Membrane instantons from a semiclassical TBA

Flavio Calvo and Marcos Mariño

Département de Physique Théorique et Section de Mathématiques, Université de Genève, Genève, CH-1211 Switzerland

marcos.marino@unige.ch

ABSTRACT: The partition function on the three-sphere of ABJM theory contains non-perturbative corrections which correspond to membrane instantons in M-theory. These corrections can be studied in the Fermi gas approach to the partition function, and they are encoded in a system of integral equations of the TBA type. We study a semiclassical or WKB expansion of this TBA system in the ABJM coupling $k$, which corresponds to the strong coupling expansion of the type IIA string. This allows us to study membrane instanton corrections in M-theory at high order in the WKB expansion. Using these WKB results, we verify the conjectures for the form of the one-instanton correction at finite $k$ proposed recently by Hatsuda, Moriyama and Okuyama (HMO), which are in turn based on a conjectural cancellation of divergences between worldsheet instantons and membrane instantons. The HMO cancellation mechanism is important since it shows in a precise, quantitative way, that the perturbative genus expansion is radically insufficient at strong coupling, and that non-perturbative membrane effects are essential to make sense of the theory. We propose analytic expressions in $k$ for the full two-membrane instanton correction and for higher-order non-perturbative terms, which pass many consistency checks and provide further evidence for the HMO mechanism.
1. Introduction

The partition functions on the three-sphere of $\mathcal{N} \geq 2$ Chern–Simons–matter theories can be reduced by localization to matrix models [1, 2, 3] which have been much studied in the last years. Surprisingly, these matrix models contain an enormous amount of information. When studied in the large $N$ limit, this information can be decoded in terms of M-theory AdS duals. For example, the leading large $N$ free energy can be seen to reproduce the gravity action evaluated on-shell, as first found in [4] in the case of ABJM theory [5] (see [6] for a review and a list of relevant references).

Although most of the work done on these models has focused on the leading order contribution at large $N$, from the point of view of M-theory and quantum gravity the most interesting information is contained in the subleading corrections. For example, the subleading logarithmic correction in $N$ corresponds to a one-loop correction in quantum supergravity, as it has been shown in [7]. There are two types of corrections in the large $N$ expansion: the perturbative corrections in $1/N$, and the non-perturbative or exponentially small corrections. The perturbative corrections can be computed in closed form in a large class of $\mathcal{N} = 3$ theories [8], and they can be resummed in terms of an Airy function, as first shown in [9] for ABJM theory. The
non-perturbative corrections can be in turn divided in two types: the ones due to worldsheet instanton corrections in the AdS dual, and the ones due to more general membrane instanton corrections in M-theory [10]. The appearance of worldsheet instanton corrections was anticipated in [11], in the case of ABJM theory. They appear naturally in the ’t Hooft expansion of the matrix model, at strong ’t Hooft coupling, and were determined in a systematic, recursive way in [4], in a weak string coupling expansion.

The study of membrane instanton corrections is more challenging, since these are non-perturbative effects both in the string dual and in the large $N$ matrix model. In [12] some information about these instantons (like their action) was obtained from the study of the large genus asymptotics of the ’t Hooft expansion, but no concrete recipe was given to calculate them. In [8], a new method was introduced to study the ABJM matrix model and its close cousins, based on an equivalence with a quantum Fermi gas. In the Fermi gas approach, the Planck constant is naturally identified with the inverse string coupling, and the semiclassical limit of the gas corresponds to the strong string coupling limit. One of the main virtues of the Fermi gas approach is that it makes possible to calculate membrane instanton effects systematically, at least in the WKB expansion. This opened the way for a quantitative determination of non-perturbative effects in the M-theory duals to Chern–Simons–matter theories.

In a recent paper [13], Hatsuda, Moriyama and Okuyama (HMO) made various crucial observations on the structure of non-perturbative corrections in ABJM theory, where the inverse string coupling is essentially given by the Chern–Simons level $k$. First of all, they noticed that the worldsheet instanton contributions to the free energy can be resummed at finite $k$ by using a Gopakumar–Vafa representation. The resulting expressions are divergent for integer $k > 1$. Since the free energy is finite for any value of $k$, these divergences have to disappear in the final answer, and [13] suggested that they cancel against similar divergences in the contributions of membrane instantons, in such a way that the total sum of all non-perturbative effects at integer $k$ is finite. We will call this the HMO cancellation mechanism. This mechanism is beautiful and natural, and we believe is of deep conceptual importance for the understanding of M-theory. It shows, in a precise and quantitative way, that the genus expansion based on strings is essentially meaningless: in the non-perturbative completion of type IIA string theory at finite, integer $k$ through M-theory, only the combination of membrane instantons and worldsheet instantons makes sense.

In some cases, the HMO mechanism gives a set of constraints for the membrane instanton corrections at finite $k$. Using these constraints, as well as the first few terms of the semiclassical expansion at small $k$ obtained in [8], an expression for the one-instanton membrane correction was proposed in [13] which passes many consistency checks. It reproduces for example the low order, non-perturbative corrections to the free energy, for small integer values of $k$.

The purpose of the present note is twofold. It was pointed out in [8] that the Hamiltonian problem appearing in the Fermi gas approach to ABJM theory can be studied with a pair of TBA equations first considered in [14, 15] and studied in detail in [16]. This observation was exploited in [17, 18, 13], where the TBA equations were solved for finite $k$ but small chemical potential. This makes it possible to compute the partition function of ABJM theory for finite $k$ and small $N$. On the other hand, the Fermi gas can be studied in the WKB approximation, at small $k$, and this provided many valuable insights into the problem [8]. Our first goal in this paper is to develop a semiclassical or WKB expansion directly in the TBA equations. This leads to an algorithm which calculates the grand potential of the ABJM model systematically, as a power series in $k$, and arbitrary chemical potential. This method is more powerful than the

\footnote{This had been already noticed in 2011 in unpublished work by the second author and Pavel Putrov.}
original WKB expansion of the Fermi gas studied originally in [8], and one can easily push the computation to higher orders.

Our second goal is to use this information to further explore the cancellation mechanism proposed in [13]. We verify that the expression for the membrane one-instanton correction proposed in [13] agrees with the WKB expansion to high order, and we propose analytic expressions for the full two-instanton correction and for some higher-order terms in the non-perturbative expansion. Our proposal for the two-instanton correction passes all the consistency tests, and in particular agrees with the results for the grand potential at $k = 1, 3$ obtained in [13]. Our WKB results confirm then the HMO cancellation mechanism and are very helpful in obtaining conjectural expressions at finite $k$.

The organization of the paper is as follows. In section 2 we review some general results on non-perturbative effects in ABJM theory. In section 3 we explain how the Fermi gas approach of [8] can be reformulated in terms of the TBA system considered by Al. Zamolodchikov in [16]. In section 4 we study in detail the WKB expansion of the TBA system and explain how it leads to the semiclassical expansion for the grand potential considered in [8]. In section 5 we give an application of the WKB expansion to the calculation of membrane instantons: we test the conjecture for one-membrane instanton corrections in [13], and we propose an exact formula for the two-membrane instanton contributions. Finally, we conclude with some open problems. In the first appendix we list some useful results for the integration of generalized hypergeometric functions, and in the second appendix we list WKB expansions for some low order contributions of membrane instantons.

2. Perturbative and non-perturbative aspects of ABJM theory

The quantity we will focus on in this paper is the partition function of ABJM theory on the three-sphere, $Z(N,k)$, which is given by the matrix integral [1]

$$Z_{ABJM}(N) = \frac{1}{N!^2} \int \frac{d^N \mu}{(2\pi)^N} \frac{d^N \nu}{(2\pi)^N} \prod_{i<j} \left[ 2 \sinh \left( \frac{\mu_i - \mu_j}{2} \right) \right]^2 \left[ 2 \sinh \left( \frac{\nu_i - \nu_j}{2} \right) \right]^2 \exp \left[ \frac{ik}{4\pi} \sum_{i=1}^{N} (\mu_i^2 - \nu_i^2) \right].$$

(2.1)

This matrix integral can be calculated in the 't Hooft expansion

$$N \to \infty, \quad \lambda = \frac{N}{k} \text{ fixed}$$

(2.2)

by using techniques of matrix model theory and topological string theory [23, 4]. In particular, one can obtain explicit formulae for the genus $g$ free energies appearing in the $1/N$ expansion

$$F(\lambda, g_s) = \sum_{g=0}^{\infty} g_s^{2g-2} F_g(\lambda),$$

(2.3)

where

$$g_s = \frac{2\pi i}{k}.$$ 

(2.4)

The genus $g$ free energies $F_g(\lambda)$ obtained in this way are exact interpolating functions, and they can be studied in various regimes of the 't Hooft coupling. When expanded at strong coupling, they have the structure

$$F_g(\lambda) = F_g^p(\lambda) + F_g^{np}(\lambda).$$ 

(2.5)
The first term represents the perturbative contribution in $\alpha'$, while the second term is non-perturbative in $\alpha'$,

$$F^{np}_g(\lambda) \sim \mathcal{O}\left(e^{-2\pi\sqrt{2}\lambda}\right). \quad (2.6)$$

The type IIA dual of ABJM theory involves the space $\text{AdS}_4 \times \mathbb{CP}^3$ [5], and this geometry supports worldsheet instantons wrapping a $\mathbb{CP}^1 \subset \mathbb{CP}^3$ [11]. The non-perturbative piece (2.6) was interpreted in [4] as the contribution of these worldsheet instantons.

Besides the non-perturbative effects in $\alpha'$, one can use the connection between the large-order behavior of perturbation theory and instantons to deduce the structure of non-perturbative effects in the string coupling constant. In [12] a detailed analysis showed that these effects would have the form

$$\exp\left(-k\pi\sqrt{2\lambda}\right) \quad (2.7)$$

at large $\lambda$. It was also proposed in [12] that the source of these effects are D2-branes wrapped around three-cycles in the target space. An appropriate, explicit family of generalized Lagrangian submanifolds with the topology of $\mathbb{RP}^3 \subset \mathbb{CP}^3$ was proposed as an explicit candidate for these cycles. We will refer to these non-perturbative effects as membrane instanton effects: they can be interpreted as M2-instantons in M-theory [10], where the M2-brane wraps a three-cycle inside $S^7/Z_k$ which is the lift of the three-cycle in $\mathbb{CP}^4$. Notice that these membrane instanton effects are invisible in ordinary string perturbation theory.

Further information on the membrane instantons can be obtained by using the Fermi gas approach introduced in [8]. In this approach, one first notices (see also [19]) that the matrix integral (2.1) can be written as

$$Z(N, k) = \frac{1}{N!} \sum_{\sigma \in S_N} (-1)^{\epsilon(\sigma)} \int \frac{d^N x}{(2\pi k)^N} \frac{1}{\prod_i 2 \cosh \left(\frac{x_i}{2}\right) 2 \cosh \left(\frac{x_i-x_{\sigma^{-1}(i)}}{2k}\right)}. \quad (2.8)$$

This in turn can be interpreted as the canonical partition function of a one-dimensional Fermi gas with a non-trivial one-particle density matrix

$$\rho(x_1, x_2) = \frac{1}{2\pi k} \frac{1}{(2 \cosh \frac{x_1}{2})^{1/2}} \frac{1}{(2 \cosh \frac{x_2}{2})^{1/2}} \frac{1}{2 \cosh \left(\frac{x_1-x_2}{2k}\right)}. \quad (2.9)$$

The one-particle Hamiltonian $\hat{H}$ of this system is then defined as

$$\hat{\rho} = e^{-\hat{H}}, \quad \langle x_1 | \hat{\rho} | x_2 \rangle = \rho(x_1, x_2), \quad (2.10)$$

and the Planck constant is

$$\hbar = 2\pi k. \quad (2.11)$$

The semiclassical or WKB expansion is then around $k = 0$, and it corresponds to the strong string coupling expansion in the type IIA dual.

The Fermi gas approach suggests to look instead to the grand partition function (see also [20])

$$\Xi(\mu, k) = 1 + \sum_{N=1}^{\infty} Z(N, k) z^N, \quad (2.12)$$

where

$$z = e^\mu \quad (2.13)$$
plays the rôle of the fugacity and $\mu$ is the chemical potential. The grand potential is then defined as

$$ J(\mu, k) = \log \Xi(\mu, k). \quad \text{(2.14)} $$

The canonical partition function is recovered from the grand-canonical potential as

$$ Z(N, k) = \oint \frac{dz}{2\pi i} \frac{\Xi(\mu, k)}{z^{N+1}}. \quad \text{(2.15)} $$

At large $N$, this integral can be computed by applying the saddle-point method to

$$ Z(N, k) = \frac{1}{2\pi i} \int d\mu \exp \left[ J(\mu, k) - \mu N \right]. \quad \text{(2.16)} $$

As shown in [8], the grand potential is the sum of a perturbative and a non-perturbative piece,

$$ J(\mu, k) = J^p(\mu, k) + J^{np}(\mu, k). \quad \text{(2.17)} $$

The perturbative piece is a cubic polynomial in $\mu$:

$$ J^p(\mu, k) = C(k) \mu^3 + B(k) \mu + A(k). \quad \text{(2.18)} $$

The coefficients $C(k)$, $B(k)$ were computed in [8] for ABJM theory (in fact, the analogs of these coefficients for other $N = 3$ theories can be also computed in closed form). The coefficient $A(k)$ can be computed in a WKB expansion around $k = 0$ [8], and the all-orders result was conjectured in [21]. In this paper we will be interested in the non-perturbative piece $J^{np}(\mu, k)$. In general one has the following result:

$$ J^{np}(\mu, k) = J^{M2}(\mu, k) + J^{WS}(\mu, k) + \cdots. \quad \text{(2.19)} $$

Let us explain this structure in more detail. The first term $J^{M2}(\mu, k)$ has the following expansion for $\mu \gg 1$,

$$ J^{M2}(\mu, k) = \sum_{\ell \geq 1} \left( a_\ell(k) \mu^2 + b_\ell(k) \mu + c_\ell(k) \right) e^{-2\ell \mu}. \quad \text{(2.20)} $$

This type of contributions to the grand potential was first found in [8] and it was interpreted there as due to membrane instantons, i.e. to $M2$-branes wrapping a three-cycle $M \subset S^7/\mathbb{Z}_k$ which is a lift of a three-cycle in $\mathbb{CP}^3$. The positive integer $\ell$ is the winding number of the wrapping. We will refer to the $\ell$-th term in the infinite series (2.20) as the the contribution of the $\ell$-membrane instanton. The coefficients $a_\ell(k)$, $b_\ell(k)$ and $c_\ell(k)$ are non-trivial functions of $k$. Using the WKB expansion of the quantum Fermi gas, it was shown in [8] that they have a perturbative expansion around $k = 0$ of the form

$$ a_\ell(k) = \frac{1}{k^2} \sum_{n=0}^{\infty} a_{\ell,n} k^{2n}. \quad \text{(2.21)} $$

A similar expansion holds for $b_\ell(k)$ and $c_\ell(k)$, with coefficients $b_{\ell,n}$, $c_{\ell,n}$, respectively. The WKB expansion was obtained in [8] up to order $n = 2$. One goal of this paper will be to obtain a more efficient method to calculate the WKB expansion.

The second term in (2.19), $J^{WS}(\mu, k)$, is the contribution of worldsheet instantons wrapping $\mathbb{CP}^1 \subset \mathbb{CP}^3$. It has the following expansion for $\mu \gg 1$,

$$ J^{WS}(\mu, k) = \sum_{m=1}^{\infty} d_m(k) e^{-4m \mu}. \quad \text{(2.22)} $$
In [13] a very useful formula was proposed for $J^{WS}(\mu, k)$: by using the fact [23, 4] that the ABJM matrix integral is dual to topological string theory on local $\mathbb{P}^1 \times \mathbb{P}^1$, one can write

$$J^{WS}(\mu, k) = \sum_{n,g,d \geq 1} n_d^g \left( \sin \frac{2\pi n}{k} \right)^{2g-2} \left( -1 \right)^{dn} \frac{(-1)^{dn}}{n} e^{-\frac{4dn}{n}}. \quad (2.23)$$

In this formula, $n_d^g$ is the weighted sum of the Gopakumar–Vafa invariants [22] of local $\mathbb{P}^1 \times \mathbb{P}^1$, $n_d^g = \sum_{d_1+d_2=d} n_{d_1,d_2}^g$. \quad (2.24)

![Diagram](image)

**Figure 1:** From the M-theory point of view, there are two types of non-perturbative effects in ABJM theory: M2-branes can wrap a cycle $M \subset S^7/\mathbb{Z}_k$ which descends to an $\mathbb{R}P^3 \subset \mathbb{C}P^3$ cycle in the type IIA target; or they can wrap a cycle $W$ which descends to $\mathbb{C}P^1 \subset \mathbb{C}P^3$. The most general M2-brane configuration wraps $\ell$ times the cycle $M$ and $m$ times the cycle $W$.

Notice that, from the point of view of M-theory, both types of instantons are due to M2-branes wrapping three-cycles: membrane instantons correspond to a three-cycle $M$ which descends to a three-cycle wrapped by D2-branes in the type IIA target (if the proposal of [12] is correct, this three-cycle is an $\mathbb{R}P^3 \subset \mathbb{C}P^3$); worldsheet instantons correspond to M2-branes wrapping the three-cycle $W = S^3/\mathbb{Z}_k$, which descends to $\mathbb{C}P^1 \subset \mathbb{C}P^3$. Therefore, from the point of view of M-theory, there are two types of three-cycles, $M$ and $W$, and the most general M2-brane configuration wraps $\ell$ times the cycle $M$ and $m$ times the cycle $W$, see Fig. 1. We should then expect that the most general contribution to $J^{np}(\mu)$ is of the form

$$J^{np}(\mu, k) = \sum_{\ell,m \geq 1} f_{\ell,m}(k, \mu) \exp \left[ -\left( 2\ell + \frac{4m}{k} \right) \mu \right]. \quad (2.25)$$

The contribution $J^{M2}(\mu, k)$ corresponds to $m = 0$, while the contribution $J^{WS}(\mu, k)$ corresponds to $\ell = 0$. The existence of such bound states, labeled by two integers $(\ell, m)$, was first indicated in [13] in order to interpret their data on the grand potential. From the point of view of M-theory they are completely natural, since they correspond to the most general set of supersymmetric cycles in the geometry.

Can we compute the contribution of these non-perturbative states? In the case of worldsheet instantons, the expressions (2.23) and (2.24) reduce the problem to the determination of the
Gopakumar–Vafa invariants of local $\mathbb{P}^1 \times \mathbb{P}^1$. This can be done in many different ways, and there is no difficulty in calculating $J^{WS}(\mu, k)$ to any desired order. The calculation of $J^{M2}(\mu, k)$ can be done order by order in $k$ and at all possible orders in the membrane instanton number by using the WKB method of [8]. In the remaining of the paper, we will develop a more efficient method, based on the TBA equations of [16], to calculate this perturbative expansion around $k = 0$. However, there is no known procedure to determine the contribution of bound states appearing in (2.25), not even at small $k$. This is probably the most important open problem in this subject, as we will mention in the final section of the paper.

3. The Fermi gas/TBA approach to ABJM theory

3.1 Fredholm determinants and TBA

We will now summarize some results from [16, 24] which are needed in this paper.

In [16], the following integral kernel is considered

$$K(\theta, \theta') = \frac{1}{2\pi} \exp \left( -u(\theta) - u(\theta') \right) \frac{2 \cosh \frac{\theta - \theta'}{2}}{2 \cosh \frac{\theta_i - \theta_{i+1}}{2}}. \tag{3.1}$$

This defines a homogeneous Fredholm integral equation of the second kind, of the form

$$\int_{-\infty}^{\infty} K(\theta, \theta') f(\theta') = \lambda f(\theta), \tag{3.2}$$

Let $\lambda_n$ be the possible eigenvalues appearing in (3.2), and let us introduce the Fredholm determinant

$$\Xi(z) = \prod_n (1 + z\lambda_n). \tag{3.3}$$

We will regard $\Xi$ as a grand canonical partition function. The grand potential is then given by

$$J(z) = \log \Xi(z) \tag{3.4}$$

and it has the expansion

$$J(z) = -\sum_{\ell=1}^{\infty} \frac{(-z)^\ell}{\ell} Z_\ell \tag{3.5}$$

where $Z_\ell$ is given by the integral

$$Z_\ell = \int_{-\infty}^{\infty} \prod_{i=1}^{\ell} e^{\frac{\theta_i - \theta_{i+1}}{2} \cosh \theta_i - \theta_{i+1}} \frac{d\theta_i}{2\pi} \tag{3.6}$$

with the periodicity condition

$$\theta_{\ell+1} = \theta_1. \tag{3.7}$$

We now introduce the iterated integral

$$R_\ell(\theta) = e^{-2u(\theta)} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{-2u(\theta_1) - \cdots - 2u(\theta_\ell)} e^{\frac{\theta_i - \theta_{i+1}}{2} \cosh \theta_i - \theta_{i+1}} \frac{d\theta_1 \cdots d\theta_\ell}{2\pi}, \quad \ell \geq 1, \tag{3.8}$$

and

$$R_0(\theta) = e^{-2u(\theta)}. \tag{3.9}$$
Notice that
\[ \int_{-\infty}^{\infty} d\theta R_\ell(\theta) = (4\pi)^{\ell+1}Z_{\ell+1} \] (3.10)
and the generating series
\[ R(\theta|z) = \sum_{\ell\geq 0} \left(-\frac{z}{4\pi}\right)^\ell R_\ell(\theta) \] (3.11)
satisfies
\[ \int_{-\infty}^{\infty} \frac{d\theta}{4\pi} R(\theta|z) = \sum_{\ell\geq 0} (-z)^\ell Z_{\ell+1} = \frac{\partial J}{\partial z}. \] (3.12)

It was conjectured in [16] and proved in [24] that the function \( R(\theta|z) \) can be obtained by using TBA-like equations which first appeared in the context of two-dimensional \( \mathcal{N} = 2 \) theories [14, 15]. We first define
\[ R_+(\theta|z) = \frac{1}{2} (R(\theta|z) + R(\theta|-z)), \]
\[ R_-(\theta|z) = \frac{1}{2} (R(\theta|z) - R(\theta|-z)). \] (3.13)

Let us now consider the TBA system
\[ 2u(\theta) = \epsilon(\theta) + \int_{-\infty}^{\infty} \frac{d\theta'}{2\pi} \log \left( 1 + \eta^2(\theta') \right), \]
\[ \eta(\theta) = -z \int_{-\infty}^{\infty} \frac{d\theta'}{2\pi} \frac{e^{-\epsilon(\theta')}}{\cosh(\theta - \theta')}. \] (3.14)

Then, one has that
\[ R_+(\theta|z) = e^{-\epsilon(\theta)}, \]
\[ R_-(\theta|z) = R_+(\theta|z) \int_{-\infty}^{\infty} \frac{d\theta'}{\pi} \frac{\arctan \eta(\theta')}{\cosh^2(\theta - \theta')}. \] (3.15)

This conjecture has been proved in [24] for general \( u(\theta) \). In general, the system (3.14) has to be solved numerically, although an exact solution exists for \( u(\theta) = e^{\theta} \) in terms of Airy functions [25].

### 3.2 From the ABJM Fermi gas to TBA

Let us now find the relation between the Fermi gas approach to ABJM theory and the TBA system considered above (see also [18, 17]). We start from the expression (2.8) for the partition function. We can now regard the density matrix as an integral kernel with the structure of (3.1). The Fredholm determinant of this kernel is nothing but the grand partition function of the Fermi gas. We can then use the results of [16, 24] to write a TBA-like equation determining its grand potential. Let us consider the quantities
\[ \rho^{\ell+1}(x,x) = \langle x|\rho^{\ell+1}|x\rangle = \frac{e^{-2\nu(x)}}{2\pi k} \int_{-\infty}^{\infty} \frac{dx_1 \cdots dx_\ell}{(2\pi k)^\ell} \frac{e^{-2\nu(x_1)} \cdots e^{-2\nu(x_\ell)}}{2\cosh\left(\frac{x-x_1}{2k}\right) \cdots 2\cosh\left(\frac{x-x_\ell}{2k}\right)} \] (3.16)
where
\[ \nu(x) = \frac{1}{2} \log \left( 2 \cosh \frac{x}{2} \right). \] (3.17)
If we now change variables
\[ x_i = k \theta_i, \tag{3.18} \]
and compare with (3.8), we find
\[ \rho^{\ell+1}(x, x) = \frac{1}{(4\pi)^{\ell+1}k} R_\ell(x) \tag{3.19} \]
where we have denoted
\[ R_\ell(x) \equiv R_\ell \left( \theta = \frac{x}{k} \right). \tag{3.20} \]
The function \( R_\ell(\theta) \) is calculated with the TBA system (3.14) and the potential
\[ u(\theta) = \frac{1}{2} \log \left( 2 \cosh \frac{k \theta}{2} \right) \tag{3.21} \]
which depends explicitly on \( k \). The grand potential is given by
\[ \frac{\partial J}{\partial z} = \frac{1}{4\pi k} \int_{-\infty}^{\infty} dx R(x|z). \tag{3.22} \]
Notice that the function \( R(x|z) \) can be written as
\[ R(x|z) = \frac{4\pi k}{z} \left\langle x \left| \frac{1}{e^{H-\mu} + 1} \right| x \right\rangle. \tag{3.23} \]
The quantity appearing in (3.23) is essentially the full quantum-corrected version of the Fermi momentum \( p_F(x) \) of the Fermi gas. Indeed, (3.23) can be computed by using the Wigner map, which associates to any quantum operator \( \mathcal{O} \) a function \( \mathcal{O}_W(x, p) \) in phase space (see [8] for details):
\[ R(x|z) = \frac{1}{z} \int_{-\infty}^{\infty} dp \left( \frac{1}{e^{H-\mu} + 1} \right)_W = \frac{2}{\pi z} p_F(x). \tag{3.24} \]
There is an important property of the TBA system of [16] which is worth discussing in some detail. Notice that the functions \( \epsilon(\theta), \eta(\theta) \) make it possible to calculate both \( R(x|z) \) and \( R(x|-z) \). The last quantity is given by
\[ R(x|-z) = \frac{4\pi k}{z} \left\langle x \left| \frac{1}{e^{H-\mu} - 1} \right| x \right\rangle, \tag{3.25} \]
and it corresponds to the same one-particle Hamiltonian (2.10) but with Bose–Einstein statistics. If we now take into account the expression (3.3), we deduce that for Bose–Einstein statistics there is a physical singularity at
\[ z = \lambda_0^{-1} > 0 \tag{3.26} \]
where \( \lambda_0 \) is the largest eigenvalue of the non-negative Hilbert–Schmidt operator \( \rho(x_1, x_2) \). This singularity corresponds of course to the onset of Bose–Einstein condensation in the gas, and as a consequence the functions \( R_\pm(x|z) \) will have singularities in the \( x \)-plane for \( z \geq \lambda_0^{-1} \). But this is precisely the regime in which we are interested, since large \( N \) corresponds to \( \mu \gg 1 \). Of course, the singularity at large positive \( z \) cancels once one adds up \( R_+ \) and \( R_- \).

The appearance of this Bose–Einstein condensate has prevented the direct study of the large \( N \) limit of ABJM theory with the TBA equations (3.14), even numerically: the standard iteration of the integral equations does not converge when \( z \) is large enough. In the papers [17, 18] they study in fact the small \( z \) regime of (3.14) at finite \( k \) in order to extract the small \( z \) expansion of the grand partition function and therefore the canonical partition functions \( Z(N, k) \) for small \( N \). The large \( N \) limit is then extracted from these small \( N \) results by numerical extrapolation [18, 13]. We will now propose another approach to study the TBA equations.
4. The semiclassical TBA system

4.1 General aspects

As we showed in the last section, the TBA system is equivalent to the Fermi gas picture. In particular, after setting $\theta = x/k$, the function $R(x|z)$ is the Fermi momentum of the gas. As shown in [8], one can do a systematic computation of all quantities in the Fermi gas in a WKB expansion, i.e. in a perturbative expansion in $k$. This suggests studying the TBA equation in a scaling regime where $k$ is small and one can construct a perturbative expansion around $k = 0$.

To do this, we set $\theta = x/k$ and write (3.14) as

$$U(x) = \epsilon(x) + \int_{-\infty}^{\infty} \frac{dx'}{2\pi k} \frac{\log (1 + \eta^2(x'))}{\cosh \left( \frac{x-x'}{k} \right)},$$

$$\eta(x) = -z \int_{-\infty}^{\infty} \frac{dx'}{2\pi k} \frac{e^{-\epsilon(x')}}{\cosh \left( \frac{x-x'}{k} \right)}.$$  

where

$$U(x) = 2u\left( \frac{x}{k} \right).$$

The second equation in (3.13) becomes

$$R_-(x|z) = R_+(x|z) \int_{-\infty}^{\infty} \frac{dx'}{\pi k} \frac{\arctan \eta(x')}{\cosh^2 \left( \frac{x-x'}{k} \right)}.$$  

In the case of ABJM theory, where the potential is given by (3.21), we have

$$U(x) = \log \left( 2 \cosh \frac{x}{2} \right)$$

and it is independent of $k$. Notice that in [16, 24] one considers a general potential $u(\theta)$, with no $k$ parameter to start with, but we can study a one-parameter deformation of the problem by considering a potential $u(k\theta)$, in such a way that $U(x)$ in (4.2) is independent of $k$. The original problem is then obtained when $k = 1$.

The advantage of the $k$-dependent equations (4.1), (4.3) is that they admit a systematic, perturbative expansion around $k = 0$, where they can be solved algebraically. This can be seen from the fact that, when $k \to 0$, the kernel becomes a $\delta$-function:

$$\lim_{k \to 0} \frac{1}{2\pi k \cosh \left( \frac{x}{k} \right)} = \frac{1}{2} \delta(x),$$

and the integral equations become algebraic equations. The most convenient form of the TBA system for the small $k$ expansion involves finite difference equations, and it was already considered (for $k = 1$) in [16] and specially in [24]. In more complicated models, the system of TBA integral equations, when written in terms of difference equations, is usually called the functional $Y$ system.

This form of the TBA system can be easily obtained by Fourier transform. If we denote

$$F[f(x); p] = \hat{f}(p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dx f(x)e^{ixp},$$

we obtain, from the first equation in (4.1),

$$2 \cosh \left( \frac{\pi kp}{2} \right) \left( \hat{U}(p) - \hat{\epsilon}(p) \right) = F \left[ \log(1 + \eta^2(x)); p \right].$$
The first factor in the l.h.s. of (4.7) is the displacement operator
\[ e^{\frac{ikx}{2}} \frac{d}{d x} + e^{-\frac{ikx}{2}} \frac{d}{d x}. \] (4.8)

Doing the same thing in the second equation of (4.1) we find that the two integral equations are equivalent (provided some mild analyticity conditions are satisfied, see [24]) to the two difference equations,
\[ \epsilon \left( x + \frac{\pi ik}{2} \right) + \epsilon \left( x - \frac{\pi ik}{2} \right) = U \left( x + \frac{\pi ik}{2} \right) + U \left( x - \frac{\pi ik}{2} \right) - \log(1 + \eta^2(x)), \] (4.9)
\[ \eta \left( x + \frac{\pi ik}{2} \right) + \eta \left( x - \frac{\pi ik}{2} \right) = -ze^{-\epsilon(x)}. \]

Equivalently, in terms of \( R_+ \), we have
\[ 1 + \eta^2(x) = R_+ \left( x + \frac{\pi ik}{2} \right) R_+ \left( x - \frac{\pi ik}{2} \right) \exp \left\{ U \left( x + \frac{\pi ik}{2} \right) + U \left( x - \frac{\pi ik}{2} \right) \right\}, \] (4.10)
\[ -zR_+(x) = \eta \left( x + \frac{\pi ik}{2} \right) + \eta \left( x - \frac{\pi ik}{2} \right). \]

The equation (4.3) giving \( R_- \) reads now,
\[ \frac{R_- \left( x + \frac{\pi ik}{2} \right)}{R_+ \left( x + \frac{\pi ik}{2} \right)} - \frac{R_- \left( x - \frac{\pi ik}{2} \right)}{R_+ \left( x - \frac{\pi ik}{2} \right)} = 2ik \frac{\eta'(x)}{1 + \eta^2(x)}. \] (4.11)

When \( k = 1 \), these equations have been written down in [24].

The TBA equations, in the form (4.10), (4.11), can be solved systematically in an expansion around \( k = 0 \). Let us denote
\[ r(x) = R_+(x), \quad t(x) = \frac{R_-(x)}{R_+(x)}. \] (4.12)

We introduce the ansatz
\[ r(x) = \sum_{n=0}^{\infty} r_n(x)k^{2n}, \]
\[ \eta(x) = \sum_{n=0}^{\infty} \eta_n(x)k^{2n}, \] (4.13)
\[ t(x) = \sum_{n=0}^{\infty} t_n(x)k^{2n}. \]

We then make a Taylor expansion in \( k \) of the displaced functions in (4.10), and we solve order by order in \( k \) for \( r_n, \eta_n \). Once \( \eta \) is known, \( t(x) \) can be obtained from
\[ t(x) = \frac{2}{\pi \sin \zeta} \tan^{-1}(\eta), \quad \zeta = \frac{k\pi}{2} \theta. \] (4.14)
The solution for these functions at leading order in the $k$ expansion is immediate,

\begin{align}
  r_0 &= \frac{2}{z} \frac{\xi}{\sqrt{1 - \xi^2}}, \\
  \eta_0 &= -\frac{\xi}{\sqrt{1 - \xi^2}}, \\
  t_0 &= -\frac{2}{\pi} \sin^{-1}(\xi),
\end{align}

(4.15)

where we have introduced the variable

\[ \xi = \frac{z}{2} e^{-U}. \]

(4.16)

It is easy to see that, at each order in $k^n$, we obtain from (4.10) two linear equations for $r_n$ and $\eta_n$, which are solved in terms of derivatives of lower $r_n'$, $\eta_n'$ with $n' < n$ and of the potential $U(x)$. The procedure can be easily automatized in a computer code to obtain the functions $r_n, \eta_n$ to any desired order.

It is now straightforward to write down a power series expansion for $R_-(x), R(x)$. If we denote the coefficient of $k^n$ in this expansion as $R_-(x), R(x)$, we find

\begin{align}
  R_{-n}(x) &= r_n(x) t_0(x) + B_n(x), \\
  R_n(x) &= r_n(x) (1 + t_0(x)) + B_n(x),
\end{align}

(4.17)

where

\[ B_n(\xi) = \sum_{m=0}^{n-1} r_m(\xi) t_{n-m}(\xi). \]

(4.18)

At leading order we find, for example,

\[ R_0 = r_0 (1 + t_0) = \frac{4}{\pi z} \sqrt{\frac{\xi}{1 - \xi^2}} \arccos(\xi). \]

(4.19)

Notice that both $r_0(x)$ and $\eta_0(x)$ have a branch cut, as functions of $\xi$, along $[1, \infty)$. However, $R_0(\xi)$ is holomorphic on the half-plane $\text{Re}(\xi) > -1$, as it can be easily seen from the representation

\[ R_0(\xi) = \frac{4}{\pi z} \sqrt{\frac{\xi}{\xi^2 - 1}} \ln \left( \xi + \sqrt{\xi^2 - 1} \right). \]

(4.20)

This is the semiclassical manifestation of the phenomenon we pointed out at the end of the previous section: the functions $\eta, R_+$ have a branch cut for

\[ z > 2 \exp(\min U), \]

(4.21)

due to the appearance of a Bose–Einstein condensate. However, this branch cut cancels in the function $R_0$, since this function corresponds to Fermi statistics and there is no possible source of singularities for positive $z$ in that case.

We can test the result for $R_0$ with a semiclassical calculation of (3.24) in the Fermi gas. At leading order in the WKB expansion, we have

\[ \left( \frac{1}{e^{H-\mu} + 1} \right) \approx \frac{1}{e^{H(x,p)-\mu} + 1}, \]

(4.22)
where
\[ H(x, p) = \log \left( 2 \cosh \frac{p}{2} \right) + U(x), \quad (4.23) \]
and
\[ R_0 = \frac{1}{\pi z} \int_{-\infty}^{\infty} dp \frac{1}{\xi^{-1} \cosh \left( \frac{\xi}{2} \right) + 1}, \quad (4.24) \]
which indeed can be explicitly computed and it agrees with (4.19).

### 4.2 Semiclassical TBA for ABJM theory

So far, we have considered the semiclassical TBA equations for a general potential \( U(x) \). In the case of ABJM theory, we have
\[
\exp \left\{ U \left( x + \frac{\pi i k}{2} \right) + U \left( x - \frac{\pi i k}{2} \right) \right\} = \frac{z^2}{4\xi^2} - 4 \sin^2 \left( \frac{k\pi}{4} \right), \quad (4.25)
\]
and it is useful to express everything in terms of \( \xi \). It is then straightforward to study the semiclassical TBA system at higher order. For example, at next-to-leading order we find the equations
\[
2\eta_0 \eta_1 = \frac{z^2}{4\xi^2} \left[ \frac{\pi^2}{4} \left( (r_0')^2 - r_0 r''_0 \right) + 2r_0 r_1 \right] - \frac{\pi^2 r_0^2}{4}, \quad (4.26)
\]
with the solution
\[
r_1(\xi) = \frac{\pi^2 \xi^3 \left( z^2 (2\xi^2 + 3) - 80\xi^2 \right)}{16 z^3 (1 - \xi^2)^{7/2}},
\]
\[
\eta_1(\xi) = -\frac{\pi^2 \xi \left( z^2 (4\xi^2 + 1) + 16\xi^2 (\xi^4 - 4\xi^2 - 2) \right)}{32 z^2 (1 - \xi^2)^{7/2}}. \quad (4.27)
\]
It is easy to see from the recursion that the functions
\[
(1 - \xi^2)^{3n+1/2} r_n(\xi), \quad (1 - \xi^2)^{3n+1/2} \eta_n(\xi), \quad (1 - \xi^2)^{3n-1/2} t_n(\xi) \quad (4.28)
\]
are polynomials in \( \xi \). Moreover, we have found that the explicit solution for \( R_n(\xi) \) has no singularity at \( \xi = 1 \) and it is holomorphic in the half-plane \( \text{Re}(\xi) > -1 \), as expected from general principles. This involves a non-trivial cancellation of poles at \( \xi = 1 \) in the sum of terms (4.17) giving \( R_n(\xi) \), and it can be regarded as a check of the procedure.

Although we have not studied in detail the structure of the functions \( r_n(\xi), \eta_n(\xi) \), some patterns can be easily observed. As we will see in a moment, it is useful to use, instead of the variable \( \xi \), the variable \( u \), defined by
\[
u = \frac{4\xi}{z}. \quad (4.29)
\]
The dominant terms in \( r_n, \eta_n \) for \( z \to \infty \) and \( u \) fixed seem to have the general structure
\[
r_n(u, z) \sim \frac{z^{6n-2} u^{6n-1} \pi^{2n}}{(1 - \frac{z^2 u^2}{16})^{3n+\frac{1}{2}}} (2n)!2^{14n-2}, \quad (4.30)
\]
\[
\eta_n(u, z) \sim -\frac{z^{6n-1} u^{6n+1} \pi^{2n}}{(1 - \frac{z^2 u^2}{16})^{3n+\frac{1}{2}}} (2n)!2^{14n}. \quad (4.30)
\]
This has been guessed by looking at the first orders in the explicit solution, but we have not found an analytic proof. As we will see in the next section, however, the above structure for $r_n$ implies the conjecture for the form of the one-membrane function $a_1(k)$ conjectured in [13].

4.3 The grand potential

Our ultimate goal is to compute the WKB expansion of $J^{M2}(\mu, k)$, therefore we have to compute the grand potential. This follows from (3.22). We will denote

$$J_z \equiv \frac{\partial J}{\partial z}. \quad (4.31)$$

It is convenient to split this quantity w.r.t. the parity of $z$,

$$\frac{\partial J}{\partial z} = J^+_z + J^-_z, \quad (4.32)$$

where

$$J^\pm_z = \frac{1}{4\pi k} \int_{-\infty}^{\infty} dx R_\pm(x). \quad (4.33)$$

After changing variables from $x$ to

$$u = \frac{4\xi}{z} = \text{sech} \left( \frac{x}{2} \right), \quad (4.34)$$

as defined in (4.29), we find

$$kJ^\pm_z = \frac{1}{\pi} \int_0^1 du \frac{1}{\sqrt{1-u^2}} R_\pm \left( zu/4 \right). \quad (4.35)$$

The WKB expansion of $R_\pm(x)$ leads to the WKB expansion of the grand potential considered in [8],

$$J_z(\mu, k) = \sum_{n \geq 0} J_{z,n}(\mu) k^{2n-1}. \quad (4.36)$$

The integrals appearing in the calculation of $J_{z,n}$ can be evaluated in terms of generalized hypergeometric functions. For $J^+_{z,n}$, the answer involves the functions $2F_1$, while for $J^-_{z,n}$ we also find the functions $3F_2$. The result can then be expanded at large $z$, and from this expansion, together with (2.20), one reads the small $k$ expansion of $a_\ell(k)$, $b_\ell(k)$ and $c_\ell(k)$, for any $\ell$. In the Appendix we collect some useful results on hypergeometric integration which are needed in these calculations, as well as the resulting small $k$ expansion of these coefficients for $\ell = 1, 2, 3$.

As a simple example of this procedure, let us calculate the leading order correction to the grand potential, $J_{z,0}$. We have

$$J^+_{z,0} = \frac{1}{\pi} \int_0^1 du \frac{1}{\sqrt{1-u^2} \sqrt{1-z^2 u^2/16}} = \frac{1}{2\pi} K \left( \frac{z^2}{16} \right), \quad (4.37)$$

and

$$J^-_{z,0} = \frac{1}{\pi^2} \int_0^1 du \sin^{-1} \left( \frac{uz}{4} \right) \frac{1}{\sqrt{1-u^2} \sqrt{1-z^2 u^2/16}} = -\frac{z}{4\pi^2} \ 3F_2 \left( 1, 1, 1, \frac{3}{2}, \frac{3}{2}, \frac{z^2}{16} \right), \quad (4.38)$$

- 14 -
where we used (A.5) and (A.8). The above results for $J_{z,0}^\pm$ agree with the calculation in [8]. We have also checked that the results for $J_{z,1}$ and $J_{z,2}$ agree with the results in [8].

It is interesting to notice that, in order to read off the coefficients $a_\ell(k)$, $b_\ell(k)$, it is enough to calculate $J_z^\pm$: by using the structural result (2.20) and assuming that the branch cut of the log is along the positive real axis, we find that

$$J_{z}^{M2,+} = \sum_{\ell \geq 1} \left[ (2\pi i \log z - \pi^2) \ell a_\ell(k) + \pi i (\ell b_\ell(k) - a_\ell(k)) \right] z^{-2\ell - 1}. \quad (4.39)$$

It is easy to verify that the leading order term in the expansion of (4.39) at large $z$ comes from integrating the leading term of $r_n(u,z)$ in (4.30). By using (A.3) we find

$$\int_0^1 \frac{du}{\sqrt{1-u^2}} \frac{u^{6n-2}}{(1 - \frac{z^2u^2}{16})^{3n + \frac{1}{2}}} = \frac{1}{2} \frac{\Gamma(3n - 1/2)\Gamma(1/2)}{\Gamma(3n)} 2F_1 \left( 3n - \frac{1}{2}, 3n + \frac{1}{2}, 3n; \frac{z^2}{16} \right). \quad (4.40)$$

At large $z$ we have the logarithmic behavior,

$$2F_1 \left( 3n - \frac{1}{2}, 3n + \frac{1}{2}, 3n; \frac{z^2}{16} \right) = i(-1)^n 2^{12n+2} \frac{\Gamma(3n)}{\Gamma(-1/2)\Gamma(3n - 1/2)} z^{-6n-1} \log(-z^2) + \cdots \quad (4.41)$$

If we put everything together and we compare the result with (4.39), we find that the ansatz (4.30) leads to the following perturbative expansion for the coefficient $a_1(k)$:

$$a_1(k) = -\frac{4}{\pi^2 k} \sum_{n \geq 0} \frac{(-1)^n}{(2n)!} \left( \frac{\pi k}{2} \right)^{2n} = -\frac{4}{\pi^2 k} \cos \left( \frac{\pi k}{2} \right). \quad (4.42)$$

This is precisely the result conjectured in [13] for this coefficient.

5. Predictions for membrane instantons

5.1 The HMO cancellation mechanism

The Gopakumar–Vafa representation (2.23) of $J^{WS}(\mu, k)$ shows that it has double poles at all integer values of $k$. Since the original matrix integral is not singular for any value of $k$, there must be some way of canceling these divergences. The proposal of HMO in [13] is that, in the total non-perturbative grand potential, and order by order in $e^{-\mu}$, there should be no divergences. This means that, in the sum (2.25) over all bound states $(\ell, m)$ which contribute to a given order in $e^{-\mu}$, singularities must cancel. This is the HMO cancellation mechanism.

In general, since the contribution of generic bound states is not known, it is difficult to verify this cancellation mechanism in detail. However, some non-trivial information can be obtained by looking at low orders in the expansion. The case studied in detail in [13] is the case of one-membrane instanton, i.e. of contributions to the grand potential which go like $e^{-2\mu}$. Let us assume that $k$ is an even, positive integer. Then, the only contributions to the term of order $e^{-2\mu}$ come from

$$(\ell, m) = (1, 0) \quad \text{and} \quad (\ell, m) = (0, k/2), \quad (5.1)$$

i.e. they involve only worldsheet instantons and membrane instantons. In this case, the HMO cancellation mechanism can be studied in detail. The worldsheet instanton poles at $k = 2m$ are of the form

$$d_m(k)e^{-4\mu/k} = (-1)^{m-1} \left[ \frac{4m}{\pi^2 (k - 2m)^2} + \frac{4(\mu + 1)}{\pi^2 (k - 2m)} + \frac{2\mu^2 + 2\mu + 1}{m \pi^2} + w(m) \right] e^{-2\mu} + \cdots \quad (5.2)$$
where \( w^{(m)} \) can be calculated from the Gopakumar–Vafa invariants of local \( \mathbb{P}^1 \times \mathbb{P}^1 \). According to the HMO mechanism, the poles have to be cancelled by similar poles in the membrane instanton contribution. Using these constraints, as well as the first three terms of the WKB expansion calculated in [8], HMO were able to propose an ansatz for the values of the coefficients \( a_1(k) \), \( b_1(k) \), \( c_1(k) \) appearing in (2.20),

\[
\begin{align*}
a_1(k) &= -\frac{4}{\pi^2 k} \cos \left( \frac{\pi k}{2} \right), \\
b_1(k) &= \frac{2}{\pi} \cos^2 \left( \frac{\pi k}{2} \right) \csc \left( \frac{\pi k}{2} \right), \\
c_1(k) &= \left[ -\frac{2}{3k} + \frac{5k}{12} + \frac{k}{2} \csc^2 \left( \frac{\pi k}{2} \right) + \frac{1}{\pi} \cot \left( \frac{\pi k}{2} \right) \right] \cos \left( \frac{\pi k}{2} \right). \tag{5.3}
\end{align*}
\]

One can check that these expressions have the right singularity structure to cancel the divergences in (5.2): near \( k = 2m \), one has

\[
\begin{align*}
b_1(k) &= -\frac{4(-1)^{m-1}}{\pi^2 (k-2m)} + \mathcal{O}(k-2m), \\
c_1(k) &= (-1)^{m-1} \left[ -\frac{4m}{\pi^2 (k-2m)^2} - \frac{4}{\pi^2 (k-2m)} + \frac{1}{3m} - \frac{2m}{3} \right] + \mathcal{O}(k-2m). \tag{5.4}
\end{align*}
\]

Once the membrane instanton and the worldsheet instanton contributions are added for even integer \( k = 2m \), the poles cancel and we find a finite contribution

\[
(-1)^{m-1} \left[ \frac{4\mu^2 + 2\mu + 1}{m\pi^2} + w^{(m)} + \frac{1}{3m} - \frac{2m}{3} \right] e^{-2\mu}. \tag{5.5}
\]

This result reproduces a numerical calculation of the terms of order \( e^{-2\mu} \) in \( J(\mu, k) \) for \( k = 2, 4, 6 \), and confirms the ansatz (5.3).

The conjecture (5.3) shows that there is a simple, hidden structure in the contribution of membrane instantons to the grand potential. Thanks to our semiclassical TBA expansion, we can now compute the series expansion of these coefficients to higher order in \( k \), as listed in the Appendix. Our results confirm the expressions in (5.3). In the case of \( a_1(k) \), as we showed in the last section, this follows from the ansatz (4.30).

5.2 The second membrane instanton

In order to provide further tests of the HMO cancellation mechanism with the current techniques, we have to look at exponentially small terms which only receive contributions from states of the form \((\ell, 0)\) and of the form \((0, m)\). As we saw in the previous section, following [13], this is what happens for the terms \( \sim e^{-2\mu} \) when \( k \) is even. Similarly, when \( k \) is odd, the exponentially small term \( e^{-4\mu} \) has contributions only from the states

\[
(\ell, m) = (2, 0) \quad \text{and} \quad (\ell, m) = (0, k), \tag{5.6}
\]

and it involves the two-membrane instanton. This case was only partially analyzed in [13]. They noted that the contribution of the worldsheet instanton for \( k = m \) odd is singular, and it has the behavior

\[
d_m(k) e^{-4\mu / k} = \left[ \frac{m}{4\pi^2 (k-m)^2} + \frac{2\mu + 1}{2\pi^2 (k-m)} + \frac{2\mu^2 + \mu + 1/4}{m\pi^2} + w^{(m)} \right] e^{-4\mu} + \cdots \tag{5.7}
\]
The $v^{(m)}$ can be calculated from the Gopakumar–Vafa invariants and they are given, for $m = 1, 3$, by

$$v^{(1)} = \frac{1}{3}, \quad v^{(3)} = \frac{37}{9}. \quad (5.8)$$

The poles appearing in (5.7) were also conjectured in [13] to be cancelled by similar poles in the membrane instanton contribution. Using this cancellation mechanism, as well as our data for the WKB expansion, we propose the following exact expressions for the coefficients of the second membrane instanton:

$$a_2(k) = -\frac{1}{\pi^2 k} (8 + 10 \cos (\pi k)), \quad b_2(k) = \frac{4}{\pi^2 k} (1 + \cos (\pi k)) + \frac{1}{2\pi} \csc (\pi k) (17 + 24 \cos (\pi k) + 9 \cos (2\pi k)),$$

$$c_2(k) = \frac{4}{3k} - \frac{5 \cos(\pi k)}{3k} + k \left(\frac{49}{24} \cos(\pi k) - \frac{7}{6}\right) + \frac{\cot(\pi k)}{\pi} + 5k \csc^2(\pi k)$$

$$+ \cos(\pi k) \left(\frac{5 \cot(\pi k)}{4\pi} + \frac{21}{4} k \csc^2(\pi k)\right). \quad (5.9)$$

It is easy to check that the above expressions reproduce the WKB expansions presented in the Appendix. They also cancel the poles in (5.7). Indeed, one finds that, near $k = m$ odd,

$$b_2(k) = -\frac{1}{\pi^2 (k - m)} + O(k - m), \quad c_2(k) = -\frac{m}{4\pi^2 (k - m)^2} - \frac{1}{2\pi^2 (k - m)} + \frac{1}{3m} - \frac{2m}{3} + O(k - m). \quad (5.10)$$

The finite piece, after adding both contributions, is

$$\left[\frac{4\mu^2 + \mu + 1/4}{m\pi^2} + v^{(m)} + \frac{1}{3m} - \frac{2m}{3}\right] e^{-4\mu}. \quad (5.11)$$

For $m = 1, 3$ we find

$$\left[\frac{4\mu^2 + \mu + 1/4}{\pi^2}\right] e^{-4\mu} \quad \text{and} \quad \left[\frac{4\mu^2 + \mu + 1/4}{3\pi^2} + \frac{20}{9}\right] e^{-4\mu}, \quad (5.12)$$

respectively, which correctly reproduce the terms of order $e^{-4\mu}$ in the expression for $J(\mu, k)$ with $k = 1, 3$ presented in [13]. Notice that, although the coefficient $c_2(k)$ is very different from the coefficient $c_1(k)$, their finite parts near the relevant poles have the same form.

5.3 Higher membrane instantons

The HMO cancellation mechanism can be studied in detail for one-membrane instantons and two-membrane instantons. The contribution of higher membrane instantons mixes with the contribution of generic bound states and it is difficult to extract information about the former without a more detailed knowledge of the latter. In particular, we don’t have enough information about the singularity structure of the membrane coefficients as a function of $k$, and so far the only available information is contained in the WKB expansions.

In the case of the $a_\ell(k)$ coefficients, however, it is easy to fit these WKB expansions to a sum of trigonometric functions. We find that, for odd (even) instanton number, the $a_\ell(k)$ are given
by a sum of cosines whose argument is an odd (respectively, even) integer times $\pi k/2$. Their explicit expressions, up to instanton number $\ell = 5$, are

$$a_3(k) = -\frac{1}{\pi^2 k} \left( 88 \cos \left( \frac{\pi k}{2} \right) + \frac{124}{3} \cos \left( \frac{3\pi k}{2} \right) + 4 \cos \left( \frac{5\pi k}{2} \right) \right),$$

$$a_4(k) = -\frac{1}{\pi^2 k} \left( 364 + 560 \cos(\pi k) + 245 \cos(2\pi k) + 48 \cos(3\pi k) + 8 \cos(4\pi k) \right),$$

$$a_5(k) = -\frac{1}{\pi^2 k} \left( 6080 \cos \left( \frac{\pi k}{2} \right) + 4100 \cos \left( \frac{3\pi k}{2} \right) + \frac{9104}{5} \cos \left( \frac{5\pi k}{2} \right) + 536 \cos \left( \frac{7\pi k}{2} \right) \right) + 136 \cos \left( \frac{9\pi k}{2} \right) + 24 \cos \left( \frac{11\pi k}{2} \right) + 4 \cos \left( \frac{13\pi k}{2} \right).$$

(5.13)

Finding a natural expression which fits our data for the expansion of the $b_\ell(k)$, $c_\ell(k)$ with $\ell \geq 3$ is more challenging, and further information (including more data points in the WKB expansion) is probably needed.

6. Conclusions and open problems

In this paper we have developed a semiclassical approach to the TBA equations of [16]. When applied to ABJM theory, this reproduces in a more efficient way the WKB approach developed in [8]. Our results confirm the conjecture made in [13] for the one-instanton contribution. We have also proposed analytic expressions at finite $k$ for the membrane instantons at order two, which are in perfect accord with the HMO cancellation mechanism. In addition, we have obtained some conjectural results at higher instanton number for the coefficients $a_\ell(k)$. It is obvious from the results in this paper and its predecessors that there is a new and rich story concerning non-perturbative corrections in the M-theory dual to ABJM theory. Corrections to the partition function coming from intrinsic M-theory objects can be now computed in detail, and sometimes we can even guess their exact expression for arbitrary $k$. Moreover, the HMO cancellation mechanism shows very clearly that the genus expansion of type IIA string theory, although it can be resummed with a Gopakumar–Vafa representation, is essentially incomplete at strong coupling: only when the contribution of membrane instantons is taken into account do we find a finite answer for integer values of $k$ (precisely the values for which we believe that the theory is defined non-perturbatively).

On the other hand, it is fair to say that we are only in a preliminary exploratory period of all these non-perturbative phenomena. Although the ABJM matrix integral and the TBA system contain detailed information about these instanton effects, it is not obvious how to extract it. The techniques proposed in [18, 17, 13] work for finite $k$, but they are based on a small $z$ expansion which corresponds to small $N$. In this paper we have proposed a WKB expansion which is only valid for small $k$, but provides analytic large $N$ results. Clearly, both approaches are insufficient, and they should be combined: what we need is the large $\mu$ expansion of the grand potential at finite $k$. This will very likely require a clever analysis of the TBA system which circumvents the problem with the Bose–Einstein condensate singularity pointed out in section 3.

A method leading to results at large $N$ and finite $k$ would also determine the functions $f_{\ell,m}(\mu, k)$ appearing in (2.25) for general bound states (maybe order by order in $e^{-\mu}$). In fact, it seems to us that the most important open problem is finding an approach to calculate the contribution of bound states, about which nothing is known. Even a WKB approach to the
problem would be useful. It was suggested in [8] that worldsheet instantons appear as quantum-mechanical instantons in the Fermi gas, and bound states might then appear as exponentially small corrections in $\mu$ to a general quantum-mechanical instanton amplitude. This approach might give a first handle on this problem. Another possibility is to look for trans-series solutions of the difference equations (4.1), similar to what was done in [27] for the difference equations arising in matrix models.

In addition to these general issues, there are a multitude of more concrete questions that come to our mind when we look at the conjectural expressions for the membrane instantons. Their contribution involves trigonometric functions whose coefficients are rational numbers and, very often, integer numbers. Is there some sort of Gopakumar–Vafa formula for them? Is there some duality between membrane instantons and worldsheet instantons? Can we compute some of these membrane instanton corrections directly in M-theory, by adapting for example the framework of [26]?

Finally, it was pointed out in [8, 28] that any non-perturbative result in the context of the Fermi gas of ABJM theory can be interpreted as a non-perturbative result for topological strings in local $\mathbb{P}^1 \times \mathbb{P}^1$. In particular, the grand potential studied in this paper can be interpreted as the topological string free energy at large radius, and the membrane instantons computed here lead to corrections to this free energy of order $O(e^{-1/g_s})$. Is there an interpretation for these effects directly in topological string theory? Are there “topological” membranes which lead to this sort of effect?

As Bertolt Brecht would put it, “so many stories, so many questions.”

Acknowledgements

We would like to thank Albrecht Klemm for useful conversations. This work is supported in part by the Fonds National Suisse, subsidies 200020-126817 and 200020-137523.

A. Hypergeometric integrals

In the calculation of the grand potential one finds two types of integrals. For $J^+_{z}$, we have integrals of the form

$$
\int_0^1 \frac{du}{\sqrt{1-u^2}} \frac{u^m}{\left(1 - \frac{z^2u^2}{16}\right)^{3s+\frac{1}{2}}}.
$$

(A.1)

After the change of variable

$$
u^2 = t, \quad \frac{z^2}{16} = Z$$

(A.2)

they can be written as

$$
\frac{1}{2} \int_0^1 dt (1-t)^{-1/2} t^{-m+1/2} (1-Zt)^{-3s-1/2} = \frac{\Gamma\left(m+\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right)}{2 \Gamma\left(m+\frac{3}{2}\right)} \, _2F_1\left(3s + \frac{1}{2}, \frac{m+1}{2}; 1 + \frac{m}{2}; Z\right).
$$

(A.3)

For $J^-_{z}$, we have to do integrals of the form

$$
\int_0^1 \frac{du}{\sqrt{1-u^2}} \frac{u^{m-1} \sin^{-1}\left(\frac{u}{\sqrt{16}}\right)}{\left(1 - \frac{z^2u^2}{16}\right)^{3s+\frac{1}{2}}}.
$$

(A.4)
Using that
\[
\frac{1}{x} \sin^{-1}(x) = 2F_1 \left( 1, 1; \frac{3}{2}; x^2 \right)
\]  
(A.5)
we can write the above integral as
\[
z \frac{1}{4} \int_0^1 \frac{du}{\sqrt{1-u^2}} u^m \left( 1 - \frac{z^2 u^2}{16} \right)^{-3s} 2F_1 \left( 1, 1; \frac{3}{2}; \frac{z^2 u^2}{16} \right).
\]  
(A.6)
To calculate this integral, we need two ingredients. First of all, we use the identity 132 in p. 436 of [30], with \( n = 1 \) (we correct a minor misprint in the statement of the identity):
\[
(1-x)^{-3s} 2F_1 \left( 1, 1, \frac{3}{2}; x \right) = (1)^{3s} \left[ 2F_1 \left( 1, 3s + 1, \frac{3}{2}; x \right) - \frac{1}{6s} \sum_{k=0}^{3s-1} (1/3s-k)(1-x)^{-k-1} \right].
\]  
(A.7)
After using this identity, and the change of variables (A.2), the integral (A.6) can be reduced to an integral of the form
\[
\int_0^1 dt t^{a_{p+1}-1}(1-t)^{b_{q+1}-a_{p+1}-1} \, _pF_q(a_1, \ldots, a_p; b_1, \ldots, b_q; Zt) = \Gamma(a_{p+1})\Gamma(b_{q+1}) \Gamma(b_{q+1})
\]  
(A.8)
with \( p = 2, q = 1 \), as well as to a sum of integrals of the form (A.1) with half-integer \( s \).

The asymptotic behavior as \( z \gg 1 \) of the hypergeometric functions involved in these expressions can be easily found by using for example the Barnes representation. For the function \( 2F_1 \), it is given by (see for example [29], eq. (9.7.7)):
\[
2F_1(\alpha, \alpha + n; \gamma; z) = \frac{\Gamma(\gamma)(-z)^{-\alpha}}{\Gamma(\gamma - \alpha)\Gamma(\alpha + n)} \sum_{i=0}^{n-1} \frac{(n - i - 1)! (1 - \gamma + \alpha)^i}{i!} (-z)^{-i} 
+ \frac{\Gamma(\gamma)(-z)^{-\alpha}}{\Gamma(\gamma - \alpha - n)\Gamma(\alpha)} \sum_{i=0}^{\infty} \frac{(\alpha + n)_i(1 - \gamma + \alpha + n)^i}{i!(n + i)!} \left[ \psi(i + 1) + \psi(n + i + 1) - \psi(\alpha + n + i) - \psi(\gamma - \alpha - n - i) + \log(-z) \right] z^{-i-n}.
\]  
(A.9)

B. WKB expansions

Here we list the WKB expansion of the membrane instanton coefficients \( a_\ell(k) \), \( b_\ell(k) \), \( c_\ell(k) \) for \( \ell = 1, 2, 3 \).

\[
ka_1(k) = -\frac{4}{\pi^2} + \frac{k^2}{2} - \frac{\pi^2 k^4}{96} + \frac{\pi^4 k^6}{11520} - \frac{\pi^6 k^8}{2580480} + \frac{\pi^8 k^{10}}{928972800} - \frac{\pi^{10} k^{12}}{490497638400} + \frac{\pi^{12} k^{14}}{357082280755200} - \frac{\pi^{14} k^{16}}{342798989524992000} + O(k^{18}),
\]  
(B.1)
\[ k a_2(k) = -\frac{18}{\pi^2} + 5k^2 - \frac{5\pi^2 k^4}{12} + \frac{\pi^4 k^6}{72} - \frac{\pi^6 k^8}{4032} + \frac{\pi^8 k^{10}}{362880} - \frac{\pi^{10} k^{12}}{47900160} + \frac{\pi^{12} k^{14}}{8717829120} \]  \tag{B.2} 

\[ k a_3(k) = -\frac{400}{3\pi^2} + 70k^2 - \frac{371\pi^2 k^4}{24} + \frac{1159\pi^4 k^6}{576} - \frac{16373\pi^6 k^8}{92160} + \frac{74113\pi^8 k^{10}}{6635520} - \frac{62408051\pi^{10} k^{12}}{122624409600} + \frac{43949569\pi^{12} k^{14}}{2550587719680} - \frac{546545381\pi^{14} k^{16}}{12242821054464000} + \mathcal{O}(k^{18}) \]  \tag{B.3} 

\[ k b_1(k) = \frac{4}{\pi^2} - \frac{5k^2}{6} + \frac{67\pi^2 k^4}{1440} - \frac{19\pi^4 k^6}{48384} + \frac{247\pi^6 k^8}{38707200} + \frac{89\pi^8 k^{10}}{1226244096} + \frac{1430857\pi^{10} k^{12}}{66952927641600} + \frac{1637\pi^{12} k^{14}}{30607052636160} + \frac{118522319\pi^{14} k^{16}}{8741374238872960000} + \mathcal{O}(k^{18}) \]  \tag{B.4} 

\[ k b_2(k) = \frac{33}{\pi^2} - \frac{77k^2}{6} + \frac{119\pi^2 k^4}{72} - \frac{1199\pi^4 k^6}{15120} + \frac{251\pi^6 k^8}{12096} - \frac{607\pi^8 k^{10}}{119750400} + \frac{311813\pi^{10} k^{12}}{130767436800} + \frac{56863\pi^{12} k^{14}}{261534873600} + \frac{4740503\pi^{14} k^{16}}{213412456857600} + \mathcal{O}(k^{18}) \]  \tag{B.5} 

\[ k b_3(k) = \frac{2560}{9\pi^2} - 188k^2 + \frac{4769\pi^2 k^4}{90} - \frac{258689\pi^4 k^6}{30240} + \frac{159091\pi^6 k^8}{172800} - \frac{262188523\pi^8 k^{10}}{3832012800} + \frac{16019071489\pi^{10} k^{12}}{4184557977600} - \frac{15646173899\pi^{12} k^{14}}{133905855283200} + \frac{1471885987607\pi^{14} k^{16}}{1365839723888640000} + \mathcal{O}(k^{18}) \]  \tag{B.6} 

\[ k c_1(k) = \left( \frac{4}{\pi^2} - \frac{2}{3} \right) + \frac{1}{12} (\pi^2 - 1) k^2 + \left( -\frac{13\pi^2}{360} - \frac{\pi^4}{576} \right) k^4 + \left( \frac{55\pi^4}{96768} + \frac{\pi^6}{69120} \right) k^6 + \left( -\frac{671\pi^6}{38707200} - \frac{\pi^8}{15482880} \right) k^8 + \left( -\frac{\pi^{10}}{5573836800} - \frac{3659\pi^8}{12262440960} \right) k^{10} \]  \tag{B.7} 

\[ k c_2(k) = \left( \frac{25}{2\pi^2} - 3 \right) + \left( \frac{7}{24} + \frac{5\pi^2}{6} \right) k^2 + \left( -\frac{47\pi^2}{60} - \frac{5\pi^4}{72} \right) k^4 + \left( \frac{641\pi^4}{8640} + \frac{\pi^6}{432} \right) k^6 + \left( -\frac{3443\pi^6}{1209600} - \frac{\pi^8}{24192} \right) k^8 + \left( \frac{29\pi^8}{4561920} + \frac{\pi^{10}}{2177280} \right) k^{10} \]  \tag{B.8} 

\[ k c_3(k) = \left( \frac{1642}{27\pi^2} - \frac{200}{9} \right) + \left( \frac{437}{36} + \frac{35\pi^2}{3} \right) k^2 + \left( -\frac{1098\pi^2}{576} - \frac{371\pi^4}{144} \right) k^4 + \left( \frac{182071\pi^4}{32256} + \frac{1159\pi^6}{3456} \right) k^6 + \left( \frac{3294323\pi^6}{3686400} - \frac{16373\pi^8}{552960} \right) k^8 + \left( \frac{76955074\pi^8}{8758886400} + \frac{74113\pi^{10}}{39813120} \right) k^{10} + \left( -\frac{43661410369\pi^{10}}{7084965888000} - \frac{62408051\pi^{12}}{735746457600} \right) k^{12} + \mathcal{O}(k^{14}) \]  \tag{B.9}
References

[1] 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]].

[2] D. L. Jafferis, “The Exact Superconformal R-Symmetry Extremizes Z,” JHEP 1205, 159 (2012) [arXiv:1012.3210 [hep-th]].

[3] N. Hama, K. Hosomichi, S. Lee, “Notes on SUSY Gauge Theories on Three-Sphere,” JHEP 1103, 127 (2011). [arXiv:1012.3512 [hep-th]].

[4] N. Drukker, M. Mariño and P. Putrov, “From weak to strong coupling in ABJM theory,” Commun. Math. Phys. 306, 511 (2011) [arXiv:1007.3837 [hep-th]].

[5] O. Aharony, O. Bergman, D. L. Jafferis and J. Maldacena, “N=6 superconformal Chern-Simons-matter theories, M2-branes and their gravity duals,” JHEP 0810, 091 (2008) [arXiv:0806.1218 [hep-th]].

[6] M. Mariño, “Lectures on localization and matrix models in supersymmetric Chern-Simons-matter theories,” J. Phys. A 44, 463001 (2011) [arXiv:1104.0783 [hep-th]].

[7] S. Bhattacharyya, A. Grassi, M. Mariño and A. Sen, “A One-Loop Test of Quantum Supergravity,” arXiv:1210.6057 [hep-th].

[8] M. Mariño and P. Putrov, “ABJM theory as a Fermi gas,” J. Stat. Mech. 1203, P03001 (2012) [arXiv:1110.4066 [hep-th]].

[9] H. Fuji, S. Hirano and S. Moriyama, “Summing Up All Genus Free Energy of ABJM Matrix Model,” JHEP 1108, 001 (2011) [arXiv:1106.4631 [hep-th]].

[10] K. Becker, M. Becker and A. Strominger, “Five-branes, membranes and nonperturbative string theory,” Nucl. Phys. B 456, 130 (1995) [hep-th/9507158].

[11] A. Cagnazzo, D. Sorokin and L. Wulff, “String instanton in AdS4 × CP3,” JHEP 1005, 009 (2010) [arXiv:0911.5228 [hep-th]].

[12] N. Drukker, M. Mariño and P. Putrov, “Nonperturbative aspects of ABJM theory,” JHEP 1111, 141 (2011) [arXiv:1103.4844 [hep-th]].

[13] Y. Hatsuda, S. Moriyama and K. Okuyama, “Instanton Effects in ABJM Theory from Fermi Gas Approach,” arXiv:1211.1251 [hep-th].

[14] S. Cecotti, P. Fendley, K. A. Intriligator and C. Vafa, “A New supersymmetric index,” Nucl. Phys. B 386, 405 (1992) [hep-th/9204102].

[15] P. Fendley and H. Saleur, “N=2 supersymmetry, Painlevé III and exact scaling functions in 2-D polymers,” Nucl. Phys. B 388, 609 (1992) [hep-th/9204094].

[16] A. B. Zamolodchikov, “Painlevé III and 2-d polymers,” Nucl. Phys. B 432, 427 (1994) [hep-th/9409108].

[17] Y. Hatsuda, S. Moriyama and K. Okuyama, “Exact Results on the ABJM Fermi Gas,” JHEP 1210, 020 (2012) [arXiv:1207.4283 [hep-th]].

[18] P. Putrov and M. Yamazaki, “Exact ABJM Partition Function from TBA,” Mod. Phys. Lett. A 27, 1250200 (2012) [arXiv:1207.5066 [hep-th]].

[19] A. Kapustin, B. Willett and I. Yaakov, “Nonperturbative Tests of Three-Dimensional Dualities,” JHEP 1010, 013 (2010) [arXiv:1003.5694 [hep-th]].

[20] K. Okuyama, “A Note on the Partition Function of ABJM theory on S3,” Prog. Theor. Phys. 127, 229 (2012) [arXiv:1110.3555 [hep-th]].
[21] M. Hanada, M. Honda, Y. Honma, J. Nishimura, S. Shiba and Y. Yoshida, “Numerical studies of the ABJM theory for arbitrary N at arbitrary coupling constant,” JHEP 1205, 121 (2012) [arXiv:1202.5300 [hep-th]].

[22] R. Gopakumar and C. Vafa, “M theory and topological strings. 2.,” hep-th/9812127.

[23] M. Mariño and P. Putrov, “Exact Results in ABJM Theory from Topological Strings,” JHEP 1006, 011 (2010) [arXiv:0912.3074 [hep-th]].

[24] C. A. Tracy, H. Widom, “Proofs of two conjectures related to the thermodynamic Bethe ansatz,” Commun. Math. Phys. 179, 667-680 (1996). [solv-int/9509003].

[25] P. Fendley, “Airy functions in the thermodynamic Bethe ansatz,” Lett. Math. Phys. 49, 229 (1999) [hep-th/9906114].

[26] J. A. Harvey and G. W. Moore, “Superpotentials and membrane instantons,” hep-th/9907026.

[27] M. Mariño, “Nonperturbative effects and nonperturbative definitions in matrix models and topological strings,” JHEP 0812, 114 (2008) [arXiv:0805.3033 [hep-th]].

[28] M. Mariño, “Topological strings at strong string coupling,” talk given in november 2011 at the Banff workshop New recursion formulae and integrability for Calabi-Yau spaces, http://www.birs.ca/events/2011/5-day-workshops/11w5114/videos.

[29] N. N. Lebedev, Special functions and their applications, Dover, 1962.

[30] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, Integrals and Series, vol. 3, CRC, 1990.