Probing non-perturbative effects in M-theory

Yasuyuki Hatsuda[a] **and Kazumi Okuyama**[b]

[a] *DESY Theory Group, DESY Hamburg,*
Notkestrasse 85, D-22603 Hamburg, Germany

[b] *Department of Physics,*
Shinshu University, Matsumoto 390-8621, Japan

E-mail: yasuyuki.hatsuda@desy.de, kazumi@azusa.shinshu-u.ac.jp

ABSTRACT: The AdS/CFT correspondence enables us to probe M-theory on various backgrounds from the corresponding dual gauge theories. Here we investigate in detail a three-dimensional $\mathrm{U}(N)$ $\mathcal{N} = 4$ super Yang-Mills theory coupled to one adjoint hypermultiplet and N_f fundamental hypermultiplets, which is large N dual to M-theory on $AdS_4 \times S^7/\mathbb{Z}_{N_f}$. Using the localization and the Fermi-gas formulation, we explore non-perturbative corrections to the partition function. As in the ABJM theory, we find that there exists a non-trivial pole cancellation mechanism, which guarantees the theory to be well-defined, between worldsheet instantons and membrane instantons for all rational (in particular, physical or integral) values of N_f.

Contents

1 Introduction

In this paper, we study non-perturbative aspects of M-theory via the AdS/CFT correspondence [1]. Our analysis here is based on the belief that the AdS/CFT correspondence (or more generally, the gauge/gravity duality) is exactly true even at quantum level. This means that gauge theories, if they have gravity duals, provide us a "non-perturbative definition" of their dual string theories/M-theory on the corresponding backgrounds. Recent remarkable developments on exact understandings of gauge theories enable us to probe the non-perturbative effects in the dual string theories/M-theory, quantitatively.

We are interested in the low energy effective theories on multiple M2-branes, which have a dual M-theory description on some AdS_4 background. The most well-known example is a $U(N)_k \times U(N)_{-k}$ $\mathcal{N} = 6$ supersymmetric Chern-Simons-matter theory, known as the ABJM theory [2]. The ABJM theory describes the low energy effective theory on the N M2-branes probing a $\mathbb{C}^4/\mathbb{Z}_k$ singularity, and it is dual to M-theory on $AdS_4 \times S^7/\mathbb{Z}_k$ in the large N limit. In this paper, we pick up another example: a three-dimensional $U(N)$ $\mathcal{N} = 4$ super Yang-Mills theory coupled to one adjoint hypermultiplet and N_f fundamental hypermultiplets. This theory also describes the theory on N M2-branes probing a $\mathbb{C}^2 \times \mathbb{C}^2/\mathbb{Z}_{N_f}$ singularity [3–7]. From the Type IIA viewpoint, this is the worldvolume theory on N D2-branes in the presence of N_f D6-branes. This theory is dual to M-theory on $AdS_4 \times S^7/\mathbb{Z}_{N_f}$ in the limit $N \to \infty$ with N_f fixed. Note that the $\mathbb{Z}_{N_f}$ quotient on S^7 acts differently from the ABJM case. We can probe M-theory from these theories via the AdS/CFT correspondence. We would like to find universal (background independent) properties in M-theory through various examples.

As shown in [8–10], in many supersymmetric gauge theories on S^3, infinite dimensional path integrals for the partition function and vacuum expectation values (VEVs) of BPS Wilson loops reduce to finite dimensional matrix integrals by using the localization technique. Due to this drastic simplification, one can, in principle, evaluate the partition function and the Wilson loop VEVs beyond the perturbation theory. However, it is still non-trivial to extract their large N behaviors from the matrix integrals. The traditional matrix model technique is very helpful in the analysis in the 't Hooft limit. In the ABJM matrix model, a systematic analysis in the 't Hooft limit was done in [11, 12]. However it is not easy to access the M-theory regime in this way (see [13]). It is desirable to find more efficient ways to understand the M-theory regime systematically.

Recently, Mariño and Putrov proposed a very interesting formulation, known as the Fermi-gas approach, to analyze matrix models for a wide class of 3d Chern-Simons-matter theories [14]. This Fermi-gas approach was successfully applied to the ABJM theory and revealed a very detailed structure of the non-perturbative effects in M-theory. It turned out that the existence of two types of instantons, i.e. worldsheet instantons and membrane instantons, is crucial for the non-perturbatively complete definition of the theory. In particular, the worldsheet instanton correction diverges at every physical value of the coupling, and that divergence is precisely canceled by the similar, but opposite sign, divergence of the membrane instanton correction [15]. This pole cancellation mechanism is conceptually very important, since this mechanism guarantees that we can go smoothly from the weak coupling (Type IIA) regime to the strong coupling (M-theory) regime. More practically, this mechanism gives strong constraint on the possible form of membrane instantons. For the ABJM case, we can actually find the analytic form of a first few membrane instanton coefficients using this pole cancellation condition, together with some other input from the semi-classical expansion of Fermi-gas [15–17]. Based on these analytic results, it was finally found in [18, 19] that the membrane instantons in the ABJM theory are completely determined by the refined topological string on local $\mathbb{P}^1 \times \mathbb{P}^1$, in the Nekrasov-Shatashvili limit [20], while the worldsheet instantons are given by the standard topological string on the same manifold [12, 15]. In addition, there are bound states of membrane instantons

and worldsheet instantons, whose contributions are finally absorbed into the worldsheet instanton corrections by the effective shift of chemical potential of the Fermi-gas system [17]. In the ABJ theory [21], the similar structure was also found [22–24] based on the results [25, 26].

In the present paper, we will study the S^3 partition function $Z(N_f, N)$ of the U(N) $\mathcal{N} = 4$ super Yang-Mills theory coupled to one adjoint hypermultiplet and N_f fundamental hypermultiplets. Using the localization technique, the computation of $Z(N_f, N)$ boils down to a matrix integral, which was named as the N_f matrix model in [27]. It is known that the grand partition function of the N_f matrix model can be recast as a Fermi-gas system, and some of its properties were studied in [27, 28]. This N_f matrix model is an interesting first step beyond ABJ(M) theory to study the non-perturbative effects in M-theory. However, it turned out that it is not straightforward to apply the strategy in the previous paragraph, which was successful in the ABJM case [15], to the N_f matrix model:

1. Find the worldsheet instanton coefficients and their pole structure.

2. Determine the analytic form of the membrane instanton correction by combining the small N_f expansion and the pole cancellation condition.

At the step 1, in the case of ABJM theory, the analytic form of the worldsheet instanton correction is available thanks to the relation to the topological string on local $\mathbb{P}^1 \times \mathbb{P}^1$. On the other hand, the N_f matrix model does not seem to have a direct connection to the topological string theory, and hence the analytic form of the worldsheet instanton correction is not known, except for the genus zero part [27]. Currently, there is no systematic way to compute the worldsheet instanton corrections as analytic functions of N_f. To overcome this problem, we first compute the exact values of the partition function $Z(N_f, N)$ for various integral values of N_f up to some high N, and then guess the worldsheet instanton coefficients as functions of N_f using the exact data of $Z(N_f, N)$. In this way, we indeed find the analytic forms of worldsheet instanton coefficients up to three-instanton (3.24). We note that our conjecture gives an all-genus prediction in each instanton sector when taking the 't Hooft limit. Our conjecture passes many non-trivial checks.

The small N_f expansion at step 2 is also difficult to be carried out, since the density matrix (or Hamiltonian) of the Fermi-gas explicitly depends on N_f [27, 28]. Nevertheless, we successfully find the first few terms of the small N_f expansion of the grand potential by analyzing the so-called thermodynamic Bethe ansatz (TBA) equations. Combining the small N_f expansion and the pole cancellation condition, we determine the membrane one-instanton coefficient completely (4.34) and find a part of the membrane two-instanton coefficients (4.39), as analytic functions of N_f. As a non-trivial check, we show that our conjecture of membrane instantons is consistent with the numerical solution of the TBA equations for $0 < N_f < 1$. These results clearly show that the pole cancellation mechanism is a general phenomenon in M-theory, not the special property of the ABJ(M) theory.

This paper is organized as follows: in section 2, we review the known properties of the N_f matrix model, including the Fermi-gas approach, the TBA equations, and the 't Hooft and the M-theory limits of this model. In section 3, first we explain our algorithm

to compute the exact values of the partition functions $Z(N_f, N)$ for various integral values of N_f. Then, using these exact values, we analyze the structure of the grand potential. We find that the constant $A(N_f)$ in the grand potential (3.8) is related to the constant map contribution of the topological string. We also determine the analytic forms of the worldsheet instanton coefficients up to three-instanton. In section 4, we consider the membrane instanton corrections. First we study the small N_f expansion of the TBA equations, then we consider the pole cancellation condition between the worldsheet instantons and the membrane instantons. Finally, combining these two inputs, we determine the analytic forms of the membrane one-instanton and of a part of the membrane two-instanton. Section 5 is the conclusion. We also have three appendices A, B, and C, summarizing some results used in the main text.

2 Review of Fermi-gas approach

Let us start by reviewing the exact computation of the partition function on S^3 by using the localization [8] and the Fermi-gas approach [14]. The localization reduces the partition function to a finite dimensional matrix integral. We rewrite this matrix integral as a partition function of certain one-dimensional ideal Fermi-gas system as in [14]. This approach is quite powerful, and allows us to analyze the non-perturbative corrections to the partition function.

### 2.1 From matrix model to Fermi-gas

In this paper we consider the U(N) $\mathcal{N} = 4$ super Yang-Mills theory on S^3 with one adjoint hypermultiplet and N_f fundamental hypermultiplets. The quiver diagram for this theory is shown in figure 1(a). Following the general localization procedure in [8], one can immediately write down the partition function of this theory, and the result is given by the matrix integral

$$Z(N_f, N) = \frac{1}{N!} \int \prod_{j=1}^{N} \frac{dx_j}{4\pi} \frac{1}{(2\cosh\frac{x_j}{2})^{N_f}} \prod_{j<k} \tanh^2\left(\frac{x_j - x_k}{2}\right). \tag{2.1}$$

Our goal in this paper is to understand the large N behavior of this model, the so-called N_f matrix model, including non-perturbative corrections. We note that the matrix integral (2.1) can be evaluated for arbitrary value of N_f while the physical theory is defined only for integral N_f. It is well-known that the theory with $N_f = 1$ is equivalent to the ABJM theory at Chern-Simons level $k = 1$ via the mirror symmetry [2]. The equality of the partition functions of both theories was directly shown in [29]. Interestingly, the partition function for $N_f = 2$ is also related to the partition function of the U(N)$_k$ × U(N + 1)$_{-k}$ ABJ theory at $k = 2$

$$|Z_{\text{ABJ}}(k = 2, N, N + 1)| = \frac{1}{\sqrt{2}} Z(N_f = 2, N), \tag{2.2}$$

where the factor $1/\sqrt{2}$ comes from the partition function of U(1)$_{k=2}$ pure Chern-Simons theory. This relation can be checked by rewriting the ABJ partition function as in [23]. The quiver diagram of the ABJ(M) theory is shown in figure 1(b).
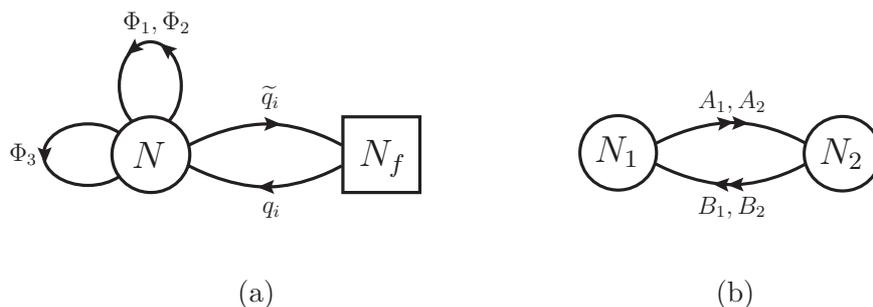
Figure 1. (a) The quiver diagram for the U(N) $\mathcal{N} = 4$ SYM with an adjoint and N_f fundamental hypermultiplets. We use $\mathcal{N} = 2$ notation. The adjoint hypermultiplet consists of two adjoint chiral multiplets (Φ_1, Φ_2), and Φ_3 is the chiral multiplet that forms the $\mathcal{N} = 4$ vector multiplet together with the $\mathcal{N} = 2$ vector multiplet. (b) The quiver for the U(N_1)$_k$ $\times$ U(N_2)$_{-k}$ ABJ(M) theory. The partition function of the former with $N_f = 1, 2$ is related to that of the latter at $k = 1, 2$, respectively.

It is not easy to perform the matrix integral (2.1) directly.[1] However, there is a very efficient way to compute the exact values of the partition function. This method was first proposed in [14] (see also [31]), and is now called the Fermi-gas approach. The key idea is to rewrite the partition function (2.1) as

$$Z(N_f, N) = \frac{1}{N!} \sum_{\sigma \in S_N} (-1)^\sigma \int d^N x \prod_{j=1}^{N} \rho(x_j, x_{\sigma(j)}), \tag{2.3}$$

where

$$\rho(x_1, x_2) = \frac{1}{2\pi} \frac{1}{(2\cosh\frac{x_1}{2})^{N_f/2}} \frac{1}{(2\cosh\frac{x_2}{2})^{N_f/2}} \frac{1}{2\cosh\left(\frac{x_1-x_2}{2}\right)}. \tag{2.4}$$

To derive (2.3), we used the Cauchy determinant formula. The partition function (2.3) can be interpreted as the partition function of an ideal Fermi-gas system described by the density matrix ρ [14]. In the following analysis, it is very convenient to introduce the grand canonical partition function and the grand potential

$$\Xi(N_f, z) = 1 + \sum_{N=1}^{\infty} z^N Z(N_f, N), \qquad J(N_f, \mu) = \log \Xi(N_f, z), \tag{2.5}$$

where $z = e^\mu$ is a fugacity with a chemical potential μ in the grand canonical ensemble. As was discussed in [14], for the partition function with the form (2.3), the grand partition function is written as a Fredholm determinant[2]

$$\Xi(N_f, z) = \det(1 + z\rho) = \exp\left[-\sum_{n=1}^{\infty} \frac{(-z)^n}{n} \operatorname{Tr} \rho^n\right], \tag{2.6}$$

where the trace of ρ is defined by

$$\operatorname{Tr} \rho^n = \int_{-\infty}^{\infty} dx_1 \cdots dx_n \, \rho(x_1, x_2)\rho(x_2, x_3) \cdots \rho(x_n, x_1). \tag{2.7}$$

---

[1]One interesting approach is to evaluate the multi-integral numerically by using the Monte Carlo method as in the (mirror) ABJM matrix model [30].

[2]We have used the well-known identity: $\log \det X = \operatorname{Tr} \log X$ for a matrix X.

Therefore the basic problem is how to compute $\text{Tr}\,\rho^n$. This is still not easy, but as will be seen in the next section, we can compute it recursively. Once we know the grand potential $J(N_f, \mu)$, it is easy to reconstruct the canonical partition function by

$$Z(N_f, N) = \int_{-\pi i}^{\pi i} \frac{d\mu}{2\pi i} e^{J(N_f, \mu) - N\mu}. \tag{2.8}$$

2.2 TBA equations

The density matrix ρ takes the form

$$\rho(x_1, x_2) = \frac{1}{2\pi} \frac{e^{-\frac{1}{2}\text{U}(x_1) - \frac{1}{2}\text{U}(x_2)}}{2\cosh\left(\frac{x_1 - x_2}{2}\right)}, \qquad \text{U}(x) = N_f \log\left[2\cosh\frac{x}{2}\right]. \tag{2.9}$$

Interestingly, for the kernel with this form, we can compute the grand potential from the TBA equations [32]. The TBA integral equations are given by

$$\begin{aligned}
\log R_+(x) &= -\text{U}(x) + \int_{-\infty}^{\infty} \frac{dx'}{2\pi} \frac{\log(1 + \eta^2(x'))}{\cosh(x - x')}, \\
\eta(x) &= -z \int_{-\infty}^{\infty} \frac{dx'}{2\pi} \frac{R_+(x')}{\cosh(x - x')}.
\end{aligned} \tag{2.10}$$

For the solutions of these equations, we also define

$$R_-(x) = R_+(x) \int_{-\infty}^{\infty} \frac{dx'}{\pi} \frac{\arctan\eta(x')}{\cosh^2(x - x')}. \tag{2.11}$$

Once these functions are determined, the grand potential is computed by

$$\frac{\partial J_\pm}{\partial z} = \frac{1}{4\pi} \int_{-\infty}^{\infty} dx\, R_\pm(x), \qquad J_\pm(z) = \frac{J(z) \pm J(-z)}{2}. \tag{2.12}$$

We note that the integral equations (2.10) and (2.11) can be recast as the following functional relations [16], called the Y-system in the literature,

$$\begin{aligned}
R_+\left(x + \frac{\pi i}{2}\right) R_+\left(x - \frac{\pi i}{2}\right) \exp\left[U\left(x + \frac{\pi i}{2}\right) + U\left(x - \frac{\pi i}{2}\right)\right] &= 1 + \eta^2(x), \\
\eta\left(x + \frac{\pi i}{2}\right) + \eta\left(x - \frac{\pi i}{2}\right) &= -zR_+(x),
\end{aligned} \tag{2.13}$$

and

$$\frac{R_-(x + \frac{\pi i}{2})}{R_+(x + \frac{\pi i}{2})} - \frac{R_-(x - \frac{\pi i}{2})}{R_+(x - \frac{\pi i}{2})} = 2i \frac{\eta'(x)}{1 + \eta^2(x)}. \tag{2.14}$$

The TBA equations (2.10) and (2.11) are powerful in the numerical computation for various values of N_f as in [33]. The functional equations (2.13) and (2.14), on the other hand, are useful in the semi-classical analysis [16]. The TBA equations for the ABJM case were studied in [16, 33, 34].

2.3 't Hooft limit and M-theory limit

We want to understand the large N behavior of the partition function (2.1) (or equivalently (2.3)). There are two interesting limits. One is the standard 't Hooft limit, in which the parameters are taken as follows:

$$N \to \infty, \qquad N_f \to \infty, \qquad \lambda = \frac{N}{N_f} : \text{fixed}, \qquad (2.15)$$

where λ is the 't Hooft coupling. In this limit, the free energy admits the perturbative genus expansion (plus non-perturbative contribution)

$$F(g_s, \lambda) = -\log Z(N_f, N) = \sum_{g=0}^{\infty} g_s^{2g-2} F_g(\lambda) + F_{\text{np}}(g_s, \lambda), \qquad g_s = \frac{1}{N_f}, \qquad (2.16)$$

where $F_g(\lambda)$ is the genus g contribution, and $F_{\text{np}}(g_s, \lambda)$ is the non-perturbative correction in g_s. In the ABJM matrix model, the genus zero contribution was computed by the standard matrix model technique, and the higher genus corrections were determined from the holomorphic anomaly equations [11, 12]. In the N_f matrix model, the genus zero contribution was computed in [27]. However, it is difficult to compute the higher genus correction.[3] The non-perturbative correction $F_{\text{np}}(g_s, \lambda)$ is also very difficult to be computed from the usual matrix model approach. Some interesting results on the non-perturbative corrections in the ABJM matrix model are found in [35, 36].[4] We emphasize that the Fermi-gas approach overcomes this difficulty, and we can predict analytic results for $F_{\text{np}}(g_s, \lambda)$. One important consequence is that the perturbative genus expansion is insufficient in the finite g_s regime, and the existence of the non-perturbative contribution $F_{\text{np}}(g_s, \lambda)$ is essential for the consistency of the theory.

We note that the genus g contribution $F_g(\lambda)$ at strong coupling contains the non-perturbative corrections in $\alpha'(\sim 1/\sqrt{\lambda})$, which has the exponentially suppressed contribution (see [27] for the genus zero contribution)

$$\mathcal{O}(e^{-2\pi\sqrt{2\lambda}}). \qquad (2.17)$$

From the dual Type IIA string point of view, such corrections are caused by the worldsheet instanton wrapping a two-cycle. On the other hand, the non-perturbative part $F_{\text{np}}(g_s, \lambda)$ contains the exponentially suppressed contribution

$$\mathcal{O}(e^{-\pi\sqrt{2\lambda}/g_s}). \qquad (2.18)$$

As discussed in [35], such non-perturbative corrections come from the D2-branes wrapping a three-cycle. In this paper, we refer these corrections to the membrane instanton corrections because these are purely non-perturbative effects in g_s.

---

[3]We thank M. Mariño for pointing out the difficulty of the higher genus computation in the N_f matrix model.

[4]In the ABJM matrix model, the perturbative genus expansion is very likely Borel summable. One of the conclusions in [36] is that the Borel resummation of the genus expansion does not present the exact result, and one needs to consider the non-perturbative contribution $F_{\text{np}}(g_s, \lambda)$. This non-perturbative contribution is caused by so-called complex instantons [35, 36], and interpreted as D2-brane instanton effects in Type IIA string theory [35].

The other interesting limit is the following one:

$$N \to \infty, \qquad N_f : \text{fixed}, \tag{2.19}$$

corresponding to a direct thermodynamic limit of the Fermi-gas system. In this limit, the gauge theory is dual to M-theory on $AdS_4 \times S^7/\mathbb{Z}_{N_f}$, and thus we call this limit as the M-theory limit here. In the following analysis, we mainly focus on the M-theory limit. In the M-theory limit, the worldsheet instanton correction (2.17) and the membrane instanton correction (2.18) have the same order

$$\mathcal{O}\Big(e^{-2\pi\sqrt{2N/N_f}}\Big), \qquad \mathcal{O}\Big(e^{-\pi\sqrt{2N_f N}}\Big). \tag{2.20}$$

This is because, in the M-theory regime, both instantons are up-lifted to M2-branes wrapping two different types of three-cycles. See figure 1 in [18] for more detail.

Before closing this section, let us comment on the large N limit in the grand canonical ensemble. From the integral transformation (2.8), the partition function can be evaluated by the saddle point approximation in the large N limit. The saddle point equation is given by

$$J'(\mu_*) - N = 0. \tag{2.21}$$

As was discussed in [27, 28], the grand potential behaves in the large μ limit as

$$J(N_f, \mu) \approx \frac{2}{3\pi^2 N_f}\mu^3 \qquad (\mu \to \infty). \tag{2.22}$$

Therefore, the saddle point is given by

$$\mu_* = \pi\sqrt{\frac{N_f N}{2}}. \tag{2.23}$$

This means that the large N limit in the canonical ensemble corresponds to the large μ limit in the grand canonical ensemble. The saddle point analysis presents a simple derivation of the $N^{3/2}$ behavior of the free energy [14]

$$F(N_f, N) \approx -J(N_f, \mu_*) + N\mu_* \approx \frac{\pi\sqrt{2N_f}}{3}N^{3/2} \qquad (N \to \infty). \tag{2.24}$$

This reproduces the matrix model result [28]. As we will see later, the Airy function behavior is also derived easily from the grand canonical analysis. Finally, the worldsheet instanton correction and the membrane instanton correction in (2.20) correspond to the exponentially suppressed corrections

$$\mathcal{O}(e^{-4\mu/N_f}), \qquad \mathcal{O}(e^{-2\mu}), \tag{2.25}$$

respectively, in the grand canonical ensemble.

3 Exploring non-perturbative effects

In this section, we investigate the non-perturbative corrections to the partition function (2.1) in the large N limit. As explained in the previous section, the large N limit corresponds to the large μ limit in the grand canonical ensemble. We thus concentrate our attention on the large μ behavior of the grand potential. We first compute the exact values of the partition function for various integral N_f by using the Fermi-gas approach developed in [15, 33, 34]. Next, using these exact data, we extract the non-perturbative corrections to the grand potential. Based on these results, we look for exact forms of the worldsheet instanton corrections for general N_f. The membrane instanton corrections are explored in the next section.

3.1 Exact computation of the partition function

In this subsection, we compute the exact values of the partition function for some integral values of N_f. Our strategy is the same as that in the ABJM theory [15, 33]. We first divide the density matrix ρ into parity even/odd part

$$\rho(x_1, x_2) = \rho_+(x_1, x_2) + \rho_-(x_1, x_2), \quad \rho_\pm(x_1, x_2) = \frac{\rho(x_1, x_2) \pm \rho(x_1, -x_2)}{2}. \tag{3.1}$$

Then, the grand partition function is factorized into two parts [33],

$$\Xi(z) = \det(1 + z\rho_+)\det(1 + z\rho_-). \tag{3.2}$$

As in [15, 33], we write $\rho_\pm$ as the forms

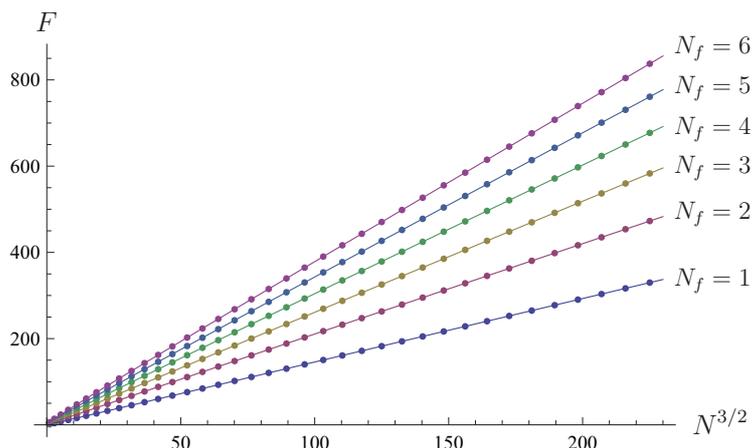$$\rho_\pm(x, y) = \frac{E_\pm(x)E_\pm(y)}{\cosh x + \cosh y}, \tag{3.3}$$

where

$$E_+(x) = \frac{\cosh \frac{x}{2}}{(2\cosh \frac{x}{2})^{N_f/2}}, \qquad E_-(x) = \frac{\sinh \frac{x}{2}}{(2\cosh \frac{x}{2})^{N_f/2}}. \tag{3.4}$$

Now we can apply the result in [37] to the kernels $\rho_\pm$. The important consequence is that we can compute $\rho_\pm^n$ from functions with one variable:

$$\rho_\pm^{2n+1}(x, y) = \frac{E_\pm(x)E_\pm(y)}{\cosh x + \cosh y} \sum_{\ell=0}^{2n} (-1)^\ell \phi_\pm^\ell(x)\phi_\pm^{2n-\ell}(y),$$

$$\rho_\pm^{2n}(x, y) = \frac{E_\pm(x)E_\pm(y)}{\cosh x - \cosh y} \sum_{\ell=0}^{2n-1} (-1)^\ell \phi_\pm^\ell(x)\phi_\pm^{2n-1-\ell}(y), \tag{3.5}$$

where the functions $\phi_\pm^\ell(x)$ are determined by the following integral equations recursively

$$\phi_+^\ell(x) = \frac{1}{\cosh \frac{x}{2}} \int_{-\infty}^\infty \frac{dx'}{2\pi} \frac{1}{2\cosh\left(\frac{x-x'}{2}\right)} \frac{\cosh \frac{x'}{2}}{(2\cosh \frac{x'}{2})^{N_f}} \phi_+^{\ell-1}(x'),$$

$$\phi_-^\ell(x) = \frac{1}{\cosh \frac{x}{2}} \int_{-\infty}^\infty \frac{dx'}{2\pi} \frac{1}{2\cosh\left(\frac{x-x'}{2}\right)} \frac{\sinh \frac{x'}{2}\tanh \frac{x'}{2}}{(2\cosh \frac{x'}{2})^{N_f}} \phi_-^{\ell-1}(x'), \tag{3.6}$$

Figure 2. The free energy $F = -\log Z$ for $N_f = 1, \ldots, 6$ as a function of $N^{3/2}$. The dots represent the exact values computed by our algorithm while the solid lines represent the perturbative part given by the Airy function (3.11).

with the initial conditions $\phi_\pm^0(x) = 1$. For a derivation of these equations, see [33]. In [15, 33], we further found that the grand partition function is expressed only in terms of the even parity part:

$$\Xi(z) = \det(1 - z^2\rho_+^2)G(z), \qquad G(z) = \sum_{\ell=0}^\infty \phi_+^\ell(0)z^\ell. \qquad (3.7)$$

This result is very useful in the practical computation. In summary, we first solve the integral equation (3.6) for $\phi_+^\ell(x)$ recursively. This can be done very efficiently by using the technique in [15, 34]. We then compute $\mathrm{Tr}\,\rho_+^{2n}$ to know $\det(1 - z^2\rho_+^2)$ by using (3.5). We finally read off the coefficient of z^N in (3.7) that is nothing but $Z(N_f, N)$.

We have computed the exact values of $Z(N_f, N)$ for $N_f = 1, 2, \ldots, 12, 14, 16$, up to certain values of N. Since the results are very complicated, we cannot write them down here. Instead, a set of ancillary files for these values readable in `Mathematica`[5] is attached to this paper on `arXiv`. In figure 2, we show the free energy $F = -\log Z$ for $N_f = 1, 2, \ldots, 6$ as a function of $N^{3/2}$. The dots represent the exact values while the solid lines represent the leading Airy function behaviors given by (3.11).[6] The exact values of the free energy indeed show the $N^{3/2}$ behaviors, and also a very good agreement with the Airy function (3.11) even for small N.

3.2 General structure of the grand potential

Next we consider the structure of the grand potential. In the large μ limit, the grand potential takes the following form [27, 28]

$$J(N_f, \mu) = \frac{C(N_f)}{3}\mu^3 + B(N_f)\mu + A(N_f) + J_{\mathrm{np}}(N_f, \mu) \qquad (\mu \to \infty), \qquad (3.8)$$

---

[5]One can import those files to `Mathematica` by `Import["file.dat", "List"]//ToExpression`.

[6]To use this formula, we need the non-trivial function $A(N_f)$. In subsection 3.3, we give an exact form of $A(N_f)$. See (3.12).

where

$$C(N_f) = \frac{2}{\pi^2 N_f}, \qquad B(N_f) = \frac{1}{2N_f} - \frac{N_f}{8}, \qquad (3.9)$$

and $A(N_f)$ is a non-trivial function of N_f. The remaining part $J_{\mathrm{np}}(N_f, \mu)$ is the exponentially suppressed correction in $\mu \to \infty$. As we have seen in the previous section, there are two such corrections, coming from the worldsheet instantons and the membrane instantons. In addition, there are also "bound states" of these two kinds of instantons [16]. Taking into account of these bound states, the non-perturbative correction has the following expansion[7]

$$J_{\mathrm{np}}(N_f, \mu) = \sum_{\substack{\ell,m=0 \\ (\ell,m)\neq(0,0)}}^{\infty} f_{\ell,m}(N_f, \mu) \exp\left[-\left(2\ell + \frac{4m}{N_f} \right)\mu \right]. \qquad (3.10)$$

The structure is very similar to the ABJM case [16], but the explicit forms of the coefficients look quite different, as we will see later. The worldsheet instanton correction corresponds to $\ell = 0$, and the membrane instanton correction to $m = 0$. The others are understood as their bound states. If we ignore the non-perturbative correction $J_{\mathrm{np}}(N_f, \mu)$, the grand potential (3.8) leads to the following canonical partition function [14]

$$Z_{\mathrm{pert}}(N_f, N) = C(N_f)^{-1/3} e^{A(N_f)} \, \mathrm{Ai}\left[C(N_f)^{-1/3}(N - B(N_f)) \right]. \qquad (3.11)$$

where $\mathrm{Ai}(z)$ is the Airy function. This result is understood as the all-genus resummation after neglecting all the exponentially suppressed corrections [39]. Using the asymptotic expansion of the Airy function, one can, of course, reproduce the $N^{3/2}$ behavior (2.24) in the large N limit.

Our remaining task is to determine the non-trivial functions $A(N_f)$ and $f_{\ell,m}(N_f, \mu)$. In the ABJ(M) theory, this program has already been done with the help of an accidental connection to the topological string on local $\mathbb{P}^1 \times \mathbb{P}^1$ (see [15–18, 22, 23]). However, in our case of the N_f matrix model, we do not know a nice connection to the topological string. Therefore we do not have any guiding principles to determine the non-perturbative corrections systematically, and it is challenging to understand $f_{\ell,m}(N_f, \mu)$. To explore the non-perturbative effects, we here take the following strategy:

- Using the exact data computed in the previous subsection, we extract the non-perturbative corrections for various integral values of N_f.

- Based on these data, we conjecture the worldsheet instanton correction $f_{0,m}(N_f, \mu)$ for general N_f order by order.

- The (conjectured) worldsheet instanton correction diverges for some values of N_f. These singularities must be canceled by the other contributions because the theory is

---

[7]Strictly speaking, the non-perturbative correction $J_{\mathrm{np}}(N_f, \mu)$ contains an additional contribution that shows oscillatory behavior [15]. However, this contribution can be removed by deforming the integration contour in (2.8) from $[-\pi i, \pi i]$ to $[-i\infty, i\infty]$. Below, we always take the deformed contour when going back to the canonical ensemble, thus we can drop this oscillatory contribution. Note that the oscillatory contribution seems to play an important role in the analysis of the "orbifold" ABJM theory [38].

always well-defined. This pole cancellation mechanism was first found in the ABJM theory [15]. Using this mechanism, we can determine the pole structure of the membrane instanton correction. Combining this information with some other inputs, we fix the analytic forms of the membrane instanton corrections.

- The obtained results can be compared with the numerical results computed from the TBA equations in section 2 for various (non-integral) values of N_f. This comparison gives a highly non-trivial test of our conjecture.

Of course, the analytic forms of higher instanton corrections become very complicated, and it gets more and more difficult to determine them in this way. In this paper, following the above strategy, we indeed determine the worldsheet instanton correction up to $m = 3$, and also the leading membrane instanton correction (and a part of the next-to-leading correction). So far, we cannot obtain any results on the bound states. This should be understood in the future work.

3.3 The constant part

Before proceeding to the instanton corrections, we give a conjecture of the exact form of $A(N_f)$. We find that the constant part $A(N_f)$ is exactly related to that in the ABJM theory as

$$A(N_f) = \frac{1}{2} \left[A_{\mathrm{const}}(N_f) + A_{\mathrm{const}}(1) N_f^2 \right], \qquad (3.12)$$

where $A_{\mathrm{const}}(k)$ is the constant part appearing in the grand potential in the ABJM Fermi-gas [14]. Although we do not have a proof of this conjecture, it passes many non-trivial tests as we will see below. Note that $A_{\mathrm{const}}(k)$ corresponds to the constant map contribution in the topological string [30]. The small k expansion of $A_{\mathrm{const}}(k)$ was first computed in [14], and then the all-loop formula and its integral expression were conjectured in [30]. As derived in appendix A, we find another simpler integral expression of $A_{\mathrm{const}}(k)$:

$$A_{\mathrm{const}}(k) = \frac{2\zeta(3)}{\pi^2 k} \left(1 - \frac{k^3}{16} \right) + \frac{k^2}{\pi^2} \int_0^\infty dx \frac{x}{e^{kx} - 1} \log(1 - e^{-2x}). \qquad (3.13)$$

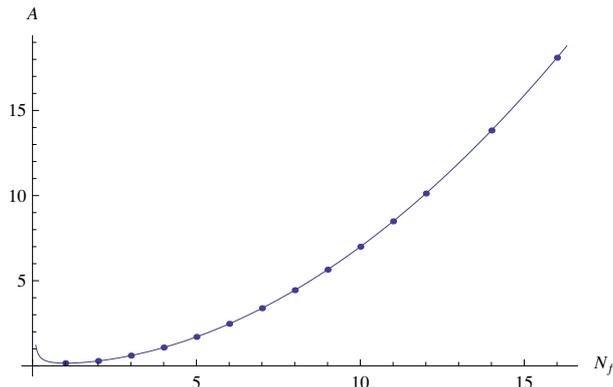In particular, we find closed form expressions for any integer k,

$$A_{\mathrm{const}}(k) = \begin{cases} -\dfrac{\zeta(3)}{\pi^2 k} - \dfrac{2}{k} \displaystyle\sum_{m=1}^{\frac{k}{2}-1} m \left(\dfrac{k}{2} - m \right) \log \left(2 \sin \dfrac{2\pi m}{k} \right) & \text{(even } k\text{)}, \\[3ex] -\dfrac{\zeta(3)}{8\pi^2 k} + \dfrac{k}{4} \log 2 - \dfrac{1}{k} \displaystyle\sum_{m=1}^{k-1} g_m(k)(k - g_m(k)) \log \left(2 \sin \dfrac{\pi m}{k} \right) & \text{(odd } k\text{)}, \end{cases} \qquad (3.14)$$

where

$$g_m(k) = \frac{k + (-1)^m (2m - k)}{4}. \qquad (3.15)$$

Our conjecture (3.12) predicts the small N_f expansion

$$A(N_f) = \frac{\zeta(3)}{\pi^2 N_f} - \frac{N_f}{24} + \left(-\frac{\zeta(3)}{16\pi^2} + \frac{\log 2}{8} \right) N_f^2 - \frac{\pi^2 N_f^3}{8640} + \mathcal{O}(N_f^5). \qquad (3.16)$$

Figure 3. The constant contribution $A(N_f)$. The numerical estimations (3.18) from exact $Z(N_f, N)$ in the previous section are shown by the dots, and our conjecture (3.12) by the solid line.

The leading term coincides with the result in [27]. As we will see in section 4, the next-to-leading correction is reproduced from the TBA analysis. Also, in the large N_f limit, we find

$$\lim_{N_f \to \infty} \frac{A(N_f)}{N_f^2} = \frac{1}{8} \left(-\frac{\zeta(3)}{\pi^2} + \log 2 \right) = 0.07141916904\ldots. \tag{3.17}$$

This result also agrees with the constant term in the genus-zero free energy found in [27]. In figure 3, we plot $A(N_f)$ for finite N_f. The solid line is our conjecture, while the dots are the numerical values of $A(N_f)$ extracted from the exact values of $Z(N_f, N)$, where we estimated them by (see (3.11))

$$A(N_f) \approx \log \left[\frac{Z^{\text{exact}}(N_f, N)}{C(N_f)^{-1/3} \operatorname{Ai} \left[C(N_f)^{-1/3}(N - B(N_f)) \right]} \right] \quad (N \gg 1), \tag{3.18}$$

for N as large as possible. For example, in the case of $N_f = 3$, we computed exact $Z(3, N)$ up to $N = 53$. Using the estimation (3.18) for $N = 53$, we find

$$A(3) \approx 0.60247027481429615744. \tag{3.19}$$

On the other hand, the exact value of $A(3)$ in (3.12) is

$$A(3) = -\frac{7\zeta(3)}{12\pi^2} + \frac{1}{6} \log \frac{512}{9}. \tag{3.20}$$

The difference is $\mathcal{O}(10^{-16})$, which is roughly the same order as the leading worldsheet instanton correction $\mathcal{O}(e^{-2\pi\sqrt{2N/N_f}})$. For other values of N_f, our conjecture (3.12) indeed show an agreement with the numerical estimation up to about $\mathcal{O}(e^{-2\pi\sqrt{2N/N_f}})$.

3.4 Non-perturbative corrections for integral N_f

Using the exact values of the partition function computed in subsection 3.1, we can determine the non-perturbative correction to the grand potential. The basic method is the

same as that in [15]. We first take an appropriate ansatz of the non-perturbative correction to the grand potential. We then pull it back to the canonical partition function. The coefficients in the ansatz are fixed by the numerical fitting of the exact values with high precision. See [15] in detail.

Since the partition function for $N_f = 1, 2$ is related to the ABJ(M) case, we can use the results in [15, 17, 22, 23]. We also find the corrections explicitly for $N_f = 3, 4, 6, 8, 12$, and numerically for other N_f's. We observe that all of these results take the form[8]

$$J_{\mathrm{np}}(N_f, \mu) = \sum_{n=1}^{\infty} f_n(N_f, \mu) e^{-\frac{4n\mu}{N_f}}, \qquad N_f \in \mathbb{N}. \tag{3.21}$$

In particular, as in the ABJM case, the terms $e^{-(4m-2)\mu}$ $(m = 1, 2, \dots)$ do not appear for odd N_f. The explicit forms of $f_n(N_f, \mu)$ are complicated, and listed in appendix B.

3.5 Worldsheet instanton corrections

Now let us consider the worldsheet instanton correction

$$J^{\mathrm{WS}}(N_f, \mu) = \sum_{m=1}^{\infty} f_{0,m}(N_f, \mu) e^{-\frac{4m\mu}{N_f}}. \tag{3.22}$$

Here we give a conjecture of the analytic form of $f_{0,m}(N_f, \mu)$, which is valid for general N_f, up to $m = 3$. To simplify the notation, we define

$$s_n(N_f) = \sin\left(\frac{2\pi n}{N_f}\right), \qquad P_n(N_f, \mu) = \frac{4n\mu + N_f}{\pi}. \tag{3.23}$$

Our conjecture of $f_{0,m}(N_f, \mu)$ up to $m = 3$ is

$$
\begin{aligned}
f_{0,1} &= -\frac{1}{2s_1} P_1, \\
f_{0,2} &= -\frac{1}{4} P_1^2 + \frac{3s_3}{8s_1 s_2} P_2 - \frac{s_4}{2s_1^2 s_2}, \\
f_{0,3} &= -\frac{s_2^2}{12s_1} P_1^3 + \frac{3s_4}{8s_1} P_1 P_2 - \frac{5s_4 s_5}{9s_1 s_2 s_3} P_3 - \frac{s_4^2}{2s_1 s_2^2} P_1 - \frac{2s_6}{s_1^2 s_3} + \frac{2s_4 s_5}{s_1^2 s_2^2}.
\end{aligned}
\tag{3.24}
$$

We have checked that this conjecture is consistent with the exact values of $Z(N_f, N)$ for $3 \le N_f \le 16$. For $N_f = 1, 2$, worldsheet instantons (3.24) have poles already at the one-instanton level. We can also see that $f_{0,2}$ have poles at $N_f = 1, 2, 4$ and $f_{0,3}$ have poles at $N_f = 1, 2, 3, 4, 6$.[9] As we will see in the next section, those poles should be canceled by the membrane instantons.

---

[8]One should not confuse this result with the worldsheet instanton correction (3.22). In general, the coefficient $f_n(N_f, \mu)$ is the sum of all the contributions from the worldsheet instantons, the membrane instantons and their bound states.

[9]More precisely, $f_{0,2}$ (resp. $f_{0,3}$) has poles at rational values of $N_f = 1/n, 2/n, 4/n$ (resp. $N_f = 1/n, 2/n, 3/n, 4/n, 6/n$) for all $n \in \mathbb{Z}$. Those poles should also be canceled by the higher membrane instantons.

One can check that our conjecture (3.24) correctly reproduces the result in appendix B. For instance, for $N_f = 8$ we find

$$f_{0,1}(8, \mu) = -\frac{4\mu + 8}{\sqrt{2}\pi},$$

$$f_{0,2}(8, \mu) = -\frac{(4\mu + 8)^2}{4\pi^2} + \frac{3(4\mu + 4)}{4\pi}, \qquad (3.25)$$

$$f_{0,3}(8, \mu) = -\frac{(4\mu + 8)^3}{6\sqrt{2}\pi^3} + 4\sqrt{2},$$

which agree with the result (B.6) in appendix B. Other cases in appendix B are also reproduced by the conjecture (3.24).[10] We also stress that, for all other cases $N_f = 5, 7, 9, \cdots, 16$, our conjecture (3.24) agrees highly non-trivially with the instanton corrections extracted from the exact partition function. To see it, let us define the non-perturbative correction to the partition function by

$$Z(N_f, N) = Z_{\text{pert}}(N_f, N)(1 + Z_{\text{np}}(N_f, N)), \qquad (3.26)$$

where $Z_{\text{pert}}(N_f, N)$ is the perturbative contribution in (3.11), neglecting all the exponentially suppressed corrections. These non-perturbative corrections are encoded in $Z_{\text{np}}(N_f, N)$. We have defined $Z_{\text{np}}(N_f, N)$ such that it decays exponentially in the large N limit. Note that for $N_f > 6$, the first three corrections come from the worldsheet instantons because $e^{-12\mu/N_f} > e^{-2\mu}$. Namely,

$$Z_{\text{np}} = Z_{\text{WS}}^{(1)} + Z_{\text{WS}}^{(2)} + Z_{\text{WS}}^{(3)} + (\text{subleading corrections}) \qquad (N_f > 6), \qquad (3.27)$$

where $Z_{\text{WS}}^{(m)}$ is the worldsheet m-instanton correction. As in the perturbative contribution, it is straightforward to translate the grand canonical result (3.22) into the canonical one $Z_{\text{WS}}^{(m)}$. We also introduce the quantity

$$\delta = e^{6\pi\sqrt{2N/N_f}} \left(\frac{Z}{Z_{\text{pert}}} - 1 - Z_{\text{WS}}^{(1)} - Z_{\text{WS}}^{(2)} - Z_{\text{WS}}^{(3)} \right). \qquad (3.28)$$

Since the worldsheet 3-instanton scales as the order $e^{-6\pi\sqrt{2N/N_f}}$, the subleading corrections in (3.27) decay faster than $e^{-6\pi\sqrt{2N/N_f}}$. Therefore the quantity δ must be exponentially suppressed in the large N limit for $N_f > 6$ if our conjecture (3.24) is correct. In figure 4, we plot δ for $N_f = 7, 9, 10, 11, 14, 16$ by using the exact data. The quantity δ indeed decays exponentially when N is large, as expected.

---

[10]More precisely, the coefficient of $e^{-\alpha\mu}$ with $\alpha < 2$ (even N_f) or $\alpha < 4$ (odd N_f) can be reproduced from (3.24). Beyond these values, we have to consider the membrane instanton and the bound state contributions.
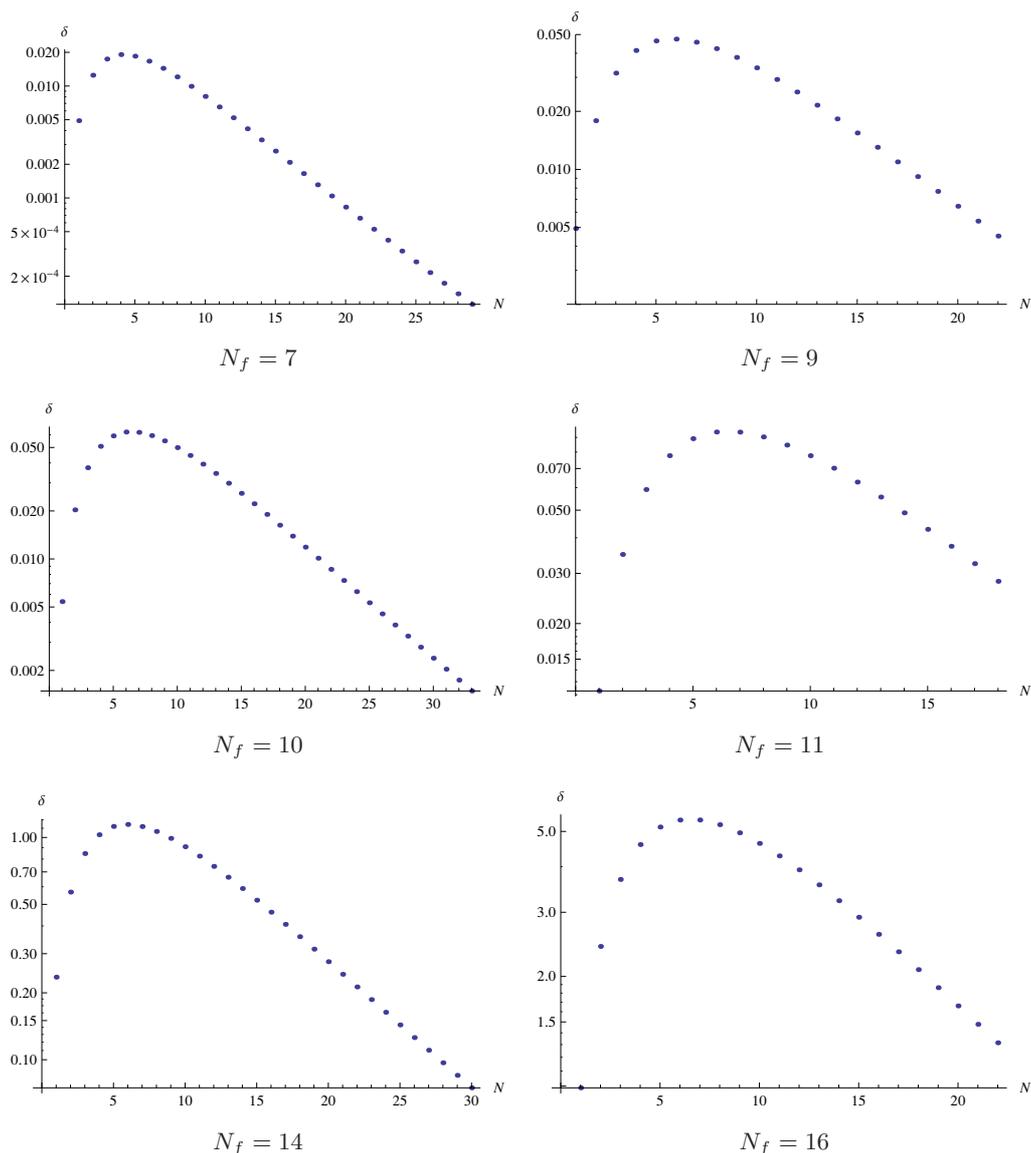
Figure 4. We plot δ defined in (3.28) for $N_f = 7, 9, 10, 11, 14, 16$. Note that the vertical axis is log scale. One can see that δ is indeed exponentially suppressed as N grows.

As a further test, we consider the large N_f limit of (3.24) with $\widehat{\mu} = \mu/N_f$ fixed. It is not difficult to see

$$
\begin{aligned}
f_{0,1}(N_f, \mu) &= \frac{N_f^2}{\pi^2}\left(-\widehat{\mu} - \frac{1}{4}\right) - \frac{2}{3}\widehat{\mu} - \frac{1}{6} + \mathcal{O}(N_f^{-2}), \\
f_{0,2}(N_f, \mu) &= \frac{N_f^2}{\pi^2}\left(-4\widehat{\mu}^2 + \frac{1}{4}\widehat{\mu} - \frac{7}{32}\right) - 6\widehat{\mu} + \frac{11}{12} + \mathcal{O}(N_f^{-2}), \\
f_{0,3}(N_f, \mu) &= \frac{N_f^2}{\pi^2}\left(-\frac{128}{3}\widehat{\mu}^3 + 16\widehat{\mu}^2 - \frac{46}{9}\widehat{\mu} + \frac{11}{27}\right) \\
&\quad + \frac{1792\widehat{\mu}^3}{9} - \frac{992\widehat{\mu}^2}{3} + \frac{356\widehat{\mu}}{3} - \frac{134}{9} + \mathcal{O}(N_f^{-2}).
\end{aligned}
\tag{3.29}
$$

The leading terms exactly coincide with the genus zero contribution of the grand potential in the 't Hooft limit computed in appendix C. We stress that our conjecture (3.24) gives an all-genus prediction in the 't Hooft expansion. It would be very interesting to confirm whether our conjecture indeed reproduces the higher genus corrections.[11]

From (3.24), we expect that the worldsheet m-instanton coefficient $f_{0,m}(N_f, \mu)$ has the following general structure:

- $f_{0,m}$ is an mth order polynomial of μ, and the highest order term is P_1^m.

- Most of the terms of $f_{0,m}$ have the same "degree" m, i.e. they have the form of $\prod_l P_l^{k_l}$ with $\sum l k_l = m$, but some remaining terms have smaller degree $\sum l k_l < m$.

- The coefficient of each term in $f_{0,m}$ is a combination of s_n $(n \le 2m)$.

It would be interesting to understand the origin of this structure and the general rule to find the coefficients.

4   Membrane instanton corrections

In the previous section, we extracted the non-perturbative corrections to the grand potential from the exact values of the partition function for some integral values of N_f. Based on these results, we proposed the analytic forms of the worldsheet instanton corrections up to $m = 3$. As mentioned before, the grand potential also receives the non-perturbative corrections from the membrane instantons. In this section, we explore analytic forms of these corrections. The membrane instanton correction corresponds to $m = 0$ in (3.10),

$$J^{\text{M2}}(N_f, \mu) = \sum_{\ell=1}^\infty f_{\ell,0}(N_f, \mu) e^{-2\ell\mu}. \qquad (4.1)$$

We want to determine the coefficient $f_{\ell,0}(N_f, \mu)$. Unfortunately, we do not have a systematic way to compute $f_{\ell,0}(N_f, \mu)$. Here, we try to fix it from many constraints. The same idea was originally used in the ABJM Fermi-gas [15]. We first investigate the expansion of the grand potential around $N_f = 0$. We then consider the singularity structure of $f_{\ell,0}(N_f, \mu)$. Using these constraints, we present exact forms of $f_{1,0}(N_f, \mu)$ and a part of $f_{2,0}(N_f, \mu)$. To fix the higher instanton corrections, we need more information.

4.1   Semi-classical analysis from TBA

We here study the expansion of the grand potential $J(N_f, z)$ around $N_f = 0$. Let us consider what kind of corrections the grand potential receives in the semi-classical limit

---

$N_f \to 0$.[12] We first observe that $Z_1 = \operatorname{Tr} \rho$ can be computed exactly [27],

$$Z_1 = \frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{dx}{(2 \cosh \frac{x}{2})^{N_f}} = \frac{1}{4\pi} \frac{\Gamma^2(N_f/2)}{\Gamma(N_f)}. \qquad (4.2)$$

This has the following semi-classical expansion

$$Z_1 = \frac{1}{\pi N_f} - \frac{\pi N_f}{24} + \frac{\zeta(3)}{4\pi} N_f^2 - \frac{\pi^3}{640} N_f^3 + \left(-\frac{\pi \zeta(3)}{96} + \frac{3\zeta(5)}{16\pi} \right) N_f^4 + \mathcal{O}(N_f^5). \qquad (4.3)$$

This observation suggests that the grand potential also has the similar semi-classical expansion

$$J(N_f, z) = \frac{1}{N_f} J_0(z) + N_f J_1(z) + N_f^2 J_{3/2}(z) + N_f^3 J_2(z) + N_f^4 J_{5/2}(z) + \mathcal{O}(N_f^5), \qquad (4.4)$$

because $J(N_f, z)$ is a kind of generating function of $Z_n = \operatorname{Tr} \rho^n$. The absence of the constant $\mathcal{O}(N_f^0)$ contribution is not obvious only from this observation, but the TBA analysis below supports this. The leading contribution $J_0(z)$ has already been computed in [27]. Note that in the ABJM Fermi-gas, the even power terms k^{2n} do not appear in the semi-classical expansion. This is a big difference between the N_f matrix model and the ABJM case. Here we compute $J_1(z)$ from the TBA equations.

What we should do is to solve the functional relations (2.13) and (2.14) in the semi-classical limit $N_f \to 0$ as in [16]. To do so, we rescale all the functions by

$$\widehat{R}_{\pm}(x) = R_{\pm}\left(\frac{x}{N_f} \right), \quad \widehat{\eta}(x) = \eta\left(\frac{x}{N_f} \right), \quad \widehat{U}(x) = U\left(\frac{x}{N_f} \right), \qquad (4.5)$$

Then the equations (2.13) and (2.14) become

$$\widehat{R}_+\left(x + \frac{\pi i N_f}{2} \right) \widehat{R}_+\left(x - \frac{\pi i N_f}{2} \right) \exp\left[\widehat{U}\left(x + \frac{\pi i N_f}{2} \right) + \widehat{U}\left(x - \frac{\pi i N_f}{2} \right) \right] = 1 + \widehat{\eta}^2(x),$$

$$\widehat{\eta}\left(x + \frac{\pi i N_f}{2} \right) + \widehat{\eta}\left(x - \frac{\pi i N_f}{2} \right) = -z \widehat{R}_+(x), \qquad (4.6)$$

and

$$\frac{\widehat{R}_-(x + \frac{\pi i N_f}{2})}{\widehat{R}_+(x + \frac{\pi i N_f}{2})} - \frac{\widehat{R}_-(x - \frac{\pi i N_f}{2})}{\widehat{R}_+(x - \frac{\pi i N_f}{2})} = 2i N_f \frac{\widehat{\eta}'(x)}{1 + \widehat{\eta}^2(x)}. \qquad (4.7)$$

These are formally the same forms as the ones in (4.10) and (4.11) of [16] with $k = N_f$. The only but big difference is the explicit form of the potential $\widehat{U}(x)$.

We assume that the functional equations (4.6) and (4.7) admit the semi-classical expansions around $N_f = 0$. The important point is that the potential part in (4.6) has the following "semi-classical" expansion

$$\widehat{U}\left(x + \frac{\pi i N_f}{2} \right) + \widehat{U}\left(x - \frac{\pi i N_f}{2} \right) = |x| + N_f \log\left(1 + e^{-\frac{2|x|}{N_f}} \right) \qquad (N_f \to 0). \qquad (4.8)$$

---

[12]The term "semi-classical limit" is a bit confusing. In the original coordinate, the commutation relation of the canonical variables (x, p) is given by $[x, p] = 2\pi i$, and thus $\hbar = 2\pi$ is a constant in this coordinate. However, as noted in [27, 28], it is more convenient to rescale the position variable by $x = q/N_f$. In this new coordinate, the commutation relation becomes $[q, p] = 2\pi i N_f$, thus the limit $N_f \to 0$ corresponds to the semi-classical limit $\hbar = 2\pi N_f \to 0$. In the TBA, this rescale corresponds to the redefinition (4.5).

The second term on the right hand side in (4.8) looks like a "non-perturbative" term in N_f. However, the integration of $e^{-2|x|/N_f}$ may potentially cause the perturbative corrections, for example,

$$\int_{-\infty}^{\infty} dx \, e^{-\frac{2|x|}{N_f}} = N_f. \tag{4.9}$$

Therefore we cannot drop the second term even in the semi-classical analysis.[13] As in [16], we formally solve (4.6) and (4.7) by the semi-classical expansions,

$$r(x) = \widehat{R}_+(x) = \sum_{n=0}^{\infty} r_n(x) N_f^n, \qquad \widehat{\eta}(x) = \sum_{n=0}^{\infty} \eta_n(x) N_f^n,$$
$$t(x) = \frac{\widehat{R}_-(x)}{\widehat{R}_+(x)} = \sum_{n=0}^{\infty} t_n(x) N_f^n. \tag{4.10}$$

This can be done systematically up to any desired order in principle. In appendix B, we give explicit forms up to $n = 3$.

Once the solutions of (4.6) and (4.7) are found, we can compute the grand potential from (2.12). Let us define

$$I_{\pm,n} \equiv \int_0^{\infty} \frac{dx}{2\pi} \widehat{R}_{\pm,n}(x), \qquad \widehat{R}_{\pm}(x) = \sum_{n=0}^{\infty} N_f^n \widehat{R}_{\pm,n}(x), \tag{4.11}$$

where $\widehat{R}_{\pm,n}(x)$ are related to $r_n(x)$ and $t_n(x)$ in (4.10). Using the solutions in appendix B, one can check that $I_{\pm,n}$ up to $n = 3$ has the following expansions around $N_f = 0$:

$$I_{+,0} = \frac{2}{\pi z} \arcsin\left(\frac{z}{2}\right), \qquad I_{-,0} = -\frac{2}{\pi^2 z} \arcsin^2\left(\frac{z}{2}\right),$$
$$I_{\pm,1} = N_f I_{\pm,1}^{(1)} + N_f^2 I_{\pm,1}^{(2)} + \mathcal{O}(N_f^3),$$
$$I_{\pm,2} = I_{\pm,2}^{(0)} + N_f I_{\pm,2}^{(1)} + \mathcal{O}(N_f^2), \tag{4.12}$$
$$I_{\pm,3} = N_f^{-1} I_{\pm,3}^{(-1)} + I_{\pm,3}^{(0)} + \mathcal{O}(N_f),$$

The explicit forms of these coefficients are also listed in appendix B. We also observe that the higher order corrections have the following expansions:

$$I_{\pm,2n} = N_f^{-2n+3} I_{\pm,2n}^{(-2n+3)} + \mathcal{O}(N_f^{-2n+4}) \quad (n = 2, 3, \cdots),$$
$$I_{\pm,2n+1} = N_f^{-2n+2} I_{\pm,2n+1}^{(-2n+2)} + \mathcal{O}(N_f^{-2n+3}) \quad (n = 2, 3, \cdots). \tag{4.13}$$

Combining all the above results, the grand potential is given by

$$J_z^{\pm}(z) = \frac{1}{2\pi N_f} \int_0^{\infty} dx \, \widehat{R}_{\pm}(x) = \frac{1}{N_f} \sum_{n=0}^{\infty} N_f^n I_{\pm,n}$$
$$= \frac{I_{\pm,0}}{N_f} + N_f(I_{\pm,1}^{(1)} + I_{\pm,2}^{(0)} + I_{\pm,3}^{(-1)}) + N_f^2(I_{\pm,1}^{(2)} + I_{\pm,2}^{(1)} + I_{\pm,3}^{(0)} + \cdots) + \mathcal{O}(N_f^3). \tag{4.14}$$

where $J_z^{\pm}(z) = \partial J^{\pm}(z)/\partial z$.

---

[13]This term makes the problem much harder than that in the ABJM case. We will comment on the difficulty of the higher order computation later.

The leading term is given by

$$J_{0,z}^+(z) = \frac{2}{\pi z} \arcsin\left(\frac{z}{2}\right), \qquad J_{0,z}^-(z) = -\frac{2}{\pi^2 z} \arcsin^2\left(\frac{z}{2}\right). \qquad (4.15)$$

This result indeed agrees with (3.73) in [27]. The next-to-leading term is also given by

$$J_{1,z}^+(z) = -\frac{\pi(16 - z^2)}{12(4 - z^2)^{5/2}},$$
$$J_{1,z}^-(z) = \frac{z(8 - z^2)}{8(4 - z^2)^2} + \frac{16 - z^2}{6(4 - z^2)^{5/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right), \qquad (4.16)$$

After integrating over z, we finally obtain

$$J_1(z) = \frac{z^2}{24(4 - z^2)} - \frac{z(16 - 3z^2)}{48(4 - z^2)^{3/2}} \left[\pi - 2\arctan\left(\frac{z}{\sqrt{4 - z^2}}\right)\right], \qquad (4.17)$$

where we have fixed the integration constant such that $J_1(0) = 0$. Note that $J_1^+(z)$ and $J_1^-(z)$ have branch cuts for $|z| > 2$. However, the branch cut along $z > 2$ disappears in the total potential $J_1(z)$, and there is no discontinuity in $J_1(z)$ along the positive real axis in the z-plane. In the large z (or μ) limit, $J_0(z)$ and $J_1(z)$ behave as

$$J_0(z) = \frac{2\mu^3}{3\pi^2} + \frac{\mu}{2} + \frac{\zeta(3)}{\pi^2} + \frac{2\mu + 1}{\pi^2} e^{-2\mu} + \frac{12\mu - 1}{8\pi^2} e^{-4\mu} + \mathcal{O}(\mu e^{-6\mu}),$$
$$J_1(z) = -\frac{\mu}{8} - \frac{1}{24} - \frac{2\mu + 1}{24} e^{-2\mu} + \frac{12\mu - 19}{48} e^{-4\mu} + \mathcal{O}(\mu e^{-6\mu}). \qquad (4.18)$$

The cubic and linear terms in μ are consistent with the large μ behavior in (3.8). The constant terms also match the expansion of $A(N_f)$ in (3.16). The exponentially suppressed terms will be used to fix the membrane instanton correction $f_{\ell,0}(N_f, \mu)$.

Let us remark on the higher order corrections. The next-to-next-to-leading correction is given by

$$J_{3/2,z}^{\pm}(z) = I_{\pm,1}^{(2)} + I_{\pm,2}^{(1)} + I_{\pm,3}^{(0)} + \cdots = \sum_{n=1}^{\infty} I_{\pm,n}^{(-n+3)}. \qquad (4.19)$$

Thus to compute $J_{3/2}(z)$, we need the infinite series of corrections $I_{\pm,n}^{(-n+3)}$ ($n \geq 1$). This means that it is very difficult to compute the semi-classical expansion beyond this order in this approach. We need a more efficient way to resolve this problem.

One possible way is to expand the density matrix around $N_f = 0$ from the beginning. Let us see it briefly. We use the identity

$$\frac{1}{(2\cosh\frac{x}{2N_f})^{N_f/2}} = \exp\left[-\frac{N_f}{2}\log\left(2\cosh\frac{x}{2N_f}\right)\right] = e^{-\frac{|x|}{4}} \sum_{k=0}^{\infty} \frac{N_f^k}{k!} \ell^k(x), \qquad (4.20)$$

where

$$\ell(x) = -\frac{1}{2}\log(1 + e^{-|x|/N_f}). \qquad (4.21)$$

We consider the rescaled density matrix

$$
\begin{aligned}
\widetilde{\rho}(x_1, x_2) &= \frac{1}{N_f} \rho\left(\frac{x_1}{N_f}, \frac{x_2}{N_f}\right) \\
&= \frac{1}{2\pi N_f} \frac{1}{(2\cosh\frac{x_1}{2N_f})^{N_f/2}} \frac{1}{(2\cosh\frac{x_2}{2N_f})^{N_f/2}} \frac{1}{2\cosh(\frac{x_1-x_2}{2N_f})}.
\end{aligned}
\tag{4.22}
$$

It is easy to check the equality $\operatorname{Tr}\widetilde{\rho}^n = \operatorname{Tr}\rho^n$. Then, the density matrix is expanded as

$$
\begin{aligned}
\widetilde{\rho}(x_1, x_2) &= \frac{1}{2\pi N_f} \frac{e^{-\frac{|x_1|}{4} - \frac{|x_2|}{4}}}{2\cosh(\frac{x_1-x_2}{2N_f})} \sum_{k=0}^{\infty} \frac{N_f^k}{k!} (\ell(x_1) + \ell(x_2))^k \\
&= \rho_0(x_1, x_2) + N_f \rho_1(x_1, x_2) + N_f^2 \rho_2(x_1, x_2) + \cdots,
\end{aligned}
\tag{4.23}
$$

where

$$
\rho_0(x_1, x_2) = \frac{1}{2\pi N_f} \frac{e^{-\frac{|x_1|}{4} - \frac{|x_2|}{4}}}{2\cosh(\frac{x_1-x_2}{2N_f})}, \qquad \rho_k(x_1, x_2) = \frac{1}{k!} \rho_0(x_1, x_2)(\ell(x_1) + \ell(x_2))^k. \tag{4.24}
$$

The trace of $\widetilde{\rho}$ can be computed as follows:

$$
\begin{aligned}
Z_n = \operatorname{Tr}\widetilde{\rho}^n &= \operatorname{Tr}[(\rho_0 + N_f \rho_1 + N_f^2 \rho_2^2 + \cdots)^n] \\
&= \operatorname{Tr}\rho_0^n + n N_f \operatorname{Tr}\rho_0^{n-1}\rho_1 + n N_f^2 \operatorname{Tr}\rho_0^{n-1}\rho_2 \\
&\quad + N_f^2[\operatorname{Tr}\rho_0^{n-2}\rho_1^2 + (\text{permutations})] + \mathcal{O}(N_f^3).
\end{aligned}
\tag{4.25}
$$

Thus we find

$$
\begin{aligned}
J(N_f, z) = -\sum_{n=1}^{\infty} \frac{(-z)^n}{n} Z_n &= -\sum_{n=1}^{\infty} \frac{(-z)^n}{n} [\operatorname{Tr}\rho_0^n + n N_f \operatorname{Tr}\rho_0^{n-1}\rho_1 + \cdots] \\
&= J_{\mathrm{ud}}(N_f, z) - N_f \sum_{n=1}^{\infty} (-z)^n \operatorname{Tr}\rho_0^{n-1}\rho_1 + \cdots
\end{aligned}
\tag{4.26}
$$

where $J_{\mathrm{ud}}(N_f, z)$ is the grand potential for the undeformed kernel ρ_0. The second term is rewritten as

$$
\begin{aligned}
\sum_{n=1}^{\infty} (-z)^n \operatorname{Tr}\rho_0^{n-1}\rho_1 &= 2\sum_{n=1}^{\infty} (-z)^n \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 \, \rho_0^{n-1}(x_1, x_2)\rho_0(x_2, x_1)\ell(x_1) \\
&= 2\sum_{n=1}^{\infty} (-z)^n \int_{-\infty}^{\infty} dx \, \rho_0^n(x, x)\ell(x) \\
&= -2\int_{-\infty}^{\infty} dx \frac{z\rho_0}{1 + z\rho_0}(x, x)\ell(x).
\end{aligned}
\tag{4.27}
$$

The semi-classical expansion of $\rho_0^n(x, x)$ can be computed from the TBA for $\rho_0(x)$.

An advantage of this approach is that the undeformed kernel ρ_0 does not contain the "non-perturbative" term in (4.8). It is interesting to consider whether this approach resolves the difficulty of the computation of $J_{3/2}(z)$.

4.2 Pole cancellation mechanism

The worldsheet instanton correction conjectured in section 3 has singularities for some (in particular, integral) values of N_f. These singularities must be canceled by the other contributions because the partition function itself is always finite. This pole cancellation mechanism was first found in the ABJM theory [15]. As emphasized in [16], this mechanism is conceptually important because it implies that the 't Hooft expansion breaks down at some finite values of the string coupling g_s. One needs to consider a non-perturbative completion to cure the divergences. Our conjecture (3.24) shows that this mechanism also exists in the N_f matrix model.

A technical merit of this mechanism is that we can know the pole structure of the membrane instanton correction from that of the worldsheet instanton correction. As an example, let us consider the order $\mathcal{O}(e^{-2\mu})$ correction for $N_f = 2$. Looking at (3.10), there are two contributions at this order. One is the worldsheet one-instanton correction $f_{0,1}$, and the other is the membrane one-instanton correction $f_{1,0}$. It is easy to see that the worldsheet one-instanton correction in (3.24) has the following singularity at $N_f = 2$.

$$\lim_{N_f \to 2} f_{0,1}(N_f, \mu) e^{-\frac{4\mu}{N_f}} = -\frac{2(2\mu+1)}{\pi^2(N_f - 2)} e^{-2\mu} - \frac{2(2\mu^2 + 2\mu + 1)}{\pi^2} e^{-2\mu} + \mathcal{O}(N_f - 2). \quad (4.28)$$

This singularity must be canceled by the membrane one-instanton correction. This means that $f_{1,0}$ must behave as

$$\lim_{N_f \to 2} f_{1,0}(N_f, \mu) = \frac{2(2\mu+1)}{\pi^2(N_f - 2)} + \mathcal{O}(1). \quad (4.29)$$

Similarly, the singularities of $f_{1,0}$ at $N_f = 4$ and $N_f = 6$ are determined by the $f_{0,2}$ and $f_{0,3}$, respectively. It is easy to find

$$\lim_{N_f \to 4} f_{1,0}(N_f, \mu) = \frac{6(2\mu+1)}{\pi^2(N_f - 4)} + \mathcal{O}(1),$$
$$\lim_{N_f \to 6} f_{1,0}(N_f, \mu) = \frac{20(2\mu+1)}{\pi^2(N_f - 6)} + \mathcal{O}(1). \quad (4.30)$$

We will use these results to fix $f_{1,0}(N_f, \mu)$.

4.3 Fixing the leading membrane instanton correction

Now we are in position to conjecture an exact form of $f_{1,0}(N_f, \mu)$. The large μ expansion of the semi-classical results $J_0(z)$ and $J_1(z)$ suggests that the membrane ℓ-instanton coefficient $f_{\ell,0}$ is a linear function of μ for all ℓ

$$f_{\ell,0}(N_f, \mu) = a_\ell(N_f)\mu + b_\ell(N_f). \quad (4.31)$$

To fix $f_{1,0}(N_f, \mu)$, we use the following three constraints:

- The singularity structure at $N_f = 2, 4, 6$ is given by (4.29) and (4.30).

- As noted in subsection 3.4, $f_{1,0}(N_f, \mu)$ must vanish for odd N_f.
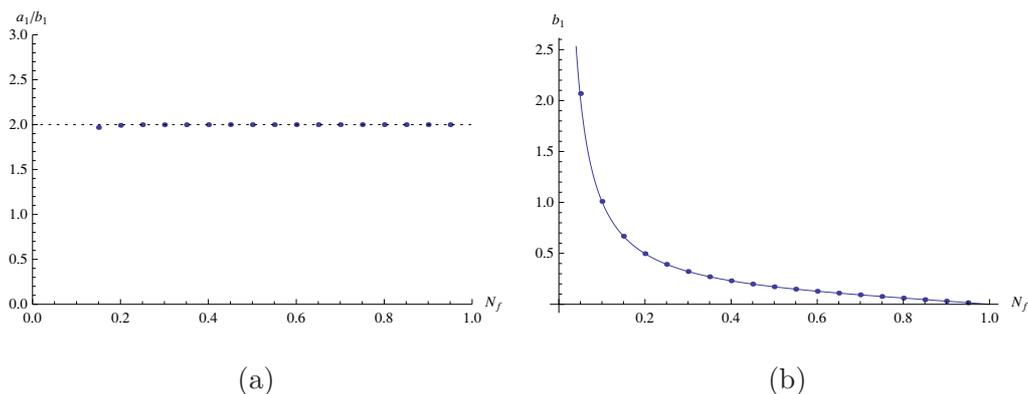
Figure 5. (a) We plot the ratio $a_1(N_f)/b_1(N_f)$ in (4.33) against N_f. These values are estimated by using the TBA equations (2.10) and (2.11). The ratio does not depend on N_f. (b) We compare the analytic conjecture (4.34) with the numerically estimated values from TBA.

- The semi-classical expansion of $f_{1,0}(N_f, \mu)$ is given by

$$f_{1,0}(N_f, \mu) = \frac{2\mu + 1}{\pi^2 N_f} - \frac{2\mu + 1}{24} N_f + \mathcal{O}(N_f^2) \qquad (N_f \to 0). \qquad (4.32)$$

The results (4.29), (4.30) and (4.32) strongly suggest that $f_{1,0}(N_f, \mu)$ has the following universal form for any N_f,

$$f_{1,0}(N_f, \mu) = b_1(N_f)(2\mu + 1), \qquad a_1(N_f) = 2b_1(N_f). \qquad (4.33)$$

We can check that this is indeed the case. In figure 5(a), we show the ratio $a_1(N_f)/b_1(N_f)$ computed numerically from the TBA equations[14] (2.10) and (2.11) for $N_f = 3/20, 4/20, \ldots, 19/20$. The ratio does not depend on N_f, and is very close to 2.

Combining all the above results, we arrive at a conjecture of $b_1(N_f)$,

$$b_1(N_f) = -\frac{\Gamma^2(-N_f/2)}{4\pi^2 \Gamma(-N_f)}. \qquad (4.34)$$

The small N_f expansion reads

$$b_1(N_f) = \frac{1}{\pi^2 N_f} - \frac{N_f}{24} - \frac{\zeta(3)N_f^2}{4\pi^2} - \frac{\pi^2 N_f^3}{640} + \left(\frac{\zeta(3)}{96} - \frac{3\zeta(5)}{16\pi^2} \right) N_f^4 + \mathcal{O}(N_f^5). \qquad (4.35)$$

Of course, this reproduces (4.32). One non-trivial check of the conjecture (4.34) is to compute the finite parts for $N_f = 2, 4, 6$. For example, in the limit $N_f \to 2$, one finds

$$\lim_{N_f \to 2} (f_{1,0}(N_f, \mu)e^{-2\mu} + f_{0,1}(N_f, \mu)e^{-\frac{4\mu}{N_f}}) = -\frac{4\mu^2 + 2\mu + 1}{\pi^2} e^{-2\mu}. \qquad (4.36)$$

---

[14]We first compute $\mathrm{Tr}\, \rho^n$ numerically from the TBA equations as in [33], and read off the numerical values of the partition function. Then, we evaluate the values of the coefficients $a_1(N_f)$ and $b_1(N_f)$ in the same way in subsection 3.4.
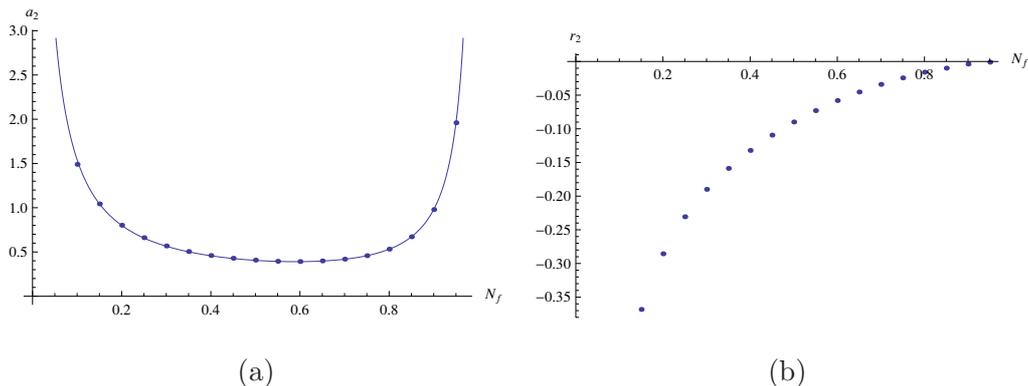
Figure 6. We plot (a) $a_2(N_f)$ and (b) $r_2(N_f) = b_2(N_f) - a_2(N_f)/4$ against N_f.

This result indeed reproduces the correct coefficient of $e^{-2\mu}$ in (B.2). In this way, one can also check that the coefficients of $e^{-2\mu}$ in (B.4) and (B.5) are reproduced by the combinations $f_{1,0}e^{-2\mu} + f_{0,2}e^{-\frac{8\mu}{N_f}}$ ($N_f \to 4$) and $f_{1,0}e^{-2\mu} + f_{0,3}e^{-\frac{12\mu}{N_f}}$ ($N_f \to 6$), respectively. As another check, we compare the numerical values computed from TBA with our conjecture (4.34) for various values of N_f. The result is shown in figure 5(b). These tests present strong supports for our conjecture (4.34).

4.4 Remark on higher instanton corrections

Let us remark on the higher instanton corrections. So far, we do not know how to determine $a_\ell(N_f)$ and $b_\ell(N_f)$ systematically. Let us consider the next simplest case: $\ell = 2$. From (4.18), these must have the small N_f expansions,

$$a_2(N_f) = \frac{3}{2\pi^2 N_f} + \frac{1}{4}N_f + \mathcal{O}(N_f^2), \qquad b_2(N_f) = -\frac{1}{8\pi^2 N_f} - \frac{19}{48}N_f + \mathcal{O}(N_f^2). \quad (4.37)$$

Furthermore, $f_{2,0}(N_f, \mu)$ must cancel the poles of $f_{0,1}(N_f, \mu)$ at $N_f = 1$ and $f_{0,3}(N_f, \mu)$ at $N_f = 3$. These conditions give the constraints

$$\lim_{N_f \to 1} a_2(N_f) = -\frac{1}{\pi^2(N_f - 1)} + \mathcal{O}(1), \qquad \lim_{N_f \to 1} b_2(N_f) = -\frac{1}{4\pi^2(N_f - 1)} + \mathcal{O}(1),$$
$$\lim_{N_f \to 3} a_2(N_f) = -\frac{10}{\pi^2(N_f - 3)} + \mathcal{O}(1), \qquad \lim_{N_f \to 3} b_2(N_f) = -\frac{5}{2\pi^2(N_f - 3)} + \mathcal{O}(1). \quad (4.38)$$

We find an analytic form of $a_2(N_f)$ satisfying all these constraints

$$a_2(N_f) = -\frac{1}{4\pi^2}\left(1 + \frac{2}{\cos \pi N_f}\right)\frac{\Gamma^2(-N_f)}{\Gamma(-2N_f)}. \quad (4.39)$$

This guess correctly reproduces the finite parts for $N_f = 1, 3$ in appendix B. It also shows a good agreement with the numerical values from TBA as shown in figure 6(a).

We also want to fix $b_2(N_f)$. Unfortunately, the ratio $b_2(N_f)/a_2(N_f)$ is no longer a constant unlike the one-instanton correction. Instead, it is convenient to define a new function by

$$r_\ell(N_f) = b_\ell(N_f) - \frac{a_\ell(N_f)}{2\ell}. \quad (4.40)$$

In other words, we parametrize the membrane ℓ-instanton coefficient as

$$f_{\ell,0}(N_f, \mu) = \left(\mu + \frac{1}{2\ell}\right) a_\ell(N_f) + r_\ell(N_f). \tag{4.41}$$

For $\ell = 1$, we found in (4.33) that the last term is absent: $r_1(N_f) = 0$. For $\ell \geq 2$, it turns out that $r_\ell(N_f)$ is a non-trivial function of N_f. From the pole structure (4.38), one finds that the function $r_2(N_f)$ is regular at $N_f = 1, 3$. Furthermore, in order to reproduce the finite parts for $N_f = 1, 3$ in (B.1) and (B.3) correctly, $r_2(N_f)$ must vanish in the limit $N_f \to 1, 3$:

$$\lim_{N_f \to 1} r_2(N_f) = \lim_{N_f \to 3} r_2(N_f) = 0. \tag{4.42}$$

Also, $r_2(N_f)$ must have the following semi-classical expansion

$$r_2(N_f) = -\frac{1}{2\pi^2 N_f} - \frac{11}{24} N_f + \mathcal{O}(N_f^2) \qquad (N_f \to 0). \tag{4.43}$$

In figure 6(b), we plot $r_2(N_f)$ for $0 < N_f < 1$ as a function of N_f. As expected, $r_2(N_f)$ goes to zero as $N_f \to 1$. So far, we have not found an exact form of $r_2(N_f)$. To fix it, we might need more information.

Interestingly, the combination $(\mu + \frac{1}{2n}) a_n(N_f)$ seems to cancel the poles of the worldsheet instantons when there are no bound state contributions. For example, for $N_f = 2/n$ $(n = 1, 2, 3, \dots)$, the worldsheet 1-instanton correction has the following pole structure:

$$\lim_{N_f \to 2/n} f_{0,1}(N_f, \mu) e^{-\frac{4\mu}{N_f}} = (-1)^n \frac{4(\mu + \frac{1}{2n})}{n^2 \pi^2 (N_f - \frac{2}{n})} e^{-2n\mu} + \mathcal{O}(1). \tag{4.44}$$

This pole must be canceled by the membrane n-instanton correction because there are no bound state contributions at this order. Thus, focusing on the $\mu e^{-2n\mu}$ term, one finds that $a_n(N_f)$ must behave at $N_f = 2/n$ as

$$\lim_{N_f \to 2/n} a_n(N_f) = (-1)^{n-1} \frac{4}{n^2 \pi^2 (N_f - \frac{2}{n})} + \mathcal{O}(1), \tag{4.45}$$

and that the combination $(\mu + \frac{1}{2n}) a_n(N_f)$ indeed cancels the pole in (4.44). Therefore we conclude that $r_n(N_f)$ is regular at $N_f = 2/n$:

$$\lim_{N_f \to 2/n} r_n(N_f) = \mathcal{O}(1). \tag{4.46}$$

It would be interesting to clarify the analytic structure of $r_n(N_f)$ in more detail.

5 Conclusions

In this paper we have studied the large N non-perturbative effects in the N_f matrix model. Combining the exact computation of the partition function and the analysis of TBA, we have successfully determined a first few terms of both the worldsheet instanton corrections and the membrane instanton corrections, as analytic functions of N_f. These analytic results

show that the pole cancellation mechanism, originally found in the ABJM model [15], works also for the N_f matrix model. As emphasized in [27], this clearly shows that the 't Hooft expansion alone is incomplete and the existence of both membrane instantons and worldsheet instantons is necessary for the non-perturbative definition of the theory and the free energy to be finite at the physical value of the coupling. We would like to emphasize that this mechanism is invisible as long as one focuses on the small/large N_f limit. One needs to consider the M-theory (or finite coupling) regime to see it. It is desirable to check that this mechanism works more generally, by computing the higher instanton coefficients of worldsheet instantons, membrane instantons, and their bound states. It is also interesting to apply the Fermi-gas approach to BPS Wilson loops along the lines of [40–42]. Moreover, it would be interesting to explore the non-perturbative structure for other examples [28, 43].

To proceed further, we need to develop a systematic way to study this model. In particular, it is very interesting to find an efficient method to compute the small N_f expansion of the grand potential, which gives important clues to determine the analytic form of the membrane instantons. As for the worldsheet instantons, using the technique of ordinary matrix models we can in principle compute the genus expansion in the 't Hooft limit, order by order in $1/N_f$. However, to discuss the pole cancellation, we need to resum this series à la Gopakumar-Vafa, which is a formidable task. Also, currently we do not have enough information to study the effect of bound states of worldsheet instantons and membrane instantons. We leave them as interesting future problems.

We found that the structure of the instanton corrections in the N_f matrix model is quite different from that of the ABJM model. In the ABJM case, the membrane instanton correction and the worldsheet instanton correction have the following form [14, 15]

$$
\begin{aligned}
J_{\mathrm{ABJM}}^{\mathrm{M2}}(k,\mu) &= \sum_{\ell=1}^{\infty} (a_\ell(k)\mu^2 + b_\ell(k)\mu + c_\ell(k))e^{-2\ell\mu}, \\
J_{\mathrm{ABJM}}^{\mathrm{WS}}(k,\mu) &= \sum_{m=1}^{\infty} d_m(k)e^{-\frac{4m\mu}{k}},
\end{aligned}
\tag{5.1}
$$

where all the coefficients $a_\ell(k)$, $b_\ell(k)$, $c_\ell(k)$ and $d_m(k)$ are expressed in terms of a combination of the trigonometric functions. On the other hand, in the case of N_f matrix model, we conjecture that the membrane instanton correction and the worldsheet instanton correction have the following structure

$$
\begin{aligned}
J^{\mathrm{M2}}(N_f,\mu) &= \sum_{\ell=1}^{\infty} (a_\ell(N_f)\mu + b_\ell(N_f))e^{-2\ell\mu}, \\
J^{\mathrm{WS}}(N_f,\mu) &= \sum_{m=1}^{\infty} \left(\sum_{\sum lk_l \le m} d_m^{(k_1,\cdots,k_m)}(N_f) \prod_{l=1}^{m} P_l^{k_l}(N_f,\mu) \right) e^{-\frac{4m\mu}{N_f}},
\end{aligned}
\tag{5.2}
$$

where $P_l(N_f,\mu)$ is defined in (3.23). We found that the membrane instanton coefficients in (4.34) and (4.39) are expressed in terms of the gamma function (and the trigonometric functions). As another difference, the pre-factor of the worldsheet instanton correction depends on μ, unlike the ABJM case. More precisely, the worldsheet m-instanton coefficient

is given by an mth order polynomial of μ, and the coefficient $d_m^{(k_1, \cdots, k_m)}(N_f)$ appearing in (5.2) is a combination of the trigonometric functions (3.24). This difference between the N_f matrix model and the ABJM model may be related to the difference of the orbifolding on the bulk side: the orbifold $\mathbb{C}^2 \times \mathbb{C}^2/\mathbb{Z}_{N_f}$, corresponding to the N_f matrix model, has a family of A_{N_f-1} ALE singularity parametrized by the first factor $\mathbb{C}^2$, while the singularity of $\mathbb{C}^4/\mathbb{Z}_k$ for the ABJM case is isolated. Perhaps, the structure of the worldsheet instantons in the N_f matrix model might be understood as the effect of a non-isolated family of rational curves (see [44] for such worldsheet instanton effects in a heterotic string compactification).

The spectral problem in the Fermi-gas is also important. The spectrum of the one-dimensional Fermi-gas that we are considering is determined by the Fredholm integral equation of the first kind [14]:

$$\int_{-\infty}^{\infty} dx' \rho(x, x') \phi(x') = e^{-E} \phi(x). \tag{5.3}$$

Since the density matrix $\rho(x_1, x_2)$ is a non-negative Hilbert-Schmidt operator, the integral equation (5.3) has a positive discrete spectrum. In the ABJ(M) Fermi-gas, the spectrum is determined by the exact WKB quantization condition [19, 24] (see also [45]), in which one has to consider not only the perturbative contribution but also the non-perturbative contribution in the Planck constant $\hbar = 2\pi k$ to reproduce the correct spectrum. Since all the information of the (grand) partition function is encoded in the Fermi-gas spectrum, it is important to find the exact WKB quantization condition in the N_f matrix model.

In the case of ABJ(M) model, we have a very detailed understanding of the instanton corrections thanks to the relation to the refined topological string on local $\mathbb{P}^1 \times \mathbb{P}^1$ [18, 19, 22–24]. This relation to the topological string is widely viewed as an accident of the ABJ(M) model. However, in view of the non-trivial relation (3.12) between the constant term $A(N_f)$ of the N_f matrix model and the constant map contribution $A_{\text{const}}(k)$ of the topological string, it is tempting to speculate that the N_f matrix model also has a hidden connection to the topological string on some background. It would be interesting to see if such a hidden connection to the topological string really exists, or not.

Finally, in this paper, we have probed the non-perturbative effects in M-theory from its gauge theory dual. Recently, there are some interesting progress on the gravity side [46, 47]. It would be very significant to confirm our pole cancellation mechanism directly in M-theory in the future.

A A simpler expression of the constant map

In this appendix, we derive the integral expression (3.13) of the constant map $A_{\mathrm{const}}(k)$ in the ABJM Fermi-gas (or equivalently in the topological string on local $\mathbb{P}^1 \times \mathbb{P}^1$). A similar integral expression was found in [30], but our expression is much simpler than theirs. Our starting point is the all-order small k expansion found in [30],

$$A_{\mathrm{const}}(k) = \frac{2\zeta(3)}{\pi^2 k} + \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} B_{2n} B_{2n-2} \pi^{2n-2} k^{2n-1}. \tag{A.1}$$

To rewrite this as an integral form, we use the identity for the Bernoulli number:

$$B_{2n} = (-1)^{n-1} 4n \int_0^{\infty} dx \frac{x^{2n-1}}{e^{2\pi x} - 1} \quad (n \geq 1). \tag{A.2}$$

This identity is simply obtained from the integral expression of $\zeta(z)$ by setting $z = 2n$. Then, we get

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} B_{2n} B_{2n-2} \pi^{2n-2} k^{2n-1} = -\frac{2}{\pi} \int_0^{\infty} dx \frac{1}{e^{2\pi x} - 1} \sum_{n=1}^{\infty} \frac{B_{2n-2}}{(2n-1)!} (\pi k x)^{2n-1}. \tag{A.3}$$

The sum can be performed,

$$\sum_{n=1}^{\infty} \frac{B_{2n-2}}{(2n-1)!} z^{2n-1} = \frac{z^2}{4} + \mathrm{Li}_2(1 - e^{-z}). \tag{A.4}$$

Therefore we obtain the integral form

$$A_{\mathrm{const}}(k) = \frac{2\zeta(3)}{\pi^2 k} \left(1 - \frac{k^3}{16} \right) - \frac{2}{\pi} \int_0^{\infty} dx \frac{\mathrm{Li}_2(1 - e^{-\pi k x})}{e^{2\pi x} - 1}. \tag{A.5}$$

After integration by parts, we finally get (3.13).

B Some explicit results

B.1 Corrections for integral N_f

Here we summarize the explicit forms of $J_{\mathrm{np}}(N_f, \mu)$ for $N_f = 1, 2, 3, 4, 6, 8, 12$. For $N_f = 1$, the partition function is equivalent to the one in the ABJM theory at $k = 1$. We can use the result in [17],[15]

$$\begin{aligned}
J_{\mathrm{np}}(1, \mu) = &\left[\frac{4\mu^2 + \mu + 1/4}{\pi^2} \right] e^{-4\mu} + \left[-\frac{52\mu^2 + \mu/2 + 9/16}{2\pi^2} + 2 \right] e^{-8\mu} \\
&+ \left[\frac{736\mu^2 - 152\mu/3 + 77/18}{3\pi^2} - 32 \right] e^{-12\mu} \\
&+ \left[-\frac{2701\mu^2 - 13949\mu/48 + 11291/768}{\pi^2} + 466 \right] e^{-16\mu} \\
&+ \left[\frac{161824\mu^2 - 317122\mu/15 + 285253/300}{5\pi^2} - 6720 \right] e^{-20\mu} \\
&+ \left[-\frac{1227440\mu^2 - 2686522\mu/15 + 631257/80}{3\pi^2} + \frac{292064}{3} \right] e^{-24\mu} + \mathcal{O}(e^{-28\mu}).
\end{aligned} \tag{B.1}$$

---

[15]There is a typo in version 2 of [17]. The term $-292064/3$ in the coefficient of $e^{-24\mu}$ must be $+292064/3$.

Similarly, in the case of $N_f = 2$, we can use the ABJ result in [22],[16]

$$
\begin{aligned}
J_{\mathrm{np}}(2,\mu) = {} & \left[-\frac{4\mu^2 + 2\mu + 1}{\pi^2}\right]e^{-2\mu} + \left[-\frac{52\mu^2 + \mu + 9/4}{2\pi^2} + 2\right]e^{-4\mu} \\
& + \left[-\frac{736\mu^2 - 304\mu/3 + 154/9}{3\pi^2} + 32\right]e^{-6\mu} \\
& + \left[-\frac{2701\mu^2 - 13949\mu/24 + 11291/192}{\pi^2} + 466\right]e^{-8\mu} \\
& + \left[-\frac{161824\mu^2 - 634244\mu/15 + 285253/75}{5\pi^2} + 6720\right]e^{-10\mu} \\
& + \left[-\frac{1227440\mu^2 - 5373044\mu/15 + 631257/20}{3\pi^2} + \frac{292064}{3}\right]e^{-12\mu} + \mathcal{O}(e^{-14\mu}).
\end{aligned} \tag{B.2}
$$

For $N_f = 3, 4, 6, 8, 12$, we find new results

$$
\begin{aligned}
J_{\mathrm{np}}(3,\mu) = {} & -\frac{4\mu + 3}{\sqrt{3}\pi}e^{-4\mu/3} + \left[-\frac{(4\mu+3)^2}{4\pi^2} + \frac{2}{3}\right]e^{-8\mu/3} \\
& + \left[-\frac{(4\mu+3)^3}{8\sqrt{3}\pi^3} + \frac{76\mu^2}{3\pi^2} + \frac{47(4\mu+1)}{24\pi^2} - \frac{4\mu+3}{\sqrt{3}\pi} - 8\right]e^{-4\mu} \\
& + \left[-\frac{(4\mu+3)^4}{32\pi^4} - \frac{(4\mu+3)^2}{2\pi^2} + \frac{166\mu + 133/8}{\sqrt{3}\pi} + \frac{2}{3}\right]e^{-16\mu/3} \\
& + \left[-\frac{\sqrt{3}(4\mu+3)^5}{160\pi^5} - \frac{(4\mu+3)^3}{\sqrt{3}\pi^3} + \frac{332\mu^2 + 1129\mu/4 + 399/16}{\pi^2}\right. \\
& \left. + \frac{6(4\mu+3)}{\sqrt{3}\pi} - 48\right]e^{-20\mu/3} + \mathcal{O}(e^{-8\mu}),
\end{aligned} \tag{B.3}
$$

$$
\begin{aligned}
J_{\mathrm{np}}(4,\mu) = {} & -\frac{4\mu+4}{2\pi}e^{-\mu} + \left[-\frac{10\mu^2 + 7\mu + 7/2}{\pi^2} + 1\right]e^{-2\mu} - \frac{88\mu + 52/3}{3\pi}e^{-3\mu} \\
& + \left[-\frac{269\mu^2 + 193\mu/4 + 265/16}{\pi^2} + 58\right]e^{-4\mu} \\
& - \frac{4792\mu + 1102/5}{5\pi}e^{-5\mu} + \mathcal{O}(e^{-6\mu}),
\end{aligned} \tag{B.4}
$$

$$
\begin{aligned}
J_{\mathrm{np}}(6,\mu) = {} & -\frac{4\mu+6}{\sqrt{3}\pi}e^{-2\mu/3} + \left[-\frac{(4\mu+6)^2}{4\pi^2} + \frac{2}{3}\right]e^{-4\mu/3} \\
& + \left[-\frac{(4\mu+6)^3}{8\sqrt{3}\pi^3} - \frac{76\mu^2}{3\pi^2} - \frac{47(2\mu+1)}{6\pi^2} - \frac{4\mu+6}{\sqrt{3}\pi} + 8\right]e^{-2\mu} \\
& + \left[-\frac{(4\mu+6)^4}{32\pi^4} - \frac{(4\mu+6)^2}{2\pi^2} - \frac{166\mu + 133/4}{\sqrt{3}\pi} + \frac{2}{3}\right]e^{-8\mu/3} \\
& + \left[-\frac{\sqrt{3}(4\mu+6)^5}{160\pi^5} - \frac{(4\mu+6)^3}{\sqrt{3}\pi^3} - \frac{332\mu^2 + 1129\mu/2 + 399/4}{\pi^2}\right. \\
& \left. + \frac{6(4\mu+6)}{\sqrt{3}\pi} + 48\right]e^{-10\mu/3} + \mathcal{O}(e^{-4\mu}),
\end{aligned} \tag{B.5}
$$

---

[16]For $N_f = 1, 2$, one can check that the coefficients of $J_{\mathrm{np}}(\mu)$ are reproduced from the general formula in [23] by plugging in the explicit values of the refined BPS invariants $N^{d_1, d_2}_{j_L, j_R}$ of local $\mathbb{P}^1 \times \mathbb{P}^1$.

$$J_{\text{np}}(8,\mu) = -\frac{4\mu+8}{\sqrt{2}\pi}e^{-\mu/2} + \left[-\frac{(4\mu+8)^2}{4\pi^2} + \frac{3(4\mu+4)}{4\pi}\right]e^{-\mu} \tag{B.6}$$

$$+ \left[-\frac{(4\mu+8)^3}{6\sqrt{2}\pi^3} + 4\sqrt{2}\right]e^{-3\mu/2} + \mathcal{O}(e^{-2\mu}),$$

$$J_{\text{np}}(12,\mu) = -\frac{4\mu+12}{\pi}e^{-\mu/3} + \left[-\frac{(4\mu+12)^2}{4\pi^2} + \frac{\sqrt{3}(8\mu+12)}{2\pi} - 2\right]e^{-2\mu/3} \tag{B.7}$$

$$+ \left[-\frac{(4\mu+12)^3}{8\pi^3} + \frac{3\sqrt{3}(4\mu+12)(8\mu+12)}{8\pi^2}\right.$$

$$\left. - \frac{32\mu+56}{3\pi} + \frac{8}{\sqrt{3}}\right]e^{-\mu} + \mathcal{O}(e^{-4\mu/3}).$$

B.2 Semi-classical solutions of TBA

Here we give the solutions (4.10) of the TBA equations in the semi-classical limit. For simplicity, we introduce

$$L(x) = \log(1 + e^{-\frac{2|x|}{N_f}}), \tag{B.8}$$

Note that the solutions below are valid for $x > 0$. Since the solutions are invariant under $x \to -x$, it is easy to know the solutions for $x < 0$. The solutions up to $n = 3$ are given by

$$r_0(x) = \frac{2}{\sqrt{4e^x - z^2}}, \qquad \eta_0(x) = -\frac{z}{\sqrt{4e^x - z^2}},$$

$$t_0(x) = -\frac{2}{\pi}\arctan\left(\frac{z}{\sqrt{4e^x - z^2}}\right), \tag{B.9}$$

$$r_1(x) = -\frac{4e^x L(x)}{(4e^x - z^2)^{3/2}}, \quad \eta_1(x) = \frac{2e^x z L(x)}{(4e^x - z^2)^{3/2}},$$

$$t_1(x) = \frac{zL(x)}{\pi\sqrt{4e^x - z^2}}, \tag{B.10}$$

$$r_2(x) = \frac{\pi^2 z^2 e^x(6e^x + z^2)}{2(4e^x - z^2)^{7/2}} + \frac{2e^x(2e^x + z^2)}{(4e^x - z^2)^{5/2}}L^2(x),$$

$$\eta_2(x) = -\frac{2\pi^2 z e^{2x}(e^x + z^2)}{(4e^x - z^2)^{7/2}} - \frac{z e^x(2e^x + z^2)}{(4e^x - z^2)^{5/2}}L^2(x), \tag{B.11}$$

$$t_2(x) = -\frac{\pi z e^x(16e^x + 11z^2)}{12(4e^x - z^2)^{5/2}} - \frac{z e^x}{\pi(4e^x - z^2)^{3/2}}L^2(x),$$

$$r_3(x) = -\frac{\pi^2 z^2 e^x(36e^{2x} + 22e^x z^2 + z^4)}{2(4e^x - z^2)^{9/2}}L(x) - \frac{2e^x(4e^{2x} + 10e^x z^2 + z^4)}{3(4e^x - z^2)^{7/2}}L^3(x)$$

$$+ \frac{\pi^2 z^2 e^x(6e^x + z^2)}{(4e^x - z^2)^{7/2}}L'(x) - \frac{\pi^2 e^x(4e^x + z^2)}{2(4e^x - z^2)^{5/2}}L''(x),$$

$$\eta_3(x) = \frac{2\pi^2 z e^{2x}(2e^{2x} + 9e^x z^2 + 2z^4)}{(4e^x - z^2)^{9/2}}L(x)\frac{z e^x(4e^{2x} + 10e^x z^2 + z^4)}{3(4e^x - z^2)^{7/2}}L^3(x) \tag{B.12}$$

$$- \frac{4\pi^2 z e^{2x}(e^x + z^2)}{(4e^x - z^2)^{7/2}}L'(x) + \frac{2\pi^2 z e^{2x}}{(4e^x - z^2)^{5/2}}L''(x),$$

$$t_3(x) = \frac{\pi z e^x (32 e^{2x} + 98 e^x z^2 + 11 z^4)}{12(4e^x - z^2)^{7/2}} L(x) + \frac{z e^x (2 e^x + z^2)}{3\pi (4 e^x - z^2)^{5/2}} L^3(x)$$
$$- \frac{\pi z e^x (16 e^x + 11 z^2)}{6(4 e^x - z^2)^{5/2}} L'(x) - \frac{\pi z (-28 e^x + z^2)}{24(4 e^x - z^2)^{3/2}} L''(x).$$

Using these solutions, one can compute $I_{\pm,n}$ in (4.11). The coefficients of the small N_f expansions (4.12) are given by

$$I^{(1)}_{+,1} = -\frac{\pi}{12\left(4 - z^2\right)^{3/2}}, \qquad I^{(0)}_{+,2} = \frac{\pi z^2}{4(4 - z^2)^{5/2}}, \qquad I^{(-1)}_{+,3} = -\frac{\pi\left(4 + z^2\right)}{4\left(4 - z^2\right)^{5/2}},$$

$$I^{(1)}_{-,1} = \frac{z}{24(4 - z^2)} + \frac{1}{6(4 - z^2)^{3/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right),$$

$$I^{(0)}_{-,2} = -\frac{z}{24(4 - z^2)^2} \left[8 + z^2 + \frac{12 z}{\sqrt{4 - z^2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right)\right], \tag{B.13}$$

$$I^{(-1)}_{-,3} = \frac{z(28 - z^2)}{24(4 - z^2)^2} + \frac{4 + z^2}{2(4 - z^2)^{5/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right),$$

and

$$I^{(2)}_{+,1} = \frac{3\zeta(3)\left(2 + z^2\right)}{8\pi\left(4 - z^2\right)^{5/2}}, \qquad I^{(1)}_{+,2} = \frac{\zeta(3)(2 + z^2)}{8\pi(4 - z^2)^{5/2}}, \qquad I^{(0)}_{+,3} = \frac{\pi\left(2 + z^2\right)\log 2}{4\left(4 - z^2\right)^{5/2}},$$

$$I^{(2)}_{-,1} = -\frac{3\zeta(3)}{8\pi^2} \left[\frac{3z}{(4 - z^2)^2} + \frac{2(2 + z^2)}{(4 - z^2)^{5/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right)\right],$$

$$I^{(1)}_{-,2} = -\frac{\zeta(3)}{8\pi^2} \left[\frac{3z}{(4 - z^2)^2} + \frac{2(2 + z^2)}{(4 - z^2)^{5/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right)\right], \tag{B.14}$$

$$I^{(0)}_{-,3} = -\frac{\log 2}{4} \left[\frac{3z}{(4 - z^2)^2} + \frac{2(2 + z^2)}{(4 - z^2)^{5/2}} \arctan\left(\frac{z}{\sqrt{4 - z^2}}\right)\right].$$

C 't Hooft expansion of the grand potential

Here we compute the 't Hooft expansion of the grand potential. As explained in [14, 27], the 't Hooft limit in the grand canonical ensemble corresponds to

$$\mu \to \infty, \qquad N_f \to \infty, \qquad \widehat{\mu} = \frac{\mu}{N_f} : \text{fixed}. \tag{C.1}$$

In this limit, the grand potential has the following "genus" expansion

$$J(N_f, \mu) = \sum_{g=0}^{\infty} N_f^{2-2g} \mathcal{J}_g(\widehat{\mu}). \tag{C.2}$$

As noted in [27], the genus zero contribution $\mathcal{J}_0(\widehat{\mu})$ is given by the Legendre transformation of the planar free energy $F_0(\lambda)$. Thus we have the relations

$$\mathcal{J}_0(\widehat{\mu}) = F_0(\lambda) - \lambda F_0'(\lambda), \qquad \widehat{\mu} = -F_0'(\lambda). \tag{C.3}$$

The planar free energy was computed in [27]. The result is expressed by the elliptic integral

$$F_0''(\lambda) = -2\pi \frac{K(k)}{K(\sqrt{1-k^2})}, \qquad \widehat{\lambda} = \lambda + \frac{1}{8} = \frac{(1+k)^2}{8\pi^2} K^2(\sqrt{1-k^2}), \qquad \text{(C.4)}$$

where k is the elliptic modulus. Using these relations, we find the large $\widehat{\mu}$ expansion of $\mathcal{J}_0(\widehat{\mu})$,

$$
\begin{aligned}
\mathcal{J}_0(\widehat{\mu}) = {}& \frac{2}{3\pi^2}\widehat{\mu}^3 - \frac{\widehat{\mu}}{8} + \frac{1}{\pi^2}\left[\left(-\widehat{\mu} - \frac{1}{4} \right) e^{-4\widehat{\mu}} + \left(-4\widehat{\mu}^2 + \frac{1}{4}\widehat{\mu} - \frac{7}{32} \right) e^{-8\widehat{\mu}} \right. \\
& + \left(-\frac{128}{3}\widehat{\mu}^3 + 16\widehat{\mu}^2 - \frac{46}{9}\widehat{\mu} + \frac{11}{27} \right) e^{-12\widehat{\mu}} \\
& + \left(-\frac{2048}{3}\widehat{\mu}^4 + \frac{1408}{3}\widehat{\mu}^3 - 178\widehat{\mu}^2 + \frac{513}{16}\widehat{\mu} - \frac{2005}{768} \right) e^{-16\widehat{\mu}} \\
& \left. + \left(-\frac{40960}{3}\widehat{\mu}^5 + \frac{40960}{3}\widehat{\mu}^4 - \frac{20096}{3}\widehat{\mu}^3 + 1808\widehat{\mu}^2 - \frac{20303}{75}\widehat{\mu} + \frac{6593}{375} \right) e^{-20\widehat{\mu}} + \cdots \right].
\end{aligned}
\qquad \text{(C.5)}
$$

This should be compared with the worldsheet instanton correction (3.29).

References

[1] J.M. Maldacena, *The large-N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113 [hep-th/9711200] [INSPIRE].

[2] O. Aharony, O. Bergman, D.L. Jafferis and J. Maldacena, $N = 6$ *superconformal Chern-Simons-matter theories, M2-branes and their gravity duals*, *JHEP* **10** (2008) 091 [arXiv:0806.1218] [INSPIRE].

[3] N. Seiberg, *IR dynamics on branes and space-time geometry*, *Phys. Lett.* **B 384** (1996) 81 [hep-th/9606017] [INSPIRE].

[4] N. Seiberg and E. Witten, *Gauge dynamics and compactification to three-dimensions*, hep-th/9607163 [INSPIRE].

[5] K.A. Intriligator and N. Seiberg, *Mirror symmetry in three-dimensional gauge theories*, *Phys. Lett.* **B 387** (1996) 513 [hep-th/9607207] [INSPIRE].

[6] J. de Boer, K. Hori, H. Ooguri and Y. Oz, *Mirror symmetry in three-dimensional gauge theories, quivers and D-branes*, *Nucl. Phys.* **B 493** (1997) 101 [hep-th/9611063] [INSPIRE].

[7] F. Benini, C. Closset and S. Cremonesi, *Chiral flavors and M2-branes at toric CY4 singularities*, *JHEP* **02** (2010) 036 [arXiv:0911.4127] [INSPIRE].

[8] A. Kapustin, B. Willett and I. Yaakov, *Exact results for Wilson loops in superconformal Chern-Simons theories with matter*, *JHEP* **03** (2010) 089 [arXiv:0909.4559] [INSPIRE].

[9] D.L. Jafferis, *The exact superconformal R-symmetry extremizes Z*, *JHEP* **05** (2012) 159 [arXiv:1012.3210] [INSPIRE].

[10] N. Hama, K. Hosomichi and S. Lee, *Notes on SUSY gauge theories on three-sphere*, *JHEP* **03** (2011) 127 [arXiv:1012.3512] [INSPIRE].

[11] M. Mariño and P. Putrov, *Exact results in ABJM theory from topological strings*, *JHEP* **06** (2010) 011 [arXiv:0912.3074] [INSPIRE].

[12] N. Drukker, M. Mariño and P. Putrov, *From weak to strong coupling in ABJM theory*, *Commun. Math. Phys.* **306** (2011) 511 [arXiv:1007.3837] [INSPIRE].

[13] C.P. Herzog, I.R. Klebanov, S.S. Pufu and T. Tesileanu, *Multi-matrix models and tri-Sasaki Einstein spaces*, *Phys. Rev.* **D 83** (2011) 046001 [arXiv:1011.5487] [INSPIRE].

[14] M. Mariño and P. Putrov, *ABJM theory as a Fermi gas*, *J. Stat. Mech.* **1203** (2012) P03001 [arXiv:1110.4066] [INSPIRE].

[15] Y. Hatsuda, S. Moriyama and K. Okuyama, *Instanton effects in ABJM theory from fermi gas approach*, *JHEP* **01** (2013) 158 [arXiv:1211.1251] [INSPIRE].

[16] F. Calvo and M. Mariño, *Membrane instantons from a semiclassical TBA*, *JHEP* **05** (2013) 006 [arXiv:1212.5118] [INSPIRE].

[17] Y. Hatsuda, S. Moriyama and K. Okuyama, *Instanton bound states in ABJM theory*, *JHEP* **05** (2013) 054 [arXiv:1301.5184] [INSPIRE].

[18] Y. Hatsuda, M. Mariño, S. Moriyama and K. Okuyama, *Non-perturbative effects and the refined topological string*, *JHEP* **09** (2014) 168 [arXiv:1306.1734] [INSPIRE].

[19] J. Kallen and M. Mariño, *Instanton effects and quantum spectral curves*, arXiv:1308.6485 [INSPIRE].

[20] N.A. Nekrasov and S.L. Shatashvili, *Quantization of integrable systems and four dimensional gauge theories*, arXiv:0908.4052 [INSPIRE].

[21] O. Aharony, O. Bergman and D.L. Jafferis, *Fractional M2-branes*, *JHEP* **11** (2008) 043 [arXiv:0807.4924] [INSPIRE].

[22] S. Matsumoto and S. Moriyama, *ABJ fractional brane from ABJM Wilson loop*, *JHEP* **03** (2014) 079 [arXiv:1310.8051] [INSPIRE].

[23] M. Honda and K. Okuyama, *Exact results on ABJ theory and the refined topological string*, *JHEP* **08** (2014) 148 [arXiv:1405.3653] [INSPIRE].

[24] J. Kallen, *The spectral problem of the ABJ Fermi gas*, arXiv:1407.0625 [INSPIRE].

[25] H. Awata, S. Hirano and M. Shigemori, *The partition function of ABJ theory*, *Prog. Theor. Exp. Phys.* (2013) 053B04 [arXiv:1212.2966] [INSPIRE].

[26] M. Honda, *Direct derivation of "mirror" ABJ partition function*, *JHEP* **12** (2013) 046 [arXiv:1310.3126] [INSPIRE].

[27] A. Grassi and M. Mariño, *M-theoretic matrix models*, arXiv:1403.4276 [INSPIRE].

[28] M. Mezei and S.S. Pufu, *Three-sphere free energy for classical gauge groups*, *JHEP* **02** (2014) 037 [arXiv:1312.0920] [INSPIRE].

[29] A. Kapustin, B. Willett and I. Yaakov, *Nonperturbative tests of three-dimensional dualities*, *JHEP* **10** (2010) 013 [arXiv:1003.5694] [INSPIRE].

[30] M. Hanada et al., *Numerical studies of the ABJM theory for arbitrary N at arbitrary coupling constant*, *JHEP* **05** (2012) 121 [arXiv:1202.5300] [INSPIRE].

[31] K. Okuyama, *A note on the partition function of ABJM theory on S^3*, *Prog. Theor. Phys.* **127** (2012) 229 [arXiv:1110.3555] [INSPIRE].

[32] A.B. Zamolodchikov, *Painleve III and 2D polymers*, *Nucl. Phys.* **B 432** (1994) 427 [hep-th/9409108] [INSPIRE].

[33] Y. Hatsuda, S. Moriyama and K. Okuyama, *Exact results on the ABJM Fermi gas*, *JHEP* **10** (2012) 020 [arXiv:1207.4283] [INSPIRE].

[34] P. Putrov and M. Yamazaki, *Exact ABJM partition function from TBA*, *Mod. Phys. Lett.* **A 27** (2012) 1250200 [arXiv:1207.5066] [INSPIRE].

[35] N. Drukker, M. Mariño and P. Putrov, *Nonperturbative aspects of ABJM theory*, *JHEP* **11** (2011) 141 [arXiv:1103.4844] [INSPIRE].

[36] A. Grassi, M. Mariño and S. Zakany, *Resumming the string perturbation series*, arXiv:1405.4214 [INSPIRE].

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

[38] M. Honda and S. Moriyama, *Instanton effects in orbifold ABJM theory*, *JHEP* **08** (2014) 091 [arXiv:1404.0676] [INSPIRE].

[39] H. Fuji, S. Hirano and S. Moriyama, *Summing up all genus free energy of ABJM matrix model*, *JHEP* **08** (2011) 001 [arXiv:1106.4631] [INSPIRE].

[40] A. Klemm, M. Mariño, M. Schiereck and M. Soroush, *ABJM Wilson loops in the Fermi gas approach*, arXiv:1207.0611 [INSPIRE].

[41] Y. Hatsuda, M. Honda, S. Moriyama and K. Okuyama, *ABJM Wilson loops in arbitrary representations*, *JHEP* **10** (2013) 168 [arXiv:1306.4297] [INSPIRE].

[42] S. Hirano, K. Nii and M. Shigemori, *ABJ Wilson loops and Seiberg duality*, arXiv:1406.4141 [INSPIRE].

[43] L. Anderson and K. Zarembo, *Quantum phase transitions in mass-deformed ABJM matrix model*, *JHEP* **09** (2014) 021 [arXiv:1406.3366] [INSPIRE].

[44] C. Beasley and E. Witten, *New instanton effects in string theory*, *JHEP* **02** (2006) 060 [hep-th/0512039] [INSPIRE].

[45] M.-x. Huang and X.-f. Wang, *Topological strings and quantum spectral problems*, *JHEP* **09** (2014) 150 [arXiv:1406.6178] [INSPIRE].

[46] S. Bhattacharyya, A. Grassi, M. Mariño and A. Sen, *A one-loop test of quantum supergravity*, *Class. Quant. Grav.* **31** (2014) 015012 [arXiv:1210.6057] [INSPIRE].

[47] A. Dabholkar, N. Drukker and J. Gomes, *Localization in supergravity and quantum AdS_4/CFT_3 holography*, arXiv:1406.0505 [INSPIRE].