1909.09118].

[342] M. Hanada, G. Ishiki and H. Watanabe, *Partial deconfinement in gauge theories*, *PoS* **LATTICE2019** (2019) 055 [1911.11465].

[343] C. Lazaroiu, *Holomorphic matrix models*, *JHEP* **05** (2003) 044 [hep-th/0303008].

[344] P. Rossi, M. Campostrini and E. Vicari, *The Large N expansion of unitary matrix models*, *Phys. Rept.* **302** (1998) 143 [hep-lat/9609003].

[345] J. G. Russo and M. Tierz, *Multiple phases in a generalized Gross-Witten-Wadia matrix model*, *JHEP* **09** (2020) 081 [2007.08515].

[346] J. G. Russo, *Phases of unitary matrix models and lattice QCD2*, *Phys. Rev. D* **102** (2020) 105019 [2010.02950].

[347] F. Haake, M. Kus, H.-J. Sommers, H. Schomerus and K. Zyczkowski, *Secular determinants of random unitary matrices*, *J. Phys. A* **29** (1996) 3641 [chao-dyn/9603006].

[348] Y. V. Fyodorov, G. A. Hiary and J. P. Keating, *Freezing Transition, Characteristic Polynomials of Random Matrices, and the Riemann Zeta Function*, *Phys. Rev. Lett.* **108** (2012) 170601 [1202.4713].

[349] Y. V. Fyodorov and J. P. Keating, *Freezing transitions and extreme values: random matrix theory, and disordered landscapes*, *Phil. Trans. R. Soc. A* **372** (2014) 20120503 [1211.6063].

[350] J. Hallin and D. Persson, *Thermal phase transition in weakly interacting, large N(C) QCD*, *Phys. Lett. B* **429** (1998) 232 [hep-ph/9803234].

[351] M. Adler and P. van Moerbeke, *Integrals over classical groups, random permutations, Toda and Toeplitz lattices*, *Commun. Pure Appl. Math.* **54** (2001) 153 [math/9912143].

[352] L. Faybusovich and M. Gekhtman, *On Schur flows*, *J. Phys. A* **32** (1999) 4671.

[353] A. Mukaihira and Y. Nakamura, *Schur flow for orthogonal polynomials on the unit circle and its integrable discretization*, *J. Comput. Appl. Math.* **139** (2002) 75.

[354] I. Nenciu, *CMV matrices in random matrix theory and integrable systems: a survey*, *J. Phys. A* **39** (2006) 8811 [math-ph/0510045].

[355] I. Bars, *U(N) Integral for Generating Functional in Lattice Gauge Theory*, *J. Math. Phys.* **21** (1980) 2678.

[356] G. Mandal, *Phase Structure of Unitary Matrix Models*, *Mod. Phys. Lett. A* **5** (1990) 1147.

[357] G. Bonnet, F. David and B. Eynard, *Breakdown of universality in multicut matrix models*, *J. Phys. A* **33** (2000) 6739 [cond-mat/0003324].

[358] H. Widom, *Toeplitz determinants with singular generating functions*, *Amer. J. Math.* **95** (1973) 333.

[359] J. G. Russo, *Deformed Cauchy random matrix ensembles and large N phase transitions*, *JHEP* **11** (2020) 014 [2006.00672].

[360] M. Hanada and B. Robinson, *Partial-Symmetry-Breaking Phase Transitions*, *Phys. Rev. D* **102** (2020) 096013 [1911.06223].

[361] K. Demeterfi and C.-I. Tan, *Periodic regularization, multiband structure and orthogonal polynomials*, *Phys. Rev. D* **43** (1991) 2622.

[362] S. Elitzur, *Impossibility of Spontaneously Breaking Local Symmetries*, *Phys. Rev. D* **12** (1975) 3978.

[363] P. M. Hernandez, *Lattice field theory fundamentals*, in *Modern perspectives in lattice QCD: Quantum field theory and high performance computing. Proceedings, International School, 93rd Session, Les Houches, France, August 3-28, 2009*, (Oxford, UK), pp. 1–91, Oxford University Press, 8, 2009.

[364] M. Mariño, *Nonperturbative effects and nonperturbative definitions in matrix models and topological strings*, *JHEP* **12** (2008) 114 [0805.3033].

[365] H. Neuberger, *Nonperturbative Contributions in Models With a Nonanalytic Behavior at Infinite N*, *Nucl. Phys. B* **179** (1981) 253.

[366] R. Dijkgraaf and C. Vafa, *Matrix models, topological strings, and supersymmetric gauge theories*, *Nucl. Phys. B* **644** (2002) 3 [hep-th/0206255].

[367] L. Chekhov, A. Marshakov, A. Mironov and D. Vasiliev, *Complex geometry of matrix models*, *Proc. Steklov Inst. Math.* **251** (2005) 254 [hep-th/0506075].

[368] G. Álvarez, L. M. Alonso and E. Medina, *Determination of S-curves with applications to the theory of non-Hermitian orthogonal polynomials*, *J. Stat. Mech.* **2013** (2013) P06006 [1305.3028].

[369] A. Bilal and S. Metzger, *Special geometry of local Calabi-Yau manifolds and superpotentials from holomorphic matrix models*, *JHEP* **08** (2005) 097 [hep-th/0503173].

[370] G. Álvarez, L. Martínez-Alonso and E. Medina, *Complex saddles in the Gross-Witten-Wadia matrix model*, *Phys. Rev. D* **94** (2016) 105010 [1610.09948].

[371] C. Copetti, A. Grassi, Z. Komargodski and L. Tizzano, *Delayed Deconfinement and the Hawking-Page Transition*, 2008.04950.

[372] K. Kodaira, *On compact analytic surfaces II*, *Ann. Math.* **77** (1963) 563.

[373] N. Seiberg and E. Witten, *Monopoles, duality and chiral symmetry breaking in N=2 supersymmetric QCD*, *Nucl. Phys. B* **431** (1994) 484 [hep-th/9408099].

[374] D. Kaledin, *Symplectic singularities from the Poisson point of view*, *J. Reine Angew. Math.* **2006** (2006) 135 [math/0310186].

[375] P. C. Argyres and M. Martone, *Towards a classification of rank r $\mathcal{N}$ = 2 SCFTs. Part II. Special Kahler stratification of the Coulomb branch*, *JHEP* **12** (2020) 022 [2007.00012].

[376] A. Bourget, S. Cabrera, J. F. Grimminger, A. Hanany, M. Sperling, A. Zajac and Z. Zhong, *The Higgs mechanism – Hasse diagrams for symplectic singularities*, *JHEP* **01** (2020) 157 [1908.04245].

[377] J. F. Grimminger and A. Hanany, *Hasse diagrams for 3d $\mathcal{N}$ = 4 quiver gauge theories – Inversion and the full moduli space*, *JHEP* **09** (2020) 159 [2004.01675].

[378] T. Claeys, T. Grava and K. D. T.-R. McLaughlin, *Asymptotics for the partition function in two-cut random matrix models*, *Commun. Math. Phys.* **339** (2015) 513 [1410.7001].

[379] A. A. Migdal, *Recursion Equations in Gauge Theories*, *Sov. Phys. JETP* **42** (1975) 413.

[380] B. E. Rusakov, *Loop averages and partition functions in U(N) gauge theory on two-dimensional manifolds*, *Mod. Phys. Lett. A* **5** (1990) 693.

[381] V. Pestun, *Localization of the four-dimensional N=4 SYM to a two-sphere and 1/8 BPS Wilson loops*, *JHEP* **12** (2012) 067 [0906.0638].

[382] S. Giombi and V. Pestun, *Correlators of local operators and 1/8 BPS Wilson loops on S**2 from 2d YM and matrix models*, *JHEP* **10** (2010) 033 [0906.1572].

[383] Y. Wang, *Taming defects in $\mathcal{N}$ = 4 super-Yang-Mills*, *JHEP* **08** (2020) 021 [2003.11016].

[384] S. Cordes, G. W. Moore and S. Ramgoolam, *Lectures on 2-d Yang-Mills theory, equivariant cohomology and topological field theories*, *Nucl. Phys. Proc. Suppl.* **41** (1995) 184 [hep-th/9411210].

[385] D. J. Gross and A. Matytsin, *Instanton induced large N phase transitions in two-dimensional and four-dimensional QCD*, *Nucl. Phys.* **B429** (1994) 50 [hep-th/9404004].

[386] M. R. Douglas, S. H. Katz and C. Vafa, *Small instantons, Del Pezzo surfaces and type I-prime theory*, *Nucl. Phys. B* **497** (1997) 155 [hep-th/9609071].

[387] L. Bhardwaj, *Dualities of 5d gauge theories from S-duality*, *JHEP* **07** (2020) 012 [1909.05250].

[388] R. Gopakumar and C. Vafa, *M theory and topological strings. 2.*, hep-th/9812127.

[389] D. Gaiotto, $\mathcal{N} = 2$ *dualities*, *JHEP* **08** (2012) 034 [0904.2715].

[390] D. Gaiotto, G. W. Moore and A. Neitzke, *Wall-crossing, Hitchin Systems, and the WKB Approximation*, *Adv. Math.* **234** (2013) 239 [0907.3987].

[391] A. Gadde, L. Rastelli, S. S. Razamat and W. Yan, *The 4d Superconformal Index from q-deformed 2d Yang-Mills*, *Phys. Rev. Lett.* **106** (2011) 241602 [1104.3850].

[392] M. R. Douglas, *On D=5 super Yang-Mills theory and (2,0) theory*, *JHEP* **02** (2011) 011 [1012.2880].

[393] N. Lambert, C. Papageorgakis and M. Schmidt-Sommerfeld, *M5-Branes, D4-Branes and Quantum 5D super-Yang-Mills*, *JHEP* **01** (2011) 083 [1012.2882].

[394] Y. Fukuda, T. Kawano and N. Matsumiya, *5D SYM and 2D q-Deformed YM*, *Nucl. Phys. B* **869** (2013) 493 [1210.2855].

[395] N. Drukker, T. Okuda and F. Passerini, *Exact results for vortex loop operators in 3d supersymmetric theories*, *JHEP* **07** (2014) 137 [1211.3409].

[396] T. Kawano and N. Matsumiya, *5D SYM on 3D Deformed Spheres*, *Nucl. Phys. B* **898** (2015) 456 [1505.06565].

[397] L. Griguolo, D. Seminara, R. J. Szabo and A. Tanzini, *Black holes, instanton counting on toric singularities and q-deformed two-dimensional Yang-Mills theory*, *Nucl. Phys. B* **772** (2007) 1 [hep-th/0610155].

[398] J. Qiu and M. Zabzine, *5D Super Yang-Mills on $Y^{p,q}$ Sasaki-Einstein manifolds*, *Commun. Math. Phys.* **333** (2015) 861 [1307.3149].

[399] J. Qiu and M. Zabzine, *Factorization of 5D super Yang-Mills theory on $Y^{p,q}$ spaces*, *Phys. Rev. D* **89** (2014) 065040 [1312.3475].

[400] D. Gaiotto, L. Rastelli and S. S. Razamat, *Bootstrapping the superconformal index with surface defects*, *JHEP* **01** (2013) 022 [1207.3577].

[401] J. Kinney, J. M. Maldacena, S. Minwalla and S. Raju, *An Index for 4 dimensional super conformal theories*, *Commun. Math. Phys.* **275** (2007) 209 [hep-th/0510251].

[402] F. A. H. Dolan, V. P. Spiridonov and G. S. Vartanov, *From 4d superconformal indices to 3d partition functions*, *Phys. Lett. B* **704** (2011) 234 [1104.1787].

[403] A. Gadde and W. Yan, *Reducing the 4d Index to the S^3 Partition Function*, *JHEP* **12** (2012) 003 [1104.2592].

[404] Y. Imamura, *Relation between the 4d superconformal index and the S^3 partition function*, *JHEP* **09** (2011) 133 [1104.4482].

[405] Y. Imamura and D. Yokoyama, *N=2 supersymmetric theories on squashed three-sphere*, *Phys. Rev. D* **85** (2012) 025015 [1109.4734].

[406] A. Gadde, L. Rastelli, S. S. Razamat and W. Yan, *Gauge Theories and Macdonald Polynomials*, *Commun. Math. Phys.* **319** (2013) 147 [1110.3740].

[407] Y. Tachikawa, *4d partition function on $S^1 \times S^3$ and 2d Yang-Mills with nonzero area*, *PTEP* **2013** (2013) 013B01 [1207.3497].

[408] M. Aganagic, H. Ooguri, N. Saulina and C. Vafa, *Black holes, q-deformed 2d Yang-Mills, and non-perturbative topological strings*, *Nucl. Phys. B* **715** (2005) 304 [hep-th/0411280].

[409] N. Caporaso, M. Cirafici, L. Griguolo, S. Pasquetti, D. Seminara and R. J. Szabo, *Topological strings and large N phase transitions. I. Nonchiral expansion of q-deformed Yang-Mills theory*, *JHEP* **01** (2006) 035 [hep-th/0509041].

[410] S. S. Razamat and B. Willett, *Star-shaped quiver theories with flux*, *Phys. Rev. D* **101** (2020) 065004 [1911.00956].

[411] J. Yagi, *3d TQFT from 6d SCFT*, *JHEP* **08** (2013) 017 [1305.0291].

[412] S. Lee and M. Yamazaki, *3d Chern-Simons Theory from M5-branes*, *JHEP* **12** (2013) 035 [1305.2429].

[413] D. Gang, N. Kim, M. Romo and M. Yamazaki, *Aspects of Defects in 3d-3d Correspondence*, *JHEP* **10** (2016) 062 [1510.05011].

[414] C. Cordova and D. L. Jafferis, *Complex Chern-Simons from M5-branes on the Squashed Three-Sphere*, *JHEP* **11** (2017) 119 [1305.2891].

[415] K. Ohta and Y. Yoshida, *Non-Abelian Localization for Supersymmetric Yang-Mills-Chern-Simons Theories on Seifert Manifold*, *Phys. Rev. D* **86** (2012) 105018 [1205.0046].

[416] G. Festuccia and N. Seiberg, *Rigid Supersymmetric Theories in Curved Superspace*, *JHEP* **06** (2011) 114 [1105.0689].

[417] C. Closset, T. T. Dumitrescu, G. Festuccia and Z. Komargodski, *Supersymmetric Field Theories on Three-Manifolds*, *JHEP* **05** (2013) 017 [1212.3388].

[418] L. F. Alday, D. Martelli, P. Richmond and J. Sparks, *Localization on Three-Manifolds*, *JHEP* **10** (2013) 095 [1307.6848].

[419] Y. Pan, *Rigid Supersymmetry on 5-dimensional Riemannian Manifolds and Contact Geometry*, *JHEP* **05** (2014) 041 [1308.1567].

[420] L. F. Alday, P. Benetti Genolini, M. Fluder, P. Richmond and J. Sparks, *Supersymmetric gauge theories on five-manifolds*, *JHEP* **08** (2015) 007 [1503.09090].

[421] C. Closset and H. Kim, *Three-dimensional $\mathcal{N} = 2$ supersymmetric gauge theories and partition functions on Seifert manifolds: A review*, *Int. J. Mod. Phys. A* **34** (2019) 1930011 [1908.08875].

[422] J. Qiu and M. Zabzine, *On twisted N=2 5D super Yang-Mills theory*, *Lett. Math. Phys.* **106** (2016) 1 [1409.1058].

[423] M. Blau and G. Thompson, *Lectures on 2-d gauge theories: Topological aspects and path integral techniques*, in *Summer School in High-energy Physics and Cosmology (Includes Workshop on Strings, Gravity, and Related Topics 29-30 Jul 1993)*, 10, 1993, hep-th/9310144.

[424] J. Qiu and M. Zabzine, *Review of localization for 5d supersymmetric gauge theories*, *J. Phys. A* **50** (2017) 443014 [1608.02966].

[425] N. Caporaso, M. Cirafici, L. Griguolo, S. Pasquetti, D. Seminara and R. J. Szabo, *Topological Strings, Two-Dimensional Yang-Mills Theory and Chern-Simons Theory on Torus Bundles*, *Adv. Theor. Math. Phys.* **12** (2008) 981 [hep-th/0609129].

[426] C. Closset, H. Kim and B. Willett, *Supersymmetric partition functions and the three-dimensional A-twist*, *JHEP* **03** (2017) 074 [1701.03171].

[427] M. F. Atiyah, *Elliptic Operators and Compact Groups*, vol. 401. Springer-Verlag, Berline, Germany, 1974, 10.1007/BFb0057821.

[428] J. R. Quine, S. H. Heydari and R. Y. Song, *Zeta regularized products*, *Trans. Amer. Math. Soc.* **338** (1993) 213.

[429] C. Closset, T. T. Dumitrescu, G. Festuccia and Z. Komargodski, *The Geometry of Supersymmetric Partition Functions*, *JHEP* **01** (2014) 124 [1309.5876].

[430] J. Schmude, *Localisation on Sasaki-Einstein manifolds from holomorphic functions on the cone*, *JHEP* **01** (2015) 119 [1401.3266].

[431] D. Baraglia and P. Hekmati, *Moduli Spaces of Contact Instantons*, *Adv. Math.* **294** (2016) 562 [1401.5140].

[432] A. Bawane, G. Bonelli, M. Ronzani and A. Tanzini, $\mathcal{N} = 2$ *supersymmetric gauge theories on* $S^2 \times S^2$ *and Liouville Gravity*, *JHEP* **07** (2015) 054 [1411.2762].

[433] N. Banerjee, B. de Wit and S. Katmadas, *The Off-Shell 4D/5D Connection*, *JHEP* **03** (2012) 061 [1112.5371].

[434] J. G. Russo, *A Note on perturbation series in supersymmetric gauge theories*, *JHEP* **06** (2012) 038 [1203.5061].

[435] A. Brini, L. Griguolo, D. Seminara and A. Tanzini, *Chern-Simons theory on L(p,q) lens spaces and Gopakumar-Vafa duality*, *J. Geom. Phys.* **60** (2010) 417 [0809.1610].

[436] R. J. Szabo and M. Tierz, *q-deformations of two-dimensional Yang-Mills theory: Classification, categorification and refinement*, *Nucl. Phys.* **B876** (2013) 234 [1305.1580].

[437] T. Friedrich and I. Kath, *Einstein manifolds of dimension five with small first eigenvalue of the Dirac operator*, *J. Diff. Geom.* **29** (1989) 263.

[438] J. P. Gauntlett, D. Martelli, J. Sparks and D. Waldram, *Sasaki-Einstein metrics on* $S^2 \times S^3$, *Adv. Theor. Math. Phys.* **8** (2004) 711 [hep-th/0403002].

[439] J. C. Geipel, O. Lechtenfeld, A. D. Popov and R. J. Szabo, *Sasakian quiver gauge theories and instantons on the conifold*, *Nucl. Phys. B* **907** (2016) 445 [1601.05719].

[440] Y. Jiang, *A pedagogical review on solvable irrelevant deformations of 2D quantum field theory*, *Commun. Theor. Phys.* **73** (2021) 057201 [1904.13376].

[441] J. Cardy, *The $T\overline{T}$ deformation of quantum field theory as random geometry*, *JHEP* **10** (2018) 186 [1801.06895].

[442] S. Dubovsky, V. Gorbenko and M. Mirbabayi, *Asymptotic fragility, near AdS_2 holography and $T\overline{T}$*, *JHEP* **09** (2017) 136 [1706.06604].

[443] S. Dubovsky, V. Gorbenko and G. Hernández-Chifflet, *$T\bar{T}$ Partition Function from Topological Gravity*, *JHEP* **09** (2018) 158 [1805.07386].

[444] R. Conti, S. Negro and R. Tateo, *The $T\overline{T}$ perturbation and its geometric interpretation*, *JHEP* **02** (2019) 085 [1809.09593].

[445] T. D. Brennan, C. Ferko, E. Martinec and S. Sethi, *Defining the $T\overline{T}$ deformation on* AdS_2, 2005.00431.

[446] L. McGough, M. Mezei and H. Verlinde, *Moving the* CFT *into the bulk with $T\overline{T}$*, *JHEP* **04** (2018) 010 [1611.03470].

[447] A. Giveon, N. Itzhaki and D. Kutasov, *A solvable irrelevant deformation of AdS_3/CFT_2*, *JHEP* **12** (2017) 155 [1707.05800].

[448] T. Hartman, J. Kruthoff, E. Shaghoulian and A. Tajdini, *Holography at finite cutoff with a T^2 deformation*, *JHEP* **03** (2019) 004 [1807.11401].

[449] P. Kraus, J. Liu and D. Marolf, *Cutoff* AdS_3 *versus the $T\overline{T}$ deformation*, *JHEP* **07** (2018) 027 [1801.02714].

[450] P. Caputa, S. Datta and V. Shyam, *Sphere partition functions & cut-off* AdS, *JHEP* **05** (2019) 112 [1902.10893].

[451] M. Guica and R. Monten, *$T\bar{T}$ and the mirage of a bulk cutoff*, *SciPost Phys.* **10** (2021) 024 [1906.11251].

[452] A. Lewkowycz, J. Liu, E. Silverstein and G. Torroba, *$T\overline{T}$ and EE, with implications for* (A)dS *subregion encodings*, *JHEP* **04** (2020) 152 [1909.13808].

[453] L. Apolo, S. Detournay and W. Song, *TsT, $T\overline{T}$ and black strings*, *JHEP* **06** (2020) 109 [1911.12359].

[454] Y. Li and Y. Zhou, *Cutoff AdS_3 versus $T\overline{T}$ CFT_2 in the large central charge sector: correlators of energy-momentum tensor*, *JHEP* **12** (2020) 168 [2005.01693].

[455] M. Baggio and A. Sfondrini, *Strings on NS–NS backgrounds as integrable deformations*, *Phys. Rev. D* **98** (2018) 021902 [1804.01998].

[456] G. Bonelli, N. Doroud and M. Zhu, *$T\overline{T}$-deformations in closed form*, *JHEP* **06** (2018) 149 [1804.10967].

[457] M. Baggio, A. Sfondrini, G. Tartaglino-Mazzucchelli and H. Walsh, *On $T\overline{T}$ deformations and supersymmetry*, *JHEP* **06** (2019) 063 [1811.00533].

[458] C.-K. Chang, C. Ferko and S. Sethi, *Supersymmetry and $T\overline{T}$ deformations*, *JHEP* **04** (2019) 131 [1811.01895].

[459] H. Jiang, A. Sfondrini and G. Tartaglino-Mazzucchelli, *$T\overline{T}$ deformations with $\mathcal{N} = (0, 2)$ supersymmetry*, *Phys. Rev. D* **100** (2019) 046017 [1904.04760].

[460] C.-K. Chang, C. Ferko, S. Sethi, A. Sfondrini and G. Tartaglino-Mazzucchelli, *$T\overline{T}$ flows and $(2, 2)$ supersymmetry*, *Phys. Rev. D* **101** (2020) 026008 [1906.00467].

[461] C. Ferko, H. Jiang, S. Sethi and G. Tartaglino-Mazzucchelli, *Non-linear supersymmetry and $T\overline{T}$-like flows*, *JHEP* **02** (2020) 016 [1910.01599].

[462] S. He, J.-R. Sun and Y. Sun, *The correlation function of $(1, 1)$ and $(2, 2)$ supersymmetric theories with $T\overline{T}$ deformation*, *JHEP* **04** (2020) 100 [1912.11461].

[463] B. Le Floch and M. Mezei, *Solving a family of $T\overline{T}$-like theories*, 1903.07606.

[464] R. Conti, S. Negro and R. Tateo, *Conserved currents and $T\overline{T}_s$ irrelevant deformations of $2D$ integrable field theories*, *JHEP* **11** (2019) 120 [1904.09141].

[465] G. Hernández-Chifflet, S. Negro and A. Sfondrini, *Flow equations for generalised $T\overline{T}$ deformations*, *Phys. Rev. Lett.* **124** (2020) 200601 [1911.12233].

[466] M. Taylor, *TT deformations in general dimensions*, 1805.10287.

[467] A. Belin, A. Lewkowycz and G. Sarosi, *Gravitational path integral from the T^2 deformation*, *JHEP* **09** (2020) 156 [2006.01835].

[468] S. Datta and Y. Jiang, *$T\bar{T}$ deformed partition functions*, *JHEP* **08** (2018) 106 [1806.07426].

[469] O. Aharony, S. Datta, A. Giveon, Y. Jiang and D. Kutasov, *Modular invariance and uniqueness of $T\bar{T}$ deformed CFT*, (2018) [1808.02492].

[470] J. Cardy, *$T\overline{T}$ deformations of non-Lorentz invariant field theories*, (2018) [1809.07849].

[471] Y. Jiang, *Expectation value of $T\overline{T}$ operator in curved spacetimes*, *JHEP* **02** (2020) 094 [1903.07561].

[472] S. He and Y. Sun, *Correlation functions of CFTs on a torus with a $T\overline{T}$ deformation*, *Phys. Rev. D* **102** (2020) 026023 [2004.07486].

[473] E. Beratto, M. Billò and M. Caselle, *$T\overline{T}$ deformation of the compactified boson and its interpretation in lattice gauge theory*, *Phys. Rev. D* **102** (2020) 014504 [1912.08654].

[474] R. Conti, L. Iannella, S. Negro and R. Tateo, *Generalised Born-Infeld models, Lax operators and the* $\mathrm{T\overline{T}}$ *perturbation*, *JHEP* **11** (2018) 007 [1806.11515].

[475] A. Ireland and V. Shyam, *$T\overline{T}$ deformed YM_2 on general backgrounds from an integral transformation*, *JHEP* **07** (2020) 058 [1912.04686].

[476] E. Witten, *Topological sigma models*, *Commun. Math. Phys.* **118** (1988) 411.

[477] E. Witten, *Mirror manifolds and topological field theory*, *AMS/IP Stud. Adv. Math.* **9** (1998) 121 [hep-th/9112056].

[478] R. Malik, *New topological field theories in two-dimensions*, *J. Phys. A* **34** (2001) 4167 [hep-th/0012085].

[479] E. A. Coleman, J. Aguilera-Damia, D. Z. Freedman and R. M. Soni, *$T\overline{T}$-deformed actions and $(1,1)$ supersymmetry*, *JHEP* **10** (2019) 080 [1906.05439].

[480] A. J. Tolley, *$T\overline{T}$ deformations, massive gravity and non-critical strings*, *JHEP* **06** (2020) 050 [1911.06142].

[481] M. F. Atiyah, *Topological quantum field theories*, *Inst. Hautes Études Sci. Publ. Math.* **68** (1989) 175.

[482] G. B. Segal, *The definition of conformal field theory*, in *Differential geometrical methods in theoretical physics*, pp. 165–171. Springer, 1988.

[483] E. A. Mazenc, V. Shyam and R. M. Soni, *A $T\overline{T}$ deformation for curved spacetimes from 3d gravity*, 1912.09179.

[484] O. Ganor, J. Sonnenschein and S. Yankielowicz, *The String theory approach to generalized 2-D Yang-Mills theory*, *Nucl. Phys. B* **434** (1995) 139 [hep-th/9407114].

[485] T. D. Brennan, C. Ferko and S. Sethi, *A non-abelian analogue of DBI from $T\overline{T}$*, *SciPost Phys.* **8** (2020) 052 [1912.12389].

[486] M. Aganagic and K. Schaeffer, *Refined black hole ensembles and topological strings*, *JHEP* **01** (2013) 060 [1210.1865].

[487] C. Vafa, *Two-dimensional Yang-Mills, black holes and topological strings*, hep-th/0406058.

[488] G. W. Moore and N. Seiberg, *Taming the conformal zoo*, *Phys. Lett. B* **220** (1989) 422.

[489] N. Caporaso, M. Cirafici, L. Griguolo, S. Pasquetti, D. Seminara and R. J. Szabo, *Topological strings and large N phase transitions. II. Chiral expansion of q-deformed Yang-Mills theory*, *JHEP* **01** (2006) 036 [hep-th/0511043].

[490] D. J. Gross and W. Taylor, *Two-dimensional QCD is a string theory*, *Nucl. Phys. B* **400** (1993) 181 [hep-th/9301068].

[491] M. R. Douglas, *Conformal field theory techniques for large N group theory*, hep-th/9303159.

[492] M. R. Douglas, *Conformal field theory techniques in large N Yang-Mills theory*, in *NATO Advanced Research Workshop on New Developments in String Theory, Conformal Models and Topological Field Theory Cargese, France*, 1993, hep-th/9311130.

[493] R. Dijkgraaf, R. Gopakumar, H. Ooguri and C. Vafa, *Baby universes in string theory*, *Phys. Rev. D* **73** (2006) 066002 [hep-th/0504221].

[494] D. Jafferis and J. Marsano, *A DK phase transition in q-deformed Yang-Mills on S**2 and topological strings*, hep-th/0509004.

[495] J. A. Minahan and A. P. Polychronakos, *Classical solutions for two-dimensional QCD on the sphere*, *Nucl. Phys.* **B422** (1994) 172 [hep-th/9309119].

[496] E. Witten, *Two-dimensional gauge theories revisited*, *J. Geom. Phys.* **9** (1992) 303 [hep-th/9204083].

[497] M. Blau and G. Thompson, *Derivation of the Verlinde formula from Chern-Simons theory and the G/G model*, *Nucl. Phys. B* **408** (1993) 345 [hep-th/9305010].

[498] X. Arsiwalla, R. Boels, M. Mariño and A. Sinkovics, *Phase transitions in q-deformed 2-D Yang-Mills theory and topological strings*, *Phys. Rev. D* **73** (2006) 026005 [hep-th/0509002].

[499] R. J. Szabo, *Instantons, topological strings and enumerative geometry*, *Adv. Math. Phys.* **2010** (2010) 107857 [0912.1509].

[500] A. Gorsky, D. Pavshinkin and A. Tyutyakina, *$T\bar{T}$-deformed 2D Yang-Mills at large N: collective field theory and phase transitions*, *JHEP* **03** (2021) 142 [2012.09467].

[501] Z. Kökényesi, A. Sinkovics and R. J. Szabo, *Refined Chern-Simons theory and (q, t)-deformed Yang-Mills theory: Semi-classical expansion and planar limit*, *JHEP* **10** (2013) 067 [1306.1707].