

RECEIVED: November 9, 2018

REVISED: January 31, 2019

ACCEPTED: March 11, 2019

PUBLISHED: March 18, 2019

Cutting the cylinder into squares: the square form factor

Juan Miguel Nieto

*Departamento de Física Teórica, Universidad Complutense de Madrid,
28040 Madrid, Spain*

E-mail: juanieto@ucm.es

ABSTRACT: In this article we present a method for constructing two-point functions in the spirit of the hexagon proposal, which leads us to propose a “square form factor”. Since cutting the square gives us two squares, we can write a consistency condition that heavily constrains such form factors. In particular, we are able to use this constraint to reconstruct the Gaudin through the forest expansion of the determinant appearing in its definition. We also use this procedure to compute the norm of off-shell Bethe states for some simple cases.

KEYWORDS: AdS-CFT Correspondence, Integrable Field Theories

ARXIV EPRINT: [1810.11430](https://arxiv.org/abs/1810.11430)

Contents

1	Introduction	1
2	The SU(2) algebraic bethe ansatz	3
3	From the Gaudin determinant to the square and back	5
3.1	The square bootstrap and the Gaudin determinant	5
3.2	Reconstructing the Gaudin determinant: some particular cases	8
3.3	Reconstructing the Gaudin determinant: the general asymptotic square	11
4	Computing the off-shell square	13
4.1	The off-shell square with two excitations	13
4.2	The off-shell square with three excitations	16
4.3	The general off-shell norm-like square	18
5	Conclusions	19
A	Kirchhoff's matrix-tree theorem and Gaudin determinant	21
B	More on breaking factors	22
C	Reconstruction of the spacetime dependence	24
D	Crossing properties of the square	25
E	Leading mirror corrections to the empty and one excitation squares	26

1 Introduction

A consistent method to define a quantum field theory is to construct observables starting from form factors. Form factors are matrix elements of local observables computed in the basis of asymptotic ingoing and outgoing states. Such matrix elements fulfil the *Smirnov's Axioms* [1], a set of restrictions on their analytical properties. This construction is non-perturbative and restricts the reconstructed theory to be local and a Wightman field theory.

This construction has been successfully applied to different systems like the Sine-Gordon model [2, 3], the Thirring model [4, 5] and the SU(N) Lieb-Lininger model [6]. Concerning the AdS_5/CFT_4 correspondence, we can find the first steps towards understanding form factors in [7, 8], where the authors focus on the spin chain description and the near-plane-wave limit of the string description.

Regarding direct computations of correlation functions, the Quantum Spectral Curve has already solved (at least formally) the spectral problem in $\mathcal{N} = 4$ SYM [9–18] and thus

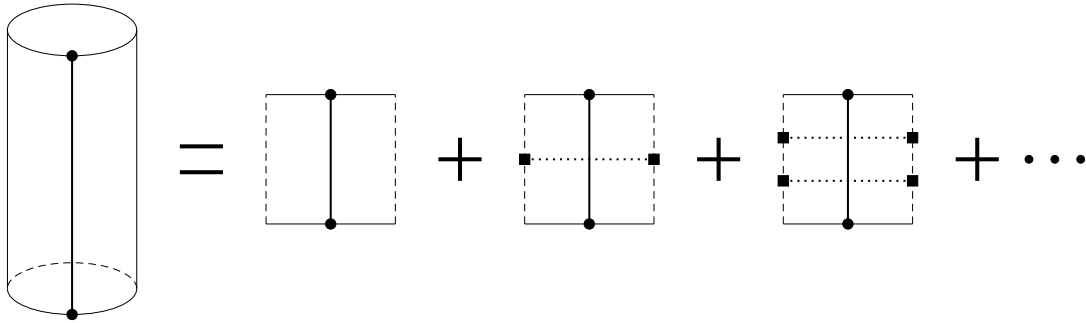


Figure 1. Pictorial representation of the cutting process applied to the cylinder. Note that cutting the cylinder just once imposes both mirror sides to have the same content.

the computation of two-point functions due to conformal symmetry. As a consequence, the focus has partially shifted to three-point functions. Although neither the interest nor the use of integrability are new in this context (see, for example, [19–22]), recently a very simple and successful proposal for an all-loop computation of three-point functions based on cutting the pair of pants into two separated hexagons has been put forward [23–25]. Interestingly, these hexagons can be understood as a form factor involving a defect operator. Since the operator involved is non-local, these form factor fulfil a slightly modified version Smirnov’s axioms [26]. It has been noticed that the hexagon form factors alone are enough to construct higher point functions [27–30] due to the feasibility of obtaining form factors with a larger number of sides just by gluing hexagons.

In this article we argue that, together with the hexagon form factor, we can define a square “form factor”, more akin to usual form factors, related to the cutting of the cylinder (see figure 1). We write form factor between quotation marks because it would correspond to the form factor of the identity, which is sometimes not considered truly a form factor but still satisfy Smirnov’s axioms. However, here we will construct the square starting from states constructed using the Algebraic Bethe Ansatz instead of states constructed with Zamolodchikov-Faddeev (ZF) operators, hence it does not fulfil the usual Smirnov’s axioms but a different ones. Despite their differences there exists a relationship between both squares, we will postpone the analysis of ZF squares to future works.

One of the most important differences between the square and the hexagon form factor refers to the processes of gluing and cutting: when we glue together two squares we get again a square and vice-versa when cutting. We can exploit this feature to write down restrictions on the analytic properties of squares and reconstruct the scalar product of states. In particular, we will explicitly reconstruct the Gaudin determinant, as we are going to focus mainly on a system with $SU(2)$ symmetry.

Another interesting characteristic of the square is how mirror excitations appear in the gluing procedure. In contrast with the hexagon form factor, the $\mathcal{N} = 4$ SYM square “form factor” is invariant under the full $PSU(2|2)_{c.e.}^2$ symmetry. As the central extension includes the dilatation operator, some configurations have to vanish and, in particular, the number of excitations in the upper and lower edges has to be the same when no mirror excitations are present. Thus, the gluing of two squares should involve only a finite sum due to symmetry

instead of the infinite sum that appears when gluing of two hexagons. This makes the square a simpler setting where we can get a better understanding of this procedure.

It is worth pointing that a different square form factor from the one proposed here has appeared in the context of correlation functions of three separated operators connected by Wilson lines [31]. Despite the different nature of both squares, we expect them to fulfil similar consistency conditions.

The article is structured as follows. In section 2 we review some concepts related to the Algebraic Bethe Ansatz. In section 3 we present the consistency condition for the square and use it to reconstruct the normalization of on-shell Bethe states of the SU(2) spin chain. In section 4 we extend the analysis to compute the normalization of off-shell Bethe states for the case of two and three excitations. We also compute the three highest orders in the large length expansion of the normalization of general SU(2) off-shell states. In section 5 we close with summary, conclusions and comments on some points that should be clarified in further works. In appendix A we review Kirchhoff’s matrix-tree theorem and its application to the Gaudin determinant. In appendix B we comment a little more on the form of the consistency condition for SU(N) spin chains. In appendix C we explain how to recover the space-time dependence of the two point function. In appendix D we expand on the constraints imposed by crossing symmetry. In appendix E we sketch the computation of the leading mirror corrections to the squares with zero and one physical excitation.

2 The SU(2) algebraic bethe ansatz

In this section we present the basis of the Algebraic Bethe Ansatz (ABA) for SU(2) models in order to fix the notation we will use through this article. We are going to work with the rational 6-vertex model R-matrix

$$\begin{aligned}
 R_{1,2}(\lambda) &= \alpha(u, v) \sum_{a=1}^2 E_{aa} \otimes E_{aa} + \beta(u, v) \sum_{1 \leq a \neq b \leq 2} E_{aa} \otimes E_{bb} \\
 &\quad + \sum_{1 \leq a \neq b \leq 2} \gamma_{ab}(u, v) E_{ab} \otimes E_{ba} \\
 &= \lambda(\mathbb{I}_a \otimes \mathbb{I}_b) + \frac{i\eta}{2} \sum_j (\sigma_a^j \otimes \sigma_b^j), \tag{2.1}
 \end{aligned}$$

but our results can be easily generalized to the trigonometric 6-vertex model by an appropriate substitution of the functions α , β and γ . Here E_{ab} is the matrix with zeros in every position except the (a, b) position, which contains a 1.

The monodromy matrix, written as a matrix on the auxiliary space, has the following structure

$$T_a(\lambda) = \prod_{j=1}^{\overrightarrow{L}} \mathcal{L}_{a,j}(u) = \mathcal{L}_{a,1}(u) \mathcal{L}_{a,2}(u) \dots \mathcal{L}_{a,L}(u) = \begin{pmatrix} A(\lambda) & B(\lambda) \\ C(\lambda) & D(\lambda) \end{pmatrix}, \tag{2.2}$$

where $\mathcal{L}_{a,j}(u)$ is the Lax matrix associated to the specific model. The action of these four operators over the reference state is given by

$$A(\lambda) |0\rangle = a(\lambda) |0\rangle, \quad B(\lambda) |0\rangle \neq 0, \quad (2.3)$$

$$D(\lambda) |0\rangle = d(\lambda) |0\rangle, \quad C(\lambda) |0\rangle = 0, \quad (2.4)$$

while their commutation relations are fixed by the RTT relation

$$R_{1,2}(\lambda - \mu) T_1(\lambda) T_2(\mu) = T_2(\mu) T_1(\lambda) R_{1,2}(\lambda - \mu). \quad (2.5)$$

Of the relations codified in this equation, we will only need the following seven of them

$$[B(\lambda), B(\mu)] = [C(\lambda), C(\mu)] = 0, \quad (2.6)$$

$$A(\lambda)B(\mu) = f(\mu, \lambda)B(\mu)A(\lambda) + g(\lambda, \mu)B(\lambda)A(\mu), \quad (2.7)$$

$$D(\lambda)B(\mu) = f(\lambda, \mu)B(\mu)D(\lambda) + g(\mu, \lambda)B(\lambda)D(\mu), \quad (2.8)$$

$$C(\mu)A(\lambda) = f(\mu, \lambda)A(\lambda)C(\mu) + g(\lambda, \mu)A(\mu)C(\lambda), \quad (2.9)$$

$$C(\mu)D(\lambda) = f(\lambda, \mu)D(\lambda)C(\mu) + g(\mu, \lambda)D(\mu)C(\lambda), \quad (2.10)$$

$$[C(\lambda), B(\mu)] = g(\lambda, \mu) [A(\lambda)D(\mu) - A(\mu)D(\lambda)], \quad (2.11)$$

where, for convenience, we have introduced the functions¹

$$\begin{aligned} f(\lambda, \mu) &= \frac{\alpha(\lambda, \mu)}{\beta(\lambda, \mu)}, & f(\lambda, \mu) &= \frac{\lambda - \mu + i}{\lambda - \mu}, & f(\mu, \lambda) &= \frac{\lambda - \mu - i}{\lambda - \mu}, \\ g(\lambda, \mu) &= \frac{\gamma_{12}(\lambda, \mu)}{\beta(\lambda, \mu)}, & g(\lambda, \mu) &= \frac{i}{\lambda - \mu}, & g(\mu, \lambda) &= \frac{-i}{\lambda - \mu}. \end{aligned}$$

Since the trace of the monodromy matrix, called transfer matrix, commute with itself for different values of the spectral parameter λ , its power expansion produces $L - 1$ commuting operators, any of which can be chosen as Hamiltonian. We can construct the Hilbert space by applying a stack of B operators over the reference state, but such space is overcomplete as not all the states created by this process are eigenstates of the transfer matrix. The states that are not eigenstates of the transfer matrix are not physical, hence we will call them *off-shell Bethe states*. Only the states constructed with B operators with some precise sets of rapidities are eigenstates of the transfer matrix and hence true physical states. We will call such states *on-shell Bethe states*. The admissible sets of rapidities are those that fulfil the Bethe Ansatz Equations

$$\frac{a(u_j)}{d(u_j)} = \prod_{k \neq j} \frac{f(u_j, u_k)}{f(u_k, u_j)}. \quad (2.12)$$

As most of the time we will work with sets of rapidities, in the remainder of the article the following shorthand notation will prove useful as it reduces the amount of products on

¹Using these functions instead of their explicit expressions allow us to generalize our results to the trigonometric model just by substituting their expression by the corresponding one without altering any formula. We also want to remark that we are defining $f(u, v) = 1 + g(u, v)$ in contrast with e.g. [32], where $f(u, v) = 1 - g(u, v)$.

the formulas

$$f(\{\bar{\alpha}\}, \{\alpha\}) = \prod_{\substack{u_i \in \alpha \\ v_j \in \bar{\alpha}}} f(v_j, u_i), \quad f^\neq(\{\alpha\}, \{\alpha\}) = \prod_{\substack{u_i, u_j \in \alpha \\ u_i \neq u_j}} f(u_i, u_j), \quad a(\{\bar{\alpha}\}) = \prod_{v_j \in \bar{\alpha}} a(v_j).$$

3 From the Gaudin determinant to the square and back

In this section we propose the *square form factor* and a consistency condition for it. After proposing this consistency equation, we verify that the Gaudin norm fulfils it as a check for our proposal. Reversing the argument, we are able to reconstruct the Gaudin determinant from the consistency condition and some initial conditions and analytic considerations.

3.1 The square bootstrap and the Gaudin determinant

Inspired by the BKV idea of cutting the three-point correlation function into two hexagon form factors [23], we apply the same procedure to two-point functions, i.e. a cylinder worldsheet. In this case we will obtain two squares after cutting two times, each with two physical sides and two mirror sides, being the last the new sides created by this process. We decided to define our square by cutting two times the cylinder instead of one time because in the later case the two mirror sides of the square are constrained to have the same number of excitations (see figure 1).

The most general square is characterized by four sets of rapidities, one for each of the physical sides and one for each of the mirror sides. For the most part of this article the two mirror sides will be empty, so we only have to indicate the content of the two physical sides. Therefore we denote the square by $S_L(\{u\} \rightarrow \{w\})$, where the subindex indicate the length of the physical sides. If both sets of rapidities are equal, we will just write $S_L(\{u\}) = S_L(\{u\} \rightarrow \{u\})$ to alleviate notation. With this notation the relationship between the scalar product of Bethe states and the square is given by

$$\left\langle 0 \left| \prod_i C(v_i) \prod_j B(u_j) \right| 0 \right\rangle = S_L(\{u\} \rightarrow \{v\}) + \text{mirror corrections} . \quad (3.1)$$

The superposition of two Bethe states gives us the two-point function of an operator at zero with an operator at infinity, i.e., the normalization of the two-point functions. Despite this being an unphysical quantity, it is important to compute it to correctly normalize our states. In this article we will consider the $L \gg 1$ limit, so we can ignore mirror corrections. In appendix E we will comment about how to include mirror corrections.

Although the construction of the hexagon form factor from symmetry and crossing arguments is very simple and elegant, a similar construction for the square presents some obstructions. Since here we are addressing the problem from the weak coupling limit, we do not have access to the crossing transformation. This transformation entails shifting the rapidity of an excitation in such a way that it crosses some particular cuts in the rapidity plane. However these cuts have a width of $4g$, thus they close at weak coupling and the mirror and crossing transformations are inaccessible perturbatively [36]. Nevertheless, an

all-loop construction of the square should behave nicely under crossing transformations, e.g. we expect that $S_L(\{u^{\text{cross}}\} \rightarrow \{w^{\text{cross}}\}) = S_L(\{w\} \rightarrow \{u\})$.

Among the several differences that exist between this square and the hexagon form factor, we are interested in the one concerning the cutting and the gluing of form factors. The cutting operation applied to the square remains well defined after we employ it once,² which means that we can continue cutting a square into smaller ones. We can reverse the argument and state that gluing two squares reconstructs one square. This idea can be used as a “bootstrap procedure” to construct such square. If we consider only the asymptotic part of the square (which means directly gluing them without adding excitations to the mirror edges), the reconstruction formula will be given by

$$S_L(\{u\} \rightarrow \{v\}) = \sum_{\substack{\alpha \cup \bar{\alpha} = \{u\} \\ \beta \cup \bar{\beta} = \{v\}}} w(\{\alpha\}, \{\bar{\alpha}\}, \{\beta\}, \{\bar{\beta}\}) S_{l_1}(\{\alpha\} \rightarrow \{\beta\}) S_{l_2}(\{\bar{\alpha}\} \rightarrow \{\bar{\beta}\}), \quad (3.2)$$

where the weight $w(\{\alpha\}, \{\bar{\alpha}\}, \{\beta\}, \{\bar{\beta}\})$, which we will call *breaking factor*, takes into account the cutting of the physical states into two. We have also defined $L = l_1 + l_2$. In this article we will work in the $L \sim l_1 \sim l_2 \gg 1$, so we will not consider any potential problems that could arise from the discrete nature of the states when we cut them. As we said before, this limit also allow us to ignore the mirror corrections as they are exponentially suppressed.

To construct the breaking factors we proceed in a similar manner as [38], i.e. we write the monodromy matrix as the product of two monodromy matrices, each associated to a subchain of the original chain. Using the definition of the monodromy matrix given in equation (2.2) we can rewrite the product of L Lax matrices into two separate products with l_1 and l_2 Lax matrices each. Equating the entries of the auxiliary space we can prove that

$$T_a^{(L)}(u) = T^{(l_1)}(u) \otimes T^{(l_2)}(u) \implies \begin{cases} B_L(u) = B_{l_1}(u) \otimes D_{l_2}(u) + A_{l_1}(u) \otimes B_{l_2}(u) \\ C_L(u) = C_{l_1}(u) \otimes A_{l_2}(u) + D_{l_1}(u) \otimes C_{l_2}(u) \end{cases} . \quad (3.3)$$

These formulas provide us with the recipe to break the creation and annihilation operators for the case of one excitation. For the general square we need the expression of several creation and annihilation operators. As the RTT relations involve at most two rapidities, it is easy to argue that the structure of the weight functions can be fixed from the breaking factor for two excitations. For the case of creation operators we have

$$\begin{aligned} B(u)B(v) &= B_{l_1}(u)B_{l_1}(v) \otimes D_{l_2}(u)D_{l_2}(v) + B_{l_1}(u)A_{l_1}(v) \otimes D_{l_2}(u)B_{l_2}(v) \\ &\quad + A_{l_1}(u)B_{l_1}(v) \otimes B_{l_2}(u)D_{l_2}(v) + A_{l_1}(u)A_{l_1}(v) \otimes B_{l_2}(u)B_{l_2}(v) \\ &= B_{l_1}(u)B_{l_1}(v) \otimes D_{l_2}(u)D_{l_2}(v) + f(u, v)B_{l_1}(u)A_{l_1}(v) \otimes B_{l_2}(v)D_{l_2}(u) \\ &\quad - g(u, v)B_{l_1}(u)A_{l_1}(v) \otimes B_{l_2}(u)D_{l_2}(v) + f(v, u)B_{l_1}(v)A_{l_1}(u) \otimes B_{l_2}(u)D_{l_2}(v) \\ &\quad + g(u, v)B_{l_1}(u)A_{l_1}(v) \otimes B_{l_2}(u)D_{l_2}(v) + A_{l_1}(u)A_{l_1}(v) \otimes B_{l_2}(u)B_{l_2}(v) . \end{aligned}$$

²Although this is true for cutting the square vertically, horizontal cuts should behave differently. We will not address that kind of cuts in this article.

In the second equality we have used RTT equations (2.7) and (2.8) to move all A and D operators to the right of the B operators. Applying this expression to the reference state gives us the contribution from B operators to the weight in the case of two excitations. As argued before, the case of more excitations is just an extension of this computation and gives us an equivalent structure. A similar procedure can be performed for annihilation operators by instead moving the A and D operators to the left of the C operators and applying the expression to the dual reference state. The final expression for the breaking factor is the product of the weights associated to B and C operators

$$\begin{aligned}
 w(\{\alpha\}, \{\bar{\alpha}\}, \{\beta\}, \{\bar{\beta}\}) &= w_B(\{\alpha\}, \{\bar{\alpha}\})w_C(\{\beta\}, \{\bar{\beta}\}) \\
 w_B(\{\alpha\}, \{\bar{\alpha}\}) &= f(\alpha, \bar{\alpha})a_{l_1}(\bar{\alpha})d_{l_2}(\alpha), \\
 w_C(\{\beta\}, \{\bar{\beta}\}) &= f(\bar{\beta}, \beta)d_{l_1}(\bar{\beta})a_{l_2}(\beta).
 \end{aligned}
 \tag{3.4}$$

Despite the simplicity of the result, this computation becomes rather messy when we move to $SU(N)$ for $N > 2$. A small discussion about it has been collected in appendix B, together with the computation of the breaking factor for $SU(2)$ when we work with operators that satisfy the Zamolodchikov-Faddeev algebra.

Let us perform now a first check of our proposal. The Gaudin determinant provides us the normalization of states of the XXX spin chain written using the Algebraic Bethe Ansatz. Its explicit expression is

$$G(\{u\}) = i^N f^\neq(\{u\}, \{u\}) \det \Phi, \quad \Phi_{ab} = -\frac{\partial}{\partial u_b} \left(\frac{a(u_a)}{d(u_a)} \prod_{k \neq a} \frac{f(u_c, u_a)}{f(u_a, u_c)} \right). \tag{3.5}$$

The proof that such determinant fulfils the consistency condition we have proposed can already be found in the literature, see for example eq. (36) in [33]. Instead, for later convenience it proves better to use the Kirchhoff's matrix-tree theorem [34] to rewrite the determinant expression as a sum over graphs. We have summarized this construction in appendix A and we refer to [35] and references therein for more details. In this forest expansion all the dependence on the length of the chain appears only through the factors associated to the root nodes of the trees. If we break the length L into $l_1 + l_2$, we have to add an extra sum over trees having either $l_1 p'$ or $l_2 p'$ as root factor. Grouping terms by root nodes and rapidity labels involved in each set of trees allows us to reconstruct the two factors $S_{l_1}(\{\alpha\})$ and $S_{l_2}(\{\bar{\alpha}\})$, with the corresponding sum coming from the forest expansion and choices of root nodes becoming the sum of equation (2.2). The appearance of the breaking factor comes not from separating the determinant into trees, but from separating the prefactor $f^\neq(\{u\}, \{u\})$.

It might seem strange that the Gaudin determinant directly fulfils this relation because the square should be an off-shell construction while the states involved in the Gaudin determinant are on-shell Bethe states. This is so because the terms dependent on the Bethe Ansatz Equations appear as

$$S_L(\{u\}) = \{\text{On-shell part}\} + \{\text{Bethe eq.}\} \cdot \{\text{Off-shell contributions}\}, \tag{3.6}$$

so the on-shell part should also fulfils the same consistency equation. In the next section we will check for some simple examples that the square has this structure.

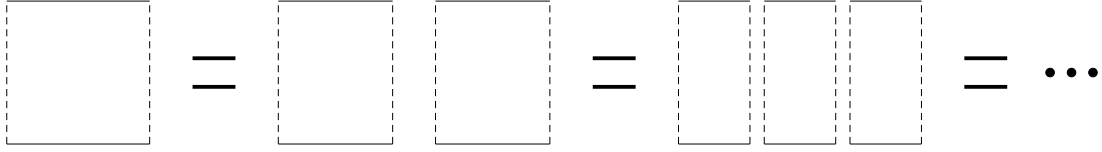


Figure 2. Pictorial representation of the cutting of an empty square.

3.2 Reconstructing the Gaudin determinant: some particular cases

After checking that the Gaudin determinant fulfils the consistency condition (3.2), we are going to try to reconstruct it from the consistency condition. First we present the reconstruction for the particular cases from zero up to three excitations in order to clarify the procedure. After that we will provide the fully reconstructed expression for a general number of excitations, up to some functions that cannot be fixed just using the consistency condition. We will fix them using comparison with explicit ABA computation and analytic considerations.

Basic idea of this reconstruction is to use the length structure in the consistency condition to impose constraints. While S_{l_1} and S_{l_2} depend explicitly on l_1 and l_2 respectively and the weight factors depend on those lengths implicitly though the momentum factors a and d , S_L cannot depend on these lengths but on the form $(l_1 + l_2)$.

Apart from the consistency condition, we will need two additional properties of the square to fix it. In particular we will need that

- The square is symmetric under the exchange of two rapidities, i. e.

$$S_L(u_1, \dots, u_i, \dots, u_j, \dots) = S_L(u_1, \dots, u_j, \dots, u_i, \dots) . \quad (3.7)$$

- The limit of the norm exists and is well defined (finite and non-vanishing)

$$\lim_{v_1 \rightarrow u_1} \lim_{v_2 \rightarrow u_2} \dots S_L(\{u\} \rightarrow \{v\}) = S_L(\{u\}) . \quad (3.8)$$

These are not the usual Smirnov's axioms for form factors because we are constructing the consistency condition from the ABA instead of from ZF operators, so the properties satisfied by both are different. In addition, because we are working in the weak coupling regime, we will not deal in this article with the constraints imposed by crossing transformations to this square.

The first case we should compute is the square with no excitations, which is fixed up to an unphysical normalization. The computation relies on the particular fact that such square can be cut as many times as we want (as long as we are far from the discrete limit) involving no other structures. As we can see from figure 2, the consistency condition for it is given by

$$S_L(\emptyset) = S_{l_1}(\emptyset)S_{l_2}(\emptyset) = S_{l'_1}(\emptyset)S_{l'_2}(\emptyset)S_{l'_3}(\emptyset) = \dots , \quad (3.9)$$

whose only physical solution is $S_L(\emptyset) = 1$.³

³Actually we can choose a more general solution $S_L(\emptyset) = \exp(\rho L)$ with ρ being a constant. However, we are going to choose $\rho = 0$ because there is no physical parameters with inverse length units we can use for that role.

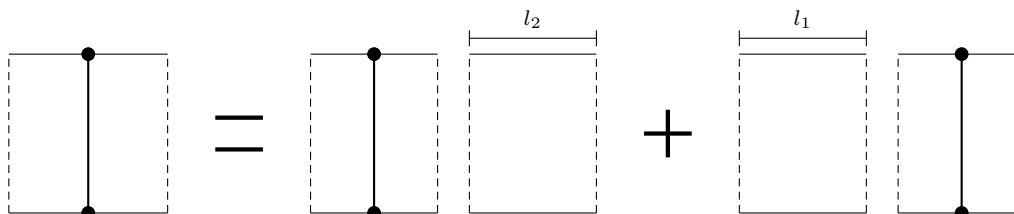


Figure 3. Pictorial representation of the cutting of a square with one excitation.

With the information about the empty square, we can move now to the computation of the square with one excitation. As we can see from its pictorial representation in figure 3, equation (3.2) reflects the fact that the excitation can end in either of the two new squares we get after we cut the original one. Applied to this case, this equation reduces to

$$S_L(u) = S_{l_1}(u)S_{l_2}(\emptyset)a_{l_2}(u)d_{l_2}(u) + S_{l_1}(\emptyset)S_{l_2}(u)a_{l_1}(u)d_{l_1}(u) . \quad (3.10)$$

No other term can appear as we are not including mirror excitations. We can see that this equation simplifies by means of the rewriting $S_L(u) = \tilde{S}_L(u)a_L(u)d_L(u)$, reducing to

$$\tilde{S}_L(u) = \tilde{S}_{l_1}(u) + \tilde{S}_{l_2}(u) . \quad (3.11)$$

The only way to solve this equation is to have $\tilde{S}_L(u) = \sigma_1^{(1)}(u)L$, where $\sigma_1^{(1)}$ is a general function not fixed by the constraint. Note that we cannot include a term independent of L as the equation automatically fixes it to zero.

We can fix the structure of the $\sigma_1^{(1)}(u)$ function if we compute the case in which the rapidity of the initial (lower-edge) excitation is different from the final (upper-edge) one and impose some analytic considerations. The consistency condition in this case reads

$$S_L(u \rightarrow v) = S_{l_1}(u \rightarrow v)S_{l_2}(\emptyset)a_{l_2}(v)d_{l_2}(u) + S_{l_1}(\emptyset)S_{l_2}(u \rightarrow v)a_{l_1}(u)d_{l_1}(v) . \quad (3.12)$$

To solve this equation we require the more general ansatz

$$S_L(u \rightarrow v) = \tilde{S}_L(u \rightarrow v) [a(u)d(v) - a(v)d(u)] , \quad (3.13)$$

which solves the consistency equation for any function $\tilde{S}_L(u \rightarrow v)$.

In order to get a non-vanishing norm we have to impose that $\tilde{S}_L(u \rightarrow v)$ has a pole of order one at $u = v$. This function can be further fixed through two different procedures. The first one is the direct comparison with the ABA matrix element $\langle 0|C(v)B(u)|0\rangle$ using (2.11), which directly fixes $\tilde{S}_L(u \rightarrow v) = g(u, v)$. The second one is to use crossing symmetry. However, as we commented above, at weak coupling the crossing transformation is not accessible, so in order to extract the function from this symmetry we have to take the weak coupling limit after performing the transformation. As this involves dealing with the non-perturbative square, we have relegated the discussion on this topic to appendix D.

With the information from $S_L(u \rightarrow v)$ we can now extract the Gaudin norm associated to the one-excitation state

$$\tilde{S}_L(u) = \frac{S_L(u)}{a(u)d(u)} = \lim_{v \rightarrow u} \tilde{S}_L(u \rightarrow v) \frac{d(v)}{a(u)} (e^{ip(v)L} - e^{ip(u)L}) = \text{const.}(-ip'L) , \quad (3.14)$$

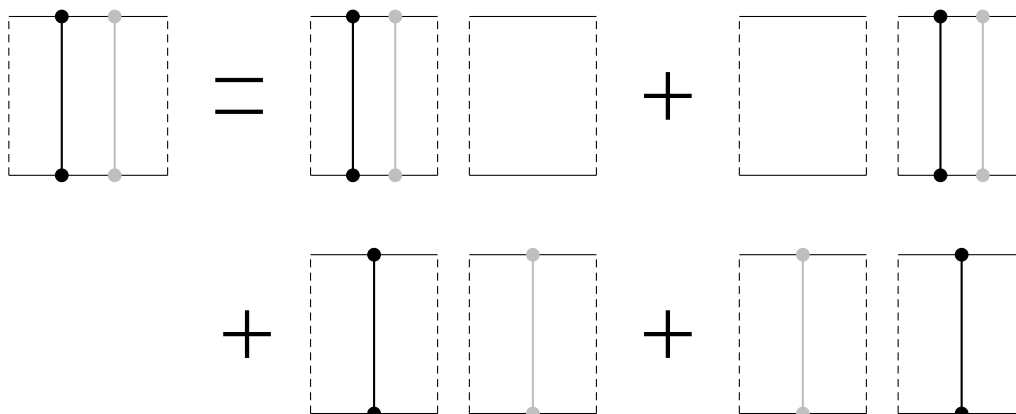


Figure 4. Pictorial mnemonic representation of the cutting of a square with two excitations. Each color represents a different rapidity.

where the momentum is defined as $ip(u)L = \log[a(u)/d(u)]$ and $p' = \frac{dp(u)}{du}$. We can see that the expression we have obtained agrees with the Gaudin norm (3.5) for one excitation.

Moving now to the case of two excitations, equation (3.2) (pictorially represented in figure 4) now reads

$$\begin{aligned}
 S_L(u, v) &= S_{l_1}(u, v)S_{l_2}(\emptyset)a_{l_2}(u)a_{l_2}(v)d_{l_2}(u)d_{l_2}(v) + S_{l_1}(\emptyset)S_{l_2}(u, v)a_{l_1}(u)a_{l_1}(v)d_{l_1}(u)d_{l_1}(v) \\
 &+ S_{l_1}(u)S_{l_2}(v)f(u, v)f(v, u)a_{l_1}(v)d_{l_1}(v)a_{l_2}(u)d_{l_2}(u) \\
 &+ S_{l_1}(v)S_{l_2}(u)f(u, v)f(v, u)a_{l_1}(u)d_{l_1}(u)a_{l_2}(v)d_{l_2}(v) .
 \end{aligned}
 \tag{3.15}$$

One might argue that there are two terms we have not taken into account, as we have not included terms involving squares whose excitations in the upper and lower sides have different rapidities. Although this is true in general, in this section we are trying to reconstruct the Gaudin norm and these terms vanish because $S_L(\{u\} \rightarrow \{v \neq u\}) = 0$ when the Bethe equations are imposed.

Following the same procedure we used for the case of one excitation, we can factor out the momentum dependence, encoded in the functions a and d , by defining $S_L(u, v) = \tilde{S}_L(u, v)a(u)a(v)d(u)d(v)$. However, in this case the ansatz for $\tilde{S}_L(u, v)$ has to be generalized to

$$\tilde{S}_L(u, v) = \sigma_2^{(2)}(u, v)L^2 + \sigma_1^{(2)}(u, v)L .
 \tag{3.16}$$

Note again that adding an extra term with no dependence in L is forbidden by the consistency condition. Substituting this ansatz into (3.15) fixes the first function to

$$\sigma_2^{(2)} = f(u, v)f(v, u)\sigma_1^{(1)}(u)\sigma_1^{(1)}(v),
 \tag{3.17}$$

but it does not give us any information about the second one. We can reconstruct the function $\sigma_1^{(2)}$ using the same trick as before, i.e., computing the square $S(\{u, v\} \rightarrow \{w, v\})$ and taking the limit $w \rightarrow u$. Nevertheless, we are going to postpone this calculation to section 4.1, where we compute the off-shell square with two excitations, because such limit should be taken without imposing Bethe equations.

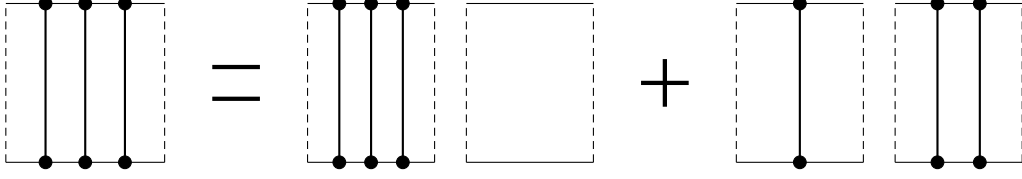


Figure 5. Schematic representation of the cutting of a square with three excitations. The other structures are obtained by exchanging l_1 and l_2 and the labeling of rapidities, giving a total of 8 contributions.

The last particular case we are going to dissect before presenting the general reconstruction is the one involving three excitations. In this case we can express equation (3.2), schematically represented in figure 5, as

$$\begin{aligned}
S_L(u, v, w) &= S_{l_1}(u, v, w)S_{l_2}(\emptyset)a_{l_2}(u)a_{l_2}(v)a_{l_2}(w)d_{l_2}(u)d_{l_2}(v)d_{l_2}(w) \\
&\quad + S_{l_1}(\emptyset)S_{l_2}(u, v, w)a_{l_1}(u)a_{l_1}(v)a_{l_1}(w)d_{l_1}(u)d_{l_1}(v)d_{l_1}(w) \\
&\quad + [S_{l_1}(u)S_{l_2}(v, w)f(u, v)f(u, w)f(v, u)f(w, u) \\
&\quad \quad \cdot a_{l_1}(v)d_{l_1}(v)a_{l_1}(w)d_{l_1}(w)a_{l_2}(u)d_{l_2}(u) \\
&\quad + S_{l_1}(v, w)S_{l_2}(u)f(u, v)f(u, w)f(v, u)f(w, u) \\
&\quad \quad \cdot a_{l_1}(u)d_{l_1}(u)a_{l_2}(v)d_{l_2}(v)a_{l_2}(w)d_{l_2}(w)] \\
&\quad + [u \leftrightarrow v] + [u \leftrightarrow w], \tag{3.18}
\end{aligned}$$

where the last two permutations are taken only over the terms in brackets.

To solve this consistency condition we need to substitute a polynomial of degree three in L as ansatz. The reason behind it is that the equation involves the product of a square with one excitation (linear in L) and a square with two excitations (quadratic in L)

$$S_L(u, v) = \left(\sigma_3^{(3)}(u, v)L^3 + \sigma_2^{(3)}(u, v)L^2 + \sigma_1^{(3)}(u, v)L \right) a(u)a(v)a(w)d(u)d(v)d(w). \tag{3.19}$$

Similarly to the case of two excitations, the consistency condition fixes $\sigma_3^{(3)}$ and $\sigma_2^{(3)}$ but the function $\sigma_1^{(3)}$ remains unconstrained. In particular

$$\sigma_3^{(3)} = f(u, v)f(u, w)f(v, w)f(w, v)f(w, u)f(v, u)\sigma_1^{(1)}(u)\sigma_1^{(1)}(v)\sigma_1^{(1)}(w), \tag{3.20}$$

$$\sigma_2^{(3)} = f(u, w)f(v, w)f(w, v)f(w, u)\sigma_1^{(2)}(u, v)\sigma_1^{(1)}(w) + [u \leftrightarrow w] + [v \leftrightarrow w]. \tag{3.21}$$

This structure is very reminiscent of the Gaudin determinant rewritten using the Kirchhoff's matrix-tree theorem if we understand the $\sigma_1^{(N)}$ as functions associated to spanning trees of the complete graph with N nodes. The relation between our results and the forest expansion will be clarified the following subsection.

3.3 Reconstructing the Gaudin determinant: the general asymptotic square

With the experience we have gathered from the previous particular cases, we are in a position to derive the full Gaudin determinant for the on-shell Bethe state with rapidities

$\{u\} = \{u_1, \dots, u_N\}$. The ansatz compatible with the consistency equation for the general case is

$$S_L(\{u\}) = a(\{u\})d(\{u\}) \sum_{n=1}^N \sigma_n^{(N)}(\{u\})L^n. \quad (3.22)$$

After substituting this ansatz into equation (3.2), the resulting constraints can be solved by

$$\sigma_i^{(N)} = f^\neq(\{u\}, \{u\}) \sum_{\substack{\alpha_1 \cup \alpha_2 \cup \dots \cup \alpha_i = \{u\} \\ \alpha_j \cap \alpha_k = \emptyset \\ \alpha_j \neq \emptyset}} \prod_{j=1}^i \frac{\sigma_1^{(|\alpha_j|)}(\{\alpha_j\})}{f^\neq(\{\alpha_j\}, \{\alpha_j\})}. \quad (3.23)$$

To clarify the formula, the explicit expressions for the two highest orders in length are

$$\sigma_N^{(N)} = f^\neq(\{u\}, \{u\}) \prod_{u_i \in \{u\}} \sigma_1^{(1)}(u_i), \quad (3.24)$$

$$\sigma_{N-1}^{(N)} = \sigma_N^{(N)} \sum_{\substack{u_i, u_j \in \{u\} \\ u_i \neq u_j}} \frac{\sigma_1^{(2)}(u_i, u_j)}{f(u_i, u_j)f(u_j, u_i)\sigma_1^{(1)}(u_i)\sigma_1^{(1)}(u_j)}. \quad (3.25)$$

As we said in the previous subsection, we can interpret this formula as the tree expansion of the Gaudin determinant. However, to be able to make this connection we still have to make another assumption. At this point, the $\sigma_i^{(N)}$ functions are the expressions associated to dividing the complete vertex-labelled graph with N vertices into i disconnected objects, but the individual $\sigma_1^{(|\alpha_i|)}$ still cannot be interpreted as a tree.

In order to proceed, we are going to assume that the $\sigma_1^{(n)}$ functions inherit certain factorization properties from the factorized scattering of the theory. In particular, we will assume that they can be written as the sum of the product of $n - 1$ factors of $\sigma_1^{(2)}$ over non-repeated pairings of the n arguments. This implies that we only have to fix the functions $\sigma_1^{(1)}$ and $\sigma_1^{(2)}$ to completely fix all squares, a task achievable just by using the ABA computations of the norm for one and two excitations as initial conditions. Even more, in the following subsection we will see that we can compute $\sigma_1^{(2)}$ just from the consistency condition, so we actually only need $\sigma_1^{(1)}$. A second consequence is that the sum over all possible pairings of the arguments can be understood as the sum over all possible spanning trees of the complete n -graph. The reason for that lies in the fact that the products involve $n - 1$ non-repeated pairs, which is equivalent to the definition of a tree, can be mapped to the tree graph constructed with each vertex and each edge representing respectively a rapidity variable and a function $\frac{\sigma_1^{(2)}(u,v)}{\sigma_1^{(1)}(u)+\sigma_1^{(1)}(v)}$ whose arguments are the vertices connected by the edge. Thus, allowing us to finally connect our formula with a forest expansion. We have chosen to divide $\sigma_1^{(2)}$ by the sum of the derivative of the momenta because it appears in the square with two magnons accompanied by a factor of the length, so it should contain a factor of the momentum and it should appear in a symmetric way. In the next section we will see that $\sigma_1^{(2)}(u, v) \propto [p'(u) + p'(v)]\partial_u \log S(u, v)$. We can check from

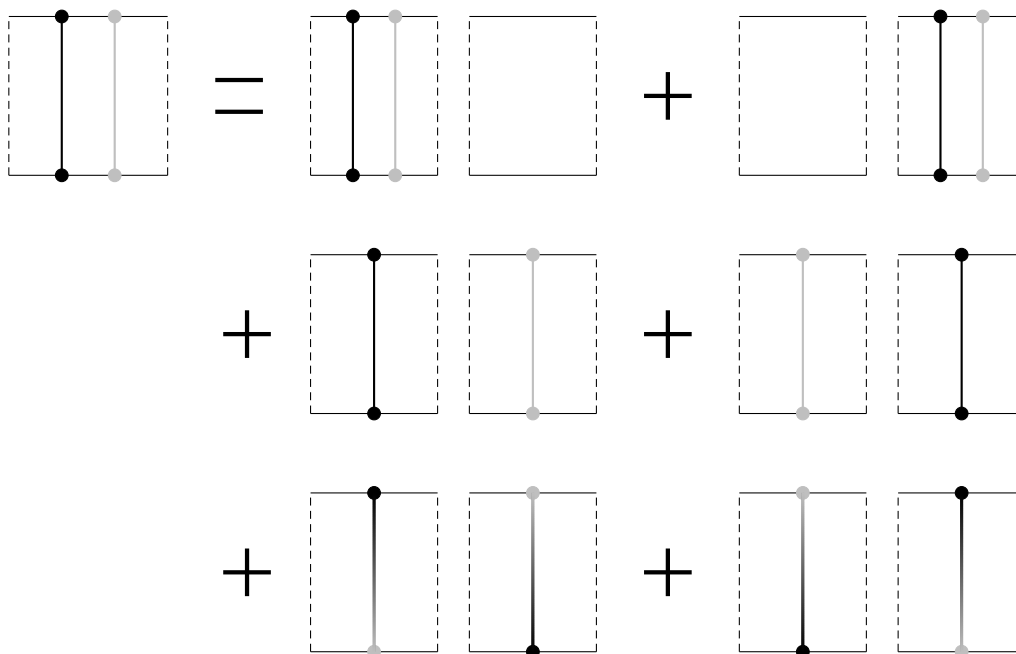


Figure 6. Pictorial representation of the cutting process applied to a square with two excitations. Each color represents a different rapidity.

the expression (A.2) that our naive construction for the $\sigma_1^{(n)}$ just lacks a factor of the sum of the momenta of all excitations.⁴

We still have to reconstruct the dependence of the two-point function with the space-time. We discuss two different approaches to it in appendix C.

4 Computing the off-shell square

After being able to reconstruct the Gaudin norm using the consistency condition 3.2, in this section we are going to include all the possible asymptotic terms appearing on the equation to reconstruct the norm of off-shell Bethe states. We will detail the computations for the cases of two and three excitations. We will also compute the first non-Gaudin term appearing in the square with a general number of excitations. The case of the square with one excitation require no special treatment as no new terms appear in the consistency equation, hence it remains unchanged

$$S_L(u \rightarrow v) = g(u, v) [a(u)d(v) - a(v)d(u)] , \quad S_L(u) = \text{const.}(-ip'L)e^{ip'L} . \quad (4.1)$$

4.1 The off-shell square with two excitations

The consistency condition for the case of two excitations, pictorially represented in figure 6, needs to include the two extra terms involving “no norm-like” one-excitation squares we

⁴The sum over momenta appearing here can be made explicit in (A.2) by separating the sum over directed forests into a sum over undirected forests and the contribution from the roots, being this last sum the sum over momenta.

neglected in the previous section, being its complete expression

$$\begin{aligned}
S_L(u, v) = & S_{l_1}(u, v)a_{l_2}(u)a_{l_2}(v)d_{l_2}(u)d_{l_2}(v) + S_{l_2}(u, v)a_{l_1}(u)a_{l_1}(v)d_{l_1}(u)d_{l_1}(v) + \\
& + S_{l_1}(u)S_{l_2}(v)f(u, v)f(v, u)a_{l_1}(v)d_{l_1}(v)a_{l_2}(u)d_{l_2}(u) \\
& + S_{l_1}(v)S_{l_2}(u)f(u, v)f(v, u)a_{l_1}(u)d_{l_1}(u)a_{l_2}(v)d_{l_2}(v) \\
& + S_{l_1}(v \rightarrow u)S_{l_2}(u \rightarrow v)f(v, u)^2 a_L(u)d_L(v) \\
& + S_{l_1}(u \rightarrow v)S_{l_2}(v \rightarrow u)f(u, v)^2 a_L(v)d_L(u) .
\end{aligned} \tag{4.2}$$

As the Gaudin norm solves this equation except for the last two terms, we are going to look for a solution of the form

$$S_L(u, v) = \tilde{S}_L^{(\text{Gaudin})}(u, v)a(u)a(v)d(u)d(v) + \text{extra terms} . \tag{4.3}$$

It is important to remark that these new terms do not even have to have the same momentum structure as the on-shell terms. Indeed

$$\begin{aligned}
S_{l_1}(u \rightarrow v)S_{l_2}(v \rightarrow u)a_L(u)d_L(v) = & -g(u, v)^2 \left[a_L(v)d_L(u) + a_L(v)d_L(u) \right. \\
& \left. - a_{l_1}(u)a_{l_2}(v)d_{l_2}(u)d_{l_1}(v) - a_{l_2}(u)a_{l_1}(v)d_{l_1}(u)d_{l_2}(v) \right] a_L(v)d_L(u) ,
\end{aligned} \tag{4.4}$$

cannot be compensated by such momentum structure. We can upgrade our ansatz to the following one

$$\begin{aligned}
S_L(u, v) = & \alpha_{u,v}[a_L(u)d_L(v)]^2 + \alpha_{v,u}[a_L(v)d_L(u)]^2 \\
& + \left[\sigma_2^{(2)}L^2 + \sigma_1^{(2)}(u, v)L + \sigma_0^{(2)} \right] a_L(u)a_L(v)d_L(u)d_L(v) ,
\end{aligned} \tag{4.5}$$

where the subindexes of the function α indicates the structure of the momentum terms accompanying them. Thus, the consistency condition can be divided into seven different equations according to the different momentum factor accompanying each term⁵

$$\begin{aligned}
\alpha_{u,v}^{(L)} = f(v, u)^2 g(u, v)^2 , \quad \alpha_{u,v}^{(l_1)} = f(u, v)^2 g(u, v)^2 , \quad \alpha_{u,v}^{(l_2)} = f(u, v)^2 g(u, v)^2 , \\
\alpha_{v,u}^{(L)} = f(u, v)^2 g(v, u)^2 , \quad \alpha_{v,u}^{(l_1)} = f(v, u)^2 g(v, u)^2 , \quad \alpha_{v,u}^{(l_2)} = f(v, u)^2 g(v, u)^2 , \\
\sigma_2^{(2)}L^2 + \sigma_1^{(2)}L + \sigma_0^{(2)} = \sigma_2^{(2)}l_1^2 + \sigma_0^{(2)} + \sigma_2^{(2)}l_2^2\sigma_1^{(2)}l_1 + \sigma_1^{(2)}l_2 + \sigma_0^{(2)} \\
+ 2l_1l_2\sigma_1^{(1)}(u)\sigma_1^{(1)}(v)f(u, v)f(v, u) \\
+ g(u, v)^2[f(u, v)^2 + f(v, u)^2] .
\end{aligned} \tag{4.6}$$

In contrast with the Gaudin square, here the term σ_0 is not only allowed, but needed and completely fixed. These equations are solved by

$$\begin{aligned}
\alpha_{u,v} = f(v, u)^2 g(u, v)^2 , \quad \sigma_2^{(2)} = f(u, v)f(v, u)\sigma_1^{(1)}(u)\sigma_1^{(1)}(v) , \\
\alpha_{v,u} = f(u, v)^2 g(u, v)^2 , \quad \sigma_0^{(2)} = -g(u, v)^2[f(u, v)^2 + f(v, u)^2] .
\end{aligned} \tag{4.7}$$

Again the function $\sigma_1^{(2)}(u, v)$ is not fixed by this procedure.

⁵Note that the first six equations reduce to only two and the condition that $\alpha_{u,v}$ and $\alpha_{v,u}$ are independent of the length.

Working with off-shell states we can constrain the function $\sigma_1^{(2)}(u, v)$ in a similar way as we constrained the function $\sigma_1^{(1)}(u)$ in the previous section, that is, by computing the square $S_L(\{u, v\} \rightarrow \{w, v\})$ and taking the limit $w \rightarrow v$. Although we only have to compute the terms proportional to $p'(v)$ to extract this function, we will compute the full square as it will appear in the consistency equation for a square with three excitations. The consistency condition for this square is

$$\begin{aligned}
S_L(\{u, v\} \rightarrow \{w, v\}) &= S_{l_1}(\{u, v\} \rightarrow \{w, v\})a_{l_2}(w)a_{l_2}(v)d_{l_2}(u)d_{l_2}(v) \\
&\quad + S_{l_2}(\{u, v\} \rightarrow \{w, v\})a_{l_1}(u)a_{l_1}(v)d_{l_1}(w)d_{l_1}(v) \\
&\quad + S_{l_1}(u \rightarrow w)S_{l_2}(v)a_{l_2}(w)a_{l_1}(v)d_{l_2}(u)d_{l_1}(v)f(u, v)f(v, w) \\
&\quad + S_{l_1}(v)S_{l_2}(u \rightarrow w)a_{l_1}(u)a_{l_2}(v)d_{l_1}(w)d_{l_2}(v)f(w, v)f(v, u) \\
&\quad + S_{l_1}(u \rightarrow v)S_{l_2}(v \rightarrow w)a_{l_1}(v)a_{l_2}(v)d_{l_2}(u)d_{l_1}(w)f(u, v)f(w, v) \\
&\quad + S_{l_1}(v \rightarrow w)S_{l_2}(u \rightarrow v)a_{l_1}(u)a_{l_2}(w)d_{l_1}(v)d_{l_2}(v)f(v, u)f(v, w) .
\end{aligned} \tag{4.8}$$

Here we have to choose the following ansatz

$$\begin{aligned}
S_L(\{u, v\} \rightarrow \{w, v\}) &= \beta_u(v)a(u)a(v)d(v)d(w) + \beta_w(v)a(w)a(v)d(v)d(u) \\
&\quad + \alpha_a a(u)a(w)[d(v)]^2 + \alpha_d [a(v)]^2 d(u)d(w) .
\end{aligned} \tag{4.9}$$

Substituted into the consistency condition it gives us nine equations, which can be reduced to only five equations and the condition that α_a and α_d are independent of the length. These equations are solved by

$$\begin{aligned}
\beta_u(v) &= g(u, w)f(w, v)f(v, u)\sigma_1^{(1)}(v)L + g(u, v)g(v, w) - [g(u, v)g(v, w)]^2 + k(u, v, w) , \\
\beta_w(v) &= g(w, u)f(u, v)f(v, w)\sigma_1^{(1)}(v)L + g(u, v)g(v, w) - [g(u, v)g(v, w)]^2 - k(u, v, w) , \\
\alpha_a &= -g(u, v)g(v, w)f(v, u)f(v, w) , \\
\alpha_d &= -g(u, v)g(v, w)f(u, v)f(w, v) .
\end{aligned} \tag{4.10}$$

Note that there is a length-independent contribution to the functions β , which we label here as $k(u, v, w)$, that cannot be fixed from the consistency condition because these relations only involves the combination $\beta_{u, l_1}(v) + \beta_{w, l_2}(v)$. We may think that going one step further and computing $S_L(\{u_1, u_2\} \rightarrow \{v_1, v_2\})$ would give us an equation that fixes this unknown function, but only an equivalent combination appears and the function cannot be fixed.

Part of $\sigma_1^{(2)}$ is extracted from this result by performing the limit $w \rightarrow u$ of the functions β with their corresponding momentum factor, which needs some care as the $g(u, w)$ factors will transform into derivatives. Focusing only in these factors, divided by $\sigma_1^{(1)}(v)L a(v)d(v)$, we have

$$\begin{aligned}
&\lim_{w \rightarrow u} g(u, w) \left[f(w, v)f(v, u)a(u)d(w) - f(u, v)f(v, w)a(w)d(u) \right] \\
&= \left(\frac{\partial f(u, v)}{\partial u} d(u) + f(u, v) \frac{\partial d(u)}{\partial u} \right) f(v, u)a(u) \\
&\quad - \left(\frac{\partial f(v, u)}{\partial u} a(u) + f(v, u) \frac{\partial a(u)}{\partial u} \right) f(u, v)d(u) ,
\end{aligned} \tag{4.11}$$

here the first term of each parenthesis together will generate half of $\sigma_1^{(2)}$ while the second ones generate $\sigma_2^{(2)}$. We can rewrite the two terms contributing to $\sigma_1^{(2)}$ as a single term

$$\sigma_1^{(2)} \propto [f(v, u)]^2 \frac{\partial \left(\frac{f(u, v)}{f(v, u)} \right)}{\partial u}, \quad (4.12)$$

which highlights that it is just the derivative of the S-matrix, as expected from the known expression of the Gaudin norm. The other half of $\sigma_1^{(2)}$ comes from the same limit applied to the α terms, giving the same result but dressed with $\sigma_1^{(1)}(u)$ instead.

To end the subsection, we want to show that the non-Gaudin terms contain the two-excitation Bethe equations. This can be made explicit if we perform the rewriting

$$\begin{aligned} & [a(u)d(v)f(v, u)g(u, v)]^2 + [a(v)d(u)f(u, v)g(v, u)]^2 \\ & \quad + a(u)a(v)d(u)d(v)g(u, v)g(v, u)[f(u, v)^2 + f(v, u)^2] \\ & = a(u)a(v)d(u)d(v)[g(u, v)]^2 f(u, v)f(v, u) \left[\frac{a(u)}{d(u)} \frac{d(v)}{a(v)} \frac{f(v, u)}{f(u, v)} - \frac{f(u, v)}{f(v, u)} \right. \\ & \quad \left. - \frac{f(v, u)}{f(u, v)} + \frac{d(u)}{a(u)} \frac{a(v)}{d(v)} \frac{f(u, v)}{f(v, u)} \right]. \end{aligned} \quad (4.13)$$

We can see that, apart from the vanishing inherited from the one-magnon square, these terms also vanish when the two-magnon Bethe equations (2.12)

$$\frac{a(u)}{d(u)} = \frac{d(v)}{a(v)} = \frac{f(u, v)}{f(v, u)}, \quad (4.14)$$

are fulfilled, proving that the reconstruction is consistent with the Bethe Ansatz.

4.2 The off-shell square with three excitations

The computation of the square for the case of three excitations follows the same pattern as the previous case, but the large number of different combinations of momenta obscures the consistency condition. However, our experience from isolating the different momentum contributions in the two-excitation case had shown us that most of the equations we get are redundant. If we substitute the ansatz

$$\begin{aligned} S_L(u, v, w) & = \alpha_{u,v} [a_L(u)d_L(v)]^2 a_L(w)d_L(w) + \alpha_{v,u} [a_L(v)d_L(u)]^2 a_L(w)d_L(w) \\ & \quad + (w \leftrightarrow u) + (w \leftrightarrow v) \\ & \quad + \left[\sigma_3^{(3)} L^3 + \sigma_2^{(3)} L^2 + \sigma_1^{(3)} L + \sigma_0^{(3)} \right] a_L(u)a_L(v)a_L(w)d_L(u)d_L(v)d_L(w), \end{aligned} \quad (4.15)$$

into the consistency condition, only three sets of equations are relevant while the others can be extracted from symmetry under exchange of rapidities or independence of the length.

One set of equations we have to study is the set associated to the term $a(\{u\})d(\{u\})$, which codifies most of the length-dependent structure of the square. It is also important to study these equations because of the presence in the consistency condition of terms with the structure $\sigma_1^{(1)}\sigma_0^{(2)}L$, which were absent in the consistency condition for ‘‘Gaudin

squares” and would make the equation inconsistent. Indeed, since the equation for $\sigma_1^{(3)}$ takes the form

$$\sigma_1^{(3)}L = \sigma_1^{(3)}l_1 + \sigma_1^{(3)}l_2 + h(u, v, w)L \quad (4.16)$$

a solution does not exist unless the function $h(u, v, w)$ vanishes.⁶ The extra contribution in this equation is given by

$$\begin{aligned} & f(u, w)f(v, u)[f(v, w)]^2g(v, w)\beta_v(u) + f(u, v)f(w, u)[f(w, v)]^2g(w, v)\beta_w(u) \\ & + f(u, v)f(u, w)f(v, u)f(w, u)\sigma_1^{(1)}(u)\sigma_0^{(2)}(v, w) + (u \leftrightarrow v) + (u \leftrightarrow w), \end{aligned} \quad (4.17)$$

which we can be shown to vanish when the explicit expression for β and σ from equations (4.7) and (4.10) are substituted, as requires by consistency. Thus the equation we get just shows us that the function $\sigma_1^{(3)}$ is not fixed by the consistency condition.

We should remark two additional points regarding this set of equations. First, the orders L^3 and L^2 receive no contributions from the no norm-like squares, hence $\sigma_3^{(3)}$ and $\sigma_2^{(3)}$ are not modified with respect to the Gaudin case. And second, as the equation that fixes $\sigma_0^{(3)}$ involves the function $k(u, v, w)$, appearing from the length-independent structure of the functions β , we need a larger set of inputs to completely fix the general square than for the Gaudin case.

The other equations we need to solve are the ones obtained from terms involving the momentum factors $a(\{u\})d(\{u\})\frac{a(u)}{d(u)}\frac{d(v)}{a(v)}$ and $a(\{u\})d(\{u\})\frac{d(u)}{a(u)}\frac{a(v)}{d(v)}$. Any other combination of rapidities is obtained by means of the symmetry inherited from the exchange symmetry of the B and C operators (2.6). There equations are

$$\alpha_{u,v} = g(u, v)[f(v, u)]^2f(w, u)f(v, w)\beta_u(w), \quad \alpha_{v,u} = g(u, v)[f(u, v)]^2f(u, w)f(w, v)\beta_v(w).$$

Substituting the expression for the functions β from equation (4.10) we get

$$\begin{aligned} \alpha_{u,v} &= [g(u, v)f(v, u)f(w, u)f(v, w)]^2\sigma_1^{(1)}(w)L \\ &\quad - g(u, w)g(w, v)g(v, u)[f(v, u)]^2f(w, u)f(v, w) \\ &\quad + [g(u, w)g(w, v)]^2g(v, u)[f(v, u)]^2f(w, u)f(v, w) \\ &\quad + g(u, v)[f(v, u)]^2f(w, u)f(v, w)k(u, w, v), \\ \alpha_{v,u} &= -[g(u, v)f(u, v)f(u, w)f(w, v)]^2\sigma_1^{(1)}(w) \\ &\quad + g(v, w)g(w, u)g(u, v)[f(u, v)]^2f(u, w)f(w, v) \\ &\quad - [g(v, w)g(w, u)]^2g(u, v)[f(u, v)]^2f(u, w)f(w, v) \\ &\quad - g(u, v)[f(u, v)]^2f(u, w)f(w, v)k(u, w, v), \end{aligned} \quad (4.18)$$

which matches the norm computed using the commutation relations for the B and C operators at order L .

⁶This is true for any power of L , as the equation $\sigma_j^{(n)}L^j = \sigma_j^{(n)}l_1^j + \sigma_j^{(n)}l_2^j + \sum_{i=1}^{j-1} \binom{j}{i}\alpha(\{u\})l_1^{j-i}l_2^i + h(u, v, w)L^j$ cannot be solved unless $h(u, v, w)$ vanishes.

4.3 The general off-shell norm-like square

As we have seen in the previous case, the difficulty in computing the off-shell square and the normalization of off-shell Bethe states increases heavily with respect to the on-shell case. We are not able to give a full general expression for the off-shell square in this article, but we can at least give the exact expression for the first no norm-like contribution, which appears in the third highest order term in length. In order to do so it is more convenient to rewrite equation (3.2) as

$$S_L(\{u\}) = \sum_{\alpha \cup \beta \cup \gamma \cup \delta = \{u\}} w\left(\{\alpha \cup \gamma\}, \{\beta \cup \delta\}, \{\alpha \cup \delta\}, \{\beta \cup \gamma\}\right) \cdot S_{l_1}\left(\{\alpha \cup \gamma\} \rightarrow \{\alpha \cup \delta\}\right) S_{l_2}\left(\{\beta \cup \delta\} \rightarrow \{\beta \cup \gamma\}\right). \quad (4.19)$$

A factor of L appears for each pair of upper- and lower-edge excitations with the same rapidity, hence in the large length limit we have $S_{l_1}\left(\{\alpha \cup \gamma\} \rightarrow \{\alpha \cup \delta\}\right) \propto l_1^{|\{\alpha\}|}$ and similarly for S_{l_2} and $\{\beta\}$. The terms in the consistency condition involving the highest powers of the length come from the terms of the sum with coinciding outgoing and incoming rapidities, i.e. from the terms with $\delta = \gamma = \emptyset$. As a consequence, to get the highest and second highest contributions in length we only have to take into account norm-like squares, and thus these contributions coincide with the corresponding terms appearing in Gaudin norm. The third and fourth orders get extra contributions apart from the ones coming from the Gaudin norm. In particular, these extra contributions come from the terms in the consistency condition in which both squares have one excitation different in the upper edge with respect to the lower edge ($|\delta| = |\gamma| = 1$, so its highest power has two powers of L less than the highest power of the Gaudin contribution, one from each square).

Nevertheless, since the consistency condition involves the expression for the leading order in length of the square with one excitation different in the outgoing state with respect to the ingoing one, $S_L(\{u_i \cup \alpha\} \rightarrow \{u_j \cup \alpha\})$, we have to find its expression first. The correct ansatz in this case is

$$S_L(\{u\} \rightarrow \{v \cup \bar{u}_1\}) = a(\{u\})d(\{u\}) \left[\beta_{1,v}(\{\bar{u}_1\}) \frac{d(v)}{d(u_1)} + \beta_{v,1}(\{\bar{u}_1\}) \frac{a(v)}{a(u_1)} + \mathcal{O}(L^{N-1}) \right], \quad (4.20)$$

where $\{\bar{u}_1\}$ stands for the set $\{u\}$ without the element u_1 . We have already computed this square for two excitations and we have seen that it is proportional to the Gaudin norm of the excitation that is present in both physical sides. Inspired by this, we are going to propose that the functions $\beta_{1,v}(\{\bar{u}_1\})$ and $\beta_{v,1}(\bar{u}_1)$ are proportional to $\tilde{S}^{(\text{Gaudin})}(\{\bar{u}_1\})$ and check it by induction.

The leading order in length of the equation in the consistency condition associated to the ratio $\frac{a(v)}{a(u_1)}$ is

$$\begin{aligned} \beta_{v,1}^{(L)}(\bar{u}_1) &= \beta_{v,1}^{(l_1)}(\bar{u}_1) + g(v, u_1) f(u_1, \bar{u}_1) f(\bar{u}_1, v) \tilde{S}_{l_2}(\bar{u}_1) \\ &+ \sum_{\substack{\alpha \cup \beta = \bar{u}_1 \\ \alpha \neq \emptyset}} f(u_1, \beta) f(\beta, v) f(\alpha, \beta) f(\beta, \alpha) \beta_{v,1}^{l_1}(\alpha) \tilde{S}_{l_2}(\beta). \end{aligned} \quad (4.21)$$

Thus, we can check that assuming the following structure for $|\alpha| < |\bar{u}_1|$

$$\begin{aligned}\beta_{1,v}^{(L)}(\alpha) &= f(\{\alpha\}, u_1) f(v, \{\alpha\}) g(u_1, v) \tilde{S}_L^{(Gaudin)}(\{\alpha\}), \\ \beta_{v,1}^{(L)}(\alpha) &= -f(\{\alpha\}, v) f(u_1, \{\alpha\}) g(u_1, v) \tilde{S}_L^{(Gaudin)}(\{\alpha\}).\end{aligned}\tag{4.22}$$

gives an equivalent structure for $\beta_{1,v}^{(L)}(\bar{u}_1)$ and $\beta_{v,1}^{(L)}(\bar{u}_1)$, as the consistency condition can be recast into the consistency condition of the Gaudin square.

Now we can compute the first non-Gaudin contribution to third order in length of the off-shell square. The correction is generated from the terms $S_{l_1}(\{u_i \cup \alpha\} \rightarrow \{u_j \cup \alpha\}) \cdot S_{l_2}(\{u_j \cup \beta\} \rightarrow \{u_i \cup \beta\})$ and the one obtained by exchanging l_1 and l_2 , which can be divided into the following three momentum contributions

$$\begin{aligned}a(u_i) d(u_j) [f(u_i, u_j)]^2 f(u_i, \bar{u}_{ij}) f(\bar{u}_{ij}, u_j) f(\alpha, \beta) \\ \cdot \left[a(u_j) d(u_i) \beta_{j,i}^{(l_1)}(\alpha) \beta_{j,i}^{(l_2)}(\beta) + a(u_i) d(u_j) \beta_{i,j}^{(l_1)}(\alpha) \beta_{i,j}^{(l_2)}(\beta) \right] \\ + a(u_j) d(u_i) [f(u_j, u_i)]^2 f(u_j, \bar{u}_{ij}) f(\bar{u}_{ij}, u_i) \\ \cdot \left[a(u_i) d(u_j) \beta_{i,j}^{(l_1)}(\beta) \beta_{i,j}^{(l_2)}(\alpha) + a(u_j) d(u_i) \beta_{j,i}^{(l_1)}(\beta) \beta_{j,i}^{(l_2)}(\alpha) \right].\end{aligned}\tag{4.23}$$

For each different contribution we can reconstruct the consistency condition for $\tilde{S}^{(Gaudin)}$, allowing us to write

$$\begin{aligned}S_L(\{u\}) &= S_L^{(Gaudin)}(\{u\}) + \sum_{i \neq j} \left\{ [g(u_i, u_j) f(u_j, u_i) f(\bar{u}_{ij}, u_j) f(u_i, \bar{u}_{ij})]^2 \right. \\ &\quad + [g(u_i, u_j) f(u_i, u_j) f(\bar{u}_{ij}, u_i) f(u_j, \bar{u}_{ij})]^2 + \left[\frac{a(u_i) d(u_j)}{a(u_j) d(u_i)} [f(u_i, u_j) g(u_i, u_j)]^2 \right. \\ &\quad \cdot f(u_i, \bar{u}_{ij}) f(u_j, \bar{u}_{ij}) f(\bar{u}_{ij}, u_i) f(\bar{u}_{ij}, u_j) + i \leftrightarrow j \left. \right\} a(\{u\}) d(\{u\}) \tilde{S}^{(Gaudin)}(\bar{u}_{ij}) \\ &\quad + \mathcal{O}(L^{|\{u\}|-4}).\end{aligned}\tag{4.24}$$

5 Conclusions

In this article we have studied how we can apply a cutting operation akin to the one used in the hexagon proposal to a cylindrical worldsheet. As the cutting operator is well defined after being applied once, we can define a square “form factor” through a consistency condition coming from subsequent applications of the cutting procedure. We have proven that, applying this construction to an integrable $SU(2)$ model, the consistency condition retrieves the Gaudin determinant. We have also computed off-shell norm of Bethe states for the $SU(2)$ spin chain for the cases of a few excitations.

There exists several possible generalization of the work presented here. The most immediate are the generalizations to the full $PSU(2, 2|4)$ symmetry and to general values of the 't Hooft coupling. There also exist two computations that provide relevant checks for the validity of the square and its consistency condition. The first computation we propose is to calculate higher loops and check the proposal for the “all-loops normalization”

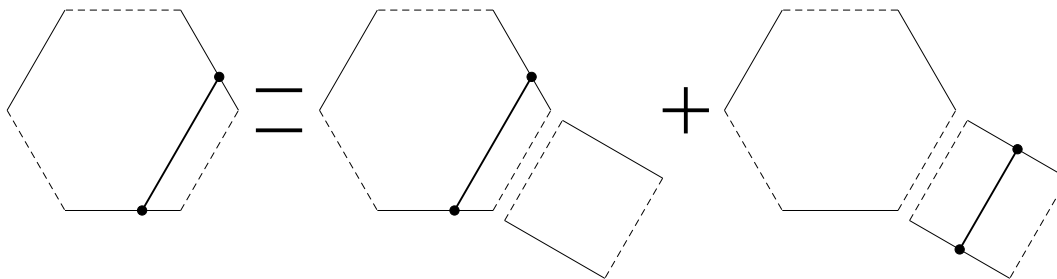


Figure 7. First non-trivial asymptotic consistency condition involving a hexagon and a square.

presented in eq. (17) of [40], which amount to replace the Heisenberg S-matrix by the all-loops S-matrix and a modification of the $f^\neq(\{u\}, \{u\})$ prefactor. The results obtained here can easily be generalized to inhomogeneous spin chains, so we can use either the θ -morphism [40] or the inhomogeneous version of the BDS spin chain [41] to check it. A second computation is to study the extremal limit of the hexagon form factor. In this limit one of the physical sides of the hexagon vanishes, as it has zero length, effectively transforming it into a square.⁷ These computations will be addressed in future works.

However, the most interesting line of work consists in integrating this square proposal with the hexagon proposal. In principle, we expect that the hexagon form factor follows a similar consistency condition to the one proposed here for the square, which in the asymptotic regime should be given by

$$\begin{aligned} & \mathcal{H}(\{u\}, \{v\}, \{w\})_{L_1, L_2, L_3} \\ &= \sum_{\substack{\alpha \cup \bar{\alpha} = \{u\} \\ \beta \cup \bar{\beta} = \{v\}}} w(\{\alpha\}, \{\bar{\alpha}\}, \{\beta\}, \{\bar{\beta}\}) \mathcal{H}_{l_1, l_2, L_3}(\{\alpha\}, \{\beta\}, \{w\}) S_l(\{\bar{\alpha}\} \rightarrow \{\bar{\beta}\}), \end{aligned} \quad (5.1)$$

where $l_1 + l = L_1$ and $l_2 + l = L_2$, and similarly with other pairs of physical sides of the hexagon. Figure 7 shows a pictorial representation of the consistency equation for the hexagon with two excitation. If we decide to include mirror contributions, the first non-trivial consistency condition we can write involves instead the hexagon with one excitation, as we can see in figure 8. The problem regarding this consistency condition is that the square constructed here has been defined using a different operator algebra than the one used for the hexagon proposal. Here we have constructed the breaking factors from the commutation relations of the B operator from the Algebraic Bethe Ansatz, while excitations in the hexagon proposal behaves according to a Zamolodchikov-Faddeev algebra.⁸ This implies that the weights associated to both constructions are different, as commented in appendix B. Despite that, we expect such consistency condition to hold when both form factors are written using the ZF algebra.

Regarding the square described in [31], we expect it to fulfil a consistency condition similar to the one we found here. We denote it by $S_{KK}^{(L)}$ or $S_{KK}^{(R)}$, depending on which

⁷The author would like to thank Ivan Kostov and Thiago Fleury for proposing this method.

⁸For a construction of the hexagon vertex using ZF operators, see section 7.3 of [39].

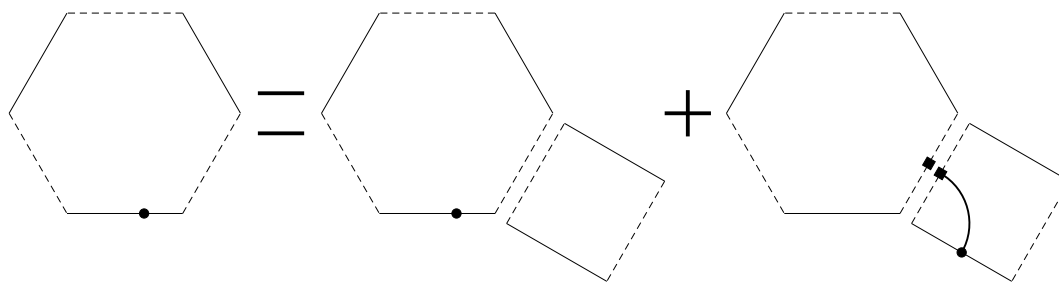


Figure 8. First non-trivial consistency condition involving a hexagon and a square if we include mirror contributions.

edge carries the contribution from the Wilson loop, to distinguish it from the square we are proposing. With this notation, we expect these two kinds of squares to satisfy the following consistency condition.

$$\begin{aligned}
 S_{KK,L}^{(L)}(\{u\} \rightarrow \{v\}) &= \sum_{\substack{\alpha \cup \bar{\alpha} = \{u\} \\ \beta \cup \bar{\beta} = \{v\}}} w(\alpha, \bar{\alpha}, \beta, \bar{\beta}) S_{KK,l_1}^{(L)}(\alpha \rightarrow \beta) S_{l_2}(\bar{\alpha} \rightarrow \bar{\beta}), \\
 S_{KK,L}^{(R)}(\{u\} \rightarrow \{v\}) &= \sum_{\substack{\alpha \cup \bar{\alpha} = \{u\} \\ \beta \cup \bar{\beta} = \{v\}}} w(\alpha, \bar{\alpha}, \beta, \bar{\beta}) S_{l_1}(\alpha \rightarrow \beta) S_{KK,l_2}^{(R)}(\bar{\alpha} \rightarrow \bar{\beta}), \quad (5.2)
 \end{aligned}$$

plus mirror corrections. In the same way that three squares S_{KK} glued to the hexagon give us the three open string interaction in the semiclassical limit, we expect the gluing of $S_{KK}^{(L)}$ with $S_{KK}^{(R)}$ to be equivalent to the propagation of an open string and the gluing of S_L with itself to be equivalent to the propagation of a closed string.

Although only sketched here, the square could be used as a tool for understanding the gluing procedure. The appeal of the square over the hexagon and octagon form factors, for which some interesting results have already been obtained [24, 42, 43], is its simplicity. As sketched in appendix E, the gluing of two squares or the gluing of a square and a hexagon form factor involve a finite amount of terms.

Acknowledgments

I would like to thank Yunfeng Jiang, Sergey Frolov and Shota Komatsu for interesting discussions and Ivan Kostov, Didina Serban, Thiago Fleury, Minkyoo Kim, Rafael Hernández and Roberto Ruiz for questions and comments on the manuscript. I thank the Institut de Physique Théorique (IPhT) for the hospitality during my visits.

A Kirchhoff’s matrix-tree theorem and Gaudin determinant

In this appendix we review the Kirchhoff’s matrix-tree theorem and the results of applying it to the Gaudin determinant.

Given a graph G with n vertices, we can define three matrices that completely define it: the degree matrix, the adjacency matrix and the Laplacian matrix, being the last one the difference of the two previous one. The degree matrix is a diagonal matrix indicating

the number of edges attaches to each vertex. The adjacency matrix is a square matrix such that its element a_{ij} with $i \neq j$ is one if there exist an edge from vertex i to vertex j and zero if there is not while its elements a_{ii} are equal to zero as we are not going to consider graphs with loops (edges that start and end in the same vertex). If there is a loop in vertex i , then we have to set $a_{ii} = 2$ instead.

To state the Kirchhoff's theorem we have first to define what a spanning tree is. On the one hand, a graph H with n vertices is said to be a tree if there is a unique path between any two vertices. This condition is equivalent to H being connected and having $n - 1$ edges. The connectivity condition can be replaced with the condition of having no cycles without adding any further restriction. A directed tree is defined as a directed acyclic graph whose underlying undirected graph is a tree. A subclass of directed trees are rooted trees, which are defined as directed trees with a natural orientation away from a vertex called root. A forest graph is just an acyclic graph, so it can be understood as a graph consisting of one or more tree graphs. On the other hand, a subgraph H of a graph G is said to be a spanning subgraph if both have the same vertex set. Finally, a spanning rooted tree is a spanning subgraph with is also a rooted tree.

Kirchhoff's matrix-tree theorem states that the number of spanning trees contained in a graph G is given by the determinant of any first minors of its Laplacian matrix with respect to diagonal elements, i.e., the Laplacian matrix with row and column j removed for any j . This theorem can be generalized to higher degree minors or more complex matrices, we refer to [35] and references within for a more comprehensive explanation.

We are interested in the Gaudin determinant, which is the determinant of a matrix of the form

$$\Phi_{ab} = D_a \delta_{ab} - K_{ab}, \tag{A.1}$$

where $D_a = L \partial_{u_a} p(u_a)$ and $iK_{a,b} = \partial_{u_a} \log S(u_a, u_b)$. The determinant of this matrix can be expressed as

$$\det \Phi = \sum_{\text{Forests}} \prod_{\text{roots}} D_a \prod_{\text{edges}} K_{ab}. \tag{A.2}$$

The only dependence of the determinant on the length of the spin chain is codified in the terms D_a , so the highest power on the length comes from the case of the forest made of isolated vertices. Furthermore, if we break the length L into $l_1 + l_2$, we have to add an extra sum over trees having either $l_1 p'$ or $l_2 p'$ as root factor.

B More on breaking factors

In this section we sketch some results about breaking factors beyond rank one. We will also comment on how to modify the construction for operators obeying the Zamolodchikov-Faddeev algebra.

Let us start with the breaking factor for an $SU(N)$ spin chain with R-matrix

$$\begin{aligned} R_{1,2}(\lambda) = & f(u, v) \sum_{a=1}^N E_{aa} \otimes E_{aa} + \sum_{1 \leq a \neq b \leq N} E_{aa} \otimes E_{bb} \\ & + g(u, v) \sum_{1 \leq a \neq b \leq N} (\delta_{a,b+1} + \delta_{a,b-1}) E_{ab} \otimes E_{ba}. \end{aligned} \tag{B.1}$$

In this case the RTT relation (2.5) can be expressed in the two following ways⁹

$$\begin{aligned} [T_{ij}(u), T_{kl}(v)] &= g(u, v) (T_{il}(u)T_{kj}(v) - T_{il}(v)T_{kj}(u)), \\ [T_{ij}(u), T_{kl}(v)] &= g(u, v) (T_{kj}(v)T_{il}(u) - T_{kj}(u)T_{il}(v)). \end{aligned} \quad (\text{B.2})$$

Writing the monodromy matrix as the product of two monodromy matrices and equating the entries of the auxiliary space we get

$$T_{ij}^{(L)} = \sum_k T_{ik}^{(l_1)} \otimes T_{kj}^{(l_2)}. \quad (\text{B.3})$$

It is more convenient to divide the sum into five different sums depending on if k is between i and j , equal to one of them, or greater or lower than both. To clarify the following computations we are going to write $T_{ii} = A_i$, $T_{ij} = B_{ij}$ if $i < j$ and $T_{ij} = C_{ij}$ if $i > j$. To compute the breaking factor we have to compute the breaking of $B_{ij}(u)B_{kl}(v)$ and $C_{kl}(v)C_{ij}(u)$. In this appendix we are only going to examine the cases “ $k = i$ and $l = j$ ” and “ $j = i + 1$ and $l = k + 1$ ” for B operators due to the length of the results. We have to take into account that B operators only commute when both indices are equal.

The breaking factor in the first case is extracted from the following expression

$$\begin{aligned} B_{ij}(v)B_{ij}(u) &= A_i(v)A_i(u) \otimes B_{ij}(v)B_{ij}(u) + B_{ij}(v)B_{ij}(u) \otimes A_j(v)A_j(u) \\ &+ f(v, u)B_{ij}(v)A_i(u) \otimes B_{ij}(u)A_j(v) + f(u, v)B_{ij}(u)A_i(v) \otimes B_{ij}(v)A_j(u) \\ &+ \sum_{k=i+1}^{j-1} \left[f(u, v)B_{ik}(u)A_i(v) \otimes B_{ij}(v)B_{kj}(u) \right. \\ &\quad + f(v, u)B_{ik}(v)A_i(u) \otimes B_{ij}(u)B_{kj}(v) \\ &\quad + f(v, u)B_{ij}(v)B_{ik}(u) \otimes B_{kj}(u)A_j(v) \\ &\quad \left. + f(u, v)B_{ij}(u)B_{ik}(v) \otimes B_{kj}(u)A_j(v) \right]. \end{aligned} \quad (\text{B.4})$$

It is worth noticing that the first three contributions are equal to the contributions we find for the $SU(2)$ case. In fact, for operators with $j = i + 1$ the sum on k vanishes and we get the same result as in the $SU(2)$ case.

Let us move to the breaking factor of $B_{i,i+1}(u)B_{j,j+1}(v)$. For simplicity, we are going to assume that $i < j$, but a similar computation can be done in the converse case.

$$\begin{aligned} B_{i,i+1}(v)B_{j,j+1}(u) &= A_i(v)A_j(u) \otimes B_{i,i+1}(v)B_{j,j+1}(u) \\ &+ B_{i,i+1}(v)B_{j,j+1}(u) \otimes A_{i+1}(v)A_{j+1}(u) \\ &+ f(v, u)B_{i,i+1}(v)A_j(u) \otimes B_{j,j+1}(u)A_{i+1}(v) \\ &+ f(u, v)B_{j,j+1}(u)A_i(v) \otimes B_{i,i+1}(v)A_{j+1}(u) \\ &+ \delta_{i+1,j}g(v, u)B_{i,i+2}(v)A_{i+1}(u) \\ &\quad \otimes [A_{i+1}(u)A_{i+2}(v) - A_{i+1}(v)A_{i+2}(u)]. \end{aligned} \quad (\text{B.5})$$

⁹The counterpart of these equations for the trigonometric case are more complex and will not be treated here. Their explicit expression can be found for example in eq. (2.4) of [38].

We should pay attention to the appearance of the composite operator $B_{i,i+2}$ in the last equation, making the construction of the weight factors (now weight matrices) more difficult.

These explicit computations show that the construction of weights is more complex in the $SU(N)$ case for $N > 2$, as obtaining the weights for higher number of excitations from the weights for two excitations involves summing over internal indices.

To end this section we should comment about ZF operators. Sometimes it is more interesting to work with operators that form a Zamolodchikov-Faddeev algebra instead of using the creation operators from the ABA, as form factors constructed with them fulfil Smirnov's axioms. First of all, we should fix an ordering of the operators, as extra S-matrices appear when they are exchanged. We are going to number the operators in the lower edge from left to right and vice-versa in the upper edge.

As before, computing the weights for this kind of operators requires first to break one single operator into two factors, which is related to the co-product operation. The co-product applied to ZF operators gives

$$\begin{aligned} \Delta[Z(u)] &= Z(u) \otimes \mathbb{I} + H(u)e^{ip(u)l_1} \otimes Z(u), \\ \Delta[Z^\dagger(u)] &= \mathbb{I} \otimes Z^\dagger(u) + Z^\dagger(u) \otimes H^{-1}(u)e^{ip(u)l_2} \end{aligned} \tag{B.6}$$

where H is a well-bred operator in the sense of [44]. In particular this kind of operators fulfil $H_1 Z_2 = S_{12} Z_2 H_1$. Assuming also that $H|0\rangle = |0\rangle$ we obtain the following weights

$$\begin{aligned} w(\{\alpha\}, \{\bar{\alpha}\}, \{\beta\}, \{\bar{\beta}\}) &= w_Z(\{\alpha\}, \{\bar{\alpha}\})w_{Z^\dagger}(\{\beta\}, \{\bar{\beta}\}) \\ w_Z(\{\alpha\}, \{\bar{\alpha}\}) &= e^{ip(\bar{\alpha})l_1} S^<(\bar{\alpha}, \alpha), \\ w_{Z^\dagger}(\{\beta\}, \{\bar{\beta}\}) &= e^{ip(\beta)l_2} S^<(\beta, \bar{\beta}). \end{aligned} \tag{B.7}$$

Here $S^<(\bar{\alpha}, \alpha)$ and $S^<(\beta, \bar{\beta})$ means that we are only keeping the terms $S(u_i, u_j)$ with $i < j$.

A relationship between the square constructed with ZF states and ABA states can be seen already at the level of these weights. If we redefine the ZF square by extracting a factor of $f^\neq(u, u)$, we can see that the weights become the ones for the ABA square, giving us a simple connection between both squares.

C Reconstruction of the spacetime dependence

In this appendix we are going to comment on how to recover the spacetime dependence of the two-point function from two different perspectives. First we are going to make use of the string bits/spin bits construction and later we are going to apply the arguments used in the hexagon proposal.

The string bits/spin bits construction was developed in [21] and [45] and the basic idea is to construct the $\mathcal{N} = 4$ SYM operators using the maximal compact subgroup of $U(1) \times SU(2) \times SU(2) \subset PSU(2, 2|4)$ and then rotate to the usual representation of the conformal group. Therefore in this language an operator is written as

$$\hat{O}(x)|0\rangle = \sum_{\pi \in \mathcal{S}_L} \text{sign}(\pi) \prod_{i=1}^L \left(\hat{O}^{\pi(i)}(x)|0\rangle^{(i)} \right), \quad \hat{O}^i(x)|0\rangle_i = e^{ixP} U O^i |0\rangle^{(i)}, \tag{C.1}$$

where $\text{sign}(\pi)$ only accounts for signs arising from commuting fermions, $|0\rangle^{(i)}$ represent a vacuum for each bit, P is the momentum operator, O is an operator constructed using such oscillator representation and $U = e^{\frac{\pi}{4}(P_0 - K_0)}$ is an rotation that relates this representation with the usual representation used for classifying fields with respect to the conformal group. This construction allow us to construct a Yangian-invariant spin vertex [45], which can be understood as the weak-coupling version of string vertex in string field theory and recovers it in the BMN limit [46].

Regarding the computation of the norm of the two-point function placed at zero and infinity, we can show that it reduces to the scalar product in the Fock space of the oscillators since the inversion of one of the states is equivalent to acting on it with U^2

$$\langle \hat{O}_j(\infty) \hat{O}_k(0) \rangle = \langle \hat{O}_j^\dagger(0) U^2 \hat{O}_k(0) \rangle = \frac{1}{L} \sum_{\pi, \sigma \in S_L} \text{sign}(\pi) \text{sign}(\sigma) \prod_{i=1}^L \langle 0 | O_j^{\dagger\pi(i)} O_k^{\sigma(i)} | 0 \rangle_i, \tag{C.2}$$

where we have used that $U^\dagger = U$ and $U^4 = 1$. If the operators were placed at a finite distance, we would have to commute the exponential of the momentum through the rotation U , giving us the tree-level spacetime dependence. The one-loop corrections to the correlation function can be interpreted in this language as the corrections in the commutation relation between the momentum and the U operator (which generate the logarithmic corrections) and from a change in the scalar product of the oscillator (which generate the correction to the mixing matrix). We refer to [21] for the proofs of these statements.

Shifting now to the hexagonalization perspective [28], we have computed a square with one operator sit at zero and other at infinity, glued with a second square where we invert the points in which the operators are inserted. The relation between these two frames should be simpler than the twisted translation needed for the hexagon from factor and might impose some constraints to an all-loop construction of the square. The spacetime factor appears as an extra overall dependence in a similar fashion as in the hexagon (see eq. (35), (36) in [28] and the discussion between them), which is consistent with the way it arises in the string bit/spin bit formalism.

D Crossing properties of the square

As we have commented before, we cannot apply the crossing transformation to quantities computed using the Heisenberg spin chain because it corresponds to the vanishing coupling constant limit, and thus the cuts in the rapidity plane necessary for the crossing transformation collapse.

The way to bootstrap a square where we can apply the crossing transformation is to instead construct the consistency condition from the ZF algebra, as in this way we would not be making any assumption on the form of the S-matrix. If we repeat the steps followed in section 3.2 using the weights (B.7), we get the same functional form for the square with one excitation

$$S_L(u \rightarrow v) = \bar{g}(u, v) \left(e^{ip(u)L} - e^{ip(v)L} \right). \tag{D.1}$$

We expect the square from factor to have the following two properties under crossing:

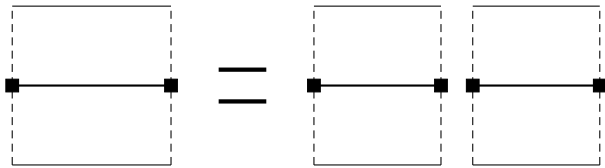


Figure 9. Pictorial representation of the cutting of the first mirror correction to the empty square.

1. Moving an excitation clockwise (or anti-clockwise) around the full square should be trivial (up to an S-matrix). We can write this statement as

$$\begin{aligned} \bar{g}(u, v) &= \bar{g}(u^c, v) = \bar{g}(v, u^c)S(u^c, v) = \bar{g}(v, u^{2c})S(u^c, v) \\ \bar{g}(u, v) &= \bar{g}(v, u^{-c}) = \bar{g}(u^{-c}, v)S(v, u^{-c}) = \bar{g}(u^{-2c}, v)S(v, u^{-c}) . \end{aligned} \quad (\text{D.2})$$

Where the superindex c means crossing across the mirror side in the clockwise direction and $-c$ means crossing in the counter-clockwise direction.

2. The crossing of the two excitations is the same as changing the excitations, i.e., $S(u^c \rightarrow v^c) = S(v \rightarrow u)$. This imposes

$$\bar{g}(u^c, v^c) = -\bar{g}(v, u)e^{i[p(u)+p(v)]L} = \bar{g}(u^{-c}, v^{-c}) . \quad (\text{D.3})$$

However we are not going to attempt to solve these equations in this article.

E Leading mirror corrections to the empty and one excitation squares

In this appendix we are going to perform computations beyond the asymptotic regime by including mirror excitations. There are two contributions we have ignored: the first one is the obvious contribution to the gluing of the mirror sides of a square to obtain the expression for the scalar product. The second one, more subtle, is the contribution to the consistency condition. We are indeed creating new mirror edges when we cut a square into two, thus we have to consider the appearance of mirror excitations in this relation. Here we will make some comments on corrections of the first kind to in the empty square and corrections of second kind to the square with one excitation.

First of all, since in this section we are going to introduce excitations in the mirror sides, we need a more general notation for the square. This notation will be $S_L(\{u\}|\{v\}|\{w\}|\{x\})$, where the different sets of rapidities are ordered starting from the lower side and continuing clockwise.

We also have to fix our notation regarding the mirror transformation. The rules regarding this transformation can be found, for example, in appendix D of [23]. In particular, we are going to define moving an excitation to the next side in the clockwise direction with the superscript γ , while moving counter-clockwise will be labelled by $-\gamma$. Applying two times this transformation corresponds to a crossing transformation, $u^{2\gamma} = u^c$, which changes the sign of momentum and energy. For simplicity and concreteness we are going to choose the

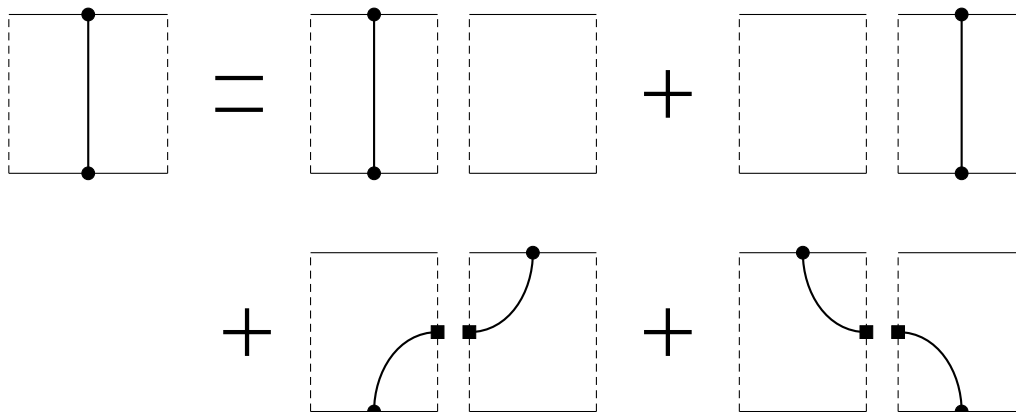


Figure 10. Pictorial representation of the cutting of the square with one excitation, including terms that involve mirror excitations.

spin chain model with $a(u) = e^{ip(u)L}$ and $d(u) = 1$, where we have

$$\tilde{a}(u) = a(u^{-\gamma}) = 1, \quad \hat{a}(u) = a(u^{+\gamma}) = e^{-E(u)R}, \quad (\text{E.1})$$

$$\tilde{d}(u) = d(u^{-\gamma}) = e^{-E(u)R}, \quad \hat{d}(u) = d(u^{+\gamma}) = 1. \quad (\text{E.2})$$

Let us consider first the case of the first mirror correction to the empty square. The consistency condition, pictorially represented in figure 9, is

$$S_L(\emptyset|u|\emptyset|w) = \int \frac{dv}{2\pi} S_{l_1}(\emptyset|u|\emptyset|v) S_{l_2}(\emptyset|v|\emptyset|w). \quad (\text{E.3})$$

If we substitute here the ansatz $S_L(\emptyset|u|\emptyset|w) = \hat{g}(u, w)[e^{-E(w)R} - e^{-E(u)R}]$, the equation reduces to the sum of the residues at the poles of the function \hat{g} . The function $\tilde{g}(u, w) \propto \frac{1}{u-w}$ we obtained for the Gaudin square gives us the correct measure and Boltzmann factor when we take the limit $w \rightarrow u$, but it is not a solution of this equation as the integrand is regular at $v = u$ and $v = w$. On the other hand, this equation can be solved if we use a square obtained from using ZF operators since, according to Smirnov's axioms, it presents the particle-antiparticle residue. This fact can be understood as the consistency condition for mirror excitations being sensible to the kind of operators we are using for constructing our squares. In particular, it forces us to use Zamolodchikov-Faddeev operators, as the form factors constructed with them fulfil the Smirnov's axioms.

Moving now to the case of one physical excitation, figure 10 represents the terms appearing in the consistency equation (3.2) when we allow for mirror corrections. We can write the consistency condition in the following way

$$\begin{aligned} S_L(u|\emptyset|w|\emptyset) &= S_{l_1}(u|\emptyset|w|\emptyset)d_{l_2}(u)a_{l_2}(w) + S_{l_2}(u|\emptyset|w|\emptyset)a_{l_1}(u)d_{l_1}(w) \\ &+ \int \frac{dv}{2\pi} S_{l_1}(u|\emptyset|\emptyset|v)S_{l_2}(\emptyset|v|w|\emptyset)d_{l_2}(u)d_{l_1}(w) \\ &+ \int \frac{dv}{2\pi} S_{l_1}(\emptyset|v|\emptyset|w)S_{l_2}(u|v|\emptyset|\emptyset)a_{l_1}(u)a_{l_2}(w). \end{aligned} \quad (\text{E.4})$$

In contrasts with the gluing of two hexagons, here only two new corrections appear. This is because we have no mirror excitations in the original mirror sides and only two physical excitations to annihilate the set of mirror excitations we add, so we are forced to put at most only one mirror excitation if we want a non-vanishing contribution due to symmetry.

To solve the consistency equation we substitute the ansatz

$$S_L(u \rightarrow w) = \alpha(u, w)a(u)a(w) + \beta(u, w)a(u)d(w) + \gamma(u, w)d(u)a(w) + \delta(u, w)d(u)d(w), \quad (\text{E.5})$$

reducing it to

$$\begin{aligned} \alpha_L &= \hat{I}_1, & \beta_L &= \beta_{l_2}, & \gamma_L &= \gamma_{l_1}, & \delta_L &= \tilde{I}_4, \\ \alpha_{l_1} &= -\hat{I}_3, & \delta_{l_1} &= -\tilde{I}_2, & \alpha_{l_2} &= -\hat{I}_2, & \delta_{l_2} &= -\tilde{I}_3, \end{aligned}$$

together with $\beta_{l_1} + \gamma_{l_2} + \tilde{I}_1 + \hat{I}_4 = 0$. The different I 's are given by the following integrals

$$\begin{aligned} \tilde{I}_1 &= \int \frac{dv}{2\pi} \left[\alpha_{l_1} \alpha_{l_2} [\tilde{a}(v)]^2 + (\alpha_{l_1} \gamma_{l_2} + \beta_{l_1} \alpha_{l_2}) \tilde{a}(v) \tilde{d}(v) + \gamma_{l_1} \gamma_{l_2} [\tilde{d}(v)]^2 \right], \\ \tilde{I}_2 &= \int \frac{dv}{2\pi} \left[\gamma_{l_1} \alpha_{l_2} [\tilde{a}(v)]^2 + (\gamma_{l_1} \gamma_{l_2} + \delta_{l_1} \alpha_{l_2}) \tilde{a}(v) \tilde{d}(v) + \delta_{l_1} \gamma_{l_2} [\tilde{d}(v)]^2 \right], \\ \tilde{I}_3 &= \int \frac{dv}{2\pi} \left[\alpha_{l_1} \beta_{l_2} [\tilde{a}(v)]^2 + (\alpha_{l_1} \delta_{l_2} + \beta_{l_1} \beta_{l_2}) \tilde{a}(v) \tilde{d}(v) + \gamma_{l_1} \delta_{l_2} [\tilde{d}(v)]^2 \right], \\ \tilde{I}_4 &= \int \frac{dv}{2\pi} \left[\gamma_{l_1} \beta_{l_2} [\tilde{a}(v)]^2 + (\gamma_{l_1} \delta_{l_2} + \delta_{l_1} \beta_{l_2}) \tilde{a}(v) \tilde{d}(v) + \delta_{l_1} \delta_{l_2} [\tilde{d}(v)]^2 \right], \\ \hat{I}_1 &= \int \frac{dv}{2\pi} \left[\alpha_{l_1} \alpha_{l_2} [\hat{a}(v)]^2 + (\gamma_{l_1} \alpha_{l_2} + \alpha_{l_1} \beta_{l_2}) \hat{a}(v) \hat{d}(v) + \gamma_{l_1} \beta_{l_2} [\hat{d}(v)]^2 \right], \\ \hat{I}_2 &= \int \frac{dv}{2\pi} \left[\beta_{l_1} \alpha_{l_2} [\hat{a}(v)]^2 + (\delta_{l_1} \alpha_{l_2} + \beta_{l_1} \beta_{l_2}) \hat{a}(v) \hat{d}(v) + \delta_{l_1} \beta_{l_2} [\hat{d}(v)]^2 \right], \\ \hat{I}_3 &= \int \frac{dv}{2\pi} \left[\alpha_{l_1} \gamma_{l_2} [\hat{a}(v)]^2 + (\gamma_{l_1} \gamma_{l_2} + \alpha_{l_1} \delta_{l_2}) \hat{a}(v) \hat{d}(v) + \gamma_{l_1} \delta_{l_2} [\hat{d}(v)]^2 \right], \\ \hat{I}_4 &= \int \frac{dv}{2\pi} \left[\beta_{l_1} \gamma_{l_2} [\hat{a}(v)]^2 + (\delta_{l_1} \gamma_{l_2} + \beta_{l_1} \delta_{l_2}) \hat{a}(v) \hat{d}(v) + \delta_{l_1} \delta_{l_2} [\hat{d}(v)]^2 \right]. \end{aligned}$$

The arguments of the functions involved in the tilde integrals are $(u, v^{-\gamma})$ and $(v^{-\gamma}, w)$ for the functions with subindices l_1 and l_2 respectively, while for the hat integrals the arguments are (v^γ, w) and (u, v^γ) for l_1 and l_2 respectively. Note that not all these integrals are independent, as the square should be invariant under the interchange of the initial and final physical rapidities. This imply the following relations between the unknown functions of the ansatz

$$\alpha(u, w) = \alpha(w, u), \quad \delta(u, w) = \delta(w, u), \quad \beta(u, w) = \gamma(w, u). \quad (\text{E.6})$$

Using these properties we can prove some consistency of this ansatz as $\alpha_{l_1} = \hat{I}_2(u, w) = \hat{I}_3(w, u) = \alpha_{l_2}$ and similarly for the tilde ones, which implies that both α and δ are independent of the length. Furthermore, as I_1 's and I_4 's transform into themselves, we can write function β and γ as

$$\beta(u, w) - \frac{\tilde{I}_1 + \hat{I}_4}{2} = \gamma(u, w) + \frac{\tilde{I}_1 + \hat{I}_4}{2} = \tilde{g}(u, v), \quad (\text{E.7})$$

where \tilde{g} is an arbitrary function with the property $\tilde{g}(u, v) = -\tilde{g}(v, u)$. Using these properties the problem reduces to a system of coupled Fredholm integral equations of second kind, which can be solved as a Liouville-Neumann series. However, we need an initial condition to apply this method. As we have seen for the case of the mirror excitation in the empty square, the initial condition we have to use is the square computed using ZF operators.

Open Access. This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

References

- [1] F. Smirnov, *Form-factors in completely integrable models of quantum field theory*, *Adv. Ser. Math. Phys.* **14** (1992) 1.
- [2] H.M. Babujian, A. Fring, M. Karowski and A. Zapletal, *Exact form-factors in integrable quantum field theories: the sine-Gordon model*, *Nucl. Phys. B* **538** (1999) 535 [[hep-th/9805185](#)] [[INSPIRE](#)].
- [3] H. Babujian and M. Karowski, *Exact form-factors in integrable quantum field theories: the sine-Gordon model. 2.*, *Nucl. Phys. B* **620** (2002) 407 [[hep-th/0105178](#)] [[INSPIRE](#)].
- [4] A. Nakayashiki and Y. Takeyama, *On form-factors of SU(2) invariant Thirring model*, *Prog. Math. Phys.* **23** (2002) 357 [[math-ph/0105040](#)] [[INSPIRE](#)].
- [5] Y. Takeyama, *Form-factors of SU(N) invariant Thirring model*, *Publ. Res. Inst. Math. Sci. Kyoto* **39** (2003) 59 [[math-ph/0112025](#)] [[INSPIRE](#)].
- [6] H.M. Babujian, A. Foerster and M. Karowski, *The form factor program: a review and new results — The nested SU(N) off-shell Bethe ansatz*, *SIGMA* **2** (2006) 082.
- [7] T. Klohe and T. McLoughlin, *Worldsheet form factors in AdS/CFT*, *Phys. Rev. D* **87** (2013) 026004 [[arXiv:1208.2020](#)] [[INSPIRE](#)].
- [8] T. Klohe and T. McLoughlin, *Comments on world-sheet form factors in AdS/CFT*, *J. Phys. A* **47** (2014) 055401 [[arXiv:1307.3506](#)] [[INSPIRE](#)].
- [9] N. Gromov, V. Kazakov, S. Leurent and D. Volin, *Quantum spectral curve for planar $\mathcal{N} = 4$ super-Yang-Mills theory*, *Phys. Rev. Lett.* **112** (2014) 011602 [[arXiv:1305.1939](#)] [[INSPIRE](#)].
- [10] N. Gromov, F. Levkovich-Maslyuk, G. Sizov and S. Valatka, *Quantum spectral curve at work: from small spin to strong coupling in $\mathcal{N} = 4$ SYM*, *JHEP* **07** (2014) 156 [[arXiv:1402.0871](#)] [[INSPIRE](#)].
- [11] N. Gromov, V. Kazakov, S. Leurent and D. Volin, *Quantum spectral curve for arbitrary state/operator in AdS₅/CFT₄*, *JHEP* **09** (2015) 187 [[arXiv:1405.4857](#)] [[INSPIRE](#)].
- [12] C. Marboe and D. Volin, *Quantum spectral curve as a tool for a perturbative quantum field theory*, *Nucl. Phys. B* **899** (2015) 810 [[arXiv:1411.4758](#)] [[INSPIRE](#)].

- [13] C. Marboe, V. Velizhanin and D. Volin, *Six-loop anomalous dimension of twist-two operators in planar $\mathcal{N} = 4$ SYM theory*, *JHEP* **07** (2015) 084 [[arXiv:1412.4762](#)] [[INSPIRE](#)].
- [14] N. Gromov, F. Levkovich-Maslyuk and G. Sizov, *Quantum spectral curve and the numerical solution of the spectral problem in AdS_5/CFT_4* , *JHEP* **06** (2016) 036 [[arXiv:1504.06640](#)] [[INSPIRE](#)].
- [15] V. Kazakov, S. Leurent and D. Volin, *T-system on T-hook: grassmannian solution and twisted quantum spectral curve*, *JHEP* **12** (2016) 044 [[arXiv:1510.02100](#)] [[INSPIRE](#)].
- [16] C. Marboe and V. Velizhanin, *Twist-2 at seven loops in planar $\mathcal{N} = 4$ SYM theory: full result and analytic properties*, *JHEP* **11** (2016) 013 [[arXiv:1607.06047](#)] [[INSPIRE](#)].
- [17] C. Marboe and D. Volin, *Fast analytic solver of rational Bethe equations*, *J. Phys. A* **50** (2017) 204002 [[arXiv:1608.06504](#)] [[INSPIRE](#)].
- [18] C. Marboe and D. Volin, *The full spectrum of AdS_5/CFT_4 I: representation theory and one-loop Q-system*, *J. Phys. A* **51** (2018) 165401 [[arXiv:1701.03704](#)] [[INSPIRE](#)].
- [19] K. Okuyama and L.-S. Tseng, *Three-point functions in $N = 4$ SYM theory at one-loop*, *JHEP* **08** (2004) 055 [[hep-th/0404190](#)] [[INSPIRE](#)].
- [20] R. Roiban and A. Volovich, *Yang-Mills correlation functions from integrable spin chains*, *JHEP* **09** (2004) 032 [[hep-th/0407140](#)] [[INSPIRE](#)].
- [21] L.F. Alday, J.R. David, E. Gava and K.S. Narain, *Towards a string bit formulation of $N = 4$ super Yang-Mills*, *JHEP* **04** (2006) 014 [[hep-th/0510264](#)] [[INSPIRE](#)].
- [22] J. Escobedo, N. Gromov, A. Sever and P. Vieira, *Tailoring three-point functions and integrability*, *JHEP* **09** (2011) 028 [[arXiv:1012.2475](#)] [[INSPIRE](#)].
- [23] B. Basso, S. Komatsu and P. Vieira, *Structure constants and integrable bootstrap in planar $N = 4$ SYM theory*, [arXiv:1505.06745](#) [[INSPIRE](#)].
- [24] B. Basso, V. Goncalves, S. Komatsu and P. Vieira, *Gluing hexagons at three loops*, *Nucl. Phys. B* **907** (2016) 695 [[arXiv:1510.01683](#)] [[INSPIRE](#)].
- [25] J. Caetano and T. Fleury, *Fermionic correlators from integrability*, *JHEP* **09** (2016) 010 [[arXiv:1607.02542](#)] [[INSPIRE](#)].
- [26] J.L. Cardy, O.A. Castro-Alvaredo and B. Doyon, *Form factors of branch-point twist fields in quantum integrable models and entanglement entropy*, *J. Statist. Phys.* **130** (2008) 129 [[arXiv:0706.3384](#)] [[INSPIRE](#)].
- [27] B. Eden and A. Sfondrini, *Tessellating cushions: four-point functions in $\mathcal{N} = 4$ SYM*, *JHEP* **10** (2017) 098 [[arXiv:1611.05436](#)] [[INSPIRE](#)].
- [28] T. Fleury and S. Komatsu, *Hexagonalization of correlation functions*, *JHEP* **01** (2017) 130 [[arXiv:1611.05577](#)] [[INSPIRE](#)].
- [29] B. Basso et al., *Asymptotic four point functions*, [arXiv:1701.04462](#) [[INSPIRE](#)].
- [30] T. Fleury and S. Komatsu, *Hexagonalization of correlation functions II: two-particle contributions*, *JHEP* **02** (2018) 177 [[arXiv:1711.05327](#)] [[INSPIRE](#)].

- [31] M. Kim and N. Kiryu, *Structure constants of operators on the Wilson loop from integrability*, *JHEP* **11** (2017) 116 [[arXiv:1706.02989](#)] [[INSPIRE](#)].
- [32] L.D. Faddeev, *How algebraic Bethe ansatz works for integrable model*, [hep-th/9605187](#) [[INSPIRE](#)].
- [33] Y. Jiang and A. Petrovskii, *Diagonal form factors and hexagon form factors*, *JHEP* **07** (2016) 120 [[arXiv:1511.06199](#)] [[INSPIRE](#)].
- [34] G. Kirchhoff, *Über die Auflösung der Gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer Ströme geführt wird*, *Ann. Phys. Chem.* **72** (1847) 497.
- [35] I. Kostov, D. Serban and D.-L. Vu, *TBA and tree expansion*, *Springer Proc. Math. Stat.* **255** (2017) 77 [[arXiv:1805.02591](#)] [[INSPIRE](#)].
- [36] B. Basso and A. Rej, *Bethe ansätze for GKP strings*, *Nucl. Phys. B* **879** (2014) 162 [[arXiv:1306.1741](#)] [[INSPIRE](#)].
- [37] Y. Wang, W. Yang, J. Cao and K. Shi, *Off-diagonal Bethe ansatz for exactly solvable models*, Springer, Berlin, Germany (2015).
- [38] A. Hutsalyuk et al., *Scalar products and norm of Bethe vectors for integrable models based on $U_q(\widehat{\mathfrak{g}}_n)$* , *SciPost Phys.* **4** (2018) 006 [[arXiv:1711.03867](#)] [[INSPIRE](#)].
- [39] J.M. Nieto, *Spinning strings and correlation functions in the AdS/CFT correspondence*, [arXiv:1711.09993](#) [[INSPIRE](#)].
- [40] N. Gromov and P. Vieira, *Tailoring three-point functions and integrability IV. Theta-morphism*, *JHEP* **04** (2014) 068 [[arXiv:1205.5288](#)] [[INSPIRE](#)].
- [41] N. Beisert, V. Dippel and M. Staudacher, *A novel long range spin chain and planar $N = 4$ super Yang-Mills*, *JHEP* **07** (2004) 075 [[hep-th/0405001](#)] [[INSPIRE](#)].
- [42] B. Basso, V. Goncalves and S. Komatsu, *Structure constants at wrapping order*, *JHEP* **05** (2017) 124 [[arXiv:1702.02154](#)] [[INSPIRE](#)].
- [43] Z. Bajnok and R.A. Janik, *From the octagon to the SFT vertex — Gluing and multiple wrapping*, *JHEP* **06** (2017) 058 [[arXiv:1704.03633](#)] [[INSPIRE](#)].
- [44] É. Ragoucy, *Vertex operators for quantum groups and application to integrable systems*, *J. Phys. A* **35** (2002) 7929 [[math/0108207](#)] [[INSPIRE](#)].
- [45] Y. Jiang, I. Kostov, A. Petrovskii and D. Serban, *String bits and the spin vertex*, *Nucl. Phys. B* **897** (2015) 374 [[arXiv:1410.8860](#)] [[INSPIRE](#)].
- [46] Y. Jiang and A. Petrovskii, *From spin vertex to string vertex*, *JHEP* **06** (2015) 172 [[arXiv:1412.2256](#)] [[INSPIRE](#)].