Semi-classical analysis of the inner product of Bethe states

Eldad Bettelheim, Ivan Kostov

Semi-classical analysis of the inner product of Bethe states

Eldad Bettelheim* and Ivan Kostov*[1]

* *Racah Inst. of Physics,*
Edmund J. Safra Campus, Hebrew University of Jerusalem,
Jerusalem, Israel 91904

⋆ *Institut de Physique Théorique, CNRS-URA 2306*
C.E.A.-Saclay,
F-91191 Gif-sur-Yvette, France

We study the inner product of two Bethe states, one of which is taken on-shell, in an inhomogeneous XXX chain in the Sutherland limit, where the number of magnons is comparable with the length L of the chain and the magnon rapidities arrange in a small number of macroscopically large Bethe strings. The leading order in the large L limit is known to be expressed through a contour integral of a dilogarithm. Here we derive the sub-leading term. Our analysis is based on a new contour-integral representation of the inner product in terms of a Fredholm determinant. We give two derivations of the sub-leading term. Besides a direct derivation by solving a Riemann-Hilbert problem, we give a less rigorous, but more intuitive derivation by field-theoretical methods. For that we represent the Fredholm determinant as an expectation value in a Fock space of chiral fermions and then bosonize. We construct a collective field for the bosonized theory, the short wave-length part of which may be evaluated exactly, while the long wave-length part is amenable to a $1/L$ expansion. Our treatment thus results in a systematic $1/L$ expansion of structure factors within the Sutherland limit.

[1] *Associate member of the Institute for Nuclear Research and Nuclear Energy, Bulgarian Academy of Sciences, 72 Tsarigradsko Chaussée, 1784 Sofia, Bulgaria*

1 Introduction and summary

The computation of structure factors, matrix elements of operators between eigenstates, analytically in exactly integrable systems remains a challenging task. In very small systems one may obtain results by employing determinant formulas derived from the algebraic Bethe ansatz. The determinants are of matrices whose size increase with the number of particles, such that fully analytical computations go quickly out of hand. In the thermodynamical limit of large number of particles, the computation of such determinants becomes intractable, except in special limits, usually accompanied by a phenomenon which in physical terms may be viewed as a condensation of excitations.

The most familiar cases are the condensation of magnons into bound complexes with large spin in the Heisenberg ferromagnet as discovered by Sutherland [1] (hence the limit is sometimes called the 'Sutherland limit'), the condensation of solitons in the quantum Sine-Gordon model to quasi-periodic solutions of the KdV [2] equation, or the condensation of Cooper pairs in a superconductor [3], to form either single or multiple condensates, the latter being described by the Richardson model (a particular example of Gaudin magnets).

More recently, bound complexes of magnons has been studied in the context of the integrability in gauge and string theories [4] [5] (see also the review [6]). Some correlation functions in supersymmetric Yang-Mills theories can be expressed in terms of inner products of Bethe states in a chain of spins [7] and can be cast in the form of a determinant [8]. The thermodynamical limit here is the limit of 'heavy' fields in the Yang-Mills theory, which correspond, by the AdS/CFT duality, to classical solutions of the string-theory sigma model. The three-point function of heavy fields is exponentially small and can be thought of as a process of semi-classical tunelling [9].

The leading order computations performed in [9–11] gave an explicit expression of the exponent as a contour integral of a dilogarithm. In the present paper we give a method to compute the higher orders of the semi-classical expansion and give an explicit formula for the pre-exponential factor.

We focus on the XXX spin chain (the isotropic Heiseberg magnet), where the thermodynamical limit corresponding to long-wavelength excitations above the ferromagnetic vacuum. In view of the applications, we consider the more general case of an inhomogeneous spin chain with twisted periodic boundary condition. We will consider M-magnon Bethe states in a chain of length L in the thermodynamical limit where $M, L \to \infty$ and $M/L \sim 1$. Our goal is to propose a systematic method for computing the $1/L$ expansion for the (logarithm of the) inner product of a Bethe eigenstate and an off-shell Bethe state.

In particular we obtained an explicit expression for the subleading term, given below. Let $|\mathbf{u}\rangle$ and $|\mathbf{v}\rangle$ be two M-magnon Bethe states in a XXX spin chain of length L, characterized by the rapidities $\mathbf{u} = \{u_1, \dots, u_M\}$ and $\mathbf{v} = \{v_1, \dots, v_M\}$. One of the two states is required to be on-shell in the sense that its rapidities satisfy the Bethe equations. The two Bethe states are characterised by their pseudo-momenta

$$p_{\mathbf{u}}(x) = \sum_{j=1}^{M} \frac{1}{x - u_j} - \frac{L}{2x}, \qquad p_{\mathbf{v}}(x) = \sum_{j=1}^{M} \frac{1}{x - v_j} - \frac{L}{2x}. \tag{1.1}$$

In the semi-classical (thermodynamical) limit the root distributions are described by continuous densities along one or several line segments in the rapidity plane. We call these line segments arcs because of the typical form they take. Each arc represents a branched cut of the pseudo-momentum [4,5]. The inner product can be considered as the amplitude for semi-classical tunelling with $\hbar = 1/L$ and as such is expected to have a $1/L$ expansion of the form

$$\langle \mathbf{u}|\mathbf{v}\rangle = e^{\mathcal{F}_0 + \mathcal{F}_1 + \dots}, \qquad \mathcal{F}_n \sim L^{1-n}. \tag{1.2}$$

We obtained for the first two terms the following expressions in terms of contour integrals:

$$\mathcal{F}_0 = \oint_{\mathcal{C}} \frac{dx}{2\pi} \, \mathrm{Li}_2[e^{ip_{\mathbf{u}}(x)+ip_{\mathbf{v}}(x)}], \tag{1.3}$$

$$\mathcal{F}_1 = -\tfrac{1}{2} \oint_{\mathcal{C}\times\mathcal{C}} \frac{dx\,dy}{(2\pi)^2} \, \frac{\log\left[1-e^{ip_{\mathbf{u}}(x)+ip_{\mathbf{v}}(x)}\right] \, \log\left[1-e^{ip_{\mathbf{u}}(y)+ip_{\mathbf{v}}(y)}\right]}{(x-y)^2}, \tag{1.4}$$

where the contour of integration $\mathcal{C}$ encircles the roots $\mathbf{u}$ and $\mathbf{v}$. This expression is valid, after redefinition of the quasimomenta, for an inhomogeneous twisted XXX spin chain. The first term, containing a contour integral of the dilogarithmic function, is of order L, because the typical size of the cuts is of order L. It was first derived in [9] for the special case when the rapidities $\mathbf{u}$ are sent to infinity, and for general $\mathbf{u}$ and $\mathbf{v}$ in [10, 11]. The expression of the subleading term, which is of order L^0, is the main result of this paper.

Our method is an improvement of the semi-classical computations in [10, 11], which used a representation of the Slavnov's determinant [12] in terms of a simpler quantity, the $\mathscr{A}$-functional[1]. The most symmetric form of such a representation was found in [13]. The present computation is based on a new representation of the $\mathscr{A}$-functional as a Fredholm determinant, where the integration kernel is defined for a specific contour in the complex plane.

We compute the semi-classical limit of this Fredholm determinant in two different ways. The first, rigorous, method consists in solving the Riemann-Hilbert problem for the Fredholm kernel. The second, less rigorous but more intuitive, method uses field-theoretical formulation of the Fredholm determinant in terms of free chiral fermions. After bosonization, we solve exactly the resulting field theory at small distances to obtain an effective infrared field theory. The semiclassical expansion of the effective infrared theory can be also thought of as Mayer expansion for a gas of dipole charges living on certain contour in the rapidity plane. The leading and the sub-leading order are encoded in a saddle-point equation, which resembles the 'TBA-like' equations considered in [14].

The paper is structured as follows. In Section 2 we recall the basics of the Algebraic Bethe Ansatz for the XXX spin chain and the expression of the inner product in terms of the $\mathscr{A}$-functional. In this section we also derive the determinantal representation of the $\mathscr{A}$-functional, which is the starting point for our semi-classical analysis. In Section 3 we develop the Riemann-Hilbert approach and find an explicit expression for the subleading term. In Section 4 we derive the same result by field-theoretical methods.

2 Inner product in the inhomogeneous $\mathbf{XXX_{1/2}}$ spin chain

2.1 Algebraic Bethe Ansatz

We first recollect some well known facts about the (twisted) periodic XXX spin chain. The inhomogeneous XXX spin chain of length L is defined by the monodromy matrix

$$M_\alpha(u) = \prod_{k=1}^{L} R_{\alpha k}(u - z_k) \tag{2.1}$$

where the auxiliary space is denoted by the index α. The rational R-matrix can be taken in the form

$$R_{\alpha\beta}(u) = \frac{u}{u+\varepsilon} \, I_{\alpha\beta} + \frac{\varepsilon}{u+\varepsilon} \, P_{\alpha\beta}, \tag{2.2}$$

[1]The $\mathscr{A}$-functional generalises a quantity defined in [7], whose thermodynamical limit was computed in [9].

with the operator $P_{\alpha\beta}$ acting as a permutation of the spins in the spaces α and β. The monodromy matrix depends of a set of L variables $\mathbf{z} = \{z_1, \ldots, z_L\}$ called inhomogeneities, associated with the sites of the chain. Sometimes one uses the notation

$$z_l = \theta_l + \varepsilon/2, \quad l = 1, \ldots, L. \tag{2.3}$$

The isotropic Heisenberg Hamiltonian describes the homogeneous point $\theta_k = 0$ or $z_k = \varepsilon/2$. The standard normalization of the rapidity variable u is such that $\varepsilon = i$, but we prefer to keep ε as a free parameter. The monodromy matrix obeys the Yang-Baxter equation

$$R_{\alpha\alpha'}(u-u')M_\alpha(u)M_{\alpha'}(u') = M_{\alpha'}(u')M_\alpha(u)R_{\alpha\alpha'}(u-u'). \tag{2.4}$$

Its diagonal matrix elements are traditionally denoted by

$$M_\alpha(u) = \begin{pmatrix} A(u) & B(u) \\ C(u) & D(u) \end{pmatrix}_\alpha . \tag{2.5}$$

The operators $A(u)$ and $D(u)$ act on the pseudo-vacuum $|\Omega\rangle = |\uparrow\uparrow \ldots \uparrow\rangle$ as

$$A(u)|\Omega\rangle = a(u)|\Omega\rangle, \qquad D(u)|\Omega\rangle = d(u)|\Omega\rangle, \tag{2.6}$$

where the eigenvalues $a(u)$ and $d(u)$ are given, in the normalization (2.2) of the R-matrix, by

$$a(u) = 1\ , \qquad d(u) = \frac{Q_{\mathbf{z}}(u)}{Q_{\mathbf{z}}(u+\varepsilon)}. \tag{2.7}$$

Here and below we will systematically denote by $Q_{\mathbf{w}}$ the monic polynomial with roots $\mathbf{w}$:

$$Q_{\mathbf{w}}(u) \equiv \prod_{i=1}^{M}(u-w_i), \qquad \mathbf{w} \equiv \{w_1, \ldots, w_M\}\,. \tag{2.8}$$

Besides the inhomogeneities $\mathbf{z}$ is convenient to introduce another deformation parameter κ by choosing twisted-periodic boundary condition at length L. The transfer matrix for the twisted chain,

$$T(u) = \mathrm{tr}_a\left[\left(\begin{smallmatrix}1 & 0\\ 0 & \kappa\end{smallmatrix}\right)M_a(u)\right] = A(u) + \kappa\, D(u), \tag{2.9}$$

commutes with itself for any value of the spectral parameter, and the algebra of the matrix elements is the same as for the homogeneous XXX model.

The Hilbert space is a Fock space spanned by states obtained from the pseudo vacuum by acting with the 'raising operators' $B(u)$:

$$|\mathbf{u}\rangle = B(u_1)\ldots B(u_M)|\Omega\rangle\ . \tag{2.10}$$

If the rapidities $\mathbf{u} = \{u_1, \ldots, u_M\}$ are generic, the state is called 'off-shell', and the state is called 'on-shell' if the rapidities obey the Bethe Ansatz equations. The Bethe equations for the twisted chain read

$$\frac{a(u_j)}{d(u_j)} + \kappa\,\frac{Q_{\mathbf{u}}(u_j+\varepsilon)}{Q_{\mathbf{u}}(u_j-\varepsilon)} = 1. \tag{2.11}$$

The 'on-shell' states are eigenstates of the transfer matrix $T(x)$ with the eigenvalue

$$t(x) = \frac{Q_{\mathbf{u}}(x-\varepsilon)}{Q_{\mathbf{u}}(x)} + \kappa\,\frac{d(x)}{a(x)}\frac{Q_{\mathbf{u}}(x+\varepsilon)}{Q_{\mathbf{u}}(x)}\ . \tag{2.12}$$

2.2 The inner product in terms of the $\mathscr{A}$-functional

We consider the bilinear form,

$$(\mathbf{v},\mathbf{u}) = \langle\Omega| \prod_{j=1}^{M} \mathcal{C}(v_j) \prod_{j=1}^{M} \mathcal{B}(u_j)|\Omega\rangle \tag{2.13}$$

which we will refer to as inner product and which is related to the scalar product by[2]

$$(\mathbf{u},\mathbf{v}) = (-1)^M \langle \mathbf{u}^*|\mathbf{v}\rangle. \tag{2.14}$$

The inner product of two Bethe vectors can be computed using the commutation relations (2.4) and the action of the diagonal elements of the monodromy matrices on the pseudo-vacuum (2.6). The result is written down by Korepin [15] as a double sum over partitions. It was shown by N. Slavnov [12] that if one of the two states is on-shell, the Korepin sum can be written as a determinant.

The Slavnov determinant for a twisted periodic XXX chain can be expressed in terms of a simpler quantity $\mathscr{A}_{\mathbf{w}}[f]$, which we call $\mathscr{A}$-functional, and which depends on the set of rapidities $\mathbf{w} = \{w_j\}_{j=1}^N$ and the function $f(u)$. The inner product is equal, up to a simple factor, to the $\mathscr{A}$-functional (2.16) with $\mathbf{w} = \mathbf{u} \cup \mathbf{v}$ and $f(u) = \kappa\, d(u)/a(u)$

$$(\mathbf{u},\mathbf{v}) = \prod_{j=1}^{M} d(u_j) a(v_j)\; \mathscr{A}_{\mathbf{u}\cup\mathbf{v}}[\kappa\, d/a]\,. \tag{2.15}$$

The formula (2.15) was derived by Y. Matsuo and one of the authors [13] for purely periodic chain (no twist), but the proof given there works without change also in the case of a twist.[3] This form of the inner product is particularly useful due to its symmetry in the rapidities $\mathbf{u}$ and $\mathbf{v}$.

The $\mathscr{A}$-functional is defined as the ratio of $N \times N$ determinants [11]

$$\mathscr{A}_{\mathbf{w}}[f] \equiv \det_{jk} \left(w_j^{k-1} - f(w_j)\,(w_j+\varepsilon)^{k-1} \right) / \det \left(w_j^{k-1} \right)\,. \tag{2.16}$$

Expanding the determinant, one obtains an alternative expression as a sum over the partitions of $\mathbf{w} = \{w_1,\ldots,w_N\}$ into two disjoint subsets $\mathbf{w}_\alpha = \{w_j\}_{j\in\alpha}$ and $\mathbf{w}_{\bar\alpha} = \{w_j\}_{j\in\bar\alpha}$:

$$\mathscr{A}_{\mathbf{w}}[f] = \sum_{\alpha}\; (-1)^{|\alpha|} \prod_{j\in\alpha} f(w_j) \prod_{j\in\alpha,k\in\bar\alpha} \frac{w_j - w_k + \varepsilon}{w_j - w_k}. \tag{2.17}$$

In this form the $\mathscr{A}$-functional appeared (with $f = \kappa\, d/a$) as one of the building blocks in the expression for the three-point function in a supersymmetric Yang-Mills theory [9].

In this paper we will be interested in functional argument of the the form

$$f(u) = \kappa\, \frac{d(u)}{a(u)} = \kappa\, \frac{Q_{\mathbf{z}}(u)}{Q_{\mathbf{z}}(u+\varepsilon)}, \tag{2.18}$$

which is relevant for the inhomogeneous twisted XXX chain. We will use a special notation for the $\mathscr{A}$-functional as a function of the magnon rapidities $\mathbf{w} = \{w_j\}_{j=1}^N$, the inhomogeneities $\mathbf{z} = \{z_l\}_{l=1}^L$, the twist κ and the shift parameter ε:

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{w},\mathbf{z}} \equiv \mathscr{A}_{\mathbf{w}}[\kappa\, d/a]. \tag{2.19}$$

[2] This follows from the complex Hermitian conjugation convention $B(u)^\dagger = -C(u^*)$.

[3] A twisted version of the determinant formula was also discussed by Kazama, Komatsu and Nishimura [16].

In these notations the inner product reads

$$(\mathbf{u},\mathbf{v}) = \prod_{j=1}^{M} d(u_j)a(v_j)\; \mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{w},\mathbf{z}} \qquad\qquad (\mathbf{w}=\mathbf{u}\cup\mathbf{v}). \tag{2.20}$$

2.3 The $\mathscr{A}$-functional as an $N \times N$ determinent

In this paper we will use another determinant representation,

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \det\left(\mathbb{1} - \kappa\, K\right) \tag{2.21}$$

where $\mathbb{1}$ is the the $N \times N$ identity matrix and the matrix K has matrix elements

$$K_{jk} \;=\; \frac{\varepsilon E_j}{u_j - u_k + \varepsilon} \qquad\qquad (j,k=1,\ldots,N)\,, \tag{2.22}$$

$$E_j \;\equiv\; \frac{Q_{\mathbf{z}}(u_j)}{Q_{\mathbf{z}}(u_j+\varepsilon)} \prod_{k(\neq j)} \frac{u_j - u_k + \varepsilon}{u_j - u_k}\,. \tag{2.23}$$

To prove (2.21), we write the sum over the partitions in (2.17) as a double sum in one of the subsets:

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \sum_{\alpha} (-\kappa)^{|\alpha|} \prod_{j\in\alpha} E_j \prod_{j,k\in\alpha\,;\, j\neq k} \frac{u_j - u_k}{u_j - u_k + \varepsilon} \tag{2.24}$$

and apply the Cauchy identity $(j,k\in\alpha)$

$$\prod_{j\neq k} \frac{u_j - u_k}{u_j - u_k + \varepsilon} = \det \frac{\varepsilon}{u_j - u_k + \varepsilon}. \tag{2.25}$$

The new determinant representation (2.21) has the advantage that it exponentiates in a simple way:

$$\log \mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = -\sum_{n=1}^{\infty} \frac{\kappa^n}{n} \sum_{j_1,\ldots,j_n=1}^{N} \frac{\varepsilon\, E_{j_1}}{u_{j_1} - u_{j_2} + \varepsilon}\, \frac{\varepsilon\, E_{j_2}}{u_{j_1} - u_{j_3} + \varepsilon} \cdots \frac{\varepsilon\, E_{j_n}}{u_{j_n} - u_{j_1} + \varepsilon}. \tag{2.26}$$

2.4 Semiclassical limit: from discrete data to meromorphic functions

We are going to study the semi-classical limit $L \to \infty, N \to \infty$ with $\alpha = N/L$ finite, when the roots $\mathbf{u}$ arrange in one or several arcs of macroscopic size. We can also choose an L-dependent normalisation of the rapidity variable so that $\varepsilon \sim 1/L$. Then the typical size of the arcs will be of order L^0.

For our task it is advantageous to replace the discrete data $\mathbf{u}$ and $\mathbf{z}$ by the external potential

$$\Phi(x) \equiv \log Q_{\mathbf{u}}(x) - \log Q_{\mathbf{z}}(x). \tag{2.27}$$

In the semi-classical limit the arcs condense in one or more cuts of the meromorphic function $p(u) \equiv \partial\Phi(u)$. The discontinuities across the cuts are approximated by continuous densities which change slowly at distances of order ε.

The crucial observation which will allow to reformulate the problem in terms of the external potential Φ is that factors E_j defined in (2.23) are the residues of the same meromorphic function at $x = u_j$:

$$E_j = \frac{1}{\varepsilon} \operatorname*{Res}_{x \to u_j} \mathcal{Q}_\varepsilon(x) \qquad (j = 1, \dots, N). \tag{2.28}$$

The function $\mathcal{Q}_\varepsilon$ is defined as

$$\mathcal{Q}_\varepsilon(x) = \frac{Q_{\mathbf{u}}(x+\varepsilon)}{Q_{\mathbf{u}}(x)} \frac{Q_{\mathbf{z}}(x)}{Q_{\mathbf{z}}(x+\varepsilon)} = e^{\Phi(x_j+\varepsilon)-\Phi(x_j)}. \tag{2.29}$$

With the help of (2.28) one can write the sum in the n-th term of the series (2.26) by a multiple contour integral along a contour $\mathcal{C}_{\mathbf{u}}$ which encircles all the roots $\mathbf{u}$.

The weight function $\mathcal{Q}_\varepsilon(x)$ strongly fluctuates when x approaches $\mathbf{u}$ or $\mathbf{z}$, but if x is far from both $\mathbf{z}$ and $\mathbf{u}$, it changes slowly at distances $\sim \varepsilon$. Our goal is to reformulate the inner product in terms of contour integrals where the contour of integration $\mathcal{C}$ is placed far from the singularities of the function $\mathcal{Q}_\varepsilon$, unlike the original contour $\mathcal{C}_{\mathbf{u}}$. Then the weight $\mathcal{Q}_\varepsilon$ can be replaced by

$$\mathcal{Q}(x) = \lim_{\varepsilon \to 0} \mathcal{Q}_\varepsilon(x) = e^{\varepsilon \partial \Phi(x)}. \tag{2.30}$$

We will achieve this goal by two different, and in a sense complementary approaches. The first one relies on the solution of a Rieman-Hilbert problem, while the second one uses field-theoretical concepts. In both approaches the general idea is the same as in the original computation in [9], namely to introduce a cutoff Λ such that $|\varepsilon| \ll \Lambda \ll L|\varepsilon|$ and split the problem into a fast (short-distance) and slow (large-distance) parts. The final result does not depend on the precise value of the cutoff Λ.

3 Riemann-Hilbert Approach

We will represent the linear operator with matrix (2.22) as an integral operator acting in a space of functions with given analytic properties. Then the determinant (2.21) takes the form of a Fredholm determinant. In the semi-classical limit it is possible to split the resolvent for the Fredholm kernel into slow and fast pieces. The fast piece can be evaluated exactly, while the computation of the slow piece is done by solving a standard scalar Riemann-Hilbert problem.

3.1 The $\mathscr{A}$-functional as a Fredholm determinant

We represent an N-dimensional vector $\mathbf{f} = \{f_1, \dots, f_N\}$ as a meromorphic function $f(u)$, which has poles at $u = u_j$ with residues f_j and no other singularities:

$$f(x) \equiv \sum_j \frac{f_j}{x - u_j}. \tag{3.1}$$

The functions

$$e_j(x) = \frac{1}{x - u_j} \qquad (j = 1, \dots, N). \tag{3.2}$$

form a canonical basis in the N-dimensional space of meromorphic functions analytic everywhere except on $\mathbf{u}$. The matrix (2.22) defines a linear operator in this basis.

In order to be able to compute traces, we should also give a functional representation of the dual space. The elements $\tilde{\mathbf{f}}$ of the dual space with respect to the scalar product $\tilde{\mathbf{f}} \cdot \mathbf{f} = \tilde{f}_1 f_1 + \cdots + \tilde{f}_N f_N$ can be mapped to the space of functions $\tilde{f}(x)$ which are analytic in the vicinity of $\mathbf{u}$. With such a function we associate a dual vector $\tilde{\mathbf{f}}$ with coordinates $\tilde{f}_j \equiv \tilde{f}(u_j)$. This function is of course not unique. The scalar product may then be represented by a contour integral

$$\langle \tilde{f} | f \rangle = \oint_{\mathcal{C}_{\mathbf{u}}} \frac{dx}{2\pi i}\, \tilde{f}(x) f(x), \tag{3.3}$$

where $\mathcal{C}_{\mathbf{u}}$ is a contour surrounding the $\mathbf{u}$'s and is contained in the domain of analyticity of $\tilde{f}$.

We will use Dirac notations $f(x) = \langle x|f\rangle$, $\tilde{f}(x) = \langle \tilde{f}|x\rangle$, $f_j = \langle j|f\rangle$ and $\tilde{f}_j = \langle \tilde{f}|j\rangle$, so that

$$|f\rangle = \sum_{j=1}^{N} |j\rangle\langle j|f\rangle, \quad \langle \tilde{f}| = \sum_{j=1}^{N} \langle \tilde{f}|j\rangle\langle j|. \tag{3.4}$$

The functional representations of the basis vectors in the direct and the dual spaces are

$$|j\rangle \to \langle x|j\rangle \equiv \frac{1}{x-u_j}; \qquad \langle j| \to \langle j|x\rangle : \ \langle j|u_k\rangle = \delta_{jk} \quad (k = 1, \ldots, N). \tag{3.5}$$

The functions corresponding to the elements of the dual basis are defined up to an arbitrary meromorphic function that vanish on $\mathbf{u}$.

The functional representation of the matrix K is given by an integral operator $\mathcal{K}$, which acts on the function $f(x) \equiv \langle x|f\rangle$ as

$$\langle x|\mathcal{K}|f\rangle = \frac{1}{2\pi i} \oint_{\mathcal{C}_{\mathbf{u}}} \mathcal{Q}_\varepsilon(y)\, f(y+\varepsilon) \frac{dy}{x-y}\,, \tag{3.6}$$

where the function $\mathcal{Q}_\varepsilon$ is defined by (2.29). The contour of integration $\mathcal{C}$ in this formula is chosen to encircles all u_j but leaves the points $z_l - \varepsilon$ as well as the point x outside. Note that there are no other poles of the integrand, as the poles of $f(y+\varepsilon)$ are compensated by the zeros of $\mathcal{Q}_\varepsilon$. Applying (2.28), we obtain the action in the canonical basis

$$[\mathcal{K} f]_j = \sum_k \frac{\varepsilon E_j f_k}{u_j - u_k + \varepsilon}, \tag{3.7}$$

which agrees with (2.22). Now the $N \times N$ determinant (2.21) takes the form of a Fredholm determinant

$$\mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = \det\left(1 - \kappa\, \mathcal{K}\right). \tag{3.8}$$

One may recast the definition of the Fredholm Kernel, Eq. (3.6), in the following operator form, which will be very useful in extracting the semiclassical limit,

$$\mathcal{K} = \mathcal{P}_{\mathbf{u}}\, e^{-\Phi}\, \mathbb{D}_\varepsilon\, e^{\Phi}, \tag{3.9}$$

where $\mathbb{D}_\varepsilon = e^{\varepsilon \partial}$ is the shift operator, acting as $\mathbb{D}_\varepsilon f(x) = f(x+\varepsilon)$, and $\mathcal{P}_{\mathbf{u}} = \mathcal{P}_{\mathbf{u}}^2$ is the operator projecting onto the space of functions (3.1):

$$[\mathcal{P}_{\mathbf{u}} f](x) = \oint_{\mathcal{C}} \frac{du}{2\pi i} \frac{f(u)}{x-u} \qquad (u \text{ is outside } \mathcal{C}). \tag{3.10}$$

Let us stress on the important fact that the contour $\mathcal{C}$, which encircles the set $\mathbf{u}$, can be placed at macroscopic distance from the roots in $\mathbf{u}$. Indeed, the resolvent has no other poles than $x = u_j$ and $x = z_l - \varepsilon$. Along the deformed contour $\mathcal{C}$ the factor $\mathcal{Q}_\varepsilon$ in the Fredholm kernel changes slowly at distances of order ε. We will denote functions in the image of $\mathcal{P}_\mathbf{u}$ with a '$+$' subscript and functions in the kernel of $\mathcal{P}_\mathbf{u}$ with a '$-$' subscript. Thus it will be implied that

$$\mathcal{P}_\mathbf{u}\, g_+ = g_+, \quad \mathcal{P}_\mathbf{u}\, g_- = 0. \tag{3.11}$$

We will also use the '$+$' subscript to denote functions which are in the image of $\mathcal{P}_\mathbf{u}$ up to a polynomial, namely $\mathcal{P}_\mathbf{u} g_+ = g_+ + P$, where P is a polynomial.

3.2 Resolvent of the Fredholm kernel

We proceed by writing the logarithm of the $\mathscr{A}$-functional as follows:

$$\log \mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \int_0^\kappa \frac{d\alpha}{\alpha} \mathrm{tr}\left[\mathbb{1} - (\mathbb{1} - \alpha K)^{-1}\right], \tag{3.12}$$

which leaves us with the task of computing the trace of the resolvent $(\mathbb{1} - \alpha K)^{-1}$. Here K is the $N \times N$ matrix defined in (2.22). We wish to find a functional representation of the resolvent, which we denote by $\mathcal{F}$ and define as

$$\langle x|\mathcal{F}|f\rangle = \sum_i \frac{\left[(\mathbb{1} - \alpha K)^{-1}\mathbf{f}\right]_i}{x - u_i}. \tag{3.13}$$

One can compute $\mathrm{tr}(\mathbb{1} - \alpha K)^{-1}$ in terms of $\mathcal{F}$ as follows:

$$\mathrm{tr}(\mathbb{1} - \alpha K)^{-1} = \sum_{i=1}^N \langle i|\mathcal{F}|i\rangle. \tag{3.14}$$

The function

$$F(x, u_i) \equiv \langle x|\mathcal{F}|i\rangle \tag{3.15}$$

will appear repeatedly in the following. We shall analytically continue $F(x, u_i)$ in the variable u_i, so that u_i can be thought of as a general complex variable rather than one of the roots from the set $\mathbf{u}$. This analytical continuation is not unique, but the ambiguity is arguably exponentially small and will be neglected in the following. To compute $F(x, u)$ explicitly, we note the following identity:

$$\langle x|(\mathbb{1} - \alpha\mathcal{K})|f\rangle = e^{-\Phi(x)}\,(1 - \alpha\mathbb{D}_\varepsilon)\, e^{\Phi(x)} f(x) + \alpha \sum_{l=1}^L \frac{Q_\mathbf{u}(z_l)}{Q'_\mathbf{z}(z_l)} \frac{e^{-\Phi(z_l-\varepsilon)} f(z_l)}{x - z_l + \varepsilon}, \tag{3.16}$$

which is obtained making use of (3.9) and taking the projection by removing the poles explicitly, the latter being located at the points $z_l - \varepsilon$. The function $F(x, u)$ satisfies by definition

$$(1 - \alpha\mathcal{K})F(x, u_j) = \frac{1}{x - u_j} \qquad (j = 1, \ldots, N). \tag{3.17}$$

Analytically continuing both sides away from the set $\mathbf{u}$ we obtain for the meromorphic function $F(x,u)$ the equation

$$(1-\alpha\mathcal{K})F(x,u)=\frac{1}{x-u}. \tag{3.18}$$

Substituting the function $F(x,u)$ for $f(x)$ in Eq. (3.16) leads to

$$F(x,u)=e^{-\Phi(x)}\,(1-\alpha\mathbb{D}_\varepsilon)^{-1}\,e^{\Phi(x)}\left(\frac{1}{x-u}-\alpha\sum_{l=1}^{L}e^{-\Phi(z_l-\varepsilon)}\frac{Q_{\mathbf{u}}(z_l)}{Q'_{\mathbf{z}}(z_l)}\frac{F(z_l,u)}{x-z_l+\varepsilon}\right). \tag{3.19}$$

Then the equation may be solved self-consistently by treating $F(z_l,u)$ on the right hand side as external parameters, solving for $F(x,u)$ and then requiring that by evaluating $F(x,u)$ at $x=z_l$ we recover these same parameters. Indeed, setting x to z_l in Eq. (3.19) and representing $(1-\alpha\mathbb{D}_\varepsilon)^{-1}$ as $\sum_n\alpha^n\mathbb{D}_\varepsilon^n$, one realizes that only the $n=0$ term in this sum contributes, which makes the application of the self-consistency straightforward, leading to:

$$\begin{pmatrix}F(z_1,u)\\F(z_2,u)\\.\\.\\F(z_L,u)\end{pmatrix}=(\mathbb{1}-\tilde{K})^{-1}\begin{pmatrix}\frac{1}{z_1-u}\\\frac{1}{z_2-u}\\.\\.\\\frac{1}{z_L-u}\end{pmatrix}, \tag{3.20}$$

with the $L\times L$ matrix $\tilde{K}$ given by

$$\tilde{K}_{ln}=-\frac{Q_{\mathbf{z}}(z_n-\varepsilon)Q_{\mathbf{u}}(z_n)}{Q'_{\mathbf{z}}(z_n)Q_{\mathbf{u}}(z_n-\varepsilon)}\,\frac{1}{z_l-z_n+\varepsilon}. \tag{3.21}$$

3.3 Separation into fast and slow pieces

We compute $\mathrm{tr}(1-\alpha K)^{-1}$ by splitting the rhs of Eq. (3.19) into two parts, $F=F^{\text{ fast}}+F^{\text{slow}}$, as follows:

$$F^{\text{ fast}}(x,u)=e^{-\Phi(x)}\,(1-\alpha\mathbb{D}_\varepsilon)^{-1}\,e^{\Phi(x)}\frac{1}{x-u}, \tag{3.22}$$

$$F^{\text{slow}}(x,u)=-\alpha\,e^{-\Phi(x)}\,(1-\alpha\mathbb{D}_\varepsilon)^{-1}\,e^{\Phi(x)}\sum_{l=1}^{L}e^{-\Phi(z_l-\varepsilon)}\frac{Q_{\mathbf{u}}(z_l)}{Q'_{\mathbf{z}}(z_l)}\frac{F(z_l,u)}{x-z_l+\varepsilon}. \tag{3.23}$$

We start by computing the contribution of $F^{\text{ fast}}(x,u)$ to $\mathrm{tr}\,(\mathbb{1}-\alpha K)^{-1}$:

$$\sum_{j=1}^{N}\operatorname*{Res}_{x\to u_j}F^{\text{ fast}}(x,u_j)=\sum_{j=1}^{N}\left(1+\sum_{n=1}^{\infty}\frac{\alpha^n}{n\varepsilon}e^{\Phi(u_j+n\varepsilon)}\frac{Q_{\mathbf{z}}(u_j)}{Q'_{\mathbf{u}}(u_j)}\right)=$$
$$=N-\oint_{\mathcal{C}}\frac{dx}{2\pi i}\,e^{-\Phi(x)}\log\left(1-\alpha\,\mathbb{D}_\varepsilon\right)e^{\Phi(x)}. \tag{3.24}$$

In order to find the contribution of F^{slow} to $\mathrm{tr}\,(\mathbb{1}-\alpha K)^{-1}$, we define an integral operator $\mathcal{F}^{\text{slow}}$ with the following action:

$$\langle x|\mathcal{F}^{\text{ slow}}|f\rangle=\oint_{\mathcal{C}}\frac{du}{2\pi i}\,F^{\text{slow}}(x,u)f(u). \tag{3.25}$$

We are interested in computing the trace $\sum_{j=1}^{N}\langle j|\mathcal{F}^{\text{slow}}|j\rangle$, the contribution of $\mathcal{F}^{\text{slow}}$ to (3.14). For that we introduce another complete set of states $|m\rangle$, represented by functions $\langle x|m\rangle = f_m(x)$ analytic on and inside the contour $\mathcal{C}$, and a dual set $\langle m|$, obeying $\langle m'|m\rangle = \delta_{m',m}$, represented by functions $\langle m|j\rangle = \tilde{f}_m(j)$, the domain of analyticity of which contains $\mathcal{C}$. The quantum number m is discrete if the contour $\mathcal{C}$ is compact and continuous otherwise. For example, if there exists a circle centered at the point x_0 such that the set $\mathbf{u}$ is inside the circle and the set $\mathbf{z}$ is outside the circle, then we let $\mathcal{C}$ be this circle and choose $f_m(x) = (x-x_0)^{-m}$, $\tilde{f}_m(x) = (x-x_0)^{m-1}$ for $m \geq 1$.

The definition of $|m\rangle$ and $\langle m|$ imply

$$\oint_{\mathcal{C}} \frac{dx}{2\pi i}\langle m|x\rangle\langle x|m'\rangle = \delta_{m,m'}\,, \qquad \sum_{m}\langle x|m\rangle\langle m|x'\rangle = \frac{1}{x-x'}\,, \tag{3.26}$$

which allows to write the trace as

$$\begin{aligned}\sum_{j=1}^{N}\langle j|\mathcal{F}^{\text{slow}}|j\rangle &= \sum_{j=1}^{N}\sum_{m}\langle j|\mathcal{F}^{\text{slow}}|m\rangle\langle m|j\rangle = \sum_{j,m}\langle m|j\rangle\langle j|\mathcal{F}^{\text{slow}}|m\rangle \\ &= \sum_{m}\langle m|\mathcal{F}^{\text{slow}}|m\rangle. \end{aligned}\tag{3.27}$$

The first equality follows from the relation $\langle x|j\rangle = \sum_m \langle x|m\rangle\langle m|j\rangle$ for any x, which is true by the definition of $|m\rangle$ as a complete set. To prove the last equality, we will show that $\langle x|\mathcal{F}^{\text{slow}}|m\rangle$ has only simple poles at the u_i's and has additional singularities only around the z_i's. For such functions, the sum of $|j\rangle\langle j|$ acts as the identity operator, and one can write

$$\sum_{j=1}^{N}\langle m|j\rangle\langle j|\mathcal{F}^{\text{slow}}|m\rangle = \oint_{\mathcal{C}}\frac{dx}{2\pi i}\langle m|x\rangle\langle x|\mathcal{F}^{\text{slow}}|m\rangle. \tag{3.28}$$

Indeed, the left hand side of (3.28) is the sum of the N residues of the integrand inside the contour $\mathcal{C}$. The last identity (3.28) of is then a consequence of (3.27) and (3.26).

We are left with the task of showing the above-mentioned analytical properties of $\langle x|\mathcal{F}^{\text{slow}}|m\rangle$. Namely we must show that $\langle x|\mathcal{F}^{\text{slow}}|m\rangle$ has only simple poles at the u_i's. Indeed, combining (3.23) and (3.20), we obtain

$$\begin{aligned}\langle x|\mathcal{F}^{\text{slow}}|m\rangle = \alpha e^{-\Phi(x)}\left(1-\alpha\mathbb{D}_{\varepsilon}\right)^{-1}e^{\Phi(x)}\times \\ \times\begin{pmatrix} \frac{Q_{\mathbf{z}}(z_1-\varepsilon)Q_{\mathbf{u}}(z_1)}{Q_{\mathbf{u}}(z_1-\varepsilon)Q'_{\mathbf{z}}(z_1)}\frac{1}{x-z_1+\varepsilon} \\ \cdot \\ \cdot \\ \frac{Q_{\mathbf{z}}(z_L-\varepsilon)Q_{\mathbf{u}}(z_L)}{Q_{\mathbf{u}}(z_L-\varepsilon)Q'_{\mathbf{z}}(z_L)}\frac{1}{x-z_L+\varepsilon}\end{pmatrix}^{t}(\mathbb{1}-\alpha\tilde{K}^t)^{-1}\begin{pmatrix} f_m(z_1)\\ \cdot\\ \cdot\\ f_m(z_L)\end{pmatrix},\end{aligned}\tag{3.29}$$

whereupon the required analytic properties become apparent. This concludes the proof of (3.27).

Writing (3.27) in terms of F^{slow} yields

$$\begin{aligned}\sum_{i}\langle j|\mathcal{F}^{\text{slow}}|j\rangle &= \sum_{m}\oint\frac{dx}{2\pi i}\oint\frac{du}{2\pi i}\langle m|x\rangle F^{\text{slow}}(x,u)\langle u|m\rangle = \\ &= \oint\frac{dx}{2\pi i}\oint\frac{du}{2\pi i}\frac{F^{\text{slow}}(x,u)}{u-x},\end{aligned}\tag{3.30}$$

where the contour for the the integral in u encircles the contour for the integral in x. Taking into account that $\oint \frac{F^{\text{fast}}(x,u)}{u-x} du = 0$, we can also write

$$\sum_{j=1}^{N} \langle j|\mathcal{F}^{\text{slow}}|j\rangle = \oint \frac{dx}{2\pi i} \oint \frac{du}{2\pi i} \, \frac{F(x,u)}{u-x}. \tag{3.31}$$

Combining (3.24), (3.12) and (3.31), we obtain:

$$\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = \int_0^\kappa \frac{d\alpha}{\alpha} \left[\oint_{\mathcal{C}} \frac{dx}{2\pi i} e^{-\Phi(x)} \log\left(1 - \alpha \mathbb{D}_\varepsilon\right) e^{\Phi(x)} + \oint_{\mathcal{C}} \frac{dx}{2\pi i} \oint_{\mathcal{C}} \frac{du}{2\pi i} \, \frac{F(x,u)}{x-u} \right]. \tag{3.32}$$

The representation (3.32) of $\log \mathscr{A}_{\mathbf{u},\mathbf{z}}$ is only useful if the $1/N$ expansion of F is computable. This turns out to be the case, and we undertake the task of performing this expansion in the following.

3.4 Semi-classical expansion of the slow piece

The resolvent $F(x,u)$ satisfies the defining equation (3.18), which can be written, making use of (3.9), in the form

$$\mathcal{P}_{\mathbf{u}} \left[F(x,u) - \alpha \mathcal{Q}_\varepsilon(u) F(x+\varepsilon,u) \right] = \frac{1}{x-u}. \tag{3.33}$$

We can treat ε in the argument of F as a small parameter. Indeed, since the contour $\mathcal{C}$ is at macroscopic distance from the arcs formed by the roots $\mathbf{u}$, the function $F(x,u)$ changes slowly at distances of order ε. We thus obtain the expansion

$$\mathcal{P}_{\mathbf{u}} \left[F(x,u) - \alpha \mathcal{Q}_\varepsilon(x) \left(F(x,u) + \varepsilon F'(x,u) + \dots \right) \right] = \frac{1}{x-u}. \tag{3.34}$$

This equation is solved order by order in powers of ε, $F = F^{(0)} + F^{(1)} + \dots$, with the leading order satisfying

$$\mathcal{P}_{\mathbf{u}} \left[\left(1 - \alpha \mathcal{Q}_\varepsilon(x)\right) F^{(0)}(x,u) \right] = \frac{1}{x-u}\,, \tag{3.35}$$

while the next to leading order can be easily seen to be given by

$$F^{(1)}(x,u) = -\alpha\varepsilon \oint_{\mathcal{C}} \frac{dv}{2\pi i} \, F^{(0)}(x,v) \mathcal{Q}(v) \partial_v F^{(0)}(v,u)\,, \tag{3.36}$$

where $\mathcal{Q}(x)$ is the limit as $\varepsilon \to 0$ of $\mathcal{Q}_\varepsilon(x)$, Eq. (2.30). Computing yet higher orders is likewise mechanical.

The function $F^{(0)}$ satisfies a standard problem in the theory of integral equations, Eq. (3.35). There is a standard method of solution of such equations [17]. Namely, we decompose $1 - \alpha\mathcal{Q}_\varepsilon(x)$ into two parts,

$$1 - \alpha \mathcal{Q}(x) = U_-(x) U_+(x), \tag{3.37}$$

where $U_+(x)$ is analytic away from the arcs formed by the roots $\mathbf{u}$ and behaves at $x \to \infty$ as $U_+(x) \to x^n$ for some n of order 1, while $U_-(x)$ is analytic around the arcs, having no zeros around the arcs. We give an explicit expression for $U_\pm$ in the following. We can write Eq. (3.35) as

$$(1 - \alpha \mathcal{Q}(x))\, F^{(0)} + g_- = \frac{1}{x-u} \tag{3.38}$$

for some function g_- regular around the arcs. Using the decomposition (3.37), we write

$$U_+ F^{(0)} + \frac{g_-}{U_-} = \frac{1}{(x-u)U_-}. \tag{3.39}$$

We now apply the projector $\mathcal{P}_\mathbf{u}$ to both sides of this equation. The second term on the left hand side drops out while the first term yields a polynomial of degree $n-1$ in x for n positive, and zero otherwise. We denote this polynomial as $P_{n-1}(x;u)$ as it is also a function of u. We thus have

$$\mathcal{P}_\mathbf{u}[U_+(x)F^{(0)}(x,u)] = U_+(x)F^{(0)}(x,u) - P_{n-1}(x;u), \tag{3.40}$$

from which one obtains:

$$F^{(0)}(x,u) = \frac{1}{U_+(x)} \left\{ \mathcal{P}_\mathbf{u} \left[\frac{1}{(x-u)U_-(x)} \right] + P_{n-1}(x;u) \right\}. \tag{3.41}$$

Finally, using the analytical properties of U_- and the definition of $\mathcal{P}_\mathbf{u}$ it is easy to see that

$$\mathcal{P}_\mathbf{u} \left[\frac{1}{U_-(x)(x-u)} \right] = \frac{1}{U_-(u)(x-u)}, \tag{3.42}$$

and Eq. (3.41) simplifies to

$$F^{(0)}(x,u) = \frac{P_{n-1}(x;u)}{U_+(x)} + \frac{1}{U_+(x)U_-(u)(x-u)}. \tag{3.43}$$

The coefficients of the polynomial $P_{n-1}(x,u)$ can be found by solving (3.35) for $F^{(0)}$ around infinity to order x^{-n} and using $P_{n-1}(x;u) = \left[U_+(x)F^{(0)}(x,u)\right]_>$, where the subscript '$>$' denotes the positive (polynomial in x) part of the Laurent expansion around infinity. Note that n is of order 1 and U_+ is known (to be computed below) such that the task of finding $P_{n-1}(x;u)$ is relatively simple.

Eq. (3.43) represents a solution for $F^{(0)}$ given the decomposition (3.37). Fortunately, the functions $U_\pm$ may be computed explicitly. Assume that the phase of the complex function $1 - \alpha\mathcal{Q}(x)$ winds n_a times as x moves around the a-th arc once (where we use the conventions of positive winding for counterclockwise rotation). Namely, we assume that the rational function $1 - \alpha\mathcal{Q}(x)$ has n_a more zeros than poles a microscopic distance around the a-th arc. Below we shall call n_a below simply 'the winding number'. Let $n = \sum_a n_a$. For each a we find a rational function $R_a(x)$ such that $[1 - \alpha\mathcal{Q}_\varepsilon(x)]R(x)$ has winding number 0 around all arcs, where $R(x) = \prod_a R_a(x)$. We choose the functions $R_a(x)$ as follows. If $n_a < 0$, we take $R_a(x) = \prod_{i=1}^{|n_a|}(x - u_{j_i^{(a)}})$, where $u_{j_i^{(a)}}$, $i = 1, \ldots, |n_a|$, are arbitrary roots belonging to the a-th arc. The final result at given order does not depend on this choice to the corresponding order. If $n_a > 0$, we choose $R_a(x) = \prod_{i=1}^{n_a} \frac{1}{x - \alpha_{j_i^{(a)}}}$, where $\alpha_{j_i}^{(a)}$ are a set of n_a roots of $1 - \alpha\mathcal{Q}(x)$ around the a-th arc. The functions $U_+(x)$ and $U_-(x)$ are then given by

$$U_+(x) = \frac{\exp\left\{\mathcal{P}_\mathbf{u}\left[\log\left(1 - \alpha\mathcal{Q}(x)\right) R(x)\right]\right\}}{R(x)}, \tag{3.44}$$

$$U_-(x) = \frac{1 - \alpha\mathcal{Q}(x)}{U_+(x)}. \tag{3.45}$$

3.5 Semi-classical expansion of the fast piece

To obtain the semiclasical expansion of the first term in (3.32), we need to be able to expand the expression $\oint e^{-\Phi} \log\left(1-\alpha\mathbb{D}_\varepsilon\right) e^{\Phi}$. This is possible only if the contour of integration $\mathcal{C}$ is far from both $\mathbf{u}$ and $\mathbf{z}$. At leading order

$$\oint\limits_{\mathcal{C}} \frac{dx}{2\pi}\, e^{-\Phi} \log\left(1-\alpha\mathbb{D}_\varepsilon\right) e^{\Phi} = \oint\limits_{\mathcal{C}} \frac{dx}{2\pi}\, \log\left(1-\alpha\mathcal{Q}(x)\right) + \mathcal{O}(1/L). \tag{3.46}$$

The $O(L^0)$ correction is an integral of pure derivative and vanishes. Note that the integrand on the right hand side has logarithmic branch cuts emanating from the arcs formed by the points from the set $\mathbf{u}$ whenever the winding number n_a of the function $1-\alpha\mathcal{Q}$ around the a-th arc is non-zero.

By construction, the function $\mathcal{Q}(x)$ has no winding numbers, but only cuts along the arcs. The cuts appear as the result of merging of the poles and the zeros of $\mathcal{Q}_\varepsilon(x)$ in the limit $\varepsilon \sim 1/L \to 0$. When α is small, the function $1-\alpha\mathcal{Q}_\varepsilon$ has no zeros near the cuts and the contour of integration may be drawn to simply encircle the cuts.

As α increases, a number of zeros of $1-\alpha\mathcal{Q}_\varepsilon$ on the second sheet can move through the cuts to the first sheet. The contour of integration in (3.46) should be drawn to surround those zeros. This presents no problem as moving the contour away from the arc is consistent with the expansion in (3.46). If, on the other hand, an extra zero (one that was not there at small α) of $1-\alpha\mathcal{Q}_\varepsilon$ approaches the a-th arc, as α increases, the contour of integration must be drawn between that zero and the arc. Eventually, that zero may approach the arc up to a microscopic distance, and the approximation leading to (3.46) will be invalidated. To deal with this scenario one must separate out he roots around the zero and compute their contribution to the fast piece by performing the sum in (3.24) for those roots more directly. We do not show how this is done explicitly in this paper, rather it will be the subject of future work.

3.6 The leading order result

Only the fast piece contributes to the leading order ($\sim L$) of the semiclassical expansion of $\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]}$, since the slow piece can be easily seen to be of order L^0:

$$\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = \int_0^\kappa \frac{d\alpha}{\alpha} \oint\limits_{\mathcal{C}} \frac{dx}{2\pi i}\, \log\left[1-\alpha\mathcal{Q}(x)\right], \tag{3.47}$$

where the branch cut of the logarithmic function is to be taken according to the prescription in the previous subsection. Sometimes the contour should be deformed so that part of it passes in the second sheet, as explained in [11]. In any case, the integral over α can be taken and the final result is

$$\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = -\frac{1}{\varepsilon} \oint\limits_{\mathcal{C}} \frac{dx}{2\pi i}\, \mathrm{Li}_2\left(\kappa Q(x)\right). \tag{3.48}$$

3.7 Subleading order for zero winding numbers

In this section we will assume that $n_a = 0$ to avoid the complications that arise in the case of non-vanishing winding numbers. With this assumption we will write a compact expression for the leading and the subleading orders. Combining (3.43) and (3.31), with $P_{n-1}(u) = 0$ (which is appropriate for

$n_a = 0$), we obtain at leading order

$$\sum_j \langle j|\mathcal{F}^{\text{slow}}|j\rangle = -\oint_{\mathcal{C}} \frac{dx}{2\pi i} \oint_{\mathcal{C}} \frac{du}{2\pi i} \frac{1}{U_+(x)U_-(u)(x-u)^2} =$$
$$= -\oint_{\mathcal{C}} \frac{du}{2\pi i} \frac{\partial_u U_+(u)}{U_+^2(u)U_-(u)} = -\oint_{\mathcal{C}} \frac{du}{2\pi i} \frac{\partial_u \log[U_+(u)]}{(1-\alpha Q(u))}. \tag{3.49}$$

Using the explicit definition of U_+, Eq. (3.44), where we take $R(u) = 1$, this can be further written as

$$\sum_j \langle j|\mathcal{F}^{\text{slow}}|j\rangle = \oiint \frac{dxdu}{(2\pi i)^2} \frac{1}{1-\alpha\mathcal{Q}(x)} \frac{1}{(x-u)^2} \log(1-\alpha\mathcal{Q}(u)) =$$
$$= -\frac{\alpha}{2}\partial_\alpha \oiint \frac{dxdu}{(2\pi i)^2} \log(1-\alpha\mathcal{Q}(x)) \frac{1}{(x-u)^2} \log(1-\alpha\mathcal{Q}(u)). \tag{3.50}$$

Adding the contributions from the fast and the slow pieces, we arrive at the following approximation for $\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]}$, correct to order $O(1)$:

$$\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = -\frac{1}{\varepsilon}\oint_{\mathcal{C}} \frac{dx}{2\pi i} e^{-\Phi(x)} \operatorname{Li}_2(\kappa\mathbb{D}_\varepsilon) e^{\Phi(x)} +$$
$$+ \frac{1}{2}\oiint_{\mathcal{C}\times\mathcal{C}} \frac{dxdu}{(2\pi i)^2} \log[1-\kappa\mathcal{Q}(x)] \frac{1}{(x-u)^2} \log[1-\kappa\mathcal{Q}(u)]. \tag{3.51}$$

The first term on the right hand side is easily expandable in $1/L$ as the contour of integration can be taken to be well away from the arcs formed by the points of the set $\mathbf{u}$. The second term is an $\mathcal{O}(1)$ correction. Higher order correction are straightforward to compute by incorporating the contribution of $F^{(n)}$ for $n > 0$. As mentioned above, the case when some winding numbers are non-zero implies more complicated expressions, which we do not develop here.

4 Effective field theory for the semiclassical limit

In this section we reformulate the $\mathscr{A}$-functional in terms of a chiral fermion or, after bosonization, in terms of a chiral boson with exponential interaction. The interaction is weak at large distances but becomes singular at distances of order ε, where the two-point function develops poles. Our goal is to formulate an effective field theory for the limit $\varepsilon \to 0$. For that we split the theory into a fast and a slow component and integrate with respect to the fast component.

4.1 Free fermions

This determinant (2.21) is a particular case of the τ-functions considered in section 9 of [20] and can be expressed as a Fock-space expectation value for a Neveu-Schwarz chiral fermion living in the rapidity complex plane and having mode expansion

$$\psi(u) = \sum_{r\in\mathbb{Z}+\frac{1}{2}} \psi_r\, u^{-r-\frac{1}{2}}, \quad \psi^*(u) = \sum_{r\in\mathbb{Z}+\frac{1}{2}} \psi_r^*\, u^{r-\frac{1}{2}}. \tag{4.1}$$

The fermion modes are assumed to satisfy the anticommutation relations

$$[\psi_r, \psi_s^*]_+ = \delta_{rs}\,, \tag{4.2}$$

and the left/right vacuum states are defined by

$$\langle 0|\psi_{-r} = \langle 0|\psi_r^* = 0 \;\text{ and }\; \psi_r\,|0\rangle = \psi_{-r}^*|0\rangle = 0, \quad \text{for } r > 0. \tag{4.3}$$

The operator ψ_r^* creates a particle (or annihilates a hole) with mode number r and the operator ψ_r annihilates a particle (or creates a hole) with mode number r. The particles carry charge 1, while the holes carry charge -1. The charge zero vacuum states are obtained by filling the Dirac see up to level zero.

Any correlation function of the operators (4.1) is a determinant of two-point correlators

$$\langle 0|\psi(u)\psi^*(v)|0\rangle = \langle 0|\psi^*(u)\psi(v)|0\rangle = \frac{1}{u-v}\,. \tag{4.4}$$

The expectation value of several pairs of fermions is given by the determinant of the two-point functions. Obviously the determinant (2.21) is equal to the expectation value

$$\mathrm{Det}(1-\kappa K) = \langle 0|\exp\left(\kappa\varepsilon\sum_{j=1}^{N} E_j\,\psi^*(u_j)\psi(u_j+\varepsilon)\right)|0\rangle. \tag{4.5}$$

The discrete sum of fermion bilinears in the exponent on the rhs of (4.5) can be written, with the help of (2.28), as an integral along the contour $\mathcal{C}_{\mathbf{u}}$ which encircles the points $u_1,\dots,u_N$, and the Fock space representation (4.5) takes the form

$$\mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = \langle 0|\exp\left(\kappa\oint_{\mathcal{C}_{\mathbf{u}}}\frac{dx}{2\pi i}\mathcal{Q}_\varepsilon(x)\,\psi^*(x)\psi(x+\varepsilon)\right)|0\rangle, \tag{4.6}$$

where the weight function $\mathcal{Q}_\varepsilon(x)$ is defined by Eq. (2.29).

4.2 Bosonic field with exponential interaction

Alternatively, one can express the $\mathscr{A}$-function in term of a chiral boson $\phi(x)$ with two-point function

$$\langle 0|\phi(x)\phi(y)|0\rangle = \log(x-y). \tag{4.7}$$

After bosonization $\psi(x) \to e^{\phi(x)}$ and $\psi^*(x) \to e^{-\phi(x)}$, where we assumed that the exponents of the gaussian field are normally ordered, the fermion bilinear $\psi^*(x)\psi(x+\varepsilon)$ becomes, up to a numerical factor, a chiral exponential field[4]

$$\mathcal{V}_\varepsilon(x) \equiv e^{\phi(x+\varepsilon)-\phi(x)}. \tag{4.8}$$

The numerical factor is determined by the OPE

$$e^{-\phi(x)}\,e^{\phi(u)} = \frac{1}{x-u}e^{\phi(u)-\phi(x)} \qquad (u = x+\varepsilon)\,, \tag{4.9}$$

[4]Our convention is that the exponential is normally ordered, $e^{\phi(u)-\phi(v)} \equiv :e^{\phi(u)-\phi(v)}:$, so that $\langle 0|e^{\phi(u)-\phi(v)}|0\rangle = 1$.

so that the fermion bilinear bosonizes as

$$\psi^*(x)\psi(x+\varepsilon) \quad \to \quad e^{-\phi(x)}e^{\phi(x+\varepsilon)} = -\frac{1}{\varepsilon}\,\mathcal{V}_\varepsilon(x). \tag{4.10}$$

The resulting bosonic field theory is that of a two-dimensional gaussian field $\phi(x,\bar{x})$ perturbed by a chiral interaction term $\kappa\mathcal{Q}_\varepsilon(x)\mathcal{V}_\varepsilon(x)$.

Expanding the exponential in series and using that the n-point correlator of the exponential field is a product of all two-point correlators

$$\langle 0|\mathcal{V}_\varepsilon(x)\mathcal{V}_\varepsilon(y)|0\rangle = \frac{(x-y)^2}{(x-y+\varepsilon)(x-y-\varepsilon)}\,, \tag{4.11}$$

we obtain that the expectation value (4.6) is given by the grand-canonical Coulomb-gas partition function

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \sum_{n=0}^{N}\frac{(-\kappa)^n}{n!}\prod_{j=1}^{n}\oint_{\mathcal{C}_\mathbf{u}}\frac{dx_j}{2\pi i}\;\frac{\mathcal{Q}_\varepsilon(x_j)}{\varepsilon}\;\prod_{k>j}^{n}\frac{(x_j-x_k)^2}{(x_j-x_k)^2-\varepsilon^2}\,. \tag{4.12}$$

In the contour integral representations (4.6) and (4.12) the integration contour $\mathcal{C}_\mathbf{u}$ is drawn close to the poles of $\mathcal{Q}_\varepsilon$. We would like to deform the contours away from these poles, where $\mathcal{Q}_\varepsilon$ can be considered as a sufficiently smooth. In the n-th term of the series we can deform sequentially the integration contours away from the set $\mathbf{u}$, so that the n contours form a nested configuration separating the $\mathbf{u}$-poles and the $\mathbf{z}$-poles of the function $\mathcal{Q}_\varepsilon$. If the subsequent contours are spaced by ε, then the poles at $x_j - x_k = \pm\varepsilon$ of the integrand do not contribute.

4.3 Integrating out the fast modes

The multiple contour integral can be evaluated in the semiclassical limit by splitting the integrand into slow and fast parts. We thus introduce an intermediate scale Λ such that

$$|\varepsilon| \ll \Lambda \ll N|\varepsilon| \tag{4.13}$$

and split the bosonic field into a fast and a slow components,

$$\phi = \phi_{\text{slow}} + \phi_{\text{ fast}}. \tag{4.14}$$

Up to exponential terms the two-point function of the bosonic field is approximated at small distances by that of the fast component and at large distances by that of the slow component. Below we will perform explicitly the integration with respect to $\phi_{\text{ fast}}$ to obtain an effective interaction for ϕ_{slow}. Since the two-point function of the exponential fields with ϕ replaced by ϕ_{slow} does not contain poles, the nested contours spaced by ε can be replaced by a *single* contour $\mathcal{C}$ placed sufficiently far from the sets $\mathbf{z}$ and $\mathbf{u}$ where the integrand has poles.[5] The effective interaction for the slow component is of

[5] The splitting into a fast and a slow components can be done explicitly if the contour $\mathcal{C}$ can be placed along the real axis. Introduce a cutoff Λ such that $|\varepsilon| \ll \Lambda \ll N|\varepsilon|$. Then the bosonic field has a continuum of Fourier modes α_E and the slow and fast parts can be defined as $\phi_{\text{slow}}(x) = \int_{|E|<\Lambda} dE\,\alpha_E\, e^{iEx}$, $\phi_{\text{ fast}}(x) = \int_{|E|>\Lambda} dE\,\alpha_E\, e^{iEx}$. The propagators of the slow and the fast components are $\langle 0|\partial\phi_{\text{slow}}(x),\phi_{\text{slow}}(y)|0\rangle = (1-e^{i\Lambda(x-y)})/(x-y)$, $\langle 0|\partial\phi_{\text{ fast}}(x),\phi_{\text{ fast}}(y)|0\rangle = e^{i\Lambda(x-y)}/(x-y)$. The propagator for the slow component contains a strongly oscillating term whose role is to kill the pole at $x=y$ and which can be neglected far from the diagonal, while the numerator in the propagator of the fast component can be replaced by 1 at small distances. The effects of the cutoff are thus exponentially small and do not influence the perturbative quasiclassical expansion.

the form

$$S_{\text{eff}}[\phi_{\text{slow}}] = \sum_{n\geq 1} \oint_{\mathcal{C}} \frac{dx}{2\pi i}\, V_{\text{eff}}^{(n)}(x)\,, \tag{4.15}$$

where n-th term is the contribution of the connected n-point function of the exponential field $\mathcal{V}_\varepsilon(x)$ with respect to the fast component,

$$V_{\text{eff}}^{(n)}(x) = \frac{(-\kappa/\varepsilon)^n}{n!} \oint \frac{dx_1}{2\pi i} \dots \frac{dx_n}{2\pi i}\, \delta(x-x_1) \Big\langle\!\Big\langle \prod_{j=1}^{n} \mathcal{Q}_\varepsilon(x_j)\, \mathcal{V}_\varepsilon(x_j) \Big\rangle\!\Big\rangle_{\text{fast}}. \tag{4.16}$$

$$\Xi_n(x) = \frac{(-1/i\varepsilon)^n}{n!} \oint \frac{dx_1}{2\pi i} \dots \frac{dx_n}{2\pi i}\, \delta(x-x_1) \Big\langle\!\Big\langle \prod_{j=1}^{n} \mathcal{Q}(x_j)\, \mathcal{V}(x_j) \Big\rangle\!\Big\rangle_{\text{fast}}. \tag{4.17}$$

To compute the connected correlation function $\langle\!\langle\ \rangle\!\rangle_{\text{fast}}$ we represent the product of n exponential fields in the form

$$\mathcal{V}_\varepsilon(x_1)\dots\mathcal{V}_\varepsilon(x_n) = \prod_{j<k} \frac{(x_j-x_k)^2}{(x_j-x_k)^2-\varepsilon^2}\ :\mathcal{V}_\varepsilon(x_1)\dots\mathcal{V}_\varepsilon(x_n):, \tag{4.18}$$

where : : signifies normal product. By definition the normal product of exponential fields has vacuum expectation value 1. Assuming that all distances $|x_j - x_k|$ are small compared to the scale Λ, we can interpret the operator product expansion (4.18) as

$$\big\langle \mathcal{V}_\varepsilon(x_1)\dots\mathcal{V}_\varepsilon(x_n)\big\rangle_{\text{fast}} \approx \prod_{j<k} \frac{(x_j-x_k)^2}{(x_j-x_k)^2-\varepsilon^2}\ :\mathcal{V}_\varepsilon^{\text{slow}}(x_1)\dots\mathcal{V}_\varepsilon(x_n)^{\text{slow}}:\,. \tag{4.19}$$

$$\big\langle \mathcal{V}(x_1)\dots\mathcal{V}(x_n)\big\rangle_{\text{fast}} \approx \prod_{j<k} \frac{x_{jk}^2}{x_{jk}^2+\varepsilon^2}\ :\mathcal{V}^{\text{slow}}(x_1)\dots\mathcal{V}(x_n)^{\text{slow}}:\,. \tag{4.20}$$

To extract the connected component of the n-point function we apply the Cauchy identity (2.25) and represent the Cauchy determinant as a sum over permutations. The result is the sum of the (identical) contributions of the $(n-1)!$ permutations representing maximal cycles of length n.[6] We find ($x_{jk} \equiv x_j - x_k$)

$$\begin{aligned} V_{\text{eff}}^{(n)}(x_1) &= \frac{(n-1)!}{n!} \oint \frac{\prod_{k=1}^n \mathcal{Q}_\varepsilon(x_k)\, \mathcal{V}(x_k)^{\text{slow}}\, \frac{dx_2}{2\pi i} \dots \frac{dx_n}{2\pi i}}{(\varepsilon - x_{12})\dots(\varepsilon - x_{n-1,n})(\varepsilon - x_{n,1})} \\ &= -\frac{Q_{n\varepsilon}(x_1)}{n^2\varepsilon}\, \mathcal{V}_{n\varepsilon}(x)\,, \end{aligned} \tag{4.21}$$

with

$$\mathcal{Q}_{n\varepsilon}(x) \equiv Q_\varepsilon(x)Q_\varepsilon(x+\varepsilon)\dots Q_\varepsilon(x+n\varepsilon) = e^{-\Phi(x)+\Phi(x+n\varepsilon)}\,, \tag{4.22}$$

$$\mathcal{V}_{n\varepsilon}(x) \equiv :\mathcal{V}_\varepsilon(x)\mathcal{V}_\varepsilon(x+\varepsilon)\dots\mathcal{V}_\varepsilon(x+n\varepsilon): = e^{-\phi(x)+\phi(x+n\varepsilon)}. \tag{4.23}$$

[6] This is basically the calculation done in ref. [21]. The difference is in the extra factor and in our convention to choose $x = x_1$ as collective coordinate, while in [21] $x = (x_1+\dots+x_n)/n$. Note that the contribution of the permutations with more than one cycle vanishes automatically.

(Here and below ϕ denotes ϕ_{slow}.) The resulting expression for the $\mathscr{A}$-functional in terms of the effective infrared theory is

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \langle 0| \exp\left(-\frac{1}{\varepsilon} \sum_{n=1}^{\Lambda/\varepsilon} \frac{\kappa^n}{n^2} \oint_{\mathcal{C}_\mathbf{u}} \frac{dx}{2\pi i} \mathcal{Q}_{n\varepsilon}(x)\, \mathcal{V}_{n\varepsilon}(x) \right) |0\rangle + \text{non-perturbative}. \tag{4.24}$$

By construction the spacing $n\varepsilon$ should be smaller than the scale Λ, but if the sum over n in the exponent is extended to infinity, this will introduce exponentially small terms and will not change the $1/L$ expansion. Introducing the shift operator $\mathbb{D}_\varepsilon = e^{\varepsilon \partial_x}$, the series in the exponent can be formally summed up as

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \langle 0| \exp\left(-\frac{1}{\varepsilon} \oint_{\mathcal{C}} \frac{dx}{2\pi i} \ :e^{-\Phi(x)-\phi(x)}\, \mathrm{Li}_2(\kappa\, \mathbb{D}_\varepsilon)\, e^{\Phi(x)+\phi(x)}: \right) |0\rangle \ + \text{non-perturbative}. \tag{4.25}$$

Another way to write this expression, without using the normal product and redefining $\phi + \Phi \to \phi$, is

$$\begin{gathered} \mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \langle 0| \exp\Big(-\oint_{\mathcal{C}} \frac{dx}{2\pi i}\, e^{-\phi(x)}\ \log(1-\kappa\, \mathbb{D}_\varepsilon)\, e^{\phi(x)}:\Big)|0\rangle \ + \text{non-perturbative}, \\ \langle 0|\phi(x)|0\rangle = \Phi(x), \ \ \langle 0|\phi(x)\phi(y)|0\rangle = \log(x-y). \end{gathered} \tag{4.26}$$

4.4 One-dimensional effective theory in the semiclassical limit

In the semiclassical limit

$$\hbar \equiv 1/L \to 0, \qquad \ell \equiv L\varepsilon \sim 1, \qquad \alpha = N/L \sim 1 \tag{4.27}$$

the classical field Φ grows as $1/\hbar$, but

$$\Phi(x+n\varepsilon) - \Phi(x) = n\varepsilon\, \partial\Phi(x) + \dots \tag{4.28}$$

remains finite, as well as the range of integration and size of the contour $\mathcal{C}$. Furthermore, the distribution of the roots u_j is assumed to be of the form of the finite zone solutions of the Bethe equations [5], which are described by hyperlliptic curves. The roots $\mathbf{u}$ condense into one or several arcs, which become the cuts of the meromorphic function

$$\partial\Phi(x) = \sum_{j=1}^{N} \frac{1}{x-u_j} - \sum_{l=1}^{L} \frac{1}{x-z_l}. \tag{4.29}$$

We assume that the inhomogeneities $\mathbf{z}$ are centered around the origin of the rapidity plane, but we do non make any other assumptions about them.

We will concentrate on the leading term (of order $1/\varepsilon$) and the subleading term (of order 1), and will ignore the corrections that vanish in the limit $\varepsilon \to 0$. Then, using the approximation (4.28), we expand the exponent in (4.25) as

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} = \langle 0| \exp \oint_{\mathcal{C}} \frac{dx}{2\pi i} \left(-\frac{1}{\varepsilon} : \mathrm{Li}_2(\kappa\, \mathcal{Q}\, e^{-\varepsilon\varphi}) : - : \log(1-\mathcal{Q}\, e^{-\varepsilon\varphi}) \partial\varphi : + \dots \right) |0\rangle\,, \tag{4.30}$$

where we introduced the derivative field

$$\varphi(x) = -\partial\phi(x) \tag{4.31}$$

and used the notation (2.30). We can retain only the first term on the rhs of (4.30), since the second term is a full derivative and can be neglected.

Now we can pass from Fock-space to path-integral formalism. For that we express the expectation value (4.30) as a path integral for the $(0+1)$-dimensional field $\varphi(x)$ defined on the contour $\mathcal{C}$ and having two-point function

$$G(x,u) \equiv \langle\varphi(x)\varphi(u)\rangle = \frac{1}{(x-u)^2}\,. \tag{4.32}$$

Introducing a second field ρ linearly coupled to ϕ we write the $\mathscr{A}$-functional as a path integral

$$\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} \;=\; \int [D\varphi\, D\rho]\; e^{-\mathcal{Y}[\varphi,\rho]}\,, \tag{4.33}$$

with the action functional given by

$$\mathcal{Y}[\varphi,\rho] = \oint\limits_{\mathcal{C}} \frac{dx}{2\pi i}\left(\frac{1}{\varepsilon}\mathrm{Li}_2(\kappa\mathcal{Q}(x)e^{-\varepsilon\varphi(x)}) + \varphi(x)\rho(x)\right) + \tfrac{1}{2}\oint\limits_{\mathcal{C}\times\mathcal{C}} \frac{dx\,du}{(2\pi i)^2}\;\rho(x)G(x,u)\rho(u). \tag{4.34}$$

The double integral in the second term can be understood as a principal value. Indeed, the contribution $\rho\rho'$ of the pole at $x=u$ is pure derivative and vanishes after being contour-integrated.

In the approximation we are looking for, the $\mathscr{A}$-functional is given by the saddle-point action

$$\log\mathscr{A}^{[\varepsilon,\kappa]}_{\mathbf{u},\mathbf{z}} \;= \mathcal{Y}_c + \mathcal{O}(\varepsilon), \qquad \mathcal{Y}_c = \mathcal{Y}[\varphi_c,\rho_c], \tag{4.35}$$

where the saddle point φ_c is given by a couple of TBA-like equations

$$\varphi_c(x) = \fint\limits_{\mathcal{C}} \frac{dy}{2\pi i} G(x-y)\rho_c(y), \quad \rho_c(x) = -\log\left(1-\kappa\mathcal{Q}(x)e^{-\varepsilon\varphi_c(x)}\right). \tag{4.36}$$

After solving for ρ_c, one obtains a non-linear integral equation[7] for the classical field φ_c:

$$\varphi_c(x) = \fint\limits_{\mathcal{C}} \frac{dy}{2\pi i} G(x-y)\log\left(1-\kappa\mathcal{Q}(y)e^{-\varepsilon\varphi_c(y)}\right). \tag{4.37}$$

Expanding

$$\mathcal{Y}_c \;=\; \oint\limits_{\mathcal{C}} \frac{dx}{2\pi i}\;\left[-\frac{1}{\varepsilon}\mathrm{Li}_2(\kappa\mathcal{Q}(x)e^{-\varepsilon\varphi_c(x)}) - \tfrac{1}{2}\varphi_c(x)\log\left(1-\kappa\mathcal{Q}(x)e^{-\varepsilon\varphi_c(x)}\right)\right] \tag{4.38}$$

[7] Such type of integral equations first appeared as alternative formulation of the Thermodinamic Bethe Ansatz without strings [22,23], and most recently in supersymmetric gauge theories [14,24]. If the space-time variable x scales as ε^0, there is no need to solve the non-linear integral equation, because only the leading order in ε matters. We don't exclude that the above analysis can be carried on for weaker assumptions about the distribution of the roots $\mathbf{u}$, such that x scales as ε^1, in which case the non-linear integral equation does not contain a small parameter.

up to $O(\varepsilon)$, we obtain an explicit expression for the leading and the subleading terms:

$$\log \mathscr{A}_{\mathbf{u},\mathbf{z}}^{[\varepsilon,\kappa]} = -\frac{1}{\varepsilon}\oint\limits_{\mathcal{C}} \frac{dx}{2\pi i}\,\mathrm{Li}_2[\kappa\mathcal{Q}(x)] + \tfrac{1}{2}\oint\limits_{\mathcal{C}\times\mathcal{C}} \frac{dx\,du}{(2\pi i)^2}\,\frac{\log\left[1-\kappa\mathcal{Q}(x)\right]\,\log\left[1-\kappa\mathcal{Q}(u)\right]}{(x-u)^2} + \mathcal{O}(\varepsilon), \tag{4.39}$$

where the double integral is understood as a principal value.

The expression (4.39) obtained by the field-theory method is identical to the result obtained by solving the Riemann-Hilbert problem, Eq. (3.51). Taking $\varepsilon = i$ and $\mathcal{Q} = \exp(ip_{\mathbf{u}} + ip_{\mathbf{v}})$, we obtain the expression for the leading and the sub-leading terms of the inner product, Eqs. (1.3)–(1.4). Here we neglected the trivial factors in the expression (2.20) of the inner product through the $\mathscr{A}$-functional.

The choice of the contour $\mathcal{C}$ is a subtle issue and depends on the analytic properties of the function $\mathcal{Q}(x)$, as discussed above in Section 3. In any particular case one can first find explicitly the function $\mathcal{Q}(x)$ in the limit of a small filling fractions ($\alpha \equiv N_a/L \ll 1$, where N_a is the number of roots that form the a-th arc), then place the contour $\mathcal{C}$ so that it does not cross any cuts of $\mathrm{Li}_2(\mathcal{Q}(x))$. If the fillings are not too large, this choice of the contour will remain valid also for $N_a/L \sim 1$. However, it is possible that at some critical filling that one of the the zeros of $1-\mathcal{Q}$ approaches the a-th arc. Such a situation has been analysed in [25]. If this is the case, the contour of integration should be deformed to avoid the logarithmic cut starting with this zero, possibly passing to the second sheet.

4.5 Relation to the Mayer expansion of non-ideal gas

The semi-classical limit of the $\mathscr{A}$-functional resembles the so-called Nekrasov-Shatashvili limit of instanton partition functions of deformed $\mathcal{N}=2$ supersymmetric gauge theories [14]. The methods developed to study this limit, outlined in [14] and recently worked out in great detail in [18, 19], are based on the iterated Mayer expansion for a non-ideal gas of particles confined along a contour $\mathcal{C}$. Below we are going to explain the connection between our approach and the Mayer expansion.

The exponential field (4.8) creates a pair of Coulomb charges with opposite signs spaced at distance ε. One can think of such a pair as a 'fundamental' particle with zero electric charge but non-vanishing dipole and higher charges. The sum (4.12) is the grand partition function of such 'fundamental' dipoles confined on the contour $\mathcal{C}_{\mathbf{u}}$, or equivalently, on a sequence of nested contours surrounding the set $\mathbf{u}$.

The fundamental dipoles interact with the external potential $\Phi(x)$ and pairwise among themselves. The pairwise interaction is determined by the two-point correlator (4.11). Subtracting the product of the one-point functions $\langle 0|\mathcal{V}_\varepsilon|0\rangle = 1$, one obtains for the connected correlator of two dipoles

$$\langle\!\langle \mathcal{V}_\varepsilon(x)\mathcal{V}_\varepsilon(y)\rangle\!\rangle = \frac{\varepsilon^2}{(x-y)^2-\varepsilon^2}. \tag{4.40}$$

The interaction between two dipoles depends both on the distance and on the direction. If $\varepsilon = |\varepsilon|i$, then the force between two dipoles is repulsive if they are spaced horizontally and attractive if they are spaced vertically. As the interaction rapidly decreases at large distances, one can compute the thermodynamics of the dipole gas by performing Mayer (cumulant) expansion. The poles of the pairwise interaction potential at $x-y = \pm\varepsilon$ lead to a phenomenon called in [14] clustering of instanton particles. The fundamental dipole can form 'bound states' of n fundamental dipoles, whose field-theoretical counterpart are the exponential fields (4.23). A composite particle made of n fundamental dipoles behaves as a pair of positive and negative electric charges spaced at distance $n\varepsilon$.

By the operator representation (4.25), the $\mathscr{A}$-functional is the grand partition function of a non-ideal gas made of the fundamental particles and all kinds of composite particles. The particles of this gas interact with the external potential $\Phi(x)$ and pairwise as

$$\langle\langle \mathcal{V}_{m\varepsilon}(x)\, \mathcal{V}_{n\varepsilon}(y) \rangle\rangle = \frac{\varepsilon^2\, mn}{(x-y+m\varepsilon)(x-y-n\varepsilon)}. \tag{4.41}$$

The effective one-dimensional theory (4.34) describes the limit when only the dipole charge is taken into account, while the quadruple etc. charges, small by powers of ε, are neglected. The first term in our final formula (4.39) corresponds to the dilute gas approximation, in which the charges interact only with the external potential, while the sub-leading second term takes into account the pairwise interactions.

Acknowledgments

I.K. thanks J.-E. Bourgine, D. Fioravanti, N. Gromov, S. Komatsu, Y. Matsuo and F. Ravanini for illuminating discussions. We are grateful for the hospitality at the Simons Center for Geometry and Physics, where this work has been initiated. EB is grateful for the hospitality at the University of Cologne. This work has been supported by European Programme IRSES UNIFY (Grant No 269217), the Israel Science Foundation (Grant No. 852/11) and by the Binational Science Foundation (Grant No. 2010345).

References

[1] B. Sutherland, "Low-Lying Eigenstates of the One-Dimensional Heisenberg Ferromagnet for any Magnetization and Momentum," *Phys. Rev. Lett.* **74** (Jan, 1995) 816–819.

[2] O. Babelon, D. Bernard, and F. A. Smirnov, "Quantization of Solitons and the Restricted Sine-Gordon Model," *Communications in Mathematical Physics* **182** (Dec., 1996) 319–354, hep-th/9603010.

[3] G. Gorohovsky and E. Bettelheim, "Exact expectation values within Richardson's approach for the pairing Hamiltonian in a macroscopic system," *Phys. Rev. B* **84** (Dec., 2011) 224503, 1111.1519.

[4] N. Beisert, J. A. Minahan, M. Staudacher, and K. Zarembo, "Stringing spins and spinning strings," *JHEP* **09** (2003) 010, hep-th/0306139.

[5] V. Kazakov, A. Marshakov, J. A. Minahan, and K. Zarembo, "Classical / quantum integrability in AdS/CFT," *JHEP* **05** (2004) 024, hep-th/0402207.

[6] N. Beisert *et al*, "Review of AdS/CFT Integrability: An Overview," *Letters in Mathematical Physics* **99** (Jan., 2012) 3–32, 1012.3982.

[7] J. Escobedo, N. Gromov, A. Sever, and P. Vieira, "Tailoring three-point functions and integrability," *JHEP* **9** (Sept., 2011) 28, 1012.2475.

[8] O. Foda, "$\mathcal{N} = 4$ SYM structure constants as determinants," *Journal of High Energy Physics* **3** (Mar., 2012) 96, 1111.4663.

[9] N. Gromov, A. Sever, and P. Vieira, "Tailoring Three-Point Functions and Integrability III. Classical Tunneling," *ArXiv e-prints* (Nov., 2011) 1111.2349.

[10] I. Kostov, "Classical Limit of the Three-Point Function of N=4 Supersymmetric Yang-Mills Theory from Integrability," *Physical Review Letters* **108** (June, 2012) 261604, 1203.6180.

[11] I. Kostov, "Three-point function of semiclassical states at weak coupling," *Journal of Physics A Mathematical General* **45** (Dec., 2012) 4018, 1205.4412.

[12] N. A. Slavnov, "Calculation of scalar products of wave functions and form factors in the framework of the algebraic Bethe ansatz," *Theoretical and Mathematical Physics* **79** (1989) 502–508. 10.1007/BF01016531.

[13] I. Kostov and Y. Matsuo, "Inner products of Bethe states as partial domain wall partition functions," *JHEP10(2012)168* (July, 2012) 1207.2562.

[14] N. A. Nekrasov and S. L. Shatashvili, "Quantization of Integrable Systems and Four Dimensional Gauge Theories," 0908.4052.

[15] V. E. Korepin, "Calculation of norms of Bethe wave functions," *Communications in Mathematical Physics* **86** (1982) 391–418. 10.1007/BF01212176.

[16] Y. Kazama, S. Komatsu, and T. Nishimura, "A new integral representation for the scalar products of Bethe states for the XXX spin chain," *ArXiv e-prints* (Apr., 2013) 1304.5011.

[17] N. I. Muskhelishvili, *Singular Integral Equations: Boundary Problems of Function Theory and Their Application to Mathematical Physics*. P. Noordhoff, 1953.

[18] C. Meneghelli and G. Yang, "Mayer-Cluster Expansion of Instanton Partition Functions and Thermodynamic Bethe Ansatz," *ArXiv e-prints* (Dec., 2013) 1312.4537.

[19] J.-E. Bourgine, "Confinement and Mayer cluster expansions," 1402.1626.

[20] M. Jimbo and T. Miwa, "Solitons and infinite dimensional Lie algebras," *Publ. RIMS, Kyoto Univ* **19** (1983) 943–1001.

[21] G.Moore, N.Nekrasov, and S.Shatashvili, "Integrating Over Higgs Branches," *Commun.Math.Phys.* **209** (2000) 97–121, hep-th/9712241.

[22] C. Destri and H. J. De Vega, "Unified approach to thermodynamic Bethe Ansatz and finite size corrections for lattice models and field theories," *Nucl. Phys.* **B438** (1995) 413–454, hep-th/9407117.

[23] A. Klumper, M. T. Batchelor, and P. A. Pearce, "Central charges of the 6- and 19-vertex models with twisted boundary conditions," *Journal of Physics A: Mathematical and General* **24** (1991), no. 13, 3111.

[24] G. Feverati, D. Fioravanti, P. Grinza, and M. Rossi, "On the finite size corrections of anti-ferromagnetic anomalous dimensions in N = 4 SYM," *JHEP* **05** (2006) 068, hep-th/0602189.

[25] T. Bargheer, N. Beisert, and N. Gromov, "Quantum Stability for the Heisenberg Ferromagnet," 0804.0324.