



Article scientifique

Article

2010

Accepted version

Open Access

This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

Large N duality beyond the genus expansion

Marino Beiras, Marcos; Pasquetti, Sara; Putrov, Pavel

How to cite

MARINO BEIRAS, Marcos, PASQUETTI, Sara, PUTROV, Pavel. Large N duality beyond the genus expansion. In: The journal of high energy physics, 2010, vol. 2010, n° 7, p. 41. doi: 10.1007/jhep07(2010)074

This publication URL: <https://archive-ouverte.unige.ch/unige:15778>

Publication DOI: [10.1007/jhep07\(2010\)074](https://doi.org/10.1007/jhep07(2010)074)

Large N duality beyond the genus expansion

Marcos Mariño^{a,b}, Sara Pasquetti^c and Pavel Putrov^b

^a *Département de Physique Théorique et* ^b *Section de Mathématiques,*
Université de Genève, Genève, CH-1211 Switzerland

^c *Department of Physics, CERN,*
Genève, CH-1211 Switzerland

marcos.marino@unige.ch, sara.pasquetti@cern.ch, pavel.putrov@unige.ch

ABSTRACT: We study non-perturbative aspects of the large N duality between Chern–Simons theory and topological strings, and we find a rich structure of large N phase transitions in the complex plane of the 't Hooft parameter. These transitions are due to large N instanton effects, and they can be regarded as a deformation of the Stokes phenomenon. Moreover, we show that, for generic values of the 't Hooft coupling, instanton effects are not exponentially suppressed at large N and they correct the genus expansion. This phenomenon was first discovered in the context of matrix models, and we interpret it as a generalization of the oscillatory asymptotics along anti–Stokes lines. In the string dual, the instanton effects can be interpreted as corrections to the saddle string geometry due to discretized neighbouring geometries. As a mathematical application, we obtain the $1/N$ asymptotics of the partition function of Chern–Simons theory on the lens space $L(2,1)$, and we test it numerically to high precision in order to exhibit the importance of instanton effects.

Contents

1. Introduction	1
2. Instanton corrections in matrix models and topological strings	4
3. Large N duality for Chern–Simons theory	8
3.1 Chern–Simons theory on lens spaces	9
3.2 The dual topological string	11
3.3 Two special slices	14
4. Large N phase diagram	14
4.1 Real g_s : the boundary expansion	15
4.2 Imaginary g_s : the interior expansion	17
4.3 The phase diagram for complex g_s	22
4.4 Stokes phenomenon and target geometries	24
4.5 Background independence	25
5. Stokes phenomenon and matrix models	26
5.1 Review of the Stokes phenomenon for the Airy function	26
5.2 The Stokes phenomenon and the cubic matrix model	28
6. Conclusions	31
A. Some aspects of topological string theory on the local $\mathbb{P}^1 \times \mathbb{P}^1$ geometry	32

1. Introduction

According to large N dualities, the $1/N$ expansion of different observables in $U(N)$ gauge theories can be reinterpreted as a genus expansion in an appropriate string theory. For example, the gauge theory free energy at finite volume has an expansion of the form

$$F(g_s, N) \sim \sum_{g=0}^{\infty} g_s^{2g-2} F_g(t), \quad (1.1)$$

where g_s is the gauge theory coupling constant and $t = g_s N$ is the 't Hooft parameter. In large N dualities, g_s is identified with the string coupling constant, the 't Hooft parameter becomes a geometric modulus of the string target space, and the amplitudes $F_g(t)$ are identified with free energies at genus g of a string theory.

The $1/N$ expansion in gauge theory is an asymptotic expansion, and it is expected to have corrections of the form

$$F(g_s, N) \sim \sum_{g=0}^{\infty} g_s^{2g-2} F_g(t) + \mathcal{O}(e^{-A(t)/g_s}). \quad (1.2)$$

These corrections are invisible in the traditional $1/N$ asymptotics¹, and they are associated to large N instantons. Although large N instantons are typically built upon classical gauge theory instantons, they should not be confused with them. A classical instanton is a saddle point of the classical action, while a large N instanton is a saddle point of the large N effective action; see [1, 40] and [23] for explicit examples in the $\mathbb{C}\mathbb{P}^N$ model and in two-dimensional Yang–Mills theory, respectively. In particular, the action $A(t)$ of a large N instanton is a non-trivial function of the 't Hooft parameter which includes quantum corrections at all loops. In some cases, large N instantons can be interpreted in terms of D-branes of the string theory dual, as it happens for example in matrix models of non-critical strings.

One could think that the instanton corrections in (1.2) are unimportant since they are exponentially suppressed at large N , but this is not always the case. As we move in parameter space they might become of order one and start contributing to the large N asymptotics. This scenario was advocated long ago by Neuberger [41] in order to explain the occurrence of critical points in Hermitian matrix models and the third-order large N phase transitions of unitary matrix models [24, 47]. The same mechanism explains as well the Douglas–Kazakov transition in two-dimensional Yang–Mills on the sphere [16, 23]. In these instanton-driven phase transitions, as we change the 't Hooft parameter at fixed g_s , the action of the instanton $A(t)$ vanishes at some finite value $t = t_c$. This leads in general to different $1/N$ expansions for the regions $t > t_c$ and $t < t_c$.

On the other hand, if we regard g_s as a complex variable, we should expect that the $1/N$ asymptotics changes discontinuously as we change the argument of g_s (and keep $|t|$ fixed). In classical asymptotic analysis, this discontinuous change is the well-known Stokes phenomenon. The reason for this phenomenon is in fact the same one that underlies instanton-induced phase transitions: an exponentially small quantity, due to a sub-leading saddle point, becomes less and less suppressed as we change the argument of the expansion variable. Along the so-called anti-Stokes line, the contribution of this saddle is of order one and has to be included in the asymptotics.

In general, a large N gauge theory with instanton sectors, should display a complex pattern of large N phase transitions, combining the instanton-driven phase transitions at fixed g_s with the Stokes-like transitions as we vary the argument of g_s . If the gauge theory that we are studying has a string dual in at least one of the phases, it is certainly interesting to understand what are the implications of these transitions in the string picture.

In this paper we will analyze these issues in a family of models which have well-understood large N duals, namely Chern–Simons theory on the lens space $L(p, 1)$. This gauge theory is described by topological string theory on a Calabi–Yau space given by an A_p fibration over a two-sphere [22, 3, 27, 26]. However, this large N string dual is intrinsically semiclassical, since it describes a generic but fixed saddle-point of the gauge theory. A full non-perturbative study of this duality, where we sum over all saddle-points, reveals a surprisingly rich phase structure as we move in the complex plane of the 't Hooft coupling. The main results of our analysis are the following:

1. For each complex value of t , the large N asymptotics of the non-perturbative free energy is dominated by a fixed string target geometry, i.e. by a point in the Calabi–Yau moduli space. However, this “saddle geometry” might change as we vary t .

¹By traditional or classical asymptotics, we mean asymptotic expansions in which non-analytic terms are not taken into account. There are refinements of traditional asymptotics, which go sometimes under the name of “hypercasympotics,” where these terms are included in a systematic way, see [7] for an overview.

2. For some values of t , the $1/N$ asymptotics is given by a conventional genus expansion. There are non-perturbative effects due to the contributions of neighbouring geometries, but they are exponentially suppressed. For other values of t , the asymptotics has an oscillatory behavior due to large N instantons. This type of behavior was first observed in the context of matrix models in [10], and further studied in [17, 18]. In particular, the large N asymptotics is no longer given by a genus expansion around a fixed geometry, and corrections due to neighboring geometries are crucial, already at the next-to-leading order.
3. The change of saddle geometry as we change the complex 't Hooft parameter can be regarded as a generalized or “deformed” Stokes phenomenon. In the limit of vanishing 't Hooft coupling, the asymptotics changes discontinuously as we change the argument of the string coupling constant, and we recover the classical Stokes phenomenon. For finite t the phenomenon is smoothed out, and the asymptotics changes continuously.

In practice we have focused on the simplest, nontrivial model in the family, namely $L(2, 1) = \mathbb{RP}^3$. The gauge theory saddle points are characterized by a symmetry breaking pattern $U(N) \rightarrow U(N_1) \times U(N_2)$. The dual Calabi–Yau manifold is the so-called local \mathbb{F}_0 geometry, where $\mathbb{F}_0 = \mathbb{P}^1 \times \mathbb{P}^1$ is a Hirzebruch surface. It has two Kähler parameters, corresponding at large radius to the sizes of the two \mathbb{P}^1 s in \mathbb{F}_0 . The partial 't Hooft couplings t_1, t_2 , where $t_i = g_s N_i$, parametrize the Kähler moduli space near the orbifold point described in [3], and their sum $t_1 + t_2$ is fixed to be the total 't Hooft parameter of the gauge theory. Each saddle geometry is characterized by an “equilibrium value” for t_2 . The phase diagram in the region $|t| > 0, 0 \leq \text{Arg}(g_s) \leq \pi/2$ is shown in Fig. 7.

Our analysis shows very clearly that the large N instantons analyzed in [10, 35, 36, 34, 17, 18, 43] are crucial in order to understand large N dualities. To be precise, for generic values of the 't Hooft parameter, the correct large N asymptotics of the exact gauge theory partition function involves these instanton effects already at next-to-leading order. In order to dispel any doubt about this, we study in detail the partition function of Chern–Simons theory on \mathbb{RP}^3 , for imaginary values of the coupling constant. These are precisely the “physical” or “on-shell” values of the coupling constant in Chern–Simons theory. Our analysis shows that these values correspond to an anti-Stokes line with oscillatory asymptotics, in which the traditional genus expansion has to be supplemented with instanton corrections. We test numerically our predictions for the instanton corrected asymptotics against the exact non-perturbative Chern–Simons partition function, and we obtain an impressive agreement.

Our analysis also clarifies the issue of background dependence in topological string theory. Let us focus again on the example of Chern–Simons theory on \mathbb{RP}^3 . Perturbative topological string theory depends on a choice of background (t_1, t_2) . The gauge theory partition function is background independent, in the sense that it only depends on the total 't Hooft parameter $t = t_1 + t_2$ (which is, as needed for consistency, a modular invariant in the dual topological string). However, the large N asymptotics of the partition function is “peaked” around a particular value of (t_1, t_2) which depends on t and corresponds to a particular saddle geometry. This fixed background emerges as an equilibrium value as a consequence of the large N limit, but it has no meaning at finite N . In fact, in the full theory, the value of the background (specified by, say, the value of t_2) is not fixed, since t_2 is rather an internal or fluctuating variable which we have to sum over. But as usual in statistical mechanics, in the thermodynamic limit $N \rightarrow \infty$ this fluctuating variable is peaked around an equilibrium value.

We would like to point out that the role of instanton effects in the $1/N$ expansion and their string theory interpretation has been addressed before in different contexts. For example, the

instanton-induced Douglas–Kazakov phase transition has been interpreted as a breakdown of the large N string dual [16]. In AdS/CFT at finite temperature, an incarnation of the Gross–Witten–Wadia phase transition has been argued to correspond to a breakdown of the geometric description of the string target [4]². The importance of exponentially small effects, and their effects on the string target space, has been emphasized in [37], which studies as well the Stokes phenomenon in the open string moduli space. In [42], a large N phase transition in $\mathcal{N} = 4$ super Yang–Mills theory on K3 was found, separating a phase dominated by the zero-instanton sector from a phase dominated by an instanton sector with large instanton number. In the context of matrix models, general techniques to analyze large N phase transitions have been developed along the years, culminating in the formulation in terms of Boutroux spectral curves in [9]. Of course, the fact that the genus expansion can not capture the large N asymptotics was discovered in the context of matrix models and explained in detail in [10].

This paper is organized as follows: In Section 2 we review the structure of instantons corrections in matrix models and topological strings. In Section 3 we present our model, the Chern–Simons theory on the lens space $L(p, 1)$, and its matrix model realization. We then specialize to the $p = 2$ case and review the large N duality between the perturbative expansion of the matrix model in a fixed filling fraction (N_1, N_2) configuration and the topological string on the local $\mathbb{P}^1 \times \mathbb{P}^1$ geometry, in a fixed background t_1, t_2 . In Section 4 we study the large N phase diagram. We start by considering two particular cases: real and imaginary g_s , which correspond to two phases of the gauge theory with two distinct dominant saddle configurations, each dual to a fixed background geometry. By following the analysis of Section 2, we explain the structure of the large N asymptotic of the two phases and provide very precise numerical checks of our predictions. We then complete the study of the phase diagram for generic complex g_s and interpret the large N phase transitions between the various phases as due to the Stokes phenomenon. We end this section by commenting on the issue of the background independence. In Section 5 we discuss the phase diagram of the cubic matrix model. Thanks to its close relation with the Airy function, the prototypical example of the Stokes phenomenon, this example clarifies the nature of the large N phase transitions between dominant saddles/backgrounds which we found in Chern–Simons theory. Section 6 contains conclusions and a list of open issues. Finally in the appendix we collect some details of the model for different submanifolds of moduli space.

2. Instanton corrections in matrix models and topological strings

In this section we review instanton corrections in matrix models/topological strings, following mainly [10, 36, 18].

Multi-cut matrix models and topological strings are characterized, at the perturbative level, by genus g amplitudes of the form

$$F_g(t) = F_g(t_1, \dots, t_s), \quad (2.1)$$

where, in the case of matrix models, t_i are partial 't Hooft couplings

$$t_i = g_s N_i, \quad i = 1, \dots, s, \quad (2.2)$$

²The Hawking–Page transition in AdS/CFT is a large N phase transition leading to a change of topology, but it involves effects of order $\exp(-N^2)$ and it is not driven by instantons.

N_i is the number of eigenvalues in the i -th cut and s is the number of cuts. The perturbative partition function is given, as usual, by

$$Z_p(N_1, \dots, N_s) = \exp \left\{ \sum_{g=0}^{\infty} F_g(t) g_s^{2g-2} \right\}. \quad (2.3)$$

A choice of moduli or 't Hooft parameters $t = (t_1, \dots, t_n)$ in this partition function is called a choice of *background*. In matrix models and in gauge theory, the quantity

$$t = \sum_{i=1}^n t_i = g_s N \quad (2.4)$$

where N is the rank of the $U(N)$ gauge group, is fixed. Background-independent quantities should only depend on g_s and N , while the perturbative partition function Z_p depends on a choice of t_i and is therefore a background dependent quantity.

The *non-perturbative partition function* $Z_{\text{np}}^{\mathcal{C}}$ was introduced in the context of matrix models with a polynomial potential $V(z)$ in [13, 10, 17], and it depends on a choice of contour \mathcal{C} in the complex plane. This contour is the integration path for the eigenvalues of the matrix integral. We then have

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s) = \frac{1}{N!} \int_{\mathcal{C}} \prod_{i=1}^N \frac{d\lambda_i}{2\pi} \Delta^2(\lambda) e^{-\frac{1}{g_s} \sum_{i=1}^N V(\lambda_i)} \quad (2.5)$$

where $\Delta(\lambda)$ is the Vandermonde determinant. Of course, the contour \mathcal{C} is chosen in such a way that this integral converges. By standard saddle-point techniques, we can always deform the path \mathcal{C} into a sum of paths \mathcal{C}_k which go through the critical points of $V(z)$ and are paths of steepest descent [20],

$$\mathcal{C} = \sum_{k=1}^s \zeta_k \mathcal{C}_k, \quad (2.6)$$

where we have assumed that $V(z)$ has s critical points. Therefore

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s) = \sum_{N_1 + \dots + N_s = N} \zeta_1^{N_1} \dots \zeta_s^{N_s} Z_p(N_1, \dots, N_s). \quad (2.7)$$

where

$$Z_p(N_1, \dots, N_s) = \frac{1}{N_1! \dots N_s!} \int_{\lambda_{k_1}^{(1)} \in \mathcal{C}_1} \dots \int_{\lambda_{k_s}^{(s)} \in \mathcal{C}_s} \prod_{i=1}^N \frac{d\lambda_i}{2\pi} \Delta^2(\lambda) e^{-\frac{1}{g_s} \sum_{i=1}^N V(\lambda_i)}. \quad (2.8)$$

To write this equation, we have split the N eigenvalues into s sets of N_I eigenvalues, $I = 1, \dots, s$, which are denoted by

$$\{\lambda_{k_I}^{(I)}\}_{k_I=1, \dots, N_I}, \quad I = 1, \dots, s. \quad (2.9)$$

Each of the integrals (2.8) has a small g_s asymptotic expansion given by (2.3).

Equation (2.7) expresses a background-independent quantity $Z_{\text{np}}^{\mathcal{C}}$ as a sum of background-dependent quantities $Z_p(t)$. This sum can be in turn regarded as a sum over matrix model instantons, which have been identified long ago in terms of eigenvalue tunneling [13, 46]. Formally, we can write the sum (2.7) as the perturbative partition function coming from a *fixed* background

$N_i = N_i^*$, plus an infinite sums of corrections for the remaining values of N_i . These corrections are non-perturbative in g_s . We write (2.7), schematically, as

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s) = Z_{\text{p}}(N_1^*, \dots, N_p^*) + \mathcal{O}(e^{-1/g_s}) \quad (2.10)$$

In order to be more concrete, we will restrict ourselves to models where $s = 2$, or equivalently, models with two moduli t_1, t_2 . Up to an overall normalization, we can set $\zeta_1 = 1, \zeta_2 = \zeta$. The detailed form of (2.10) depends on the choice of background. If the background is on the *boundary*, we have $(N_1^*, N_2^*) = (N, 0)$ or $(0, N)$. Otherwise, we say that the background is an *interior point*. The expression for (2.10) when the background is on the boundary was worked out in [36]. Assuming for concreteness that the boundary is at $(N, 0)$, we have

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s) = Z_{\text{p}}(N) \left(1 + \sum_{\ell \geq 1} \frac{g_s^{\ell^2/2}}{(2\pi)^{\ell/2}} G_2(\ell + 1) \zeta^\ell \hat{q}^{\frac{\ell^2}{2}} \exp\left(-\frac{\ell A}{g_s}\right) \right. \\ \left. \cdot \sum_k \sum_{m_i > 0} \sum_{g_i > 1 - \frac{m_i}{2}} \frac{g_s^{\sum_i (2g_i + m_i - 2)}}{k! m_1! \dots m_k!} \hat{F}_{g_1}^{(m_1)} \dots \hat{F}_{g_k}^{(m_k)}(-\ell)^{\sum_i l_i} \right). \quad (2.11)$$

In this equation we have introduced the following notations. $Z_{\text{p}}(N)$ is the partition function for the one-cut model where all eigenvalues sit in the first critical point. The functions $\hat{F}_g(t_1, t_2)$ are defined by

$$F_g(t_1, t_2) = F_g^{\text{G}}(t_2) + \hat{F}_g(t_1, t_2), \quad (2.12)$$

where $F_g^{\text{G}}(t)$ are the genus g free energies of the gauged Gaussian matrix model with 't Hooft parameter t . $\hat{F}_g^{(m)}$ denotes the m -th derivative of F_g w.r.t. s , which is defined as

$$s = \frac{1}{2}(t_1 - t_2). \quad (2.13)$$

All derivatives are evaluated at $t_1 = t$ and $t_2 = 0$. $G_2(\ell + 1)$ is the Barnes function

$$G_2(\ell + 1) = \prod_{n=0}^{\ell-1} n!. \quad (2.14)$$

Finally,

$$A(t) = \partial_s F_0 \quad \text{and} \quad \hat{q} = \exp\left(\partial_s^2 \hat{F}_0\right). \quad (2.15)$$

At leading order in g_s , we have

$$\sum_k \sum_{m_i > 0} \sum_{g_i > 1 - \frac{m_i}{2}} \frac{g_s^{\sum_i (2g_i + m_i - 2)}}{k! m_1! \dots m_k!} \hat{F}_{g_1}^{(m_1)} \dots \hat{F}_{g_k}^{(m_k)}(-\ell)^{\sum_i l_i} \\ = 1 - g_s \left(\ell \partial_s \hat{F}_1(t) + \frac{\ell^3}{6} \partial_s^3 \hat{F}_0(t) \right) + \mathcal{O}(g_s^2). \quad (2.16)$$

The expansion around an *interior point* was studied in [10, 17, 18]. It is given by

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s) = Z_{\text{p}}(N_1^*, N_2^*) \sum_k \sum_{m_i > 0} \sum_{g_i > 1 - \frac{m_i}{2}} \frac{g_s^{\sum_i (2g_i + m_i - 2)}}{k! m_1! \dots m_k!} F_{g_1}^{(m_1)} \dots F_{g_k}^{(m_k)} \Theta_{\mu, \nu}^{(\sum_i l_i)}(F_0'/g_s, \tau) \\ = Z_{\text{p}}(N_1^*, N_2^*) \left\{ \Theta_{\mu, \nu} + g_s \left(\Theta'_{\mu, \nu} F_1' + \frac{1}{6} \Theta'''_{\mu, \nu} F_0''' \right) + \mathcal{O}(g_s^2) \right\}. \quad (2.17)$$

The derivatives of the free energies F_g are again w.r.t. s defined in (2.13). The theta function $\Theta_{\mu,\nu}$ with characteristics (μ, ν) is defined by

$$\Theta_{\mu,\nu}(u, \tau) = \sum_{n \in \mathbb{Z}} e^{(n+\mu-N\epsilon)u} e^{\pi i(n+\mu-N\epsilon)\tau} e^{2i\pi(n+\mu)\nu} \quad (2.18)$$

and it is evaluated at

$$u = \frac{F'_0(N_i^*)}{g_s}, \quad \tau = \frac{1}{2\pi i} F''_0. \quad (2.19)$$

In the above equation, we have denoted

$$\epsilon = \frac{N_2^*}{N}, \quad \zeta = e^{2i\pi\nu} \quad (2.20)$$

in order to make contact with the notations of [18]. In (2.17), the derivatives of the theta function (2.18) are w.r.t. u , therefore each derivative introduces a factor of $n + \mu - N\epsilon$ in the sum (2.18). Notice that, as emphasized in [10], the expression (2.17) gives a large N asymptotics which goes beyond the genus expansion. In particular, the leading terms in the free energy are of the form,

$$F = g_s^{-2} F_0(N_i^*) + F_1(N_i^*) + \log \Theta_{\mu,\nu} + \dots, \quad (2.21)$$

and the theta function leads typically to an oscillatory large N asymptotics.

The above expressions, (2.11) and (2.17), are for the moment being merely formal. In particular, it is not clear if they provide reasonable asymptotic expansions of the original, non-perturbative partition function. There are again two different cases, applying to (2.11) and (2.17):

1. The expression (2.11) gives an admissible asymptotic expansion if

$$\operatorname{Re} \left(\frac{A(t)}{g_s} \right) > 0. \quad (2.22)$$

In this case, the classical, small g_s asymptotic expansion is given by the genus expansion of $Z_p(N)$. The terms with $\ell > 1$ give exponentially suppressed corrections to the asymptotics.

2. The expansion around a generic point N_i^* in (2.17) is admissible if the filling fractions, as determined from the spectral curve, are real, i.e.

$$\epsilon_i = \frac{N_i^*}{N} \in \mathbb{R}, \quad (2.23)$$

and if

$$\operatorname{Re} \left(\frac{F'_0(N_i^*)}{g_s} \right) = 0. \quad (2.24)$$

The conditions (2.23), (2.24) were first spelled out in detail in [13, 14]. As noticed in [14], they can be written as

$$\operatorname{Im} \left\{ \frac{1}{t} \oint_{\gamma} \frac{dx}{2\pi i} y(x) \right\} = 0, \quad (2.25)$$

where $y(x)$ is the spectral curve of the matrix model, and γ is any cycle on it. In writing these equations, we have followed the conventions of [32], so that

$$y^2(x) = V'(x)^2 + \dots \quad (2.26)$$

A curve with the property (2.25) is called a *Boutroux curve*. The condition (2.24) makes sure that the term involving u in the theta function (2.18) is oscillatory. This in turn guarantees that the sum over n will be peaked around

$$n + \mu - N\epsilon \approx 0 \tag{2.27}$$

so that the derivatives of the theta function involved in (2.17) are of order one, see [19] for a related discussion. The condition that the curve is Boutroux gives a nice geometrization of the search for good asymptotic expansions in matrix model theory, and it has been recently developed in much detail in [9].

The above conditions imply that, in order to have a well-defined large N asymptotics, the background can not be arbitrary. As we change the parameters of the theory, this “dominant background” will change. In particular, it can change from a boundary point to an interior point. We will see examples of this in the Chern–Simons model discussed in this paper.

The search for a dominant background is however subtle, since as it is well-known, the perturbative amplitudes $F_g(t_i)$ have a finite radius of convergence. The full space of ’t Hooft parameters has a nontrivial global structure. As we reach the boundary of the convergence region, we have to perform a *duality transformation* to a different frame. In particular, the condition (2.24) is attached to a particular frame and region, in which F_0 is well-defined. One could worry that the search for dominant backgrounds is not a well-posed problem, globally. However, we will now argue that this is not the case for two reasons. The first one is that the condition on the curve being Boutroux is invariant under duality transformations. Second, as shown in [18], each term in the $1/N$ expansion of (2.17) is invariant under duality transformations, up to a phase. This means, in particular, that

$$\left| e^{g_s^{-2} F_0 + F_1} \Theta_{\mu, \nu} \right| \tag{2.28}$$

is invariant and it is therefore a well-defined quantity in the global moduli space. Looking for the maxima of this function on the subspace of Boutroux curves in moduli space is therefore a well-posed global problem, i.e. independent of the duality frame. In each region of moduli space in which $F_0, F_1, \Theta_{\mu, \nu}$ are not singular, the maxima at large N satisfy the condition (2.24). Therefore, the problem of solving (2.23), (2.24) can be lifted globally in a consistent way.

The structure of instanton corrections that we have reviewed here was originally derived in the context of matrix models. However, it was proposed in [35, 34, 18] that, since it only depends formally on the data of the spectral curve, these corrections should be present in topological strings on local geometries, including toric geometries. In this sense, a suitable example is provided by topological string theory on A_p fibrations on \mathbb{P}^1 . These models have a non-perturbative definition in terms of $U(N)$ Chern–Simons gauge theory on a lens space [3], and they can be explicitly reformulated as a multi-cut matrix integral. They constitute a privileged arena for constructing the large N asymptotics presented above and for studying the role of instanton corrections. In the next section we present what is known about these models from the point of view of Chern–Simons theory, matrix models, and topological string theory.

3. Large N duality for Chern–Simons theory

In this paper we will study Chern–Simons theory on the lens space $L(p, 1)$. This model has an open string description as an A–type topological string defined on a Calabi–Yau threefold given by the cotangent bundle over the lens space $T^*L(p, 1)$. This is just a particular example of the

more general realization, due to Witten [49], of Chern–Simons theory on a three-manifold M as a type A topological string on T^*M .

For $M = \mathbb{S}^3$, which corresponds to the lens space with $p = 1$, the Gopakumar and Vafa large N duality [22] provides a description of $U(N)$ Chern–Simons gauge theory in terms of closed topological string theory on the resolved conifold $\mathcal{O}(-1) + \mathcal{O}(-1) \rightarrow \mathbb{P}^1$. This duality can be generalized [3] by performing a \mathbb{Z}_p quotient on both sides of the Gopakumar–Vafa duality. This leads to a duality between Chern–Simons theory on $S_3/\mathbb{Z}_p \equiv L(p, 1)$ and the closed topological A–model string on an A_p fibration over \mathbb{P}^1 . This target space, unlike the resolved conifold, has a non trivial Kähler moduli space and besides the large radius phase it has non–geometric phases. In particular, the large N duality relates the perturbative regime of the Chern–Simons gauge theory—small ’t Hooft couplings—to the orbifold regime of the string theory. This duality has been tested in [3] by comparing the perturbative expansion of Chern–Simons theory around a fixed flat connection, to the orbifold topological strings amplitudes computed by mirror symmetry. Further tests of the duality have been done in [27, 26]. In the following we will review some of these results. We will however go beyond the perturbative side of the duality and we will study the exact, non-perturbative gauge theory partition function. As we will see, in the full theory, the choice of filling fractions is not arbitrary, but it is dictated by the only parameters of the theory, the Chern–Simons level k and the rank of the gauge group N .

3.1 Chern–Simons theory on lens spaces

The lens space $L(p, 1)$ is a three–manifold that can be obtained by gluing two solid 2-tori along their boundaries after performing the $\text{SL}(2, \mathbb{Z})$ transformation,

$$U_p = \begin{pmatrix} 1 & 0 \\ p & 1 \end{pmatrix}. \quad (3.1)$$

This description makes it possible to calculate the partition function of Chern–Simons theory on these spaces [48]. $\text{SL}(2, \mathbb{Z})$ transformations lift to operators acting on $\mathcal{H}(\mathbb{T}^2)$, the Hilbert space obtained by canonical quantization of the Chern–Simons theory on the two-torus. This space is the space of integrable representations of a Wess–Zumino–Witten (WZW) model with gauge group G at level k , where G and k are respectively the Chern–Simons gauge group and the quantized coupling constant. In particular the partition function of the $U(N)$ Chern–Simons theory on the lens space $L(p, 1)$ is given by:

$$Z(L(p, 1)) = \langle 0 | \mathcal{U}_p | 0 \rangle, \quad (3.2)$$

where \mathcal{U}_p is the lift of (3.1) to an operator on $\mathcal{H}(\mathbb{T}^2)$ and $|0\rangle$ is the vacuum state, corresponding to the trivial representation. This partition function has the following explicit form (up to an overall constant) [44]:

$$Z = \frac{(-1)^{|\Delta_+|} e^{-\frac{2\pi i}{k p} \rho^2}}{(-i p \hat{k})^{N/2}} \sum_{n \in \mathbb{Z}^N / p \mathbb{Z}^N} \sum_{w \in \mathcal{W}} \epsilon(w) \exp \left\{ \frac{i\pi}{\hat{k} p} (\rho^2 - 2\rho \cdot (\hat{k}n + w(\rho)) + (\hat{k}n + w(\rho))^2) \right\}. \quad (3.3)$$

In this expression, $\hat{k} = k + y$, where y denotes the dual Coxeter number, $|\Delta_+|$ is the number of positive roots of G , and ρ is the Weyl vector. Using Weyl’s formula

$$\sum_{w \in \mathcal{W}} \epsilon(w) e^{w(\rho)} = \prod_{\alpha > 0} 2 \sinh \frac{\alpha}{2} \quad (3.4)$$

we can also write the partition function as:

$$Z = \frac{i^{|\Delta+|}}{(-ip\hat{k})^{N/2}} \sum_{n \in \mathbb{Z}^N / p\mathbb{Z}^N} e^{\frac{\pi i \hat{k} n^2}{p} - \frac{2\pi i}{p} \rho \cdot n} \prod_{1 \leq i < j \leq N} 2 \sin\left(\frac{\pi}{\hat{k}p} (\hat{k}(n_i - n_j) + j - i)\right). \quad (3.5)$$

This expression for the partition function can be understood as a sum over saddle-points. For $U(N)$ Chern–Simons theory on $L(p, 1) = \mathbb{S}_3/\mathbb{Z}_p$, these are flat connections, which can be obtained by considering embeddings of the first fundamental group into $U(N)$, modulo gauge transformations. Since $\pi_1(L(p, 1)) = \mathbb{Z}_p$ these embeddings are given by N -component vectors n , whose entries take values in \mathbb{Z}_p . However, since the residual Weyl symmetry S_N of the $U(N)$ gauge group permutes the different components of n , the flat connections are rather labeled by partitions N_j , where N_j is the number of entries in n which is equal to $j - 1$. The N_j label the choice of vacuum

$$U(N) \rightarrow U(N_1) \times \cdots \times U(N_p). \quad (3.6)$$

It is then possible [3] to rewrite the partition function as a sum of contributions of flat connections:

$$Z = \sum_{\{N_j\}} Z(N_j) \quad (3.7)$$

where

$$Z(N_j) = \frac{(-1)^{|\Delta+|}}{(-ip\hat{k})^{N/2}} \frac{1}{\prod_j N_j!} \sum_{w, w' \in \mathcal{W}} \epsilon(w) \exp\left\{\frac{i\pi}{\hat{k}p} (\rho^2 - 2\rho \cdot (\hat{k}w'(n) + w(\rho)) + (\hat{k}w'(n) + w(\rho))^2)\right\}. \quad (3.8)$$

In this expression, the vector n is any vector with N_j entries equal to $j - 1$. The sum over Weyl permutations $w' \in \mathcal{W}$ guarantees that the resulting object is gauge-invariant. As shown in [31] it is possible to rewrite (3.8) as a matrix integral

$$Z(N_k) = \frac{e^{-\frac{2i\pi}{\hat{k}p} \rho^2 - S_{\text{inst}}(N_j)/g_s}}{\prod_j N_j!} \int \prod_{i=1}^N \frac{dx_i}{2\pi} e^{-\frac{1}{2g_s} \sum_k (x_k - 2\pi i n_k/p)^2} \prod_{i < j} \left(2 \sinh \frac{x_i - x_j}{2}\right)^2, \quad (3.9)$$

where

$$g_s = \frac{2\pi i}{p\hat{k}} \quad (3.10)$$

and the pre-factor involves the gauge-theory action of each flat connection:

$$\frac{S_{\text{inst}}(N_j)}{g_s} = \frac{2\pi^2}{p^2 g_s} \sum_{j=1}^p (j-1)^2 N_j. \quad (3.11)$$

It can be easily checked that, although the expression (3.9) involves a choice of vector n , any two choices related by a Weyl permutation lead to the same matrix integral. The expression (3.9) can be regarded as a p -cut matrix model, where N_k eigenvalues sit around the point $2\pi i(k-1)/p$, i.e. it is a matrix model with fixed filling fractions. It has in particular an asymptotic large N expansion of the form

$$F_{N_k} = \log Z_{N_k} = \sum_{g=0}^{\infty} F_g(t_i) g_s^{2g-2} \quad (3.12)$$

where

$$t_i = g_s N_i \quad (3.13)$$

are the partial 't Hooft parameters. In the case $p = 2$, calculations around the Gaussian point of the matrix model give the genus zero free energy

$$F_0(t_1, t_2) = -\frac{\pi^2}{2} t_2 + \log(-4) t_1 t_2 + F_0^G(t_1) + F_0^G(t_2) + F_0^P(t_1, t_2). \quad (3.14)$$

In this equation, the first term comes from the instanton action, the second term comes from the overall measure of the matrix integral, and $F_0^G(t)$ is the Gaussian matrix model genus zero amplitude,

$$F_0^G(t) = \frac{1}{2} t^2 \left(\log t - \frac{3}{4} \right). \quad (3.15)$$

Finally, $F_0^P(t_1, t_2)$ comes from fatgraphs of genus zero, and it is given by

$$\begin{aligned} F_0^P(t_1, t_2) &= \frac{1}{288} (t_1^4 + 6t_1^3 t_2 + 18t_1^2 t_2^2 + 6t_1 t_2^3 + t_2^4) \\ &\quad - \frac{1}{345600} (4t_1^6 + 45t_1^5 t_2 + 225t_1^4 t_2^2 + 1500t_1^3 t_2^3 + 225t_1^2 t_2^4 + 45t_1 t_2^5 + 4t_2^6) + \dots \end{aligned} \quad (3.16)$$

For the genus one term one has:

$$F_1(t_1, t_2) = F_1^G(N_1) + F_1^G(N_2) + F_1^P(t_1, t_2) \quad (3.17)$$

where

$$F_1^G(N) = \zeta'(-1) + \frac{1}{12} \log(N) \quad (3.18)$$

is the Gaussian matrix model contribution, and

$$F_1^P(t_1, t_2) = -\frac{1}{288} (t_1^2 - 6t_1 t_2 + t_2^2) + \dots \quad (3.19)$$

comes from fatgraphs of genus one. Higher genus free energies can be computed analogously.

3.2 The dual topological string

We will now restrict ourselves to $p = 2$. The large N dual model of Chern–Simons theory on $L(2, 1)$ is the topological string on the anti-canonical bundle of the Hirzebruch surface \mathbb{F}_0 . This geometry has two \mathbb{P}^1 's in $H_2(Y)$ and a compact four cycle in $H_4(Y)$. We denote their associated classes by A_1, A_2 and B respectively. The class $C = A_2 - A_1$ does not have a dual cycle in $H_4(Y)$ and thus it will correspond to a non-normalizable modulus of the theory; one can then take $A = A_2$ as the normalizable modulus.

The mirror geometry is encoded in a family of elliptic curves Σ , which can be written as

$$y = \frac{z_1 x^2 + x + 1 - \sqrt{(1 + x + z_1 x^2)^2 - 4z_2 x^2}}{2}. \quad (3.20)$$

The periods of the one form $\lambda = \log y dx/x$ are given by the solutions of the associated Picard–Fuchs system:

$$\begin{aligned} \mathcal{L}_1 &= z_2(1 - 4z_2)\xi_2^2 - 4z_1^2\xi_1^2 - 8z_1z_2\xi_1\xi_2 - 6z_1\xi_1 + (1 - 6z_2)\xi_2, \\ \mathcal{L}_2 &= z_1(1 - 4z_1)\xi_1^2 - 4z_2^2\xi_2^2 - 8z_1z_2\xi_1\xi_2 - 6z_2\xi_2 + (1 - 6z_1)\xi_1, \end{aligned} \quad (3.21)$$

where

$$\xi_i = \frac{\partial}{\partial z_i}. \quad (3.22)$$

At the large radius point $z_1 = z_2 = 0$, the Picard–Fuchs system has two single logarithmic solutions, which give the mirror maps at large radius

$$\begin{aligned} T_1 &= -\log z_1 - 2z_1 - 2z_2 - 3z_1^2 - 12z_1z_2 - 3z_2^2 + \cdots, \\ T_2 &= -\log z_2 - 2z_1 - 2z_2 - 3z_1^2 - 12z_1z_2 - 3z_2^2 + \cdots. \end{aligned} \quad (3.23)$$

As expected, the quantity

$$T_1 - T_2 = \log \frac{z_2}{z_1}, \quad (3.24)$$

corresponding to the class C , does not receive any instanton corrections. It only depends on the “bare parameters” of the model z_1, z_2 and has to be regarded as a *parameter*. As we will see in a moment, it has a very natural meaning in the dual Chern–Simons theory.

In the moduli space of this model there is another point, discovered in [3], which makes contact with Chern–Simons perturbation theory. This point is called the *orbifold point*, and it is defined as the point $x_1 = x_2 = 0$ in terms of the variables:

$$x_1 = 1 - \frac{z_1}{z_2}, \quad x_2 = \frac{1}{\sqrt{z_2} \left(1 - \frac{z_1}{z_2}\right)}. \quad (3.25)$$

The periods near the orbifold point can be obtained again as solutions to the Picard–Fuchs system. They have the structure,

$$\begin{aligned} \sigma_1 &= -\log(1 - x_1) = \sum_m c_{m,0} x_1^m, \\ \sigma_2 &= \sum_{m,n} c_{m,n} x_1^m x_2^n, \\ \mathcal{F}_{\sigma_2} &= \sigma_2 \log(x_1) + \sum_{m,n} d_{m,n} x_1^m x_2^n, \end{aligned} \quad (3.26)$$

where the coefficients $c_{m,n}$ and $d_{m,n}$ are determined by the following recursions relations [3]

$$\begin{aligned} c_{m,n} &= c_{m-1,n} \frac{(n+2-2m)^2}{4(m-n)(m-1)}, \\ c_{m,n} &= \frac{1}{n(n-1)} (c_{m,n-2}(n-m-1)(n-m-2) - c_{m-1,n-2}(n-m-1)^2), \\ d_{m,n} &= \frac{d_{m-1,n}(n+2-2m)^2 + 4(n+1-2m)c_{m,n} + 4(2m-n-2)c_{m-1,n}}{4(m-n)(m-1)}, \\ d_{m,n} &= \frac{1}{n(n-1)} (d_{m,n-2}(n-m-1)(n-m-2) - d_{m-1,n-2}(n-m-1)^2 \\ &\quad + (2n-2-2m)c_{m-1,n-2} + (2m+3-2n)c_{m,n-2}). \end{aligned} \quad (3.27)$$

The orbifold point gives the large N solution of Chern–Simons theory on $L(2, 1)$, around a fixed, arbitrary flat connection, in the sense that the perturbative topological string amplitudes in this

frame, $F_g(\sigma_1, \sigma_2)$, correspond to the amplitudes in the $1/N$ expansion (3.12). The dictionary relating the variables is

$$t_1 = \frac{1}{4}(\sigma_1 + \sigma_2), \quad t_2 = \frac{1}{4}(\sigma_1 - \sigma_2). \quad (3.28)$$

Notice that the parameter of the theory

$$\log \frac{z_2}{z_1} = \sigma_1 = 2(t_1 + t_2), \quad (3.29)$$

is proportional to the *total* 't Hooft parameter of Chern–Simons theory $g_s N$. This is in fact required by the large N duality: although $t_{1,2}$, the partial 't Hooft parameters, change in a natural way under the action of the symplectic group, the total 't Hooft parameter t must be a symplectic invariant, since it is a parameter of the gauge theory. The derivative of the genus zero free energy of Chern–Simons theory w.r.t. the variable s defined in (2.13) is given by the following combination of the periods:

$$\frac{\partial F_0}{\partial s} = \frac{\pi^2}{2} + \frac{1}{2} \mathcal{F}_{\sigma_2} - \log(-4)\sigma_2. \quad (3.30)$$

By integrating w.r.t. σ_2 and expressing the result in terms of t_1, t_2 one recovers the genus zero free energy of the matrix model expansion.

In the following it will be useful to express some quantities in terms of modular forms. It turns out that [2, 25]

$$\tau = \frac{1}{2\pi i} \frac{\partial^2 F_0}{\partial s^2} \quad (3.31)$$

is indeed a modular parameter for the curve (3.20), and in particular $\text{Im } \tau > 0$. The modulus

$$u = -\frac{1}{8\sqrt{1-x_1}} x_1^2 x_2^2 + \frac{1}{2} \left(\sqrt{1-x_1} + \frac{1}{\sqrt{1-x_1}} \right), \quad (3.32)$$

is closely related to the modulus of a Seiberg–Witten curve [45] with

$$\tau_{\text{SW}} = 2\tau + 1, \quad (3.33)$$

and it can be expressed in terms of elliptic theta functions:

$$u = \frac{\vartheta_4^4 - \vartheta_2^4}{\vartheta_3^4}(2\tau) = 1 - 32q + 256q^2 + \dots \quad (3.34)$$

where $q = e^{2\pi i \tau}$. Notice that the orbifold point in the moduli space of local \mathbb{F}_0 corresponds to $u = 1$, i.e. to the monopole point of the Seiberg–Witten curve.

Finally one can check that the genus one free energy is given, in terms of modular forms, by

$$F_1 = -\log \eta(\tau_{\text{SW}}) + 2\zeta'(-1) + \frac{1}{6} \log \frac{g_s}{4}, \quad (3.35)$$

In particular, we find

$$F_1^{\text{P}}(t_1, t_2) = -\log \eta(\tau_{\text{SW}}) + \frac{1}{12} \log \left(\frac{t_1 t_2}{16} \right). \quad (3.36)$$

The partition function of Chern–Simons theory on $L(2, 1)$ can be written in terms of a two-cut matrix model [31, 3]. The cuts are centered around $x = 0$, $x = \pi i$, and they are precisely the cuts which appear in the mirror curve (3.20). If we write them as

$$(-a, a), \quad (\pi i - b, \pi i + b), \quad (3.37)$$

in terms of the endpoints a, b , it is easy to see that they are determined by the equations (see also [27])

$$\begin{aligned}\cosh a &= \frac{1}{2\sqrt{z_1}} + \sqrt{\frac{z_2}{z_1}} = 1 + 2t_1 + t_1(t_1 + t_2) + \dots, \\ \cosh b &= -\frac{1}{2\sqrt{z_1}} + \sqrt{\frac{z_2}{z_1}} = 1 + 2t_2 + t_2(t_1 + t_2) + \dots.\end{aligned}\tag{3.38}$$

3.3 Two special slices

The above results simplify considerably along particular submanifolds in moduli space, which will be relevant in the analysis of the phase structure in the next section.

The *slice* $t_2 = 0$ corresponds, from the point of view of the gauge theory, to an expansion around the trivial flat connection. Up to a trivial rescaling of the coupling constant g_s , the theory reduces to Chern–Simons theory on the three-sphere \mathbb{S}^3 . From the point of view of topological string theory, the theory is equivalent to the resolved conifold. The genus zero and genus one free energies are given by,

$$\begin{aligned}F_0^{\mathbb{S}^3}(t) &= \frac{t^3}{12} - \frac{\pi^2 t}{6} - \text{Li}_3(e^{-t}) + \zeta(3), \\ F_1^{\mathbb{S}^3}(t) &= -\frac{t}{24} - \frac{1}{12} \log(1 - e^{-t}) + \frac{1}{12} \log t + \zeta'(-1) - \frac{1}{12} \log N.\end{aligned}\tag{3.39}$$

The instanton action (2.15) on $t_1 = t, t_2 = 0$ can be evaluated in closed form (see the Appendix) and it is given by

$$A(t) = 2 \text{Li}_2(e^{-t/2}) - 2 \text{Li}_2(-e^{-t/2}).\tag{3.40}$$

The other relevant slice, for us, corresponds to the *symmetric slice* $t_1 = t_2$. In terms of the orbifold coordinates, this slice is parametrized by

$$x_2 = \sigma_2 = 0, \quad x_1 = 1 - e^{-\sigma_1}.\tag{3.41}$$

On this slice we have the following exact expressions for the prepotential and for the modular parameter in terms of the flat coordinate σ_1 :

$$F_0(\sigma_1) = -\frac{1}{8} \text{Li}_3(e^{-\sigma_1}) + \frac{\sigma_1^3}{96} + \frac{\pi i}{16} \sigma_1^2 - \frac{7\pi^2}{48} \sigma_1 + \frac{\zeta(3)}{8},\tag{3.42}$$

$$\tau = i \frac{K'(k)}{K(k)} - 1, \quad k^2 = 1 - e^{-\sigma_1},\tag{3.43}$$

where $K(k)$ is the standard elliptic integral. The proof of the above identities is given in the Appendix.

4. Large N phase diagram

So far we have just recalled the *perturbative* large N duality between Chern–Simons theory on $L(2, 1)$ expanded around a particular flat connection (corresponding to a choice of filling fractions N_1, N_2), and topological string theory on local \mathbb{F}_0 with Kähler parameters t_1, t_2 .

However, as we explained in the introduction, when studying the large N expansion of the non-perturbative partition function, the choice of filling fractions (dominant saddle) is not

arbitrary but it is dictated by the value of the total 't Hooft coupling t . We will now study the large N phase diagram for the Chern–Simons theory on $L(2, 1)$, i.e. we will determine the saddle which gives the dominant contribution to the non-perturbative partition function as we move on the complex plane of $t = g_s N$ (equivalently, on the complex plane of g_s).

We first notice two exact symmetries of the non-perturbative partition function. If we conjugate g_s , we find

$$Z(N, g_s^*) = (-1)^{c(N)} Z^*(N, g_s), \quad (4.1)$$

where $c(N)$ is a half-integer that only depends on N . This symmetry is valid for all p . The second symmetry is related to changing $g_s \rightarrow -g_s$. If we write

$$Z(N, g_s) = \sum_{N_2=0}^N Z(N, N_2, g_s) \quad (4.2)$$

we have that

$$Z(N, N_2, -g_s) = (-1)^{c'(N)} e^{-\frac{\pi^2 N^2}{2t}} (-1)^{(N-1)N_2} Z(N, N - N_2, g_s), \quad (4.3)$$

where $c'(N)$ is again a half-integer that only depends on N .

In determining the phase structure, we are interested in knowing which is the N_2 which contributes the most to the non-perturbative partition function (in absolute value). As we will see, for N large, the sum in (4.2) will be very peaked around a particular saddle value (N_1^*, N_2^*) . We see, from (4.1), that conjugating g_s does not change this saddle, while from (4.3) we see that the phase diagram is invariant under the simultaneous change

$$g_s \leftrightarrow -g_s, \quad (N_1^*, N_2^*) \leftrightarrow (N_2^*, N_1^*). \quad (4.4)$$

In practice, this means that in order to determine the phase diagram, we can restrict ourselves to the first quadrant of the complex g_s plane

$$0 \leq \text{Arg}(g_s) \leq \frac{\pi}{2}. \quad (4.5)$$

In order to proceed, we first study the limiting cases of real, positive g_s and purely imaginary g_s .

4.1 Real g_s : the boundary expansion

The classical action of the flat connection (3.11) has the value

$$S_{\text{inst}}(N_2) = \frac{\pi^2}{2} N_2. \quad (4.6)$$

When g_s is real and positive, it gives an exponential suppression for all the non-trivial flat connections $N_2 > 0$. Therefore, at least when the 't Hooft coupling is sufficiently small, the dominant saddle point is expected to be at the boundary $(N_1^*, N_2^*) = (N, 0)$. In fact, it is easy to see that for any real, positive t , this boundary point is the dominant saddle. To prove it, it suffices to show that the condition (2.22) holds on this slice, where $A(t)$ is the large N instanton action corresponding to the first non-trivial instanton sector with $N_2 = 1$. This action is given in (3.40), and it resums quantum fluctuations around the classical instanton solution. At small t it should have the general structure [23]

$$\exp\left(-\frac{A(t)}{g_s}\right) \sim \left(\frac{c_2}{g_s}\right)^{c_3 N} e^{-c_1/g_s} \quad (4.7)$$

where the prefactor is the one-loop fluctuation around the classical instanton with action $c_1 = \pi^2/2$, $c_3 N$ is the number of zero modes at large N , and c_2 is a numerical constant. Therefore, the expansion of $A(t)$ around $t = 0$ should be of the form

$$A(t) = c_1 - c_3 t \log\left(\frac{c_2}{t}\right) + \mathcal{O}(t). \quad (4.8)$$

Indeed, we find from the exact expression (3.40) that

$$A(t) = \frac{\pi^2}{2} - t \log\left(\frac{4}{t}\right) - t + \mathcal{O}(t^3). \quad (4.9)$$

This large N instanton action is real for $t \in [0, \infty)$ and it decreases monotonically from

$$A(0) = \frac{\pi^2}{2} \quad \text{to} \quad \lim_{t \rightarrow \infty} A(t) = 0. \quad (4.10)$$

In particular it does not vanish for any finite t . This can be seen in Fig. 2, where the continuous line represents $A(t)/t$ as a function of t/π . As we explain in the Appendix, this is equivalent to the absence of phase transition for q -deformed 2d Yang–Mills with $p = 2$ [12, 5, 28]. We conclude that, for real g_s , the large N asymptotics of the partition function of Chern–Simons theory on $L(2, 1)$ is simply given by the $1/N$ expansion of Chern–Simons theory on \mathbb{S}^3 ,

$$\log Z_{L(2,1)}(N, g_s) = \sum_{g=0}^{\infty} g_s^{2g-2} F_g^{\mathbb{S}^3}(t). \quad (4.11)$$

In other words, the relevant saddle geometry along this direction in moduli space is just the resolved conifold geometry.

Of course, the result (4.11) has an infinite number of exponentially suppressed instanton corrections of the form (2.11). From the point of view of *classical* asymptotic analysis, these subleading saddles are not taken into account, but of course they are important thanks, among other things, to the Stokes phenomenon that we will now uncover.

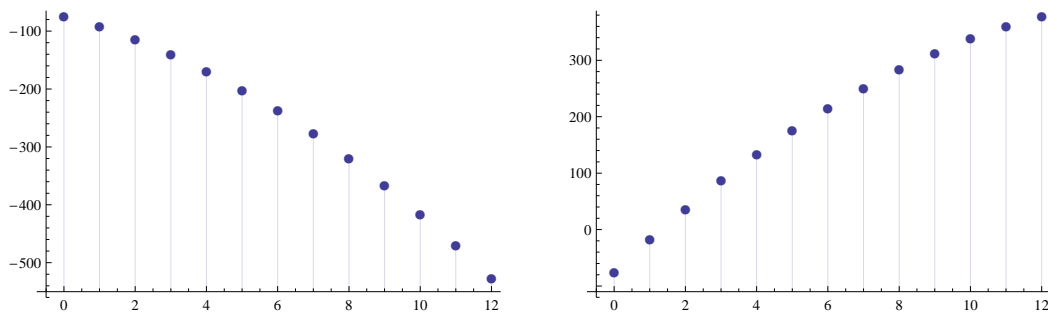


Figure 1: In this figure we plot $\log |Z(N, N_2, g_s)|$ as a function of $N_2 = 0, \dots, N$, for $N = 12$ and for $g_s = \pm\pi/24$ respectively on the left and on the right. The dominant configurations are $(N_1^*, N_2^*) = (N, 0)$ and $(0, N)$, respectively, in agreement with the symmetry (4.3).

The above analysis can be confirmed numerically for small values of N . In general, a simple way to estimate the leading saddle (N_1^*, N_2^*) for finite N is to calculate the partition functions

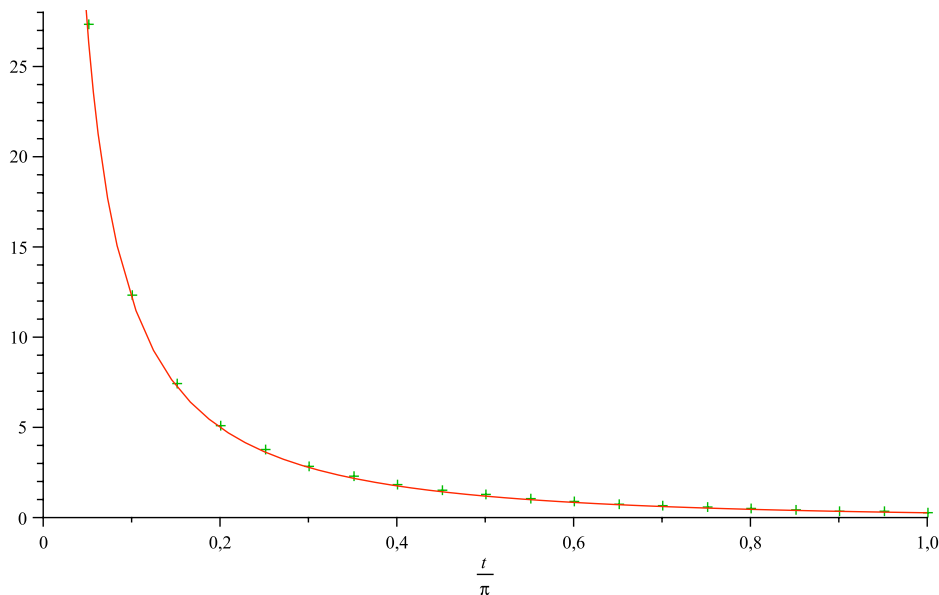


Figure 2: In this figure, the crosses represent the right hand side of (4.12), for $N = 24$ and different values of t , while the continuous line represents the value of $A(t)/t$ obtained from (3.40).

$Z(N, N_2, g_s)$ for fixed values of N_2 , and see which one is the largest in absolute value, and therefore gives the dominant contribution in (4.2). In Fig. 1 we plot $\log |Z(N, N_2, g_s)|$, for g_s real and positive. As expected, the dominant configuration is $(N_1^*, N_2^*) = (N, 0)$ ($(N_1^*, N_2^*) = (0, N)$ for g_s real and negative).

We can also confirm numerically the value of the instanton action (3.40). From (2.11) we find the large N asymptotics

$$-\frac{1}{N} \log \left(\frac{Z(N, N_2 = 1, g_s)}{Z(N, N_2 = 0, g_s)} \right) = \frac{A(t)}{t} + \mathcal{O} \left(\frac{1}{N} \right). \quad (4.12)$$

In Fig. 2 we plot the values of the r.h.s. for $N = 24$ and different values of $t = Ng_s$, and we compare them to the value of $A(t)/t$ given by (3.40). As we can see, the agreement is excellent, and it can be further improved by extracting the subdominant tails in (4.12) through the technique of Richardson transforms.

4.2 Imaginary g_s : the interior expansion

Let us analyze now the large N asymptotics along the direction $\text{Arg}(g_s) = \pi/2$. The boundary point $(N, 0)$ can not be a solution any longer, since (2.22) is no longer verified (at least for small t). This is easy to understand: for imaginary g_s , (3.11) is purely imaginary and the non-trivial instanton sectors are not suppressed anymore. We then have to look for saddles satisfying the condition (2.24), which in this case reads:

$$\text{Re} \left[\frac{1}{g_s} \left(\frac{\partial F_0}{\partial t_1} - \frac{\partial F_0}{\partial t_2} \right) \right] = 0. \quad (4.13)$$

It is easy to see that the real part of $F_0(t_1, t_2)$ is symmetric under the exchange of t_1 and t_2 , therefore the configuration

$$t_1 = t_2 = \frac{t}{2}, \quad (4.14)$$

which is an interior point, is a saddle. We claim that this gives in fact the dominant saddle when g_s is imaginary. Of course, there could be other saddles which actually dominate the asymptotics. A first indication that this is not the case, and that (4.14) is the relevant saddle, comes from the numerical analysis of $Z(N, N_2, g_s)$ for small N . As we see in Fig. 3, for N even the largest contribution comes from $N_2 = N/2$, while for N odd the dominant contributions have $N_2 = (N \pm 1)/2$.

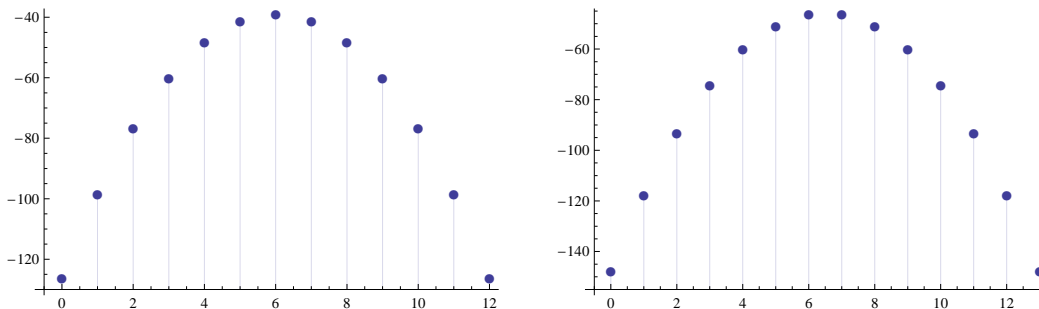


Figure 3: In this figure we plot $\log |Z(N, N_2, g_s)|$ as a function of $N_2 = 0, \dots, N$, on the left for $N = 12$ and $g_s = \pi i/48$, on the right for $N = 13$ and $g_s = \pi i/52$. The largest value is obtained when $N_2 = N/2$ for N even and when $N_2 = (N \pm 1)/2$ for N odd.

According to the general discussion in section 2, since the saddle (4.14) is an interior point, the large N asymptotics of the partition function should be given by (2.17). In our case, we have

$$\epsilon = \frac{1}{2}, \quad \mu = \nu = 0, \quad \frac{u}{g_s} = \frac{\pi^2}{2g_s} = -\frac{\pi i \hat{k}}{2}. \quad (4.15)$$

The resulting theta function depends on the value of N modulo two. If N is *even*, $N\epsilon$ is an integer which can be reabsorbed in a shift of $n \in \mathbb{Z}$, which is summed over. Therefore, (2.18) is given by

$$\vartheta_e = \sum_{n \in \mathbb{Z}} \exp \left(\pi i \tau n^2 + i \frac{\hat{k} \pi}{2} n \right). \quad (4.16)$$

Notice that, since g_s is imaginary, k is real, and the second term leads to an oscillatory behavior. If N is *odd*, $N\epsilon$ is a half-integer. We can absorb its integer part by a shift of n and we obtain for (2.18)

$$\vartheta_o = \sum_{n \in \mathbb{Z} + \frac{1}{2}} \exp \left(\pi i \tau n^2 + i \frac{\hat{k} \pi}{2} n \right). \quad (4.17)$$

We then claim that, when g_s is imaginary, the asymptotic behavior of $F(N, g_s) = \log Z(N, g_s)$, the free energy of Chern–Simons theory on $L(2, 1)$, at large N and fixed 't Hooft coupling, is given by the logarithm of (2.17). At leading order we have

$$F(N, g_s) = g_s^2 F_0 \left(\frac{t}{2}, \frac{t}{2} \right) + F_1 \left(\frac{t}{2}, \frac{t}{2} \right) + \log \Theta_{0,0}(u, \tau) + \dots \quad (4.18)$$

where $t = Ng_s$. Notice that

1. The perturbative free energies $F_g(t/2, t/2)$ are the genus g free energies of topological string theory on \mathbb{F}_0 on the slice $t_1 = t_2 = t/2$. For genus zero, it is given by the explicit expression (3.42). The higher genus amplitudes are essentially quasi-modular forms of the modular parameter τ [2, 25] which is given in (3.43) by an explicit function of

$$\sigma_1 = \frac{2\pi i N}{k + N}, \quad (4.19)$$

where we used the original Chern–Simons parameters. In particular, F_1 is given in (3.35).

2. The asymptotics depends on the parity of N through the theta function, which is given by (4.16), (4.17) for N even and odd, respectively.

In the quantum Chern–Simons gauge theory, $\hat{k} = k + N$ is an integer, and g_s is purely imaginary. Therefore, the above asymptotics is in principle the relevant one to understand the large N 't Hooft limit of the quantum Chern–Simons invariant of $L(2, 1)$.

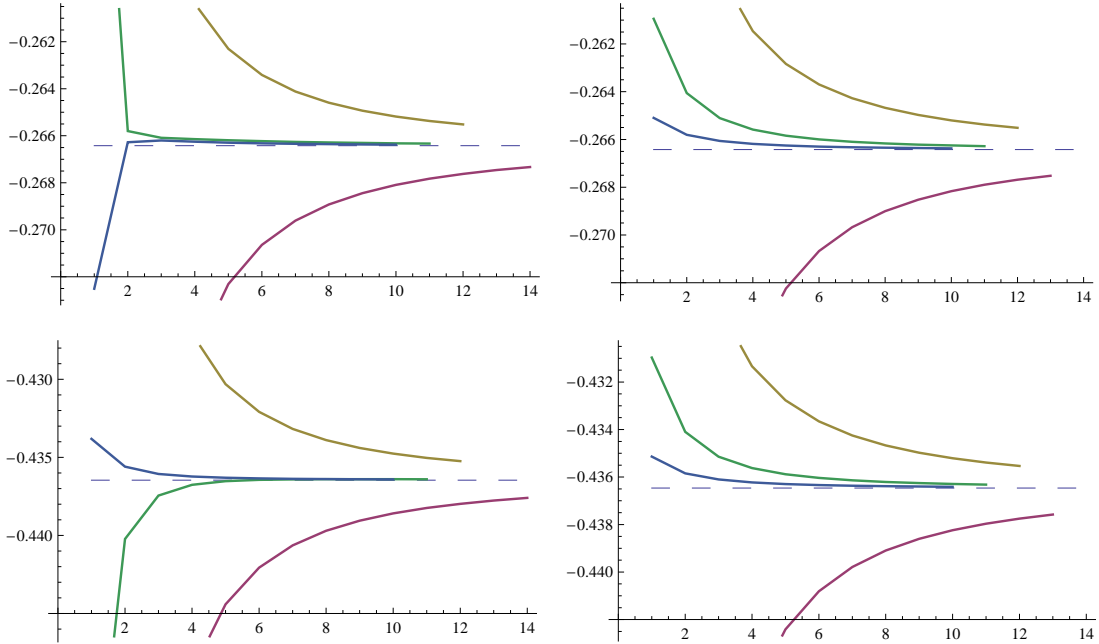


Figure 4: On the left column we plot the sequence $\text{Re } f_e(\ell, z)$ as well as its Richardson transforms $\text{Re } f_e^{(k)}(\ell, z)$ for $k = 1, 2, 3$, and for $\ell = 2 \cdots 14$. On the right column we plot the sequence $\text{Re } f_o(\ell, z)$ as well as its Richardson transforms $\text{Re } f_o^{(k)}(\ell, z)$ for $k = 1, 2, 3$ for $\ell = 2 \cdots 14$. The plots on the top are for $z = 1/4$, and the plots of the bottom are for $z = 1/8$. The dashed lines in the plots show the expected value $-\text{Re } F_0(z)/(\pi z)^2$.

Our claim about the asymptotics follows from the identification of (4.14) as the relevant saddle point, and from the general theory reviewed in section 2 concerning instanton corrections. We now provide detailed numerical evidence for our claim. Instead of working with N and g_s we will use N and z , which is defined by

$$z = \frac{N}{\hat{k}}. \quad (4.20)$$

Let us now define the sequences of complex numbers obtained from the non-perturbative partition function with $N = 2\ell, 2\ell + 1$ respectively:

$$f_e(\ell, z) = \frac{1}{(2\ell)^2} \log Z(2\ell, z), \quad f_o(\ell, z) = \frac{1}{(2\ell + 1)^2} \log Z(2\ell + 1, z). \quad (4.21)$$

According to our claim, these sequences have the large ℓ asymptotics

$$\begin{aligned} f_e(\ell, z) &= -\frac{F_0(z)}{(\pi z)^2} + \frac{1}{(2\ell)^2} (F_1(z) + \log \vartheta_e(z)) + \mathcal{O}(\ell^{-4}), \\ f_o(\ell, z) &= -\frac{F_0(z)}{(\pi z)^2} + \frac{1}{(2\ell + 1)^2} (F_1(z) + \log \vartheta_o(z)) + \mathcal{O}(\ell^{-4}) \end{aligned} \quad (4.22)$$

where $F_{0,1}(z)$ are the genus 0, 1 free energies evaluated on the slice (4.14). Notice that the theta functions have in general oscillatory behavior as a function of ℓ , but this can be fixed by choosing special values of \hat{k} . For example, if $\hat{k} = 0 \pmod{4}$ we have:

$$\vartheta_e(z) = \vartheta_3(\tau), \quad \vartheta_o(z) = \pm \vartheta_2(\tau), \quad (4.23)$$

where the sign depends on the value of $\hat{k} \pmod{8}$. In this way, the resulting large ℓ asymptotics is simply a series in inverse powers of ℓ , and we can use standard techniques of accelerated convergence to compare the actual values of $f_{e,o}(\ell, z)$, computed from the exact partition function of Chern–Simons theory for low ℓ , to the predicted asymptotics.

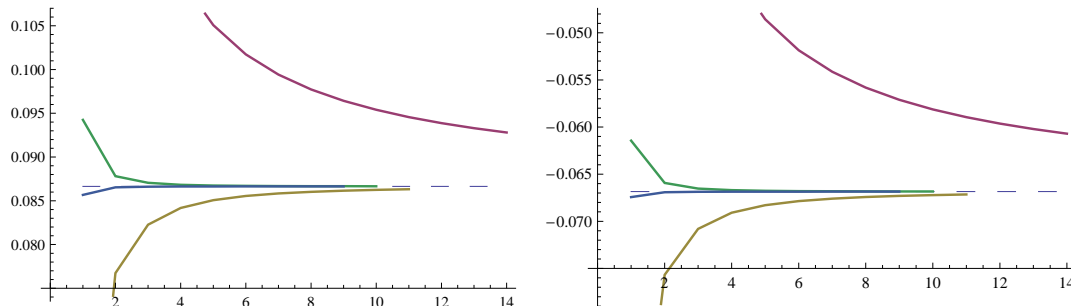


Figure 5: The sequence $\text{Re } \Theta_{eo}(z)$ for $z = 1/4, 1/8$, together with its Richardson transforms $\text{Re } \Theta_{eo}^{(k)}(z)$ for $k = 1, 2, 3$. The dashed lines in both plots show the expected value $\text{Re}(\log \vartheta_e(z) - \log \vartheta_o(z))$.

In practice, the computation of (3.5) involves generating a large number of configurations for the vectors n , and we obtained their numerical values for $N = 1, \dots, 30$ and for different values of z . This means that we obtained the values of the sequences $f_{e,o}(\ell, z)$ up to $\ell = 14$. Once these sequences are computed, we obtain their Richardson transforms (see for example [6])

$$f_{e,o}^{(r)}(\ell, z) = \sum_{k \geq 0} \frac{f_{e,o}(\ell + k, z) (\ell + k)^r (-1)^{k+r}}{k! (r - k)!} \quad (4.24)$$

in order to accelerate the convergence. In Fig. 4 we plot the sequences $\text{Re } f_{e,o}(\ell, z)$ as well as their Richardson transforms $\text{Re } f_{e,o}^{(k)}(\ell, z)$ for $k = 1, 2, 3$ for different values of z together with their expected value $-\text{Re } F_0(z)/(\pi z)^2$. As we can see, the agreement is excellent.

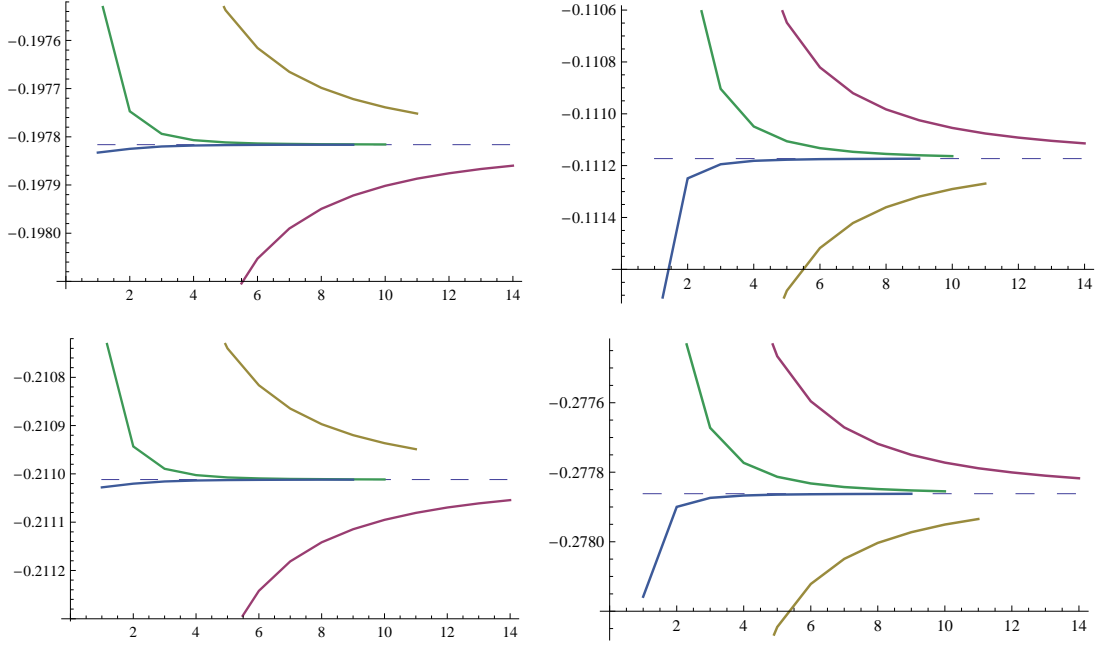


Figure 6: On the left column we plot the sequence $\text{Re } f_{e,1}(\ell, z)$ as well as its Richardson transforms $\text{Re } f_{e,1}^{(k)}(\ell, z)$ for $k = 1, 2, 3$, and for $\ell = 2 \cdots 14$. On the right column we plot the sequence $\text{Re } f_{o,1}(\ell, z)$ as well as its Richardson transforms $\text{Re } f_{o,1}^{(k)}(\ell, z)$, again for $k = 1, 2, 3$ and for $\ell = 2 \cdots 14$. The plots on the top are for $z = 1/4$, and the plots of the bottom are for $z = 1/8$. The dashed line shows the expected value $\text{Re}(F_1(z) + \log \vartheta_{e,o}(z))$.

In order to test the subleading behavior, we define the following sequences:

$$\begin{aligned}
\Theta_{eo}(\ell, z) &= (2\ell)^2 (f_e(\ell, z) - f_o(\ell, z)), \\
f_{e,1}(\ell, z) &= (2\ell)^2 \left(f_e(\ell, z) + \frac{F_0(z)}{(\pi z)^2} \right), \\
f_{o,1}(\ell, z) &= (2\ell + 1)^2 \left(f_o(\ell, z) + \frac{F_0(z)}{(\pi z)^2} \right),
\end{aligned} \tag{4.25}$$

with the expected asymptotic behavior:

$$\begin{aligned}
\lim_{\ell \rightarrow \infty} \Theta_{eo}(\ell, z) &= \log \vartheta_e(z) - \log \vartheta_o(z) \\
\lim_{\ell \rightarrow \infty} f_{e,1}(\ell, z) &= F_1(z) + \log \vartheta_e(z), \\
\lim_{\ell \rightarrow \infty} f_{o,1}(\ell, z) &= F_1(z) + \log \vartheta_o(z).
\end{aligned} \tag{4.26}$$

As we can see in Fig. 6, the agreement between the analytic prediction and the actual asymptotic behavior is again extremely good. This confirms our identification of the saddle point (4.14), as well as the formalism of [10, 17, 18] to incorporate instanton corrections in the oscillatory case. Although these corrections are subleading in the case of the free energy, they can appear at leading order when studying expectation values of operators like Wilson loops, as noticed in [10] in the context of matrix models.

4.3 The phase diagram for complex g_s

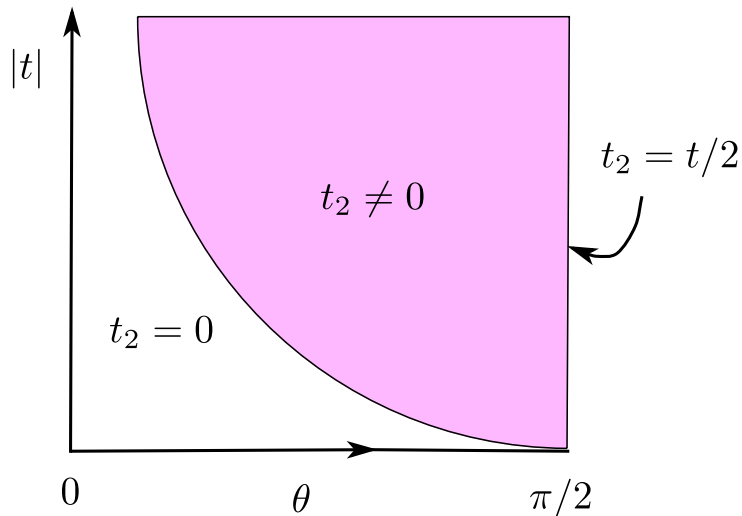


Figure 7: The phase diagram describing the large N limit of Chern–Simons theory on $L(2, 1)$ for $0 \leq \theta \leq \pi/2$, where $\theta = \text{Arg}(g_s)$. There are two phases distinguished by the equilibrium value of t_2 and separated by the critical curve (4.30). In the region $t_2 = 0$, the $1/N$ asymptotics is given by a genus expansion, and corrections due to neighbouring geometries are exponentially suppressed. In the region where $t_2 \neq 0$, neighboring geometries correct the genus expansion with oscillatory terms. Notice that, in this region, the equilibrium value of t_2 changes as we change t . The line $\theta = \pi/2$, where $t_2 = t/2$, can be regarded as an anti-Stokes line.

The qualitative features of the phase diagram for $0 < \text{Arg}(g_s) < \pi/2$ can be understood by looking at the function (3.40). Let us denote

$$\theta = \text{Arg}(g_s) \tag{4.27}$$

so that

$$t = |t|e^{i\theta}. \tag{4.28}$$

If $\text{Re}(A(t)/g_s) > 0$, when t, g_s are now complex, the dominant saddle will be $t_1 = t, t_2 = 0$, as for $\theta = 0$. For $\pi/2 > \theta > 0$, the function

$$f_\theta(|t|) = \text{Re} \left(e^{-i\theta} A \left(e^{i\theta} |t| \right) \right) \tag{4.29}$$

vanishes for a certain critical value

$$|t| = |t|_c(\theta), \tag{4.30}$$

which can be easily determined numerically. The curve (4.30) separates two different regions in the complex t plane, as shown in Fig. 7. In the region under this curve, where

$$0 \leq |t| < |t|_c(\theta), \tag{4.31}$$

the dominant saddle is $t_1 = t, t_2 = 0$. In the region above this curve, we expect the dominant saddle to have $t_2 \neq 0$.

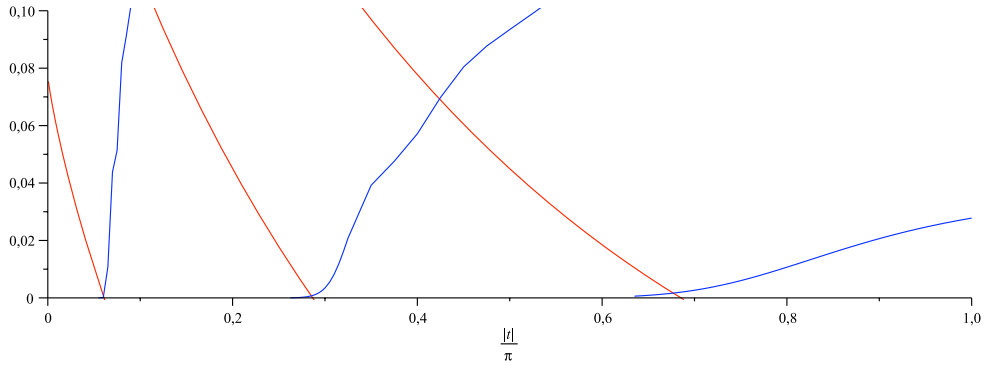


Figure 8: Plot of (4.32) (blue), calculated for $N = 22$, and the real part of the instanton action (4.29) (red), as a function of $|t|/\pi$, for $\theta = 9\pi/20$, $7\pi/20$ and $\pi/4$, from left to right. The function (4.29) decreases monotonically until it becomes zero at $|t|_c(\theta)$. At this point, the average (4.32) gets triggered and it increases with $|t|$.

In order to verify this, as well as the location of the phase boundary, we have computed numerically the average filling fraction at finite N

$$\left\langle \frac{N_2}{N} \right\rangle_\theta = \frac{1}{|Z(N, g_s)|} \sum_{N_2=0}^N \frac{N_2}{N} |Z(N, N_2, g_s)| \quad (4.32)$$

for fixed θ , as a function of $|t|$. This gives a finite N approximation to the value of $|t_2/t|$ at the dominant saddle. In Fig. 8 we plot the value of (4.32) and the real part of the instanton action (4.29), for three different values of θ , and as a function of $|t|/\pi$. Within the limits of the numerical approximation, we clearly see that N_2/N starts developing an expectation value precisely when we reach (4.30). As shown in Fig. 9, the limiting value of $|t_2/t|$ as $|t|$ increases seems to be $1/2$.

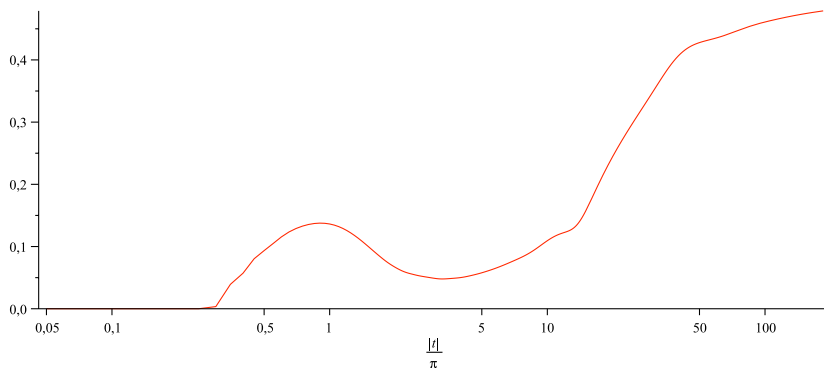


Figure 9: Plot of (4.32) as a function of $|t|/\pi$, for $N = 22$ and $\theta = 7\pi/20$.

$$\arg(g_s) = 0$$

$$Z(N, g_s) = \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \\ t \end{array} + \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \end{array} + \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \end{array} + \dots$$

$$\arg(g_s) = \pi/2$$

$$Z(N, g_s) = \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \\ T_2 \\ T_1 \end{array} + \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \end{array} + \begin{array}{c} \diagup \quad \diagdown \\ | \quad | \\ \text{---} \quad \text{---} \\ | \quad | \\ \diagdown \quad \diagup \end{array} + \dots$$

Figure 10: The Stokes phenomenon for the large N duality of $L(2,1)$. When g_s is real and positive, the saddle geometry giving the dominating contribution to the non-perturbative perturbative function is simply the resolved conifold. The corrections, which are exponentially suppressed, can be interpreted as background geometries in which $t_2 = \ell g_s$ is quantized in units of g_s . When g_s is purely imaginary, the saddle geometry is a local $\mathbb{P}^1 \times \mathbb{P}^1$ with $t_1 = t_2$. Corrections to the genus expansion contribute to the large N asymptotics. They can be regarded as a sum of infinite geometries where $t_2 = n g_s$ is quantized, and their contribution adds up to a theta function. In terms of the large radius Kähler parameters, these geometries have fixed $T_1 - T_2 = t$.

4.4 Stokes phenomenon and target geometries

We can now summarize the most important qualitative results of the analysis of the phase diagram.

First of all, for each complex value of t , the large N asymptotics of the non-perturbative free energy is dominated by a fixed filling fraction or background (t_1^*, t_2^*) . Such a background corresponds to a fixed target geometry, i.e. to a point in the Calabi–Yau moduli space. However, the dominant background changes as we vary t .

The second important point is that the structure of the large N asymptotics depend on the nature of the dominant saddle point. If the dominant saddle point is on the boundary, i.e. one of the filling fractions vanishes, the $1/N$ asymptotics is given by a conventional genus expansion. There are non-perturbative effects due to the contributions of neighbouring geometries, as shown in (2.11), but they are exponentially suppressed. However, if the dominant saddle point is an interior point, the asymptotics has an oscillatory behavior and it involves theta functions, as detailed in (2.17). In particular, the large N asymptotics is no longer given by a genus expansion around a fixed geometry, and corrections due to neighboring geometries are crucial already at the next-to-leading order. We hope that our detailed numerical analysis in the case $\theta = \pi/2$ has convinced the reader that these effects are crucial in order to obtain the correct large N asymptotics of the gauge theory partition function.

It is interesting to notice that the non-perturbative corrections are due to fluctuating target geometries whose size is *quantized* in units of g_s , see Fig. 10. In the case of the boundary

expansion, an ℓ -instanton correction corresponds to a geometry with $t_2 = \ell g_s$, $\ell = 1, 2, \dots$. In the interior case, the theta function is a discrete sum over an infinite number of geometries with $t_2 = (n - N/2)g_s$, $n \in \mathbb{Z}$. Notice that the saddle geometries for $\arg(g_s) = \pi/2$ correspond to the region in moduli space where $z_1, z_2 \rightarrow \infty$ with z_1/z_2 fixed and finite, therefore they are not in the large radius phase $z_1, z_2 \rightarrow 0$.

Finally, we point out that the change of saddle geometry as we change the complex 't Hooft parameter can be regarded as a *generalized Stokes phenomenon*. In fact, in the limit of vanishing 't Hooft coupling, the asymptotics changes *discontinuously* as we change the argument of the string coupling constant: we have a sudden jump from a dominant saddle at $t_2 = 0$ to a dominant saddle at $t_2/t = 1/2$. The jump in the asymptotics has the same origin as in the Stokes phenomenon: saddles which were subleading due to $\text{Re}(A/g_s) > 0$ are no longer suppressed exponentially, and they lead to an oscillatory asymptotics in the $\theta = \pi/2$ direction. In fact, the line $\theta = \pi/2$ plays the role of an anti-Stokes line. These analogies can be made more precise by studying a model where the connection to the standard Stokes phenomenon is manifest and can be followed in detail, namely the cubic matrix model. We will show in the next section that the phase diagram for this model is very similar to the one we have just found, and we will be able to confirm all the analogies that we have stated.

It is important to notice that the connection to the standard Stokes phenomenon is only strictly true when $t \rightarrow 0$. For finite $|t|$ the Stokes phenomenon is *smoothed out*, in the sense that for fixed $|t|$, as we increase θ from $\theta = 0$, the asymptotics changes smoothly from a phase with $t_2 = 0$ to a phase with $t_2 \neq 0$. The phase boundary (4.30) corresponds then to a second-order phase transition with t_2 as its order parameter.

4.5 Background independence

Our results on the phase structure of this model shed some light on the issue of background independence in topological string theory. As it is well-known, the genus g free energies in topological string theory, $F_g(t_i)$, are quasi-modular forms under the action of the symplectic group [2]. This can be regarded as a consequence of the holomorphic anomaly of [8], and it means in particular that the $F_g(t_i)$ are not well-defined functions on the moduli space.

In the models we are studying, thanks to the large N duality, there is a quantity which is well-defined non-perturbatively, namely the Chern–Simons theory partition function, which depends on g_s and the rank N of the gauge group. Notice that, as needed for consistency, the total 't Hooft parameter t is a symplectic invariant in the dual topological string. An important question is how this non-perturbative quantity is related to the perturbative, background dependent topological string free energies. In this paper we have provided a detailed answer to this question in the particular example of $L(2, 1)$ and its string dual.

First of all, at finite N the non-perturbative partition function (2.7) is a *sum* over all possible backgrounds $N_1 + N_2 = N$. In the gauge theory description, these backgrounds correspond to saddle points or flat connections. In the matrix model description, they correspond to different filling fractions. In the language of statistical mechanics, we can say that the background-independent, non-perturbative partition function has to be calculated in a grand-canonical ensemble where N_2 is a fluctuating variable. This is manifest in (2.7), where the parameters ζ_i play the role of fugacities. On the other hand, the background-dependent topological string partition function is computed in a canonical ensemble where N_2 is fixed.

In the thermodynamic limit (i.e. at large N), as it is well-known in statistical mechanics, the sum over backgrounds is strongly peaked around a particular one, and both ensembles are equivalent. This particular background is the dominant saddle which we have determined in

various situations. It depends on the gauge theory parameters N and g_s , and it changes as we move on the complex t plane, displaying the rich phase structure that we have analyzed. At finite N , however, the results of the two ensembles are different, and they differ precisely in the terms which go beyond the genus expansion, i.e. in the contributions of large N multi-instantons.

The resulting non-perturbative picture is very different from the picture that one obtains in perturbative topological string theory: the background-dependent quantities are *emergent* quantities and they only make sense in the large N limit, since they are defined through an asymptotic expansion.

5. Stokes phenomenon and matrix models

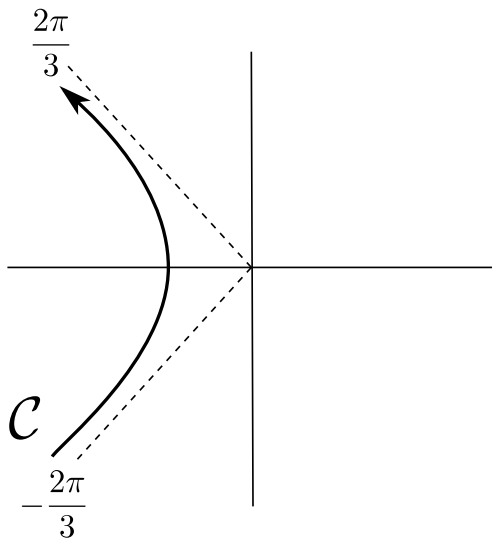


Figure 11: The integration contour \mathcal{C} for (5.2), which leads to a convergent integral.

In order to have a better understanding of the nature of the line $\text{Arg}(g_s) = \pi/2$ in the phase diagram of Chern–Simons theory, we will make a detailed analysis of the Stokes phenomenon in a simple matrix model. We will focus on the quintessential example of Stokes phenomenon, namely the Airy function, and we will study its fate as we promote the Airy integral to a matrix integral. Of course, the study of matrix models in the double-scaling limit produces differential equations which display the so-called non-linear Stokes phenomenon [39, 21]. Here we are rather interested in recovering the standard, *linear* Stokes phenomenon in the context of matrix integrals.

5.1 Review of the Stokes phenomenon for the Airy function

Let us consider the potential

$$V(x) = -e^{i\kappa}x + \frac{x^3}{3} \tag{5.1}$$

where $\kappa \in [0, 2\pi]$. The one-dimensional integral

$$Z_1(g_s) = \int_{\mathcal{C}} \frac{dx}{2\pi} e^{-\frac{1}{g_s}V(x)} \tag{5.2}$$

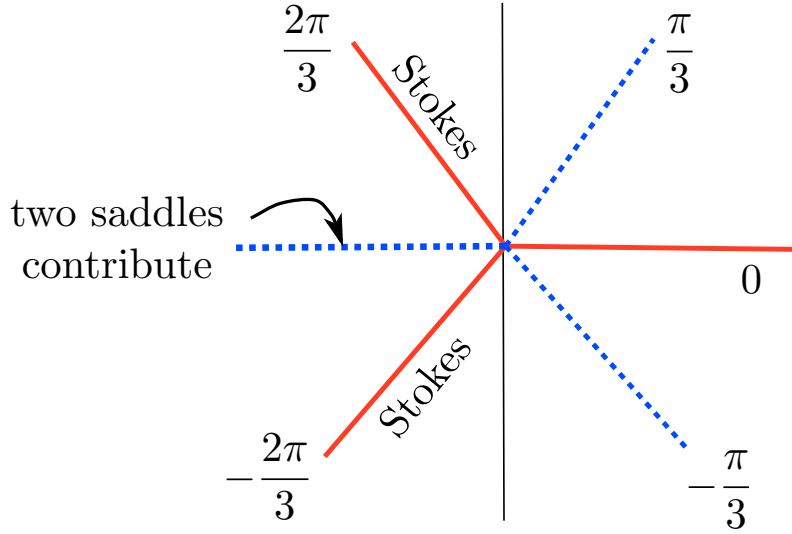


Figure 12: Saddle-point analysis of the asymptotics of the function (5.2), closely related to the Airy function $\text{Ai}(\zeta)$. Full lines (in red) are Stokes lines, while dashed lines (in blue) are anti-Stokes lines. On the Stokes lines $\kappa = \pm 2\pi/3$, a second saddle appears in the integration contour. This saddle is subdominant when $2\pi/3 \leq |\kappa| < \pi$ and does not contribute to classical asymptotics. However, at $\kappa = \pi$, the saddle is not subdominant anymore and leads to an oscillatory asymptotics.

where \mathcal{C} is the path shown in Fig. 11, and g_s is real and positive, is essentially the Airy function $\text{Ai}(w)$, where

$$w = g_s^{-2/3} e^{i\kappa}. \quad (5.3)$$

The small g_s asymptotics of this integral is given by the large w asymptotics of the Airy function. The standard analysis of this asymptotics is as follows (see [38] for a very nice discussion). There are two saddle points

$$x^{\text{L,R}} = \mp \zeta^{1/2} \quad (5.4)$$

where we have introduced the variable

$$\zeta = e^{i\kappa}. \quad (5.5)$$

These saddles have actions

$$-\frac{1}{g_s} V(x^{\text{L,R}}) = \mp \frac{2}{3g_s} \zeta^{3/2}. \quad (5.6)$$

For

$$-\frac{2\pi}{3} < \kappa < \frac{2\pi}{3} \quad (5.7)$$

the path \mathcal{C} can be deformed into a path of steepest descent through the saddle point at x^{L} . When $\kappa = 2\pi/3$, the steepest descent path coming from the saddle at x^{L} runs right into the other saddle point. At this angle we have

$$\text{Im}(V(x^{\text{L}})) = \text{Im}(V(x^{\text{R}})) \quad (5.8)$$

and the corresponding direction is called a Stokes line. This is the place where the second saddle x^{R} might start contributing to the integral. In fact, for

$$\frac{2\pi}{3} < |\kappa| < \pi \quad (5.9)$$

the contour \mathcal{C} gets deformed into a steepest descent path passing through x^L together with a steepest descent path passing through x^R . However, in this range the latter gives an exponentially suppressed contribution, and the classical asymptotics is just given by the contribution from x^L . This leads to

$$Z_1 \sim e^{-2w^{3/2}/3} \quad (5.10)$$

where w is the variable introduced in (5.3). When $\kappa = \pi$ the asymptotics is different, since both saddles have the same real part

$$\operatorname{Re}(V(x^L)) = \operatorname{Re}(V(x^R)). \quad (5.11)$$

A line where this occurs is called an anti-Stokes line. Along this line both saddles contribute to the asymptotics, and we have

$$Z_1 \sim \cos\left(\frac{2}{3}|w|^{3/2} - \frac{\pi}{4}\right). \quad (5.12)$$

The fact that different asymptotic formulae hold on different directions for the same analytic function is called the Stokes phenomenon. From the point of saddle-point analysis, what is happening is that the saddle point which appeared on the Stokes lines, at $\kappa = \pm 2\pi/3$, is no longer subdominant at $\kappa = \pi$, and it has to be included in the asymptotics. The saddle-point analysis is summarized in Fig. 12. We should mention that it is important sometimes to take into account the subleading exponentials, even in the region $\kappa \neq \pi$. This is very clearly shown in the analysis of Berry in [7], and it has led to the development of non-classical asymptotic theories. The most sophisticated of these, resurgent analysis, is applied to the Airy function in for example [15].

5.2 The Stokes phenomenon and the cubic matrix model

Let us now promote (5.2) to a full matrix integral,

$$Z_{\text{np}}^{\mathcal{C}}(N, g_s, \kappa) = \int_{\mathcal{C}} \prod_{i=1}^N \frac{dx_i}{2\pi} \Delta^2(x) e^{-\frac{1}{g_s} \sum_{i=1}^N V(x_i)}, \quad (5.13)$$

where all the eigenvalues are integrated along the contour \mathcal{C} . We can now study its small g_s asymptotics for $t = g_s N$ fixed. Of course, this is nothing but the 't Hooft expansion of this matrix model. The general saddle point is a *two-cut configuration*, labelled by (N_1, N_2) , where N_1, N_2 are the number of eigenvalues near the critical points $x^{L,R}$, respectively. Before starting the analysis, let us notice that for small t the saddle point structure should be the same as for the Airy function. The reason is that, for small t , the Vandermonde repulsion among eigenvalues is suppressed w.r.t. the potential, and the model becomes just N copies of the one-dimensional integral.

Therefore, at least for small t , we expect the following phase structure. For $|\kappa| < 2\pi/3$, the dominant saddle is a one-cut configuration where all the eigenvalues sit near $x^L = -\zeta^{1/2}$. This is the boundary saddle point $(N, 0)$. It is a one-cut configuration which can be analyzed with standard techniques. The endpoints of the cut (a, b) are determined by the equations

$$x_0(x_0^2 - \zeta) = t, \quad \delta^2 = 2(\zeta - x_0^2) \quad (5.14)$$

where

$$a = -x_0 + \delta, \quad b = -x_0 - \delta \quad (5.15)$$

and we choose the root $x_0 = \zeta^{1/2} + \dots$. The spectral curve is

$$y(x) = (x - x_0) \sqrt{x^2 + 2xx_0 + 3x_0^2 - 2\zeta} \quad (5.16)$$

and the effective potential reads

$$V_{\text{eff}}(x) = \frac{1}{3} (x(x - x_0) - 2\zeta) \sqrt{x^2 + 2xx_0 + 3x_0^2 - 2\zeta} - 2x_0 (x_0^2 - \zeta) \log \left(x_0 + x + \sqrt{x^2 + 2xx_0 + 3x_0^2 - 2\zeta} \right). \quad (5.17)$$

For $2\pi/3 \leq |\kappa| < \pi$, the second saddle should start contributing, but at least at small t it should be exponentially suppressed. In other words, the saddle configuration for the matrix integral in this region should still be the boundary saddle $(N, 0)$, but there will be corrections of the form (2.11). The instanton action is given by

$$A = V_{\text{eff}}(x_0) - V_{\text{eff}}(b). \quad (5.18)$$

The small t expansion of this action is

$$A = -\frac{4}{3}\zeta^{3/2} - t + t \log t - 3t \log(2\zeta^{1/2}) - \frac{7}{8\zeta^{3/2}}t^2 + \frac{41}{64\zeta^3}t^3 + \dots \quad (5.19)$$

Notice that, as $t \rightarrow 0$, we find

$$x_0 \rightarrow x^{\text{R}}, \quad (5.20)$$

which is the other saddle, and

$$A \rightarrow -\frac{4}{3}\zeta^{3/2} \quad (5.21)$$

which is the difference between the actions of the two saddles (5.6). This is in agreement with the expectation that for small t the saddle-point structure of the matrix model is just the one coming from the Airy function. The real part of the instanton action vanishes at a critical value of $t_c(\kappa)$, which depends on the value of κ . For example, for $\kappa = 2\pi/3$, we have

$$t_c = \frac{2}{3\sqrt{3}}. \quad (5.22)$$

This is precisely the critical point leading to pure 2d gravity. For $t > t_c(\kappa)$, a phase transition occurs to a new phase, in which we expect generically $N_2^* \neq 0$.

When $\kappa = \pi$, i.e. on the anti-Stokes line, we expect that the standard Stokes phenomenon controls the behaviour of the matrix integral for small t . Indeed, this is the case. The real part of the action $A(t)$ vanishes at $t = 0$, and it is actually negative for small t . We then have to look for a new saddle point by analyzing the behavior of F'_0 . The genus zero free energy of the cubic matrix model has been analyzed in many places [11, 30]. It has the structure

$$F_0(t_1, t_2) = F_0^{\text{np}}(t_1, t_2) + F_0^{\text{p}}(t_1, t_2), \quad (5.23)$$

where

$$F_0^{\text{np}}(t_1, t_2) = \frac{1}{2}t_1^2 \left(\ln t_1 - \frac{3}{2} \right) + \frac{1}{2}t_2^2 \left(\ln t_2 - \frac{3}{2} \right) - \frac{1}{2}(t_1^2 + t_2^2) \log(2\zeta^{1/2}), \quad (5.24)$$

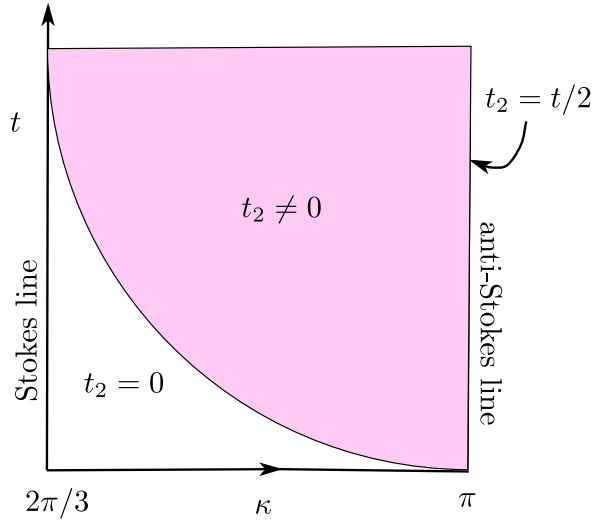


Figure 13: The phase diagram of the cubic matrix model (5.13) as a function of $2\pi/3 \leq \kappa \leq \pi$ and $t = g_s N$ (which is taken to be real), for small t . The Stokes and anti-Stokes lines of the $N = 1$ case, which occur for $\kappa = 2\pi/3$ and $\kappa = \pi$, respectively, determine to a large extent the phase diagram for small t . On the anti-Stokes line, the saddle value of t_2 is $t/2$, and the classical $1/N$ asymptotics requires corrections beyond the genus expansion.

and

$$\begin{aligned}
F_0^{\text{p}}(t_1, t_2) = & -\frac{2}{3}\zeta^{3/2}(t_1 - t_2) + 2t_1 t_2 \log(2\zeta^{1/2}) + \frac{1}{8\zeta^{3/2}} \left(-\frac{2}{3}t_1^3 - 5t_2^2 t_1 + 5t_2 t_1^2 + \frac{2}{3}t_2^3 \right) \\
& + \frac{1}{64\zeta^3} \left(\frac{8}{3}t_1^4 - \frac{91}{3}t_1^3 t_2 + 59t_1^2 t_2^2 - \frac{91}{3}t_1 t_2^3 + \frac{8}{3}t_2^4 \right) + \dots
\end{aligned} \tag{5.25}$$

On the anti-Stokes line $\arg(\kappa) = \pi$, $\zeta^{1/2}$ is purely imaginary, and it is easy to see from the structure of the genus zero free energy that its *real* part is *symmetric* in t_1, t_2 . Therefore,

$$t_1 = t_2 \tag{5.26}$$

solves the saddle-point equation

$$\text{Re} \frac{\partial F_0}{\partial s} = 0 \tag{5.27}$$

at least for small t (more precisely, (5.26) provides a solution in the domain of analyticity of the convergent expansion (5.25)). Therefore, on the anti-Stokes line and for t small enough, the partition function of the cubic matrix model (5.13) is given by an expansion of the form (2.17). This type of expansion is then the natural generalization to matrix models of the oscillatory behavior along an anti-Stokes line.

The phase diagram of the cubic matrix model, again at small t , is represented in Fig. 13. We see that it is very similar to the phase diagram of the large N Chern–Simons theory on a lens space that we analyzed in the previous section. This confirms that the line $\theta = \pi/2$, where the vacuum occurs at $t_1 = t_2 = t/2$, plays the role of an anti-Stokes line. Strictly speaking, however, the discontinuous jump in the asymptotics typical of the Stokes phenomenon occurs

only at $t = 0$. For $t \neq 0$, as we change the angle we encounter the region where t_2 goes from a zero value to a nonzero value in a smooth way. We conclude that turning on the 't Hooft parameter smooths out the Stokes discontinuity, and we pass from a first-order phase transition at $t = 0$ to a second-order phase transition for $t > 0$, where t_2 plays the role of an order parameter.

As t increases along the different directions, further phase transitions can occur which we have not explored. For example, along the $\kappa = \pi$ direction, it follows from the results of [9] that there is a transition at $t > 0$ in which the two cuts merge into a trivalent graph. Interestingly, such transitions along the anti-Stokes line are absent in the lens space partition function.

6. Conclusions

In this paper we have analyzed non-perturbative aspects of the large N expansion in a theory with a rich structure of instanton sectors and with a large N dual. There are various important conclusions that emerge from our analysis. First of all, we have seen that, for each choice of gauge theory parameters g_s , N , the large N asymptotics is dominated by a single instanton sector. This sector corresponds to a fixed geometry in the moduli space of the dual Calabi–Yau geometry, which can be characterized by the Boutroux condition (2.25). Second, there are large N phase transitions as we move in the complex t plane, and different instanton sectors dominate the asymptotics for different values of t . These transitions can be regarded as deformations of the standard Stokes phenomenon. Third, the correct large N asymptotics goes beyond the genus expansion, and subleading saddles must be incorporated.

We conclude with a list of issues left open by our work:

1. The genus g topological string free energies appear in the large N asymptotics of the “canonical” partition function $Z(N, N_2, g_s)$ with N_2 fixed. They are known to transform as quasi-modular forms when moving from one patch of the moduli space to another [2]. At the level of the gauge theory, these transformations should correspond to phase transitions within the same instanton sector, which we have not studied in detail in this paper, since we have focused on phase transitions among different instanton sectors. On the other hand, the matrix integral (3.8) gives a non-perturbative definition of this “canonical” partition function which is valid for any N_k and any complex g_s . It would be important to understand the precise connection between the modular properties of the topological string amplitudes and the large N asymptotics of the quantity (3.8), which is background independent and clearly well-defined in all the moduli space. The results obtained in this paper seem to indicate that modularity can be understood in terms of phase transitions or generalized Stokes phenomena, and that background dependence in these models is ultimately an artifact of the asymptotic large N expansion.
2. Non-perturbative effects of the form e^{-A/g_s} in gauge theories and matrix models have been identified in terms of D-branes in their string duals. In particular, the source of these effects has been shown to be Liouville branes in minimal strings, and more recently toric branes in topological strings [33, 35, 43]. It seems clear that the subleading saddles appearing in the oscillatory asymptotics (2.17) can be also interpreted in terms of D-branes, and in the example of local \mathbb{F}_0 these are also toric D-branes in this geometry. It would be very interesting to make more precise the D-brane interpretation of the non-perturbative corrections appearing in (2.17).
3. Even though in this paper we focused on a simple toy model of gauge/string duality, we believe that some of the effects we studied in this work are likely to play a role in other

large N dualities. The underlying reason for the phenomena explored in this paper is the existence of nontrivial instanton sectors which compete among them at large N . In some cases we have a single sector dominating the large N asymptotics, but in other cases, along generalized anti–Stokes lines, the standard large N asymptotics needs to be corrected, as first pointed out in [10]. This scenario seems to be quite general and it is likely to appear in other models, in particular in AdS/CFT dualities. For example, the recent work [29] indicates that matrix integrals closely related to (3.9) are relevant in the study of large N dualities for ABJM theories.

Acknowledgments

We would like to thank Bertrand Eynard, Andreas Malaspinas and Cumrun Vafa for discussions and suggestions. The work of M.M. and P.P. is supported in part by the Fonds National Suisse.

A. Some aspects of topological string theory on the local $\mathbb{P}^1 \times \mathbb{P}^1$ geometry

In this Appendix we give some further information on topological string theory on the local $\mathbb{P}^1 \times \mathbb{P}^1$ geometry, and in particular we justify some of the statements made in the bulk of the text.

The Yukawa couplings of this Calabi–Yau, when written in terms of the “bare” coordinates z_1, z_2 appearing in (3.20), are [3, 25]

$$\begin{aligned} C_{111} &= \frac{(1 - 4z_2)^2 - 16z_1(1 + z_1)}{4z_1^3\Delta}, \\ C_{112} &= \frac{16z_1^2 - (1 - 4z_2)^2}{4z_1^2z_2\Delta}, \\ C_{122} &= \frac{16z_2^2 - (1 - 4z_1)^2}{4z_1z_2^2\Delta}, \\ C_{222} &= \frac{(1 - 4z_1)^2 - 16z_2(1 + z_2)}{4z_2^3\Delta}, \end{aligned} \tag{A.1}$$

where

$$\Delta = 1 - 8(z_1 + z_2) + 16(z_1 - z_2)^2. \tag{A.2}$$

Using the Yukawa couplings (A.1), it is possible to find a closed form for the prepotential when $x_2 = 0$. The triple derivative w.r.t x_1 is given by

$$C_{x_1x_1x_1} \Big|_{x_2=0} = (C_{111}J_1^3 + 3C_{112}J_1^2J_2 + 3C_{122}J_1J_2^2 + C_{222}J_2^3)_{x_2=0} = \frac{2 - x_1}{4x_1(x_1 - 1)^3}, \tag{A.3}$$

where

$$J_1 = \frac{\partial z_1}{\partial x_1} = \frac{x_1 - 2}{x_1^3x_2^2}, \quad J_2 = \frac{\partial z_2}{\partial x_1} = -\frac{2}{x_1^3x_2^2}. \tag{A.4}$$

Since

$$\partial_{x_1}\sigma_1 = \frac{1}{1 - x_1}, \tag{A.5}$$

we obtain a closed form for the triple derivative w.r.t. the flat coordinate σ_1 :

$$C_{\sigma_1\sigma_1\sigma_1}\Big|_{x_2=0} = \frac{1 + e^{-\sigma_1}}{4(1 - e^{-\sigma_1})}, \quad (\text{A.6})$$

which can be integrated to give:

$$\mathcal{F}_0(\sigma_1, 0) = -\frac{1}{2}\text{Li}_3(e^{-\sigma_1}) + \dots \quad (\text{A.7})$$

The dots indicate a degree three polynomial. This polynomial can be fixed by comparing to the perturbative results, and one finally obtains (3.42).

It is also possible to derive closed expressions for $\partial_{\sigma_2}\mathcal{F}_{\sigma_2}$ on the slice $x_2 = 0$. To do this, one first writes the Picard–Fuchs system (3.21) in terms of the coordinates $x_{1,2}$,

$$\begin{aligned} \mathcal{L}_1 &= \frac{1}{4}(8 - 8x_1 + x_1^2)x_2\partial_{x_2} + \frac{1}{4}(-4 + (-2 + x_1)^2x_2^2)\partial_{x_2}^2 \\ &\quad + (-1 + x_1)x_1^2\partial_{x_1} - x_1(2 - 3x_1 + x_1^2)x_2\partial_{x_1}\partial_{x_2} + (-1 + x_1)^2x_1^2\partial_{x_1}^2, \\ \mathcal{L}_2 &= (2 - x_1)x_2\partial_{x_2} + (-1 + (1 - x_1)x_2^2)\partial_{x_2}^2 - x_1^2\partial_{x_1} \\ &\quad + 2(-1 + x_1)x_1x_2\partial_{x_1}\partial_{x_2} + (1 - x_1)x_1^2\partial_{x_1}^2, \end{aligned} \quad (\text{A.8})$$

and looks for a solution $\mathcal{L}_{1,2}f = 0$ in the form

$$f(x_1, x_2) = x_2G(x_1) + \mathcal{O}(x_2^3). \quad (\text{A.9})$$

The PF system leads to the following equation for $G(x)$:

$$\frac{1}{4}(x - 4)G(x) + xG'(x) - (1 - x)x^2G''(x) = 0, \quad (\text{A.10})$$

whose general solution is a linear combination of $xK(x)$ and $xK(1 - x)$. By comparing to the first terms in (3.26) and using (3.30), one immediately obtains

$$\frac{\partial\sigma_2}{\partial x_2}\Big|_{x_2=0} = \frac{2}{\pi}x_1K(x_1), \quad \frac{\partial}{\partial x_2}\left(\frac{\partial F_0}{\partial s}\right)\Big|_{x_2=0} = -x_1K(1 - x_1), \quad (\text{A.11})$$

therefore

$$\frac{\partial^2 F_0}{\partial s^2}\Big|_{x_2=0} = 4\left(\frac{\partial\sigma_2}{\partial x_2}\Big|_{x_2=0}\right)^{-1}\frac{\partial}{\partial x_2}\left(\frac{\partial F_0}{\partial s}\right)\Big|_{x_2=0} = -2\pi\frac{K(1 - x_1)}{K(x_1)}, \quad (\text{A.12})$$

and (3.43) follows. Equivalently, one can derive (A.11,A.12) by solving the recursion (3.27) for $c_{m,1}$ and $d_{m,1}$,

$$c_{m,1} = \frac{\Gamma(m - 1/2)^2}{\pi\Gamma(m)^2}, \quad m \geq 1, \quad (\text{A.13})$$

$$d_{m,1} = -2(\psi_m - \psi_{m-1/2})c_{m,1},$$

and using the series expansions

$$\begin{aligned} K(x) &= \frac{\pi}{2}\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)_k^2\frac{x^k}{(k!)^2}, \\ K(1 - x) &= -\frac{1}{\pi}\log(x)K(x) + \sum_{k=0}^{\infty}\left(\frac{1}{2}\right)_k^2\frac{x^k}{(k!)^2}(\psi_{n+1} - \psi_{n+1/2}). \end{aligned} \quad (\text{A.14})$$

Finally, we calculate the instanton action (3.40) in the slice $t_2 = 0$. This can be in principle derived from the above geometry, but in fact we can use a computation already done in [5]. As in [35, 36], the one–instanton free energy is given by the ratio of the one instanton partition function $Z_N^{(1)} \equiv Z(N, N_2 = 1, g_s)$ and the zero instanton partition function $Z_N^{(0)} \equiv Z(N, N_2 = 0, g_s)$,

$$F^{(1)} = \frac{Z_N^{(1)}}{Z_N^{(0)}} \sim e^{-A(t)/g_s}, \quad (\text{A.15})$$

We can evaluate $Z_N^{(1)}$ by taking $(N_1, N_2) = (N - 1, 1)$ in the matrix integral (3.9), as in [35], and we find

$$Z_N^{(1)} = e^{-\frac{\pi^2}{2g_s} - g_s(\rho_N^2 - \rho_{N-1}^2)} Z_{N-1}^{(0)} \int \frac{ds}{2\pi} e^{-\frac{1}{2g_s} s^2} \left\langle \prod_{j=2}^N \left(2 \sinh \frac{s + \pi i - x_j}{2} \right)^2 \right\rangle_{N-1}, \quad (\text{A.16})$$

where ρ_N is the Weyl vector in the $U(N)$ theory, and the vev is calculated in the matrix model defined by $Z_N^{(0)}$. At small g_s , the integral over s can be calculated at the saddle $s = 0$, and the v.e.v. can be calculated in the planar limit. Taking into account that

$$e^{-g_s(\rho_N^2 - \rho_{N-1}^2)} \frac{Z_{N-1}^{(0)}}{Z_N^{(0)}} \sim \exp \left\{ -\frac{1}{g_s} \partial_t \left(F_0^{\mathbb{S}^3}(t) - \frac{t^3}{12} \right) \right\} \quad (\text{A.17})$$

we find,

$$A(t) = \frac{\pi^2}{2} + \partial_t \left(F_0^{\mathbb{S}^3}(t) - \frac{t^3}{12} \right) - t \int du \rho(u) \log \left(4 \cosh^2 \frac{u}{2} \right), \quad (\text{A.18})$$

where

$$\rho(u) = \frac{1}{\pi t} \tan^{-1} \left[\frac{\sqrt{e^t - \cosh^2 \left(\frac{u}{2} \right)}}{\cosh \left(\frac{u}{2} \right)} \right] \quad (\text{A.19})$$

is the density of eigenvalues of the \mathbb{S}^3 Chern–Simons matrix model.

Fortunately we do not have to evaluate the integral in (A.18), since the result is already available in the literature. A closely related integral has been studied in [5, 12, 28] in the context of the q –deformed two–dimensional Yang–Mills theory on the two–sphere, and it calculates the large N instanton action $\gamma(A, p)$. Its explicit expression can be found in equation (4.22) of [5]. In particular, for $p = 2$ and $A = -4t$, (A.18) equals $\gamma(A, p)/4$, and using the result of [5] one finds

$$\begin{aligned} A(t) &= \frac{\pi^2}{3} - 2\text{Li}_2(1 - e^{-t/2}) + 2\text{Li}_2(e^{-t/2}) + t \log(1 - e^{-t/2}) - \text{Li}_2(e^{-t}) \\ &= 2 \left(\text{Li}_2(e^{-t/2}) - \text{Li}_2(-e^{-t/2}) \right). \end{aligned} \quad (\text{A.20})$$

For $p = 2$, the q –deformed Yang–Mills theory does not undergo a large N phase transition, as shown in [5, 12, 28]. In our context this means that there are no large N phase transitions when t is real.

References

- [1] I. Affleck, “Testing The Instanton Method,” Phys. Lett. B **92**, 149 (1980).

- [2] M. Aganagic, V. Bouchard and A. Klemm, “Topological Strings and (Almost) Modular Forms,” *Commun. Math. Phys.* **277** (2008) 771 [arXiv:hep-th/0607100].
- [3] M. Aganagic, A. Klemm, M. Mariño and C. Vafa, “Matrix model as a mirror of Chern-Simons theory,” *JHEP* **0402**, 010 (2004) [arXiv:hep-th/0211098].
- [4] L. Álvarez-Gaumé, C. Gómez, H. Liu and S. Wadia, “Finite temperature effective action, AdS(5) black holes, and 1/N expansion,” *Phys. Rev. D* **71**, 124023 (2005) [arXiv:hep-th/0502227].
- [5] X. Arsiwalla, R. Boels, M. Mariño and A. Sinkovics, “Phase transitions in q-deformed 2d Yang-Mills theory and topological strings,” *Phys. Rev. D* **73**, 026005 (2006) [arXiv:hep-th/0509002].
- [6] C.M. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*, Springer-Verlag, New York 1999.
- [7] M. V. Berry, “Asymptotics, superasymptotics, hyperasymptotics, ...”, in H. Segur *et al*, eds., *Asymptotics beyond all orders*. Plenum Press, New York, 1991.
- [8] M. Bershadsky, S. Cecotti, H. Ooguri and C. Vafa, “Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes,” *Commun. Math. Phys.* **165**, 311 (1994) [arXiv:hep-th/9309140].
- [9] M. Bertola, “Boutroux curves with external field: equilibrium measures without a minimization problem,” arXiv:0705.3062.
- [10] G. Bonnet, F. David and B. Eynard, “Breakdown of universality in multi-cut matrix models,” *J. Phys. A* **33**, 6739 (2000) [arXiv:cond-mat/0003324].
- [11] F. Cachazo, K. A. Intriligator and C. Vafa, “A large N duality via a geometric transition,” *Nucl. Phys. B* **603**, 3 (2001) [arXiv:hep-th/0103067].
- [12] 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* **0601**, 035 (2006) [arXiv:hep-th/0509041].
- [13] F. David, “Phases Of The Large N Matrix Model And non-perturbative Effects In 2-D Gravity,” *Nucl. Phys. B* **348**, 507 (1991).
- [14] F. David, “Non-perturbative effects in matrix models and vacua of two-dimensional gravity,” *Phys. Lett. B* **302**, 403 (1993) [arXiv:hep-th/9212106].
- [15] E. Delabaere, “Effective resummation methods for an implicit resurgent function,” math-ph/0602026.
- [16] M. R. Douglas and V. A. Kazakov, “Large N phase transition in continuum QCD in two-dimensions,” *Phys. Lett. B* **319**, 219 (1993) [arXiv:hep-th/9305047].
- [17] B. Eynard, “Large N expansion of convergent matrix integrals, holomorphic anomalies, and background independence,” *JHEP* **0903**, 003 (2009) [arXiv:0802.1788 [math-ph]].
- [18] B. Eynard and M. Mariño, “A holomorphic and background independent partition function for matrix models and topological strings,” arXiv:0810.4273 [hep-th].
- [19] B. Eynard and N. Orantin, “Algebraic methods in random matrices and enumerative geometry,” arXiv:0811.3531 [math-ph].
- [20] G. Felder and R. Riser, “Holomorphic matrix integrals,” *Nucl. Phys. B* **691**, 251 (2004) [arXiv:hep-th/0401191].
- [21] A. S. Fokas, A. R. Its, A. A. Kapaev and V. Yu. Novokshenov, *Painlevé Transcendents: A Riemann-Hilbert Approach*, American Mathematical Society, 2006.
- [22] R. Gopakumar and C. Vafa, “On the gauge theory/geometry correspondence,” *Adv. Theor. Math. Phys.* **3**, 1415 (1999) [arXiv:hep-th/9811131].

- [23] D. J. Gross and A. Matytsin, “Instanton induced large N phase transitions in two-dimensional and four-dimensional QCD,” Nucl. Phys. B **429**, 50 (1994) [arXiv:hep-th/9404004].
- [24] D. J. Gross and E. Witten, “Possible Third Order Phase Transition In The Large N Lattice Gauge Theory,” Phys. Rev. D **21**, 446 (1980).
- [25] B. Haghighat, A. Klemm and M. Rauch, “Integrability of the holomorphic anomaly equations,” JHEP **0810**, 097 (2008) [arXiv:0809.1674 [hep-th]].
- [26] N. Halmagyi, T. Okuda and V. Yasnov, “Large N duality, lens spaces and the Chern-Simons matrix model,” JHEP **0404**, 014 (2004) [arXiv:hep-th/0312145].
- [27] N. Halmagyi and V. Yasnov, “The spectral curve of the lens space matrix model,” JHEP **0911**, 104 (2009) [arXiv:hep-th/0311117].
- [28] D. Jafferis and J. Marsano, “A DK phase transition in q-deformed Yang-Mills on S^2 and topological strings,” arXiv:hep-th/0509004.
- [29] A. Kapustin, B. Willett and I. Yaakov, “Exact Results for Wilson Loops in Superconformal Chern-Simons Theories with Matter,” JHEP **1003**, 089 (2010) [arXiv:0909.4559 [hep-th]].
- [30] A. Klemm, M. Mariño and S. Theisen, “Gravitational corrections in supersymmetric gauge theory and matrix models,” JHEP **0303**, 051 (2003) [arXiv:hep-th/0211216].
- [31] M. Mariño, “Chern-Simons theory, matrix integrals, and perturbative three-manifold invariants,” Commun. Math. Phys. **253**, 25 (2004) [arXiv:hep-th/0207096].
- [32] M. Mariño, “Les Houches lectures on matrix models and topological strings,” arXiv:hep-th/0410165.
- [33] M. Mariño, “Open string amplitudes and large order behavior in topological string theory,” JHEP **0803** (2008) 060 [arXiv:hep-th/0612127].
- [34] M. Mariño, “Non-perturbative effects and non-perturbative definitions in matrix models and topological strings,” JHEP **0812**, 114 (2008) [arXiv:0805.3033 [hep-th]].
- [35] M. Mariño, R. Schiappa and M. Weiss, “Non-perturbative Effects and the Large-Order Behavior of Matrix Models and Topological Strings,” arXiv:0711.1954 [hep-th].
- [36] M. Mariño, R. Schiappa and M. Weiss, “Multi-Instantons and Multi-Cuts,” J. Math. Phys. **50**, 052301 (2009) [arXiv:0809.2619 [hep-th]].
- [37] J. M. Maldacena, G. W. Moore, N. Seiberg and D. Shih, “Exact vs. semiclassical target space of the minimal string,” JHEP **0410**, 020 (2004) [arXiv:hep-th/0408039].
- [38] P. Miller, *Applied asymptotic analysis*, American Mathematical Society, 2007.
- [39] G. W. Moore, “Geometry Of The String Equations,” Commun. Math. Phys. **133**, 261 (1990).
- [40] G. Münster, “The $1/N$ Expansion And Instantons In CP^{N-1} Models On A Sphere,” Phys. Lett. B **118**, 380 (1982).
- [41] H. Neuberger, “Non-perturbative Contributions In Models With A Nonanalytic Behavior At Infinite N,” Nucl. Phys. B **179**, 253 (1981).
- [42] K. Papadodimas, “S-duality and a large N phase transition in $N = 4$ SYM on K3 at strong coupling,” arXiv:hep-th/0510216.
- [43] S. Pasquetti and R. Schiappa, “Borel and Stokes non-perturbative Phenomena in Topological String Theory and $c=1$ Matrix Models,” arXiv:0907.4082 [hep-th].
- [44] L. Rozansky, “A contribution of the trivial connection to Jones polynomial and Witten’s invariant of 3-D manifolds. 1, 2,” hep-th/9401061, 9403021, Commun. Math. Phys. **175** (1996) 275, 297.

- [45] N. Seiberg and E. Witten, “Electric-Magnetic Duality, Monopole Condensation, And Confinement In $N = 2$ Supersymmetric Yang-Mills Theory,” Nucl. Phys. B **426**, 19 (1994) [Erratum-ibid. B **430**, 485 (1994)] [arXiv:hep-th/9407087].
- [46] S.H. Shenker, “The strength of non-perturbative effects in string theory,” in O. Álvarez, E. Marinari and P. Windey (eds.), *Random Surfaces and Quantum Gravity*, Plenum, New York 1992.
- [47] S. R. Wadia, “A Study Of U(N) Lattice Gauge Theory In Two-Dimensions,” EFI-79/44-CHICAGO, Jul. 1979.
- [48] E. Witten, “Quantum field theory and the Jones polynomial,” Commun. Math. Phys. **121** (1989) 351.
- [49] E. Witten, “Chern-Simons Gauge Theory As A String Theory,” Prog. Math. **133**, 637 (1995) [arXiv:hep-th/9207094].