Exact Results in QFT:
Minimal Areas and Maximal Couplings

by

Jonathan Toledo

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis is devoted to a two-pronged study of non-perturbative quantum field theory. In Part I we focus on the four-dimensional super conformal $\mathcal{N} = 4$ Yang Mills theory. We compute smooth Wilson loops and correlation functions in the strong-coupling regime of the theory using the classical integrability of the dual string theory as our main tool. In both cases the solution is given as a set of integral equations of thermodynamic bethe ansatz type. The correlation function and Wilson loop are then written in terms of the corresponding TBA free energy. The equations for the Wilson loop expectation value can be used for generic smooth contours embeddable in an $\mathbb{R}^{1,1}$ subspace of $\mathbb{R}^{1,3}$. In Part II we ask general questions about the allowed space of massive quantum field theories based only on crossing symmetry and unitarity. We approach this question in two ways. First we consider putting massive QFT into an AdS box and study the conformal boundary theory using standard conformal bootstrap tools. We call this procedure the *boundary bootstrap*. It is applicable in any dimension but takes its simplest form for a $1 + 1$ dimensional bulk QFT where we use it to obtain rigorous bounds on allowed QFT couplings. For $1 + 1$ dimensional QFT we are also able to obtain rigorous bounds directly in flat space using unitarity, crossing symmetry and analyticity of the S-matrix. The bounds obtained in this way agree perfectly with those obtained from the boundary bootstrap.

Acknowledgements

As a student I have been fortunate to have had great mentors and friends. I want to thank the many important influences without whom I may have never been writing this acknowledgement. My direction as a high school student was greatly influenced by my chemistry teacher Bobbi Henderson and my first physics instructor Jon Kolena as well as the administration of the North Carolina School of Science and Mathematics. As a university student I was lucky to have the guidance of many great scientists and instructors, most notably Sean Washburn, Dmitri Khveshenko, Jack Ng, Rich McLaughlin, Ben Widom and Paolo De Gregorio, all of whom greatly impacted me at a formative stage in my development as a student of physics. I also want to thank my undergraduate colleague and friend Anton Kibalnik who will always be like a brother to me. As a graduate student I had the great pleasure of sharing an office with Nima Duroud, Bruno Le Floch and Razieh Pourhasan all of whom taught me so much about life and physics and who I am lucky and proud to call my friends and colleagues. I especially want to thank Nima for his unwavering friendship, kindness and steadfastness which helped me through some tough times. I also want to thank my academic brother João Caetano. As a collaborator he always seemed to find a way through when I was about to give up and as a friend and traveling companion I could ask for no one better. I also owe a special thanks to Debbie Guenther whose kindness, patience and skill for navigating university bureaucracy saved me on more than one occation. As a young scientist I have been lucky to have had the guidance of many great physicists. I especially want to thank Amit Sever from whom I learned many valuable lessons. Finally I want to thank my PhD advisor Pedro Vieira whose guidance and instruction has been the most important variable in my growth as a researcher. Pedro has set a standard which I can aspire to as a scientist and when faced with a difficult research problem, I often find it useful to ask myself "what would Pedro do?" More than that, his kindness, enthusiasm for science and the care that he devotes to his students goes well beyond the role of a PhD advisor. To be a student of Pedro is to be part of his family.

Of course I would never have even started down this path if it were not for my parents Janet and Charles. Your love and your enduring patience for my stubbornness and waywardness is something that I comprehend more and more every day. You gave me the tools and confidence necessary to persevere and have given me a standard of strength and tenderness that I can live by. I am grateful to my grandfather Will James ('Papa') whose love and support has always helped me hold the course. I also want to thank Martina, Martin, Jovy and Simona who gave me a home away from home here in Waterloo. Finally I am so grateful for the love and support of my girlfriend Sabrina. Your friendship and creativity add so much color to my world and your belief in me has made the impossible possible. *Who cleared the static and made it sing? You did.*

Dedication

This thesis is dedicated to Ben Widom who introduced me to the beautiful subject of exactly solvable models.

Table of Contents

Chapter 1

Searching for a QFT Rosetta Stone

The current status of fundamental physics presents an unusual predicament. Our understanding is completely based on quantum field theory, yet our mastery of quantum field theory itself is limited in many ways. It is often as if we are given the laws of physics in a language that we cannot fully decipher. An outstanding example is the problem of computing the mass of the proton. Although the theory of QCD has been well known for many decades, and certainly contains this number within its predictions (as predicated by lattice results), given the QCD Lagrangian a theorist has no means to extract this number within a controlled approximation scheme.

This example is representative of a widespread difficulty in QFT which is that we often do not know what is the best set of variables to use in the description of a given theory. Generically when the interactions become strong the utility of the Lagrangian formulation is limited and the propagating degrees of freedom are far removed from those described by the Lagrangian (as in the example of QCD, which becomes a theory of stringlike objects in the IR). Moreover, the parameters in the Lagrangian are not the physical parameters and undergo renormalization. These issues point to two of the most important problems in modern physics:

Q1: How do we describe strongly coupled QFT?

Q2: How do we formulate QFT in terms of physical quantities?

Of course these questions are not independent, and an answer to the second may provide insight into the first.

In this thesis we will approach these questions from two perspectives. In part I we will focus on the study of very special theories with enhanced symmetries which makes them more tractable. In particular, we focus on $\mathcal{N} = 4$ Super Yang Mills, which is perhaps the simplest $3 + 1$ dimensional gauge theory. In part II, rather than focus on any specific theory, we instead do exactly the opposite and consider the space of all possible quantum field theories compatible unitarity and Lorentz invariance. These two parts are mostly disjoint and may be read independently.

1.1 Invitation to part I

When discussing such general questions as Q1 and Q2, it is useful to have toy models which can provide inspiration and a means to test ideas. For this reason, it is of paramount importance to have at least one example of a fully interacting 3+1 dimensional gauge theory which we can solve exactly. Exact solvability, or integrability, is a property which formally is only possible in two-dimensional systems and one may wonder in what sense a 4 dimensional theory may be integrable without also being trivial. What makes such an idea possible is the concept of holography [1–6]. A remarkable outcome of this principle is that certain gauge theories are exactly equivalent to theories of strings living in higher dimensional, curved space-times. Thus, some special quantum field theories can be mapped to a 2-dimensional world-sheet theory where one might hope to apply the techniques of integrability. We shall now discuss a very special theory for which this is indeed the case.

The canonical example of holography is the equivalence of the fully interacting, 4D gauge theory $\mathcal{N} = 4$ Super Yang Mills with $U(N_c)$ gauge group and the theory of type IIB strings living in the space $AdS_5 \times S^5$ [3–5]. As an important example of how this correspondence works, let us consider the computation of a wilson loop

$$W[\mathcal{C}] = \frac{1}{N_c} \mathrm{Tr}\, \mathrm{Pexp}\, i \int_{\mathcal{C}} A \tag{1.1}$$

in the gauge theory [6]. Here $\mathcal{C}$ is a closed spacetime contour, A is the gauge field and the trace is over the fundamental representation. One should think of the guage theory as living at the boundary of AdS_5 which is precisely $\mathbb{R}^{1,3}$ and thus the Wilson loop contour $\mathcal{C}$ sits at this boundary. The strings live inside the full $AdS_5 \times S^5$. The prescription for computing the Wilson loop using the string theory is

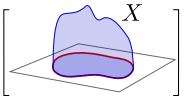
$$\langle W\,[\mathcal{C}]\rangle = \sum_{X \in \{\text{surfaces ending on } \mathcal{C}\}} e^{-S_{\text{string}}[X]} \tag{1.2}$$

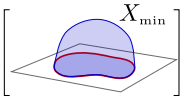
Although a remarkable equation, at first sight computation using the right hand side seems even harder than computation in the QFT. This brings us to an extremely powerful aspect of this duality: it is a so-called *weak/strong duality* in the sense that the strong coupling regime of the QFT maps to a weak coupling regime of the string theory and visa versa. Let us make this more precise. The parameters of the gauge theory and string theory are related by [3]

$$g_{YM}^2 = 2\pi\, g_{\text{string}}\,, \qquad T = \frac{\sqrt{\lambda}}{2\pi} \tag{1.3}$$

where we have introduced the t'Hooft coupling $\lambda = g_{YM}^2 N_c$ and T is the string tension. We first consider the planar limit of the theory defined as $N_c \to \infty$ keeping λ fixed. Finally we concentrate on the regime of infinite t' Hooft coupling which corresponds to $g_{\text{string}} \to 0$ and $T \to \infty$. When the string tension is large the string path integral (1.2) is localized upon the classical solution so that we have the following recipe for computing wilson loops at strong coupling[1]

$$\langle W\,[\mathcal{C}]\rangle = \sum_X e^{-\frac{\sqrt{\lambda}}{2\pi}\text{Area}}\Big[\quad\Big] \xrightarrow{\lambda\to\infty} e^{-\frac{\sqrt{\lambda}}{2\pi}\text{Area}}\Big[\quad\Big] \tag{1.4}$$





where X is any surface in AdS_5 ending on $\mathcal{C}$ and $X_{\text{min}} \in \{X\}$ is the solution of the classical string equations. Thus the completely intractable problem of computing an arbitrary wilson loop in the strongly interacting guage theory has been reduced to a simple minimal surface problem!

For the case of $\mathcal{N} = 4$ the utility of the duality goes far beyond strong coupling. It turns out that the worldsheet theory is integrable [7]. This provides a concrete realization of the possibility mentioned above that a 4D QFT can indeed be nontrivial and secretly integrable by having an integrable string dual. Using this correspondence and exploiting the techniques of integrability it has been possible to compute the full spectrum of operators in planar $\mathcal{N} = 4$ for any value of the coupling [8] and also to give an all-loop description of the scattering amplitudes in that theory [9–17].

These unprecedented results in interacting quantum field theories give hope that it might be possible to solve $\mathcal{N} = 4$ completely (at least in the planar limit, which we shall focus on exclusively in this thesis). To solve a quantum field theory completely it is useful

[1]Note that the usual duality involves the super wilson loop which also contains a coupling to the scalars of $\mathcal{N} = 4$ related to motion of the string in S^5. This coupling is through a curve $n^I(\sigma)$ on the five sphere. By choosing n^I to point in a single direction we ensure that the sting is point-like in the sphere and the usual gluonic contribution (1.1) is isolated.

to solve for a complete basis of operators in that theory. The set of all correlation functions of local operators is one sufficient basis. The set of all possible Wilson loops is another possible basis. Inspired by the achievements of [10,18] Part I of this thesis is toward an all-loop description of correlation functions and Wilson loops in $\mathcal{N} = 4$. The first step toward this end is to develop an integrable formulation of the classical worldsheet problem that emerges at strong coupling. Indeed, this was the case in both the spectrum problem [8,18] and the polygon OPE program [10, 19] where an integrable formulation of the purely *classical* worldsheet problem was a key step in solving the full *quantum* problem. Chapter 2 describes an integrability-based program for computing smooth Wilson Loops at strong coupling in $\mathcal{N} = 4$. Chapter 3 focuses on computing the strong coupling contribution to a large class of n-point correlation functions in $\mathcal{N} = 4$.

To conclude this invitation, we note that the study of $\mathcal{N} = 4$ clearly relates to the first question Q1 posed above: how to describe a strongly coupled QFT. Although the beta function of $\mathcal{N} = 4$ vanishes, the theory nonetheless mimics some behavior of QCD. In particular when the coupling is weak the theory is best described in terms of the QFT Lagrangian and its respective degrees of freedom. However, when the coupling is strong the theory is best described in terms of string theory. By solving $\mathcal{N} = 4$ for any value of the coupling, we may be able to extract some key concepts about how a weakly coupled gauge theory can transition into a theory of strings when interactions are strong. For example, can we understand in $\mathcal{N} = 4$ precisely how the holographic direction emerges from the gauge theory? It is expected that such a holographic direction is also necessary to describe the QCD string [2] – what insight can we gain from $\mathcal{N} = 4$?

1.2 Invitation to part II

In part II of this thesis we take the precisely opposite approach to that taken in part I. Rather than consider a very specific theory, such as $\mathcal{N} = 4$, we instead consider the space of possible Lorentz invariant and unitary quantum field theories. The idea of constraining theories through a minimal set of indisputable principles is what is commonly referred to as *bootstrap* philosophy. It shows up in various incarnations, the most well known being perhaps the *conformal bootstrap* and the *S-matrix bootstrap*.

The conformal bootstrap works beautifully in two dimensions [20] where it allows for analytic description of a plethora of conformal field theories. In higher dimensions, the bootstrap lay dormant for decades until the seminal work [21]. This work gave rise to a new research field where one looks for bounds on the couplings and spectra of conformal field theories by exploring crossing and reflection positivity. Using computers, one rules out

particular couplings or spectra by searching for linear functions which yield impossibilities when acting on the crossing symmetry relations. Remarkably, one can ultimately use these methods to study some specific theories (such as the 3D ising model!) which show up as distinguished points at the boundary of allowed theories.

The S-matrix bootstrap tries to completely determine S-matrix elements by exploring the analytic properties of these objects to the fullest. This program enjoyed significant attention during the sixties (see e.g. [22, 23] for nice books on the subject) and resulted in many powerful results such as the celebrated Froissart bound [24]. However, ultimately the S-matrix bootstrap fell short of its goal of being a useful tool for determining S-matrix elements or calculating observable quantities.[2] Moreover, with the development of efficient perturbative techniques and with the appearance of quantum chromodynamics, the program lost much of its original motivation. However, the basic philosophy of the S-matrix bootstrap still has undeniable appeal and one cannot help but wonder if it is awaiting a magnificent revival of its own.

Given the recent stunning success of the CFT bootstrap, the time seems ripe to return to the S-matrix bootstrap with insight gained from the former. Most importantly, the CFT bootstrap provides us with the right type of questions to ask. Our usual approach to QFT is to start with some specific theory – usually specified by a Lagrangian – and trying to compute its physical properties such as its spectrum of masses and couplings. The intuition from the CFT bootstrap seems to indicate that the correct question is the reverse of the usual: one should start with a tentative set of physical data and ask weather or not this data is compatible with the constraints of unitarity and crossing. In this way, one studies the space of allowed quantum field theories.

Part II of this thesis is devoted to exactly such a program. We aim at carving out the space of massive quantum field theories by trying to establish upper bounds on couplings given a fixed spectrum of masses. As we shall see, one may use the exact same structures appearing in the CFT bootstrap to constrain non-conformal quantum field theories. To see this, consider putting a gapped D-dimensional QFT into a large AdS box described by the metric

$$ds^2 = R^2 \frac{dz^2 + dr^2 + r^2 d\Omega_{d-1}^2}{z^2} \tag{1.5}$$

where $z > 0$ and r is a radial coordinate for $\mathbb{R}^d$ where $d = D - 1$. This idea is not a new one [26], and one may think of it as a way of introducing an IR regulator (the AdS radius R) while maintaining all of the isometries of flat space. For our purposes, this is a useful

[2] A notable exception is in the special case of integrable theories, where the program was borne out to great success starting with the work [25].

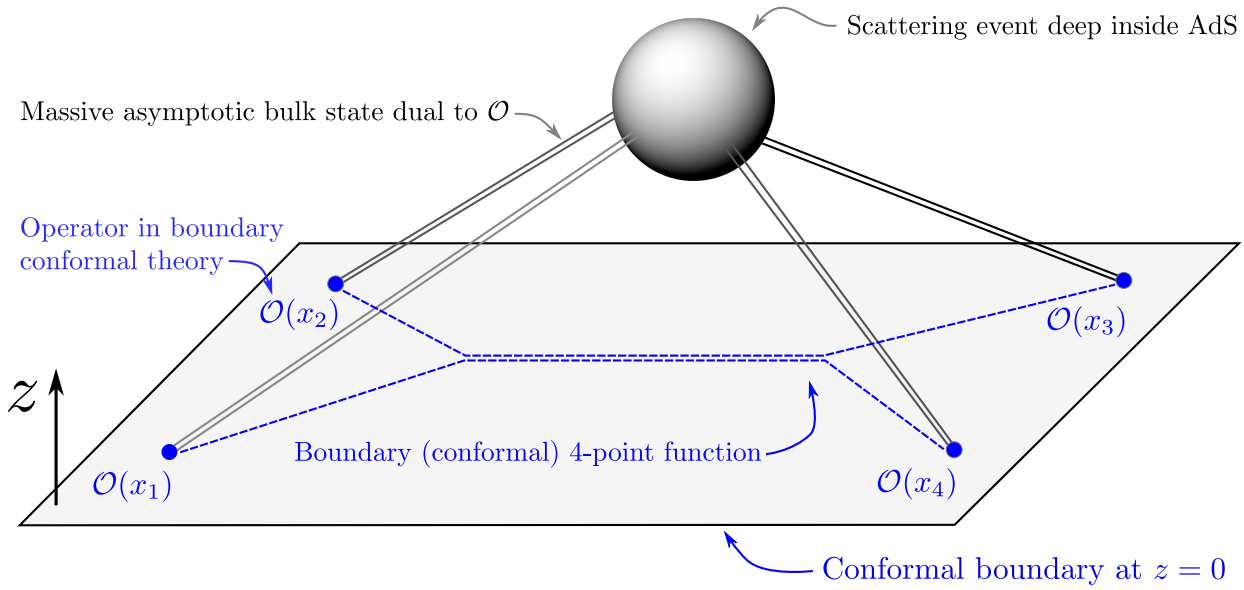


Figure 1.1: Cartoon of QFT in AdS.

box due to boundary at $z = 0$ where the isometry group $SO(D, 1)$ acts as the conformal group on $\mathbb{R}^d$. As a result bulk correlation functions whose insertion points approach this flat conformal boundary obey all the axioms of correlation functions in a d-dimensional conformal theory.[3] In other words, bulk QFTs define or "induce" conformal theories at the boundary whose correlators are defined by the boundary correlators of the bulk theory as shown in figure 1.1. Further, the AdS/CFT dictionary [3–5] tells us that a massive scalar bulk state is dual to a boundary operator with dimension

$$\Delta(\Delta - d) = m^2 R^2 \tag{1.6}$$

The space of these conformal boundary theories is constrained by the usual CFT bootstrap that we just discussed above. This, in turn, translates to constraints on the space of massive bulk theories.

The structure of QFTs in hyperbolic space forms an interesting subject by itself, but for obvious reasons it would be more interesting if we could study flat space physics within this framework. The flat space limit is achieved by taking the radius of curvature R to

[3]Except for the existence of a stress tensor. For this reason we refrain from referring to these boundary theories as conformal *field* theories, and rather call them simply conformal theories.

infinity while keeping the particle masses of the bulk theory fixed which implies that the dimensions $\Delta \sim mR$ of the dual boundary operators will also diverge. In this limit the boundary correlators of the bulk theory are naturally associated with S-matrix elements which are then constrained by the large-Δ limit of the conformal bootstrap. We refer to this procedure as the *boundary bootstrap*.

In chapter 4 we will discuss the implementation of the boundary bootstrap numerics. The large Δ limit in the conformal bootstrap is computationally rather challenging. For this reason we will focus on the the simplest setting: a $1+1$ dimensional bulk and thus a 1-dimensional conformal theory. The major simplification in this case is that we do not need to sum over spins in the OPE. In chapter 5 we shall take a more pedestrian approach and study the question of bounding couplings directly within the framework of flatspace QFT. The flatspace QFT intuition motivating the existence of such bounds is that as couplings become larger bound state masses tend to decrease and new bound-states may be pulled down from the continuum. Thus it is reasonable to expect that for a fixed spectrum the couplings cannot be arbitrarily large. We again focus on unitary, Lorentz invariant theories in 1+1 dimensions. Using only the usual crossing symmetry and unitarity of the flat space S-matrix we are able to establish bounds on the couplings which perfectly match those derived from the the boundary bootstrap.

To conclude this invitation, we would like to note that these bootstrap methods address both questions Q1 and Q2 posed in the introduction. Most notably, Q2 is directly addressed. In the formulation described here we work directly with the physical data of the QFT (particle masses and couplings) which is the input. Moreover, the formulation is completely non-perturbative and assumes nothing about the strengths of interactions, thus addressing Q1. Will the bootstrap be the Rosetta Stone that allows us to decode the secrets of nonperturbative QFT?

Part I

Wilson Loops and Correlators from Integrability

Chapter 2

Wilson Loops at strong coupling

In this chapter we focus on the computation of Wilson loops in $\mathcal{N} = 4$ at strong coupling. Our main tool is the formula (1.4) which relates the impossible QFT computation to a tractable (although still difficult) computation of minimal surfaces in AdS.

The study of minimal surfaces goes back at least as far as the time of Lagrange who in 1768 considered the problem: Find a surface of least area ending on a given closed contour [27]. This problem grew into an entire field of mathematics known as minimal surface theory and has occupied the attention of mathematicians and physicists alike for over two centuries (see [28] for a recent review).

Although historically most effort has focused on surfaces embedded in flat space, recent years have seen a shift in attention to minimal surfaces embedded in special curved spaces with the advent of the AdS/CFT correspondence [3–5]. While the mathematical statement of the problem is simple and perfectly well-posed – compute the area of the minimal surface ending on a given closed contour at the boundary of AdS – in practice this is a challenging task. A hand full of exact solutions exist in cases where the boundary curve has an exceptional amount of symmetry. For example, for closed loops in Euclidean AdS_3 one can construct solutions for a circular [29] and lens-shaped [30] boundary curve. In the case of closed spacelike loops in Minkowskian AdS the area can be computed exactly for the circle and the 4-cusp [31, 32] solution, for example. There is also a beautiful method for constructing quite general solutions parameterized by Riemann surfaces using theta-function techniques [33–37].

In a parallel development, recent years have witnessed a boom in our understanding of special types of surfaces in AdS based on the integrability of the underlying sigma model. Thus far, integrability based techniques have successfully been applied to describe

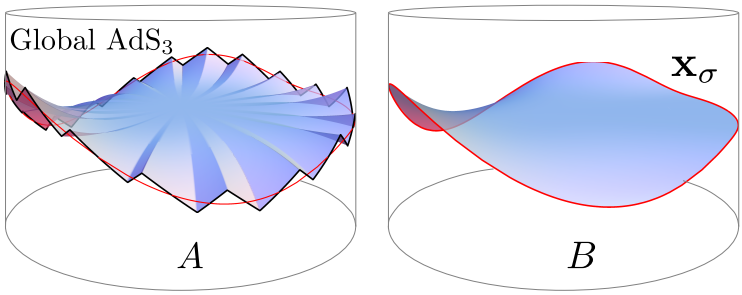


Figure 2.1: *Artistic depiction of minimal surfaces in global AdS.* Surface A ends on a null polygon at the boundary of global AdS_3, which is indicated by the gray cylinder. As the number of cusps becomes large surface A limits to surface B, which ends on the smooth curve $\mathbf{x}_\sigma$ shown in red.

surfaces which approach the boundary at spikes – relevant for the study of correlation functions [38–44] – as well as surfaces which approach the boundary along generic null polygons – relevant for the study of Wilson Loops and scattering amplitudes [19, 45–47]. These results hold only in the strong coupling limit of the theory, where the problem becomes one of classical strings moving in AdS.

Two features of the integrability-based approach should be emphasized. First, this approach is very economical in that one directly computes the minimal area without ever needing to know the shape of the embedding surface. This gives an enormous analytical and numerical advantage in the treatment of the problem. Second, and most important, they provide a manifestly integrable formulation of the purely classical worldsheet problem. As we have already mentioned in the invitation 1.1, this was the key step in determining the full finite coupling solution for the spectral problem and for scattering amplitudes.

In this chapter, we set our sights on an all-loop description of smooth Wilson loops in $\mathcal{N} = 4$ Super Yang Mills Theory. The first step toward this end is to develop a manifestly integrable formulation of the classical worldsheet problem that emerges at strong coupling. We will initiate a systematic integrability-based study of minimal surfaces in AdS which end on smooth curves at the boundary. The main observation is the simple and well-known fact that any smooth curve can be approximated to arbitrary accuracy by a sequence of null segments (see figure 2.1). Thus we can start with the results of [19] for null polygons and compute the minimal area of *any* smooth (simply connected) boundary curve by performing a careful continuum limit. The result of this continuum limit, and the main result of this chapter, is a novel set of integral equations whose solution yields the area of

minimal surfaces ending on smooth curves at the boundary of AdS. We will refer to these equations as the Continuum Thermodynamic Bethe Ansatz equations or simply CTBA equations since they are the continuum analog of TBA equations derived in [19] for null polygons.

The content of this chapter is based on the work [48]. We begin in section 2.1.1 with a brief review of the AMSV solution [19] for the case of null polygonal Wilson loops. In section 2.1.2 we develop the the necessary technology for the analytic continuation of the AMSV equations. This analytic continuation is needed in order to describe all possible configurations of a null polygonal wilson loop. Indeed, as we understand in section 2.2 an elaborate continuation is needed to pass from the original AMSV equations to a set of equations which is suitable for the continuum limit. Only after this continuation is performed do the equations describe generic smooth curves in the continuum limit. Sections 2.2.2 and 2.2.3 contains the derivation of the continuum version of the AMSV equations and present the main results of this chapter. In section 2.2.5 we study a special exact solution of the CTBA whose area can be computed exactly. Finally, in section 2.3.1 we develop a numerical implementation of our method for computing minimal areas. This allows us to demonstrate that the CTBA is not only a powerful tool for analytics, but is also a useful computational tool. It also affords the opportunity to perform a final check of the equations presented here. In particular, we show that results obtained from the CTBA agree with those obtained from direct numerical integration of the string equations of motion.

2.1 Null Polygonal Wilson Loops

2.1.1 Review of AMSV solution

In this section we briefly review the solution of AMSV for computing the area of surfaces in AdS_3 ending on null polygonal boundary contours. Recall that AdS_3 can be considered as a surface embedded in $\mathbb{R}^{2,2}$. The starting point of the AMSV construction is to reduce the AdS_3 sigma model to a set of manifestly $\mathbb{R}^{2,2}$ invariant variables whose equations of motion simultaneously encode those of the embedding coordinates and the Virasoro constraints. This is the well known Pohlmeyer reduction procedure [49–51]. In these reduced variables the $2N-6$ cross ratios of a null polygon with $2N$ sides are encoded in cycle integrals

$$Z_a = \oint_{\gamma_a} dz \sqrt{p(z)} \tag{2.1}$$

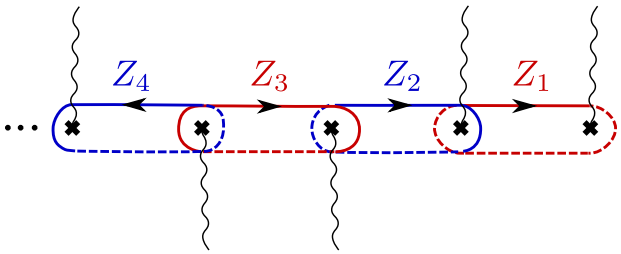


Figure 2.2: Cycles integrals Z_a. The black $\times$ indicate zeros of $p(z)$ and are arranged along the real axis. The wavy black lines indicate our convention for defining the branches of $\sqrt{p(z)}$. The branch chosen here is such that Z_2 is real and positive when the zeros are all on the real axis.

where $p(z)$ is a holomorphic polynomial of degree $N-2$ and γ_a are cycles on the riemann surface defined by $y^2 = p(z)$. This polynomial is constructed from the string embedding coordinates and z is the worldsheet coordinate.[1] Since the worldsheet has the topology of a disk the riemann surface has $N-3$ independent cycles, which gives $2N-6$ real parameters as required by the counting of cross ratios.

Of course there are many different possible choices of cycles. Careful WKB analysis of the flat connection problem associated with the Pohlmeyer reduced equations of motion selects a distinguished set of these cycles. Which set is selected depends on the precise form of p and is exhaustively explored in [52,53]. The simplest scenario – the one originally considered in AMSV – is the case in which the zeros of p are close to the real axis. In this case we have the cycles shown in figure 2.2 which are the input into the TBA equations

$$\log Y_a(\theta) = -2|Z_a|\cosh\theta - \sum_b K_{ab}(\theta-\theta') * \log(1+Y_b(\theta')) \tag{2.2}$$

where the kernel is given by

$$K_{ab}(\theta) = \frac{\langle a,b\rangle}{2\pi i \sinh(\theta + i\phi_a - i\phi_b)} \tag{2.3}$$

Here $\phi_a = \arg Z_a$ and $\langle a,b\rangle$ is the intersection number of cyles γ_a and γ_b.[2] There is one so-called Y-function $Y_a(\theta)$ corresponding to each cycle integral. The Y-functions obey an

[1] The worldsheet is euclidean and we choose the standard complex coordinates z,$\bar{z}$.

[2] The branch of arg is not important due to the periodicity of the kernel. Note that when the zeros are all along the real axis we have $\phi_{2a+1} = \pi/2$, $\phi_{2a} = 0$. This pattern of phases is easily inferred from the the contours of $\mathrm{Re}\sqrt{p}$ and $\mathrm{Im}\sqrt{p}$. Furthermore, according to figure 2.2 we have $\langle 2a+1, 2a\rangle = -\langle 2a, 2a+1\rangle = +1$. Thus we have $K_{ab} = K_{ba} = -1/\cosh\theta$ and we recover the usual form of the TBA given in [19].

important functional identity known as the Y-system

$$Y_a(\theta + i\pi/2)Y_a(\theta - i\pi/2) = (1 + Y_{a+1}(\theta))(1 + Y_{a-1}(\theta)) \tag{2.4}$$

which can be derived from the TBA (2.2) by smoothly shifting the equation to $\theta \to \theta + i\pi/2$ and adding it to the equation shifted to $\theta \to \theta - i\pi/2$.

Once the Y-functions are obtained from (2.2) (typically via numerical integration) the interesting part of the minimal area is computed as

$$\text{Area} = \frac{1}{2\pi} \sum_a \int d\theta |Z_a| \, e^{-\theta} \, \log(1 + Y_a) + ... \tag{2.5}$$

where the +... are some explicitly computable terms that we are not concerned with presently. As we have already mentioned, the geometry of the boundary curve is encoded parametrically in the polynomial $p(z)$. After solving the TBA we determine the physical cross ratios of the polygon from the Y-functions via

$$\widehat{Y}_{2a}(0) = (-1)\frac{\left(x^+_{-a} \wedge x^+_a\right)\left(x^+_{a+1} \wedge x^+_{-a-1}\right)}{\left(x^+_a \wedge x^+_{a+1}\right)\left(x^+_{-a-1} \wedge x^+_{-a}\right)} \tag{2.6}$$

$$\widehat{Y}_{2a+1}(0) = (-1)\frac{\left(x^+_{-a-1} \wedge x^+_a\right)\left(x^+_{a+1} \wedge x^+_{-a-2}\right)}{\left(x^+_a \wedge x^+_{a+1}\right)\left(x^+_{-a-2} \wedge x^+_{-a-1}\right)} \tag{2.7}$$

where we have defined

$$Y_a(\theta) = \widehat{Y}_a\left(\theta + i\phi_a\right) \tag{2.8}$$

The same expressions with $x^+ \to x^-$ are obtained by evaluating at $\theta = i\pi/2$. Equations (2.2)-(2.8) give, in principle, a complete solution to the problem of computing minimal surfaces in AdS_3 ending on an arbitrary null polygon. Several comments are in order.

Note that this solution is parametric. The input into the equations is a polynomial (or rather it's cycle integrals). The number of edges of the polygon fixes the degree of the polynomial, but other than that little is known about what specific polynomial should be chosen to describe a specific polygon. In this parametric form of the equations one would need slowly vary the parameters of the polynomial until the desired cross-ratios (2.8) are attained. We can do much better though. In fact it is possible to explicitly eliminate the Z_a in (2.2) in favour of the cross-ratios (2.8) such that the latter are the direct input into the problem as explained in [9].

Equations (2.2)-(2.8) are derived when the zeros of the polynomial are close to the real axis, and therefore are only valid for a region of the full space of null polygons which correspond to such polynomials. In order to describe the full space of polygons the equations

must be analytically continued. This issue will play an central role in our analysis of the continuum limit later in this chapter and section (2.1.2) is devoted to this point. The parametric form of the equations given above seems to be the most suitable for performing this continuation. For this reason, we work exclusively with this form of the equations, keeping in mind that we can rewrite them directly in terms of cross ratios if we wish.

2.1.2 TBA Morphing

In this section we study the analytic continuation of the AMSV solution in the space of cross ratios, which amounts to analytically continuing the TBA equations in the parameters Z_a.[3] The analytic continuation is non trivial due to fact that the poles of the kernel (2.3) move as we deform the Z_a and can cross the contours of integrations as we pass from one region of parameter space to another. We will refer to this as a "wall crossing". It will be important for later purposes to be able to easily follow the changes in the TBA equations as we wall cross through the parameter space. The goal of this section is to introduce a set of simple graphical rules which allows this process to be performed with ease.

As we perform generic deformations of the phases ϕ_a the poles of the kernel 2.3 move. When a pole of the kernel crosses the integration contour the correct prescription is to analytically continue the Y-functions by deforming the contour and picking the residue of the pole. As we will see in this section, ultimately the process of analytic continuation does not change the schematic form of the equations (2.2); the only change is in what cycles are present in the equations.

Let us begin with the AMSV equations [19] which hold when the zeros of p are sufficiently close to the real axis. The TBA is given by (2.2) with the cycles shown in figure 2.2. Our first step will be to put these equations into a graphical form. To this end, it is useful to make the following notational change. We number the zeros of p from left to right and then label the cycles according to which zeros they encircle – e.g $Z_1 \to Z_{12}$, $Z_2 \to Z_{23}$,etc. The Y-functions, and other variables are relabelled accordingly $Y_a \to Y_{\mathbf{s}}$ where we use the bold index $\mathbf{s} = 12, 23, ...$ to represent a pair of numbers. Note that in this notation the order of the indices on a given object do not matter – e.g $Y_{12} \equiv Y_{21}$, etc.

Although figure 2.2 encodes the intersection matrix, one can not see from this figure when a wall crossing will occur. For this reason, it is useful to work directly with the $Z_{\mathbf{s}}$ since it is their relative phases that control the crossings. Each $Z_{\mathbf{s}}$ is a complex number

[3] This analytic continuation presents a beautiful mathematical problem with many connections to the topic of wall-crossing [52, 53].

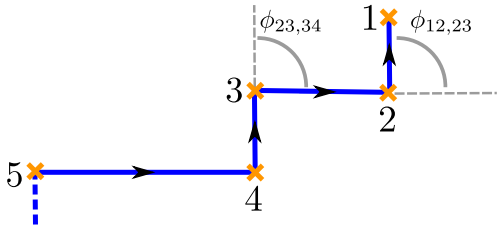


Figure 2.3: Cycles $Z_{\mathbf{s}}$ when the zeros of p are along the real axis. In this case the phases obey $\phi_{12} = \phi_{34} = \cdots = \pi/2$ and $\phi_{23} = \phi_{45} = \cdots = 0$. We represent each complex $Z_{\mathbf{s}}$ by a vector. By arranging these vectors tail-to-tip the intersection matrix is encoded as follows: if $Z_{\mathbf{s}}$ and $Z_{\mathbf{t}}$ touch at their endpoints $\langle \mathbf{s}, \mathbf{t} \rangle = \pm 1$; if they do not then $\langle \mathbf{s}, \mathbf{t} \rangle = 0$. The sign is determined by the usual right-hand-rule. We thus read off from the figure $\langle 12, 23 \rangle = \langle 34, 23 \rangle = \cdots = -1$.

which can be represented with a vector. By arranging these vectors 'tail to tip' as shown in figure 2.3 we can read off the intersection matrix as follows: if $Z_{\mathbf{s}}$ and $Z_{\mathbf{t}}$ touch at their endpoints $\langle \mathbf{s}, \mathbf{t} \rangle = \pm 1$; if they do not then $\langle \mathbf{s}, \mathbf{t} \rangle = 0$. The sign is determined by the usual right-hand-rule. Thus the TBA equations corresponding to figure 2.3 are given by 2.2 with $\phi_{\mathbf{s}}$ and $\langle \mathbf{s}, \mathbf{t} \rangle$ given in the caption. As a concrete example consider the case of a 12-sided polygon for which the corresponding TBA is given by

$$\begin{aligned}
\log Y_{12}(\theta) &= -2|Z_{12}| \cosh\theta - (-1) \int \frac{d\theta'}{2\pi} \frac{\log\left(1 + Y_{23}(\theta')\right)}{i \sinh\left(\theta - \theta' + i\phi_{12,23}\right)} \\
\log Y_{23}(\theta) &= -2|Z_{23}| \cosh\theta - \sum_{\mathbf{t}=12,34} (+1) \int \frac{d\theta'}{2\pi} \frac{\log\left(1 + Y_{\mathbf{t}}(\theta')\right)}{i \sinh\left(\theta - \theta' + i\phi_{23,\mathbf{t}}\right)} \\
\log Y_{34}(\theta) &= -2|Z_{34}| \cosh\theta - (-1) \int \frac{d\theta'}{2\pi} \frac{\log\left(1 + Y_{23}(\theta')\right)}{i \sinh\left(\theta - \theta' + i\phi_{34,23}\right)}
\end{aligned} \tag{2.9}$$

and the corresponding graphical representation is given in figure 2.4A.

Now let us consider deforming the $Z_{\mathbf{s}}$ away from the AMSV configuration. In fact, the 12-gon case given above will demonstrate all the essential features of the continuation for a general N and so we will continue to use it as a concrete example. Consider smoothy rotating the point 1 clockwise about the point 2 as shown in figure 2.4A. When Z_{12} aligns with Z_{23} the phases $\phi_{12,23} = -\phi_{23,12}$ are such that the poles in the kernels coupling Y_{12} and Y_{23} will cross the contour of integration. We analytically continue by deforming the

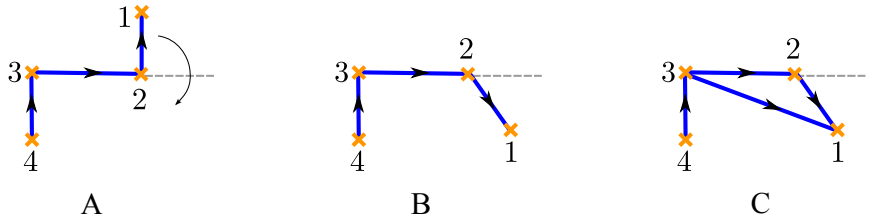


Figure 2.4: Deformation of the $Z_{\mathbf{s}}$. The cycles shown in A corresponds to the AMSV case where all zeros of p are along the real axis. Whenever two of the $Z_{\mathbf{s}}$ align the TBA must be analytically continued since a pole of the kernel is crossing the integration contour. Thus to pass from configuration in A to that in B we must analytically continue the TBA. The analytic continuation due to the the alignment of $Z_{s_1 s_2}$ and $Z_{s_2 s_3}$ results in the addition of a new Y-function associated with the cycle $Z_{s_1 s_3} = Z_{s_1 s_2} + Z_{s_2 s_3}$ as explained in the main text. This gives the TBA shown in C.

integration contour and picking the pole of the kernel which gives the modified equations

$$\begin{aligned}
\log Y_{12}(\theta) &= -2|Z_{12}|\cosh\theta + \log(1+Y_{23}(\theta+i\phi_{12,23})) - K_{12,23}*\log(1+Y_{23}) \\
\log Y_{23}(\theta) &= -2|Z_{23}|\cosh\theta + \log(1+Y_{12}(\theta+i\phi_{23,12})) - \sum_{\mathbf{t}=12,34} K_{23,\mathbf{t}}*\log(1+Y_{\mathbf{t}}) \\
\log Y_{34}(\theta) &= -2|Z_{34}|\cosh\theta - K_{34,23}*\log(1+Y_{23})
\end{aligned} \tag{2.10}$$

Since the new source terms are evaluated off of the real axis, the equations are no longer closed. One needs to introduce additional equations which compute the Y-functions at shifted argument. There is a particularly beautiful way of implementing this described in [19]. We first absorb the new source terms in the left hand side by defining the new Y-functions

$$Y'_{12}(\theta) = \frac{Y_{12}(\theta)}{1+Y_{23}(\theta+i\phi_{12,23})}, \qquad Y'_{23}(\theta) = \frac{Y_{23}(\theta)}{1+Y_{12}(\theta+i\phi_{23,12})} \tag{2.11}$$

To close the equations we must rewrite the right hand side of equations (2.10) in terms of $Y'_{12}(\theta)$ and $Y'_{23}(\theta)$. To achieve this we introduce a new Y-function Y_{13} defined by the equations[4]

$$(1+Y_{12}(\theta)) = (1+Y'_{12}(\theta))(1+Y_{13}(\theta+i\phi_{12,13})) \tag{2.12}$$

$$(1+Y_{23}(\theta)) = (1+Y'_{23}(\theta))(1+Y_{13}(\theta+i\phi_{23,13})) \tag{2.13}$$

[4]These two equations are consistent due to the identity $(1+\widehat{Y}_{12})/(1+\widehat{Y}'_{12}) = (1+\widehat{Y}_{23})/(1+\widehat{Y}'_{23})$.

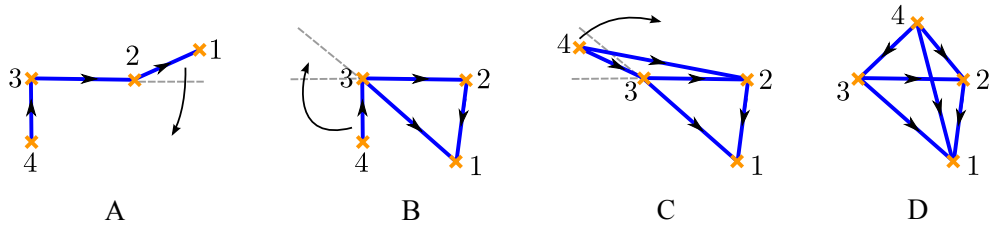


Figure 2.5: Analytic continuation of the TBA as we deform the $Z_{\mathbf{s}}$.

Plugging these into (2.10) we have

$$\begin{aligned}
\log Y'_{12}(\theta) &= -2|Z_{12}|\cosh\theta - \sum_{\mathbf{t}=23,13} K_{12,\mathbf{t}} * \log\left(1+Y'_{\mathbf{t}}\right) \\
\log Y'_{23}(\theta) &= -2|Z_{23}|\cosh\theta - \sum_{\mathbf{t}=12,34,13} K_{23,\mathbf{t}} * \log\left(1+Y'_{\mathbf{t}}\right) \\
\log Y_{34}(\theta) &= -2|Z_{34}|\cosh\theta - \sum_{\mathbf{t}=23,13} K_{34,\mathbf{t}} * \log\left(1+Y'_{\mathbf{t}}\right)
\end{aligned} \tag{2.14}$$

Now to close the equations we need only to derive the equation satisfied by Y_{13}. This is done by adding together the Y_{12} and Y_{23} equations with the appropriate shifts. The result is

$$\log Y_{13}(\theta) = -2|Z_{13}|\cosh\theta - \sum_{\mathbf{t}=12,23,34} K_{13,\mathbf{t}} * \log\left(1+Y'_{\mathbf{t}}\right) \tag{2.15}$$

where $|Z_{13}|e^{\phi_{13}} = Z_{12} + Z_{23}$. This completes the necessary analytic continuation to pass from the parameters in figure 2.4A to the those in figure 2.4B. Equations (2.14)-(2.15) provide a closed set of equations valid in the new range of parameters.

Now we would like to find a graphical representation of this procedure. To this end, consider the graph shown in figure 2.4C which contains the deformed $Z_{\mathbf{s}}$ as well as the new Z_{13}. This graph encodes the analytically continued equations (2.14)-(2.15). For example, one can check that the intersection matrix between the Y-functions is correctly encoded in this figure.[5] Thus we have a simple graphical rule for analytically continuing the equations: whenever two of the $Z_{\mathbf{s}}$, say Z_{rs} and Z_{st} (with non-zero intersection $\langle rs, st\rangle \neq 0$) cross over by aligning we simply to add a new $Z_{rt} = Z_{rs} + Z_{st}$ which represents a new Y-function in the integral equations. Thus with such diagrams we can easily keep track of the form of the equations as we perform a general analytic continuation of the TBA for a Wilson loop

[5] In other words, the intersection matrix is still just equal to the intersection matrix of the cycles.

with any number of sides. As an example, let us continue to deform the $N = 6$ example studied above. Suppose we now want to rotate the point 4 clockwise about the point 3 as shown in figure 2.5B. First Z_{34} will align with Z_{23} so we must analytically continue resulting in the new form of the TBA shown in figure 2.5C which includes the new cycle Z_{24}. From this graph we can easily read off the TBA equations in this region of parameter space. For example, the equation for Y_{13} is given by

$$\log Y_{13}(\theta) = -2|Z_{13}|\cosh\theta - \sum_{\mathbf{t}=12,23,34} \langle 13, \mathbf{t}\rangle \int \frac{d\theta'}{2\pi} \frac{\log\left(1 + Y_{\mathbf{t}}(\theta')\right)}{i\sinh\left(\theta - \theta' + i\phi_{13,\mathbf{t}}\right)}$$

with $\langle 13, 12\rangle = -1$, $\langle 13, 23\rangle = +1$,$\langle 12, 13\rangle = +1$. If we continue to rotate then Z_{34} will align with Z_{13} resulting in the equations shown in figure 2.5D with the new cycle Z_{14}. In this later figure we see a new feature of the graphs: Z_{14} and Z_{23} cross over each other rather than touching at an endpoint. Such a crossing indicates an intersection matrix $\langle rs, tu\rangle = \pm 2$ with the sign fixed again by the right hand rule.[6] In particular, in the example of figure 2.5D we have $\langle 14, 23\rangle = +2$. For example, the equation for Y_{14} is given by

$$\log Y_{14}(\theta) = -2|Z_{14}|\cosh\theta - \sum_{\mathbf{t}=12,13,34,24,23} \langle 14, \mathbf{t}\rangle \int \frac{d\theta'}{2\pi} \frac{\log\left(1 + Y_{\mathbf{t}}(\theta')\right)}{i\sinh\left(\theta - \theta' + i\phi_{14,\mathbf{t}}\right)} \tag{2.16}$$

with $\langle 14, 12\rangle = -1$, $\langle 14, 13\rangle = +1$, $\langle 14, 34\rangle = -1$, $\langle 14, 24\rangle = +1$, $\langle 14, 23\rangle = +2$.

Let us now make a few comments about the features of this example which are characteristic of the general N case. First, note that in configuration D the $Z_{\mathbf{s}}$ form a complete graph. As a result, there is no way to create new edges in the graph: all possible wall-crossings will only remove edges. For example, this can happen through a reversal of the rule in figure 2.4. Thus this region of parameter space contains the maximum number of Y-functions in its TBA and we will refer to this region of parameter space as the *maximal* region. This is in contrast with the region of parameters where the AMSV equations hold, which contains the minimal number of Y-functions and we thus refer to as the *minimal* region. For any N the TBA in the maximal region corresponds to the complete graph spanned by the (deformed) $Z_{\mathbf{s}}$ of the minimal region. In section 2.1.3 We will study the maximal region TBA in detail.

As we wall-cross in the TBA it is useful to rewrite the area (2.5) in terms of the new Y-functions in each region of parameter space. This can be done with a simple application

[6] This rule can be derived by simply repeating the steps given above for deriving the analytically continued equations.

of the identities (2.12). For example, the free energy corresponding to the TBA in figure 2.5A is given by

$$A_{\text{free}} = \frac{1}{2\pi} \sum_{\mathbf{s}=12,23,34} \int d\theta |Z_{\mathbf{s}}| \, e^{\theta} \, \log(1 + Y_{\mathbf{s}}(\theta)) \tag{2.17}$$

$$= \frac{1}{2\pi} \sum_{\mathbf{s}=12,23,34} \int d\theta Z_{\mathbf{s}} \, e^{\theta} \, \log(1 + \widehat{Y}_{\mathbf{s}}(\theta)) \tag{2.18}$$

To rewrite this in terms of the the Y-functions appropriate for the parameters of figure 2.5B we simply plug in (2.12) to obtain

$$A_{\text{free}} = \frac{1}{2\pi} \sum_{\mathbf{s}=12,23,34,13} \int d\theta Z_{\mathbf{s}} \, e^{\theta} \, \log(1 + \widehat{Y}^B_{\mathbf{s}}(\theta)) \tag{2.19}$$

$$= \frac{1}{2\pi} \sum_{\mathbf{s}=12,23,34,13} \int d\theta |Z_{\mathbf{s}}| \, e^{\theta} \, \log(1 + Y^B_{\mathbf{s}}(\theta)) \tag{2.20}$$

where in the first line we have used the fact that $Z_{13} = Z_{12} + Z_{23}$ and we are using the notation $Y^B_{\mathbf{s}}$ to indicate the Y-functions in the region of parameter space of figure 2.5B (i.e. $Y^B_{12} = Y'_{12}$ and $Y^B_{34} = Y_{34}$). Applying this procedure after each wall crossing, the free energy always maintains the form (2.5) where the sum is taken over all the Y-functions of the parameter region.

2.1.3 Maximal TBA

In this section we will perform a detailed study of the TBA in the maximal region. We will first determine what form of p corresponds to the maximal TBA. We then write the form of this TBA for general N and work out the relation between the maximal cell Y-functions and geometry. Both of these are crucial steps in obtaining the continuum TBA as we will discover in the following sections.

Let us begin by determining what form of p corresponds to the maximal region. It is useful to start with the example of $N = 6$ and consider a specific deformation of the zeros that achieves the wall-crossing sequence of figure 2.5. For example, consider a deformation like the one shown on the left side of figure 2.6. In the right panel of figure 2.6 we evaluate the cycle integrals as a function of the deformation parameters (α_1, α_2). Our objective is to analytically continue from the lower left corner $(\alpha_1, \alpha_2) = (0, 0)$ to the upper right corner $(\alpha_1, \alpha_2) = (1, 1)$ which corresponds to the configuration where the zeros are at the

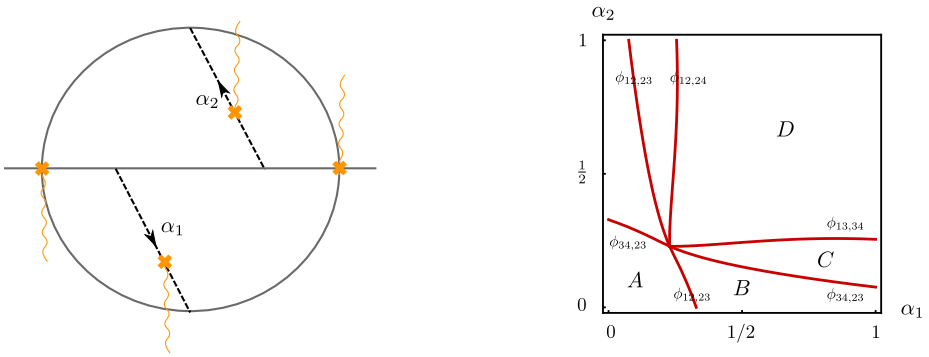


Figure 2.6: We consider a deformation parameterized by (α_1, α_2) of the polynomial p for the case $N = 6$. We start at $(\alpha_1, \alpha_2) = (0, 0)$ where all of the zeros are real and deform to $(\alpha_1, \alpha_2) = (1, 1)$ where the zeros are arranged at the 4th roots of unity. We evaluate the cycle integrals as a function of the deformation parameters (α_1, α_2) and on the right we show the contours where the relative phases of are vanishing such that a pole is crossing the integration contour (i.e two $Z_{\mathbf{s}}$ are aligning). The lower left corner corresponds to the AMSV configuration where all of the zeros are along the real axis and thus the TBA is represented by the diagram in figure 2.5A and is given in equation 2.9.

4th roots of unity. As per the TBA morphing rules of section 2.1.2 the AMSV form of the TBA will hold until two of the $Z_{\mathbf{s}}$ align and thus throughout the region labeled A. The red contour labeled $\phi_{12,23}$ corresponds to Z_{12} and Z_{23} aligning, and thus the new TBA in the region labeled B is given by that in figure 2.5B. Similarly, the TBA in regions C and D correspond to figure 2.5 C and D respectively. In particular, we see in this example that when the zeros are arranged near the roots of unity the TBA takes the maximal form described in section 2.1.2. Moreover, repeating this process for the next few values of N reveals that this is a general pattern – for any N the TBA for zeros near the roots of unity takes the maximal form. Now it is a trivial matter to write the TBA in the maximal region for any N as we can easily draw its graphical representation in that region.

Let us make a simplifying remark. Note that according to the discussion of section 2.1.2 the sign of the intersection matrix is correlated with the sign of the angle $\phi_{\mathbf{s},\mathbf{t}}$ – they are determined by the same right-hand-rule. Thus, we can use the identity $\sinh(\theta) = -\sinh(\theta \pm i\pi)$ to write the TBA in the form

$$\log Y_{\mathbf{s}}(\theta) = -2|Z_{\mathbf{s}}| \cosh\theta - \sum_{\mathbf{t}} |\langle \mathbf{s}, \mathbf{t} \rangle| \int \frac{d\theta'}{2\pi i} \frac{\log(1 + Y_{\mathbf{t}}(\theta'))}{\sinh(\theta - \theta' + i\varphi_{\mathbf{s},\mathbf{t}})} \tag{2.21}$$

where $\varphi_{\mathbf{s},\mathbf{t}} \equiv \text{Mod}\,[\phi_{\mathbf{s},\mathbf{t}}, \pi]$ so that $0 < \varphi_{\mathbf{s},\mathbf{t}} < \pi$. The great advantage of this form is that we no longer need to keep track of the orientation of the edges in the graphical representation of the TBA – i.e. we can simply drop the arrowhead on each graph. To clarify, consider again the $N = 6$ case, and examine the maximal region Y_{14} equation given in (2.16). Note that the elements of the intersection matrix that satisfy $\langle \mathbf{s}, \mathbf{t} \rangle > 0$ correspond to angles obeying $0 < \phi_{\mathbf{s},\mathbf{t}} < \pi$ (see figure 2.5D) while $\langle \mathbf{s}, \mathbf{t} \rangle < 0$ correspond to $-\pi < \phi_{\mathbf{s},\mathbf{t}} < 0$. Thus, in the later case we can absorb the minus sign from the intersection matrix into the kernel by replacing $\phi_{\mathbf{s},\mathbf{t}} \to \phi_{\mathbf{s},\mathbf{t}} + \pi$ which gives equations (2.21).

We will now examine the relationship between the maximal cell Y functions and geometry. We can easily use the morphing rules to follow the change in the Y-functions as we wall cross. Doing this for the $N = 6$ and $N = 7$ deformations discussed above suggests that the basis Y-functions transform as

$$\widehat{Y}^{\max}_{s,s+1} = \frac{\widehat{Y}^{\min}_{s,s+1}}{(1 + \widehat{Y}^{\min}_{s-1,s})(1 + \widehat{Y}^{\min}_{s+1,s+2})} = \frac{1}{(\widehat{Y}^{\min}_{s,s+1})^{\pm\pm}} \tag{2.22}$$

where in the second step we used the Y-system (2.4). For example, for $\widehat{Y}_{45}$ we have[7]

$$(\widehat{Y}^{\min}_{45})(\widehat{Y}^{\min}_{45})^{--} = (1 + \widehat{Y}^{\min}_{34})(1 + \widehat{Y}^{\min}_{56}) \tag{2.23}$$

Since the relation between the $Y^{\min}$ and geometry is given by (2.6) then equation (2.22) provides the relation between the $Y^{\max}$ and geometry. There is a simple way to derive (2.22) from the graphical rules which goes as follows. Consider the wall crossing sequence of figure 2.5. In this example we see that Y_{12} is only involved in the first wall-crossing and Y_{23} is involved in the first two. Using the morphing rules 2.11 we immediately see

$$\widehat{Y}^{\max}_{12} = \frac{\widehat{Y}_{12}}{1 + \widehat{Y}_{23}}, \qquad \widehat{Y}^{\max}_{23} = \frac{\widehat{Y}_{23}}{(1 + \widehat{Y}_{12})(1 + \widehat{Y}_{34})} \tag{2.24}$$

which is in accord with (2.22) (we now drop the label min of the RHS). The Y_{34} is involved in the second and third wall-crossing and thus it is not so simple to see that (2.22) holds. However, after some algebra we indeed see

$$\widehat{Y}^{\max}_{34} = \frac{\widehat{Y}_{34}}{\left(\frac{\widehat{Y}_{23}}{\widehat{Y}_{12}+1} + 1\right)\left(\frac{\widehat{Y}_{12}\widehat{Y}_{23}}{\widehat{Y}_{12}+\widehat{Y}_{23}+1} + 1\right)} = \frac{\widehat{Y}_{34}}{1 + \widehat{Y}_{23}} \tag{2.25}$$

[7] Since $\phi^0_{34} = \phi^0_{56} = \pi/2$ and $\phi^0_{23} = 0$ (where $\phi^0_{\mathbf{s}}$ is the value of the phases when the zeros are real) which we can see from figure 2.3.

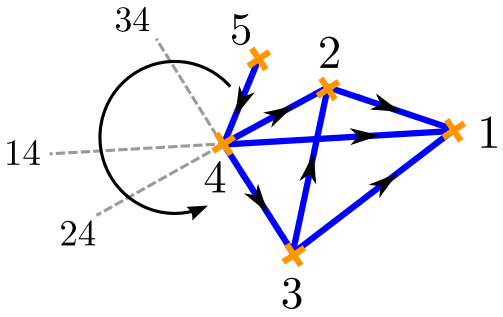


Figure 2.7: Generating the maximal graph for the $N = 7$ case from the $N = 6$ case. We start with the maximal graph for the $N = 6$ case and attach a new "leg" to the point 4. The new leg should be attached such that the angle $\phi_{34,45}$ is as it would be in the minimal region. We then rotate the point 5 counterclockwise such that all possible wall-crossings involving this leg are achieved. The result is the maximal graph for the $N = 7$ case.

However, such a simplification must occur because we could have wall-crossed to the maximal region by the alternate route in figure 2.6 where we first cross angle $\phi_{34,23}$ then $\phi_{12,23}$ and then $\phi_{12,24}$. In this case Y_{34} undergoes only one wall-crossing and it is obvious that Y_{34} and Y_{23} satisfy (2.22). Thus by considering these two deformations (2.22) is obvious for the $N = 6$ case. We will now extend this argument to any N recursively. We can generate the maximal region equations for the $N = 7$ case from the $N = 6$ case as shown in figure 2.7. Then we see that Y_{34} undergoes an additional crossing such that it is given by

$$\widehat{Y}_{34}^{\max} = \frac{\widehat{Y}_{34}}{(1+\widehat{Y}_{23})(1+\widehat{Y}_{45})} \tag{2.26}$$

in accord with 2.22 for $N = 7$. The Y_{45} however undergoes 3 crossings giving

$$\widehat{Y}_{45} = \frac{\widehat{Y}_{45}}{(X_1+1)(X_2+1)(X_3+1)} \tag{2.27}$$

where in the denominator we have defined

$$X_1 = \frac{\widehat{Y}_{34}}{\left(\frac{\widehat{Y}_{23}}{\widehat{Y}_{12}+1}+1\right)\left(\frac{\widehat{Y}_{12}\widehat{Y}_{23}}{\widehat{Y}_{12}+\widehat{Y}_{23}+1}+1\right)} \tag{2.28}$$

$$X_2 = \frac{\widehat{Y}_{23}\widehat{Y}_{34}}{(\widehat{Y}_{12}+1)\left(\frac{\widehat{Y}_{23}}{\widehat{Y}_{12}+1}+\widehat{Y}_{34}+1\right)} \tag{2.29}$$

$$X_3 = \frac{\widehat{Y}_{12}\widehat{Y}_{23}\widehat{Y}_{34}}{(\widehat{Y}_{12}+\widehat{Y}_{23}+1)\left(\frac{\widehat{Y}_{23}}{\widehat{Y}_{12}+1}+1\right)\left(\frac{\widehat{Y}_{12}\widehat{Y}_{23}}{\widehat{Y}_{12}+\widehat{Y}_{23}+1}+\frac{\widehat{Y}_{34}}{\frac{\widehat{Y}_{23}}{\widehat{Y}_{12}+1}+1}+1\right)} \tag{2.30}$$

Again one can check with a bit of algebra that this complicated denominator does indeed reduce to $1+\widehat{Y}_{34}$ such that (2.22) holds. However, we can consider a different sequence to the maximal region in which Y_{45} is involved in the first crossing only. That is, start with four points labeled $2,...,5$ and cross to the maximal region for $N=6$. Then attach the new point 1 to point 2 and generate the $N=7$ maximal equations by the same method as in figure (2.7). In the later case it is obvious that Y_{45} obeys (2.22) and thus we have also established this identity for $N=7$. We could now repeat this argument for $N=8$ by attaching a point 6 to point 5 in the maximal graph for $N=7$ and so on. Clearly this can be repeated recursively to derive (2.22) for any N. Now that we have determined the relation to geometry for a basis of the $Y_{\mathbf{s}}^{\max}$ it is simple to determine the relations for the rest of the Y-functions. Here we simply note that there is a simple symmetry relating any Y-function to a basis Y-function by an overall rotation. For example, for the $N=6$ case we have

$$\widehat{Y}_{12}(\theta)=\widehat{Y}_{24}(\theta-i\pi)=\widehat{Y}_{34}(\theta-3i\pi)=\widehat{Y}_{13}(\theta-4i\pi),\qquad \widehat{Y}_{23}(\theta)=Y_{14}(\theta-i2\pi) \tag{2.31}$$

This is all the information we will require for our purposes below.

2.2 The continuum limit

In this section we consider minimal area surfaces in AdS ending on smooth contours at the AdS boundary. We are interested in computing the area of these surfaces as a function of the boundary contour. We solve this problem by taking a continuum limit of the AMSV equations for the minimal area of surfaces ending on generic null polygonal contours. The AMSV result consists of a set of Thermodynamic Bethe Ansatz (TBA) equations (2.2),

with the minimal area computed by the TBA free energy (2.5). These TBA equations must be carefully analytically continued in order for the limit of a large number of cusps to approach a smooth boundary contour. This analytic continuation or "TBA morphing" was considered in the previous section, and the tools and intuition developed there will be essential. Once the continuum limit is performed we find a novel TBA-like system whose input is a continuous curve parametrizing the continuous boundary curve. The output is a continuous spectrum of y-functions which encoding the cross ratios of the boundary curve, as well as the area which is a continuum version of the free energy. The layout of this section is as follows. In section 2.2.1 we will discuss how the continuum limit is achieved at the level of the polynomial p which parameterizes the cross ratios. From this, using the tools we developed in section 2.1.2 we then determine the proper form of the TBA for taking the continuum limit. The limit is performed in section 2.2.2.

2.2.1 Parameterization of the continuum limit

Since the form of the TBA depends on p, we must first understand how to take the continuum limit at the level of p. Clearly we must take the number of zeros to infinity since we are taking the limit of a large number of cusps. But how these zeros should be arranged is a more subtle question. Generically it is the divergence of p which drives the string embedding to the boundary. For example, in the null polygon case p diverges at infinity, and thus infinity in the z-plane maps to the AdS boundary. On the other hand, for the continuum limit one must create a curve in the z plane where $p(z)$ passes rapidly from finite values to infinity – this divergence forces the worldsheet to the boundary along this curve and thus creates a smooth boundary contour. The way to achieve this is to let zeros of $p(z)$ accumulate uniformly along the unit circle as shown in figure 2.8. In this case $p(z)$ will limit to a polynomial p_{in} inside the unit circle and infinity outside the unit circle. This remnant polynomial p_{in} parameterizes the (smooth) boundary curve of the wilson loop. In this way we can describe a generic family of surfaces ending on smooth boundary contours.

We now see that the form of p needed for the continuum limit is radically different from that considered by AMSV. In particular, we need to understand what form of the TBA corresponds to p of the form shown in figure 2.8 for general N. For this, the tools developed in section 2.1.2 will be essential. To utilize these tools we must understand how the $Z_{\mathbf{s}}$ change as we deform the zeros from the real axis into this configuration. Then we can easily use the graphical representation of the previous section to obtain the correct form of the TBA equations in that region.

In section 2.1.3 we considered the case in figure 2.8 where there are no yellow zeros, and only zeros near the unit circle. Let us consider the case in which the zeros are arranged

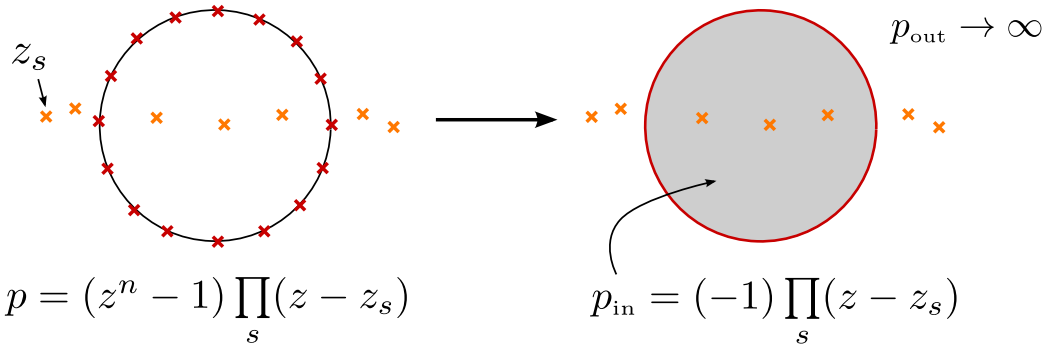


Figure 2.8: Parametrizing the continuum limit. We consider a polynomial like the one shown on the left where each $\times$ represents a zero. We take a subset of the zeros (the red ones) and place them uniformly along the unit circle, accumulating there in the continuum limit. The yellow zeros remain at arbitrary positions. As number of red zeros increases the polynomial limits to p_{in} inside the unit circle and ∞ outside. The divergence in p drives the worldsheet to the boundary along the edge of the unit circle, while the remnant polynomial p_{in} parametrizes the shape of the physical boundary curve.

precisely at the roots of unit and consider large N. In this case, the polynomial p will limit to $p_{\text{in}} = -R$ inside the unit circle. This will describe a specific family of continuous solutions (parameterized by R) and we shall study this solution in detail below. For now we would like to describe a much more general family of solutions. In fact there is a simple generalization of the above discussion which allows us to do this without re-doing the analytic continuation of section 2.1.2. The key observation is that each Y-function $Y_{\mathbf{s}}$ is strongly suppressed when $Z_{\mathbf{s}}$ is large, vanishing in the limit $Z_{\mathbf{s}} \to \infty$. Now consider adding some yellow zeros *outside* of the unit circle. For finite N the corresponding Y-functions will interact in some non-trivial way with the Y-functions corresponding to those of the red zeros thus complicating the analytic continuation discussed above. However, in the limit $N \to \infty$ since $p_{\text{out}} \to \infty$, all these Y-functions will vanish and the TBA will take the usual maximal form described above – the only remnant of the yellow zeros will be that now

$$p_{\text{in}} = R\prod_a (z - z_a) \tag{2.32}$$

where z_a are the yellow zeros sprinkled only outside the unit circle. In this way we can parameterize a huge class of solutions. Of course to discuss the most general solution one must also include yellow zeros inside of the unit disk in the large N limit. We shall not consider such solutions in this thesis. We only note that one could derive the corresponding

TBA using the TBA morphing tools developed in section 2.1.2 and following the values of the $Z_{\mathbf{s}}$) in the presence of these additional zeros.

Let us summarize the main points of this section. We first determined that the correct form of p for the continuum limit is that shown in figure 2.8 where we take the number of red zeros to infinity and the yellow zeros generate a residual polynomial p_{in} which parameterizes the shape of the continuous boundary curve that results in the limit. We then considered the case in which p_{in} is a polynomial with no zeros inside the unit disk. In this case the form of the TBA in the limit of a large number of cusps is the same as that with no yellow zeros – i.e. the maximal TBA of the red zeros – but with a nontrivial p_{in} generated by the presence of these zeros. In the next section we will derive the continuum limit of this TBA.

2.2.2 Continuum limit of maximal TBA

In this section we will study the limit $N \to \infty$ in the maximal region TBA. The first order of business is to determine the scaling of the Y-functions with N which can be ascertained from the relation between the Y-functions and their corresponding space time cross ratios. Note that each time a wall-crossing occurs involving an angle $\phi_{\mathbf{s},\mathbf{t}}$ the Y-functions $Y_{\mathbf{s}}$ and $Y_{\mathbf{t}}$ are redefined according to (2.11) and the geometrical cross ratios corresponding to these Y-functions changes accordingly. Thus the scaling of the $Y^{\min}$ can be different from that of $Y^{\max}$. Indeed, from (2.6) we see that $Y^{\min}_{s,s+1} \sim N^2$ since the distances in the denominator involve two neighbouring cusps. Thus is follows from (2.22) that $Y^{\max}_{s,s+1} \sim N^{-2}$. Since all other maximal Y-functions are just rotations of this basis, it follows that in the continuum limit the Y-functions scale like

$$Y_{\mathbf{s}}(\theta) \sim \frac{1}{N^2}\, y_{\boldsymbol{\sigma}}(\theta) \tag{2.33}$$

where $y_{\boldsymbol{\sigma}}$ is finite and independent of N and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2)$ where the $\sigma \in [0, 2\pi)$ are continuous variables.

Now note that on the right hand side of the maximal TBA we sum over $\mathcal{O}(N^2)$ y-functions, with each one scaling as N^{-2} (after expanding the log) such that it is very tempting to replace $\sum_{t_1 t_2} N^{-2} F_{t_1 t_2} \to \int d\tau_1 d\tau_2 f_{\tau_1 \tau_2}$. For this, we must be sure that $F_{t_1 t_2}$ limits to a smooth function and thus the alternating sign of $\langle \mathbf{s}; \mathbf{t} \rangle$ seems to pose a problem. However, note that according to the discussion of section 2.1.3 the sign of the intersection matrix is correlated with the sign of the angle $\phi_{\mathbf{s},\mathbf{t}}$ such that we can write the TBA in the form (2.21) where it is manifest that the integrand will limit to a smooth function of

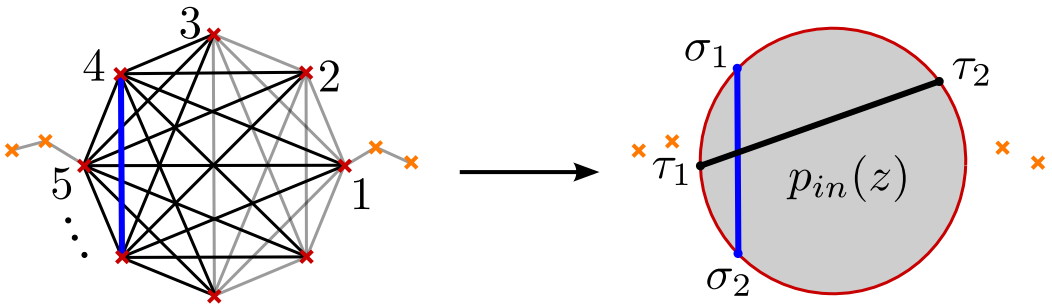


Figure 2.9: On the left we show the discrete version of the equations in the maximal cell for $N = 12$. Note that we have already dropped the couplings to the Y-functions related to the yellow zeros since they vanish in the continuum limit. The s_a and t_a label red zeros which accumulate along the unit circle. In the continuum limit these become σ_a and τ_a which parameterize the unit circle as shown on the right. On the left, the black lines indicate Y-functions with non-zero coupling to the blue Y-function while the grey lines indicate those with no coupling to this Y-function. Recall that two Y-functions have a nonzero intersection only if their lines touch $|\langle \mathbf{s}, \mathbf{t}\rangle| = 1$, or if they cross $|\langle \mathbf{s}, \mathbf{t}\rangle| = 2$. This intersection matrix results in the $\boldsymbol{\tau}$ integration limits in (2.34) as well as the "missing" $1/2$ in the continuum kernel (for a given blue line there are only $\mathcal{O}(N)$ Y-functions with $|\langle \mathbf{s}, \mathbf{t}\rangle| = 1$ and thus these do not contribute in the continuum limit).

t_1, t_2. Plugging the scaling (2.33) into (2.21) and and replacing the sum with integration we obtain

$$\log y_{\boldsymbol{\sigma}}(\theta) + \log \epsilon^2 = -4|z_{\boldsymbol{\sigma}}| \cosh\theta - \int_{\sigma_2+\epsilon}^{\sigma_1+2\pi-\epsilon} d\tau_2 \int_{\sigma_1+\epsilon}^{\sigma_2-\epsilon} d\tau_1 \int_{-\infty}^{+\infty} d\theta' \, \frac{y_{\boldsymbol{\tau}}(\theta')}{i\pi \sinh(\theta - \theta' + i\varphi_{\boldsymbol{\sigma\tau}})} \tag{2.34}$$

where now σ and τ are continuous variables that live in $[0, 2\pi)$, $\epsilon = 1/N$ and in accord with our usual notation $z_{\boldsymbol{\sigma}} = z_{\sigma_1\sigma_2}$ with[8]

$$z_{\sigma_1\sigma_2} = \int_{e^{i\sigma_1}}^{e^{i\sigma_2}} dz \sqrt{p_{\text{in}}(z)} \tag{2.35}$$

The limits of integration of τ_1 and τ_2 are the continuum version of the intersection matrix

[8]The source term in (2.34) has acquired an additional factor of 2 since $z_{\boldsymbol{\sigma}}$ is defined to be only half of the cycle integral. That is $Z_{\mathbf{s}} \to 2z_{\boldsymbol{\sigma}}$ in the continuum limit.

$|\langle s_1 s_2; t_1 t_2\rangle|$ as shown in figure 2.9.[9] The explicit log divergence that appears on the left hand side of the TBA equations comes from the region of integration where the points τ_1, τ_2 pinch down on σ_1, σ_2. Because of the local nature of the singularity, it can be regulated by a solution of (2.34) corresponding to a different z_σ. In particular we can consider the limit $|z_{\boldsymbol{\sigma}}| \to 0$ which is $R \to 0$ in figure 2.8. Then a solution is given by any θ-independent function of the form

$$y^{\text{circ}}_{\boldsymbol{\sigma}} \equiv \frac{\partial f_{\sigma_1} \partial f_{\sigma_2}}{(f_{\sigma_1} - f_{\sigma_2})^2} \tag{2.36}$$

where f_σ is any smooth, monotonic curve running from $-\infty$ to $+\infty$. Now we can write a regulated form of (2.34) by subtracting from it the equation satisfied by (2.36) to obtain

$$\log \frac{y_{\boldsymbol{\sigma}}(\theta)}{y^{\text{circ}}_{\boldsymbol{\sigma}}} = -4|z_{\boldsymbol{\sigma}}| \cosh\theta - \int_{\sigma_2}^{\sigma_1+2\pi} d\tau_2 \int_{\sigma_1}^{\sigma_2} d\tau_1 \int_{-\infty}^{+\infty} d\theta' \, \frac{y_{\boldsymbol{\tau}}(\theta') - y^{\text{circ}}_{\boldsymbol{\tau}}}{i\pi \sinh(\theta - \theta' + i\varphi_{\boldsymbol{\sigma\tau}})} \tag{2.37}$$

What is important is that because of its bilocal form (2.36) automatically has the proper short-distance singularity in σ such that the integration in (2.34) is regulated in the pinch region. Additionally, it has the correct transformation properties such that (2.37) is parameterization invariant. Of course regulating with a solution of the $R \to 0$ problem is not the only possible choice. We shall see in later sections that we can find much less trivial solutions to (2.34) and that for some purposes (such as numerical integration of (2.34)) these solutions are preferable as regulators.

Recall that the TBA equations of AMSV (2.2) imply the the Y-system (2.4). In the same way our continuum TBA equations (2.37) imply the continuum y-system

$$\log \hat{y}_{\boldsymbol{\sigma}}(\theta + i\pi) + \log \hat{y}_{\boldsymbol{\sigma}}(\theta) = -2\log\epsilon^2 - 2\int_{\sigma_2+\epsilon}^{\sigma_1+2\pi-\epsilon} d\tau_2 \int_{\sigma_1+\epsilon}^{\sigma_2-\epsilon} d\tau_1 \; \hat{y}_{\boldsymbol{\tau}}(\theta) \tag{2.38}$$

where we have defined $\hat{y}_{\boldsymbol{\sigma}}(\theta + i\varphi_{\boldsymbol{\sigma}}) = y_{\boldsymbol{\sigma}}(\theta)$.[10] To see the full implication of this equation, we note that the (2.37) further implies that the y-functions are $i\pi$-periodic in θ.[11] This allows us to drop the shift by $i\pi$ on the left hand side of the y-system. The resulting equation is a constraint on the $\boldsymbol{\sigma}$-structure of the y-functions. It requires them to have the

[9] The factor of 2 that cancels the 2 in $1/2\pi i$ also comes from the intersection matrix as explained in the caption of figure 2.9.

[10] Note that we could eliminate the ϵ's in (2.38) in the same way as we did in (2.34) – by dividing through by the y-system for y^{circ} which is identical to (2.38) but without the θ-dependence.

[11] To see this, one needs to shift by $i\pi$ in (2.37) making sure to pick all of the residues on the righthand side. The end result is that the equation for $y_{\boldsymbol{\sigma}}(\theta + i\pi)$ is identical to that for $y_{\boldsymbol{\sigma}}(\theta)$. Since the integral equation uniquely determines the y-functions, we conclude that $y_{\boldsymbol{\sigma}}(\theta) = y_{\boldsymbol{\sigma}}(\theta + i\pi)$.

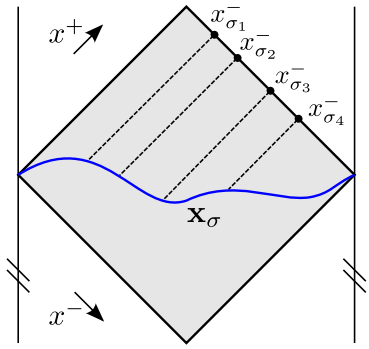


Figure 2.10: Boundary curve $\mathbf{x}_\sigma$ and cross ratios. To go back to global AdS one simply imagines wrapping the Poincaré patch (in grey) on the cylinder. We indicate the lightcone directions $x^\pm$ and show four points along the x^- direction. From four such points one can form the cross ratios $X^-_{\sigma_1\sigma_2\sigma_3\sigma_4}$. The regularized area depends only on the conformal invariant quantities $X^\pm_{\sigma_1\sigma_2\sigma_3\sigma_4}$. Here we will exclusively work in terms of the bi-local cross ratios formed by taking the limit $\sigma_2 \to \sigma_1$, $\sigma_4 \to \sigma_3$. These bi-local cross ratios are an (over)complete basis of the full set of cross ratios.

form of a bi-local cross-ratio

$$\hat{y}_{\boldsymbol{\sigma}}(\theta) = \frac{\partial \hat{x}_{\sigma_1} \partial \hat{x}_{\sigma_2}}{(\hat{x}_{\sigma_1} - \hat{x}_{\sigma_2})^2} \tag{2.39}$$

for some function $\hat{x}_\sigma = \hat{x}_\sigma(\theta)$ and we use the notation $\partial f_\sigma = \partial_s f_\sigma$. This is the general solution to (2.38) once the shift in θ is dropped.[12] This same structure can be deduced by studying the continuum limit of the Y-functions defined in terms of special solutions of the linear problem associated with the Pohlmeyer-reduced equations of motion. The important point is that at the physical values $\theta = 0, i\pi$ we can identify $\hat{x}_s$ with the physical boundary curve x_σ so that we have

$$\hat{y}_{\boldsymbol{\sigma}}(0) = \frac{\partial x^+_{\sigma_1} \partial x^+_{\sigma_2}}{(x^+_{\sigma_1} - x^+_{\sigma_2})^2}, \qquad \hat{y}_{\boldsymbol{\sigma}}(i\pi/2) = \frac{\partial x^-_{\sigma_1} \partial x^-_{\sigma_2}}{(x^-_{\sigma_1} - x^-_{\sigma_2})^2} \tag{2.40}$$

which allows us to reconstruct the (continuous) cross ratios of the (continuous) boundary curve once we have solved for the y-functions. These bi-local cross ratios are related to the usual definition of a cross ratio (involving four points) as shown in figure 2.10. Equations (2.40) are of course the continuum analog of the relations (2.6).

We call (2.37) the Continuum Thermodynamic Bethe Ansatz (CTBA) equations. They are the continuum analog of the TBA equations derived in [19] for null polygons and the

[12] Once we drop the shift in θ the y-system (2.38) is an integral form of the liouville equation.

main result of this section. In the next section we will derive the continuum analog of the area formula. Finally, in section 2.2.5 we will consider an exact solution of these equations.

2.2.3 Continuum area formula

The goal of this section is to write the continuum analog of the area formula (2.5). In the case of the TBA equations (2.2) it turned out to be possible to easily take the continuum limit starting with null polygons, once the proper analytic continuation was identified. Once this analytic continuation has been performed, the area formula (2.5) also has a nice continuum limit, although with an additional subtlety due to regulation. This issue arises because the nature of the divergence in the area is very different in the case of null polygons and smooth curves. In particular, it is not clear apriori how/if the cusp divergences of the former become the arc-length divergence of the later. To address this issue we must momentarily return to the flat connection and recall some aspects of Pohlmeyer reduction that are relevant for the computation and regulation of the area. Once the issue of regulation is dealt with, the continuum area emerges naturally from the continuum limit of the AMSV result.

In the Pohlmeyer reduced formalism we define the $SO(2,2)$ invariant combination $\alpha(z,\bar{z})$ in terms of the worldsheet metric as

$$2e^{2\alpha} = \partial X \cdot \bar{\partial} X \tag{2.41}$$

where $X(z,\bar{z})$ are the AdS_3 embedding coordinates. The full area (including cusp and/or arclength divergences) is then given as

$$A_{\text{full}} = 4\int dz d\bar{z} e^{2\alpha} \tag{2.42}$$

It is possible to show (after some calculation) that the equations of motion for the embedding coordinates and the virasoro constraints imply that α obeys the equation

$$\partial\bar{\partial}\alpha = e^{2\alpha} - p\bar{p}e^{-2\alpha} \tag{2.43}$$

Generically p can be any holomorphic function. For real surfaces $\bar{p}$ is the complex conjugate of p. In the null polygon case p is the holomorphic polynomial discussed in the introduction and the cusp divergences in A_{full} come from the large z region where $\alpha \sim 1/2 \log \sqrt{p\bar{p}}$. For this reason [19] (following [46]) separate the area as

$$A_{\text{full}} = 4\int dz d\bar{z}\sqrt{p\bar{p}} + 4\int dz d\bar{z}(e^{2\alpha} - \sqrt{p\bar{p}}) \tag{2.44}$$

The first term contains the cusp divergences as well as various finite contributions (see [46] for a detailed discussion). The second term is now manifestly finite and is given by [19]

$$\int dzd\bar{z}(e^{2\alpha} - \sqrt{p\bar{p}}) = \frac{1}{2\pi}\sum_{s_2>s_1}\int d\theta |Z_{s_2s_1}|\, e^{-\theta}\, \log(1 + Y_{s_2s_1}) \tag{2.45}$$

We would now like to consider the continuum limit of (2.44).

In the the continuum limit discussed in section 2.2.1 we replace p with the residual function p_{in} and the domain reduces to the unit disk. The boundary conditions on α are that the solution should look locally like the straight line solution at any point near the boundary. The strait line solution corresponds to $p_{\text{in}} \to 0$ and thus satisfies the equation

$$\partial\bar{\partial}\alpha^{\text{circ}} = e^{2\alpha^{\text{circ}}} \tag{2.46}$$

The solution obeying the proper boundary conditions for the strait line is then given by

$$e^{2\alpha^{\text{circ}}(z,\bar{z})} = (1 - z\bar{z})^{-2} \tag{2.47}$$

We use the notation α^{circ} to denote the strait line solution since it is conformally equivalent to the circular solution. Thus we see that the second term in (2.44) develops a new divergence, which is just the arc-length divergence of the continuous wilson loop. We are interested in computing the finite part of the area that remains after removing this divergence. To this end we write (2.42) as

$$A_{\text{full}} = A_{\text{div}} + A_{\Sigma} + A_{\text{free}} \tag{2.48}$$

where

$$\begin{aligned} A_{\text{div}} &= 4\int_D dzd\bar{z}e^{2\alpha^{\text{circ}}} \\ A_{\Sigma} &= 4\int_D dzd\bar{z}|p_{\text{in}}| \\ A_{\text{free}} &= 4\int_D dzd\bar{z}(e^{2\alpha} - |p_{\text{in}}| - e^{2\alpha^{\text{circ}}}) \end{aligned} \tag{2.49}$$

In writing (2.48) we have added and subtracted the boundary behavior of the integrand $e^{2\alpha} \sim e^{2\alpha^{\text{circ}}}$. We have also discarded all contributions outside of the unit disk D which is like cutting the surface where it approaches the boundary along the AdS boundary curve $\mathbf{x}_\sigma$. Note that in doing this, we have discarded the cusp divergences stemming from the large z integration of the first term in (2.44).

Let us now study (2.48) term by term. The contribution $A_{\rm div}$ can be evaluated by adapting appendix B of [34] to obtain[13]

$$A_{\rm div} = \frac{L}{\mathcal{E}} - 2\pi \tag{2.51}$$

where L is the arc-length of the boundary contour and $\mathcal{E}$ is the cutoff distance from the boundary. The term A_Σ is already written in terms of the continuum quantity $p_{\rm in}$. Thus we are left only with the last term in (2.48) which we must write in terms of y-functions. The (divergent) integral over $(e^{2\alpha} - |p_{\rm in}|)$ is given by the continuum limit of (2.45). Plugging the scaling (2.33) into the area formula (2.45) where now the sum is over all of the Y-functions of the maximal cell we obtain[14]

$$4\int_D dzd\bar{z}(e^{2\alpha} - |p_{\rm in}|) = \lim_{\epsilon\to 0}\frac{1}{2\pi}\int_0^{2\pi} d\sigma_2 \int_{0+\epsilon}^{\sigma_2-\epsilon} d\sigma_1 \int_{-\infty}^{+\infty} d\theta\, |z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, y_{\boldsymbol{\sigma}}(\theta) \tag{2.52}$$

As expected, this expression is divergent as $\epsilon \to 0$ due to the region of integration $\sigma_1 \sim \sigma_2$ where $y_{\boldsymbol{\sigma}} \sim (\sigma_1 - \sigma_2)^{-2}$. This divergence is just the manifestation of the arc-length divergence of the area, and we should regulated as explained above by subtracting the integral over $e^{2\alpha^{\rm circ}}$. This last integral is just (by definition) the $p_{\rm in} \to 0$ limit of (2.52). This limit is subtle and we must carefully consider the behaviour of the solutions of (2.37) when the sources become small. We provide a detailed analysis in appendix A. The end result is that the $p_{\rm in} \to 0$ limit of (2.52) is given by the same formula, but with the y-function replaced by its limiting form

$$y^{\rm kink}_{\boldsymbol{\sigma}}(\theta) = z^2_{\boldsymbol{\sigma}}(\partial_{\sigma_1}\partial_{\sigma_2}\log z_{\boldsymbol{\sigma}})\, e^{-2(\theta+i\varphi_{\boldsymbol{\sigma}})}\mathrm{csch}^{-2}(e^{-\theta}|z_{\boldsymbol{\sigma}}|) \tag{2.53}$$

Finally, $A_{\rm free}$ is the difference between (2.52) and the same formula with y replaced by $y^{\rm kink}$ giving the finite result

$$A_{\rm free} = \frac{1}{2\pi}\int_0^{2\pi} d\sigma_2 \int_0^{\sigma_2} d\sigma_1 \int_{-\infty}^{+\infty} d\theta\, |z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, (y_{\boldsymbol{\sigma}} - y^{\rm kink}_{\boldsymbol{\sigma}}) \tag{2.54}$$

[13] More specifically, we use the equations of motion for $\alpha^{\rm circ}$ to write

$$+4\int_D dzd\bar{z}e^{2\alpha^{\rm circ}} = +4\int_D dzd\bar{z}\partial\bar{\partial}\alpha^{\rm circ} = \frac{L}{\mathcal{E}} - 2\pi \tag{2.50}$$

The last equality is proven in appendix B of [34]. Note that we are using the fact that we can replace $\alpha^{\rm circ} \to \alpha$ in the boundary term.

[14] See the discussion at the end of section (2.1.2) for the area formula after wall crossing.

Physically the subtraction of y^{kink} corresponds to regulating by subtracting the area of a surface ending on a circular boundary contour whose circumference is equal to the arc-length L of the boundary contour $\mathbf{x}_\sigma$. Of course we should add back the same quantity, which is nothing other than A_{div} (2.51) in (2.49).

To summarize the results of this section we have

$$A_{\text{full}} = \frac{L}{\mathcal{E}} - 2\pi + A_\Sigma + A_{\text{free}} \tag{2.55}$$

where A_Σ is given explicitly in (2.49) and A_{free} is given by (2.54) in terms of the y-functions which are in turn obtained by solving the CTBA (2.37) and y^{kink} which is given explicitly by (2.53). Equation (2.55) is the continuum generalization of the AMSV area formula. This equation together with (2.37) are the main results of this chapter. They provide an integrability-based solution to the problem of computing minimal surfaces in AdS_3 ending on smooth boundary contours.

2.2.4 Numerical check

Given the length and technicality of the derivation of the CTBA system (2.37), (2.55), (2.54) we find it useful to pause momentarily and perform a check that makes contact with our usual notion of minimal area computations. We consider the special (but highly non-trivial) case $p_{\text{in}} = R$. As we shall see below, although p_{in} has rotation symmetry, the corresponding worldsheet curve generically does not.

Our strategy in this section is to numerically compute the minimal area in two ways. First, we solve the CTBA by iteration and evaluate the area using (2.55) and (2.54). We will describe how to implement this iterative procedure in section 2.3 (for general p_{in}). Second, we perform a direct numerical integration of the Pohlmeyer reduced string equations of motion. These later numerics are quite simple for the case at hand so we describe them in a bit more detail. Due to the rotational symmetry the equation of motion (2.43) reduces to the ODE

$$1/4\left(\partial_r^2 + r^{-1}\partial_r\right)\alpha = e^{2\alpha} - R^2 e^{-2\alpha} \tag{2.56}$$

with the boundary condition

$$a \underset{r\to 1}{\sim} -\log(1-r^2) \tag{2.57}$$

This ODE can be solved by a number of methods (e.g. shooting point or relaxation). We solve it for a few values of R using a standard relaxation method and directly evaluate each part of the area according to (2.49). In table 2.1 we compare the results of these numerics with the results of the CTBA numerics. We see that the results obtained from the CTBA agree well with the results of the direct computation of the minimal area!

R	CTBA	Relaxation
1/2	-5.292	-5.293
1	-2.781	-2.782
2	4.431	4.429

Table 2.1: Comparison of the area obtained by numerically integrating the CTBA and the area obtained from solving (2.56) by relaxation.

2.2.5 The Mathieu solution

In the previous section we performed a numerical test of the CTBA equations for the special case $p_{\text{in}} = R$. In this section we observe that, quite surprisingly, the CTBA can actually be solved exactly in this case! The exact y-function is given by (2.39) with

$$\hat{x}_\sigma(\theta) = \frac{Mc(i\theta + \sigma)}{Ms(i\theta + \sigma)} \tag{2.58}$$

where Mc and Ms are Mathieu cos and sin functions.[15] We can extract the area of the minimal area surface ending on this family of boundary curves from the exact y-function and formula (2.55). The result is

$$A(R) = -2\pi + 2\pi\,(1/4 - a(R)) \tag{2.59}$$

where $a(R)$ is the Mathieu Characteristic (see footnote 15).

It is interesting to consider the boundary curve of this solution. Although p_{in} has rotational symmetry, the spacetime surface enjoys a rotational symmetry only when $R \to 0$ which corresponds to the circle. On the other hand, as $R \to \infty$ the boundary curve approaches a null square. For intermediate values of R the boundary curve is some non-trivial closed curve on the cylinder which interpolates between these limiting cases as shown in figure 2.11. Note that while the finite R curve is specific to this example, its limiting behaviour is a demonstration of a generic feature of all solutions. Indeed, for any p_{in} (2.32)

[15] In Mathematica we have the following implementation:

$$\begin{aligned} Mc(z) &= \text{MathieuC}[a(R), R, z] \\ Ms(z) &= \text{MathieuS}[a(R), R, z] \\ a(R) &= \text{MathieuCharacteristicA}[1/2, R] \end{aligned}$$

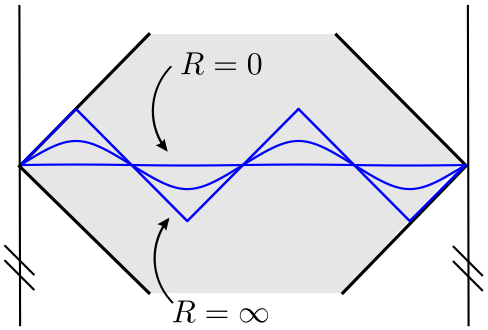


Figure 2.11: Boundary curves for the Mathieu solution for $R = 0, 3/4, \infty$ plotted in the Poincaré Patch. For this plot we use the convenient conformal frame $x^+_\sigma = x^{\theta=0}_\sigma$, $x^-_\sigma = (-1+x^{\theta=i\pi/2}_\sigma)/(1+x^{\theta=i\pi/2}_\sigma)$ which differs from the one in the main text by a simple conformal transformation.

with no zeros inside the unit circle the boundary curve approaches the circle or null square the limits $R \to 0$ and $R \to \infty$ respectively.[16]

Unfortunately we are not able to obtain this solution directly by solving the CTBA which seems rather complicated. Rather we engineered the solution by a different method that we describe below and then numerically verify that it satisfies the CTBA. For example, one can directly substitute (2.58) into the CTBA and verify using numerical integration that it is indeed a solution.[17] A stronger check is to solve the CTBA numerically by iteration. We will describe how to implement such a procedure in section 2.3. In figure 2.12 we plot the exact y-function against the results of these numerics and in figure (2.13) we compare the numerically computed area with the exact area (2.59). One can see that the numerical y-functions and the resulting area converge to the exact results.

Finally, let us now comment on how the solution (2.58) was obtained. Underlying the integrability construction described above is the linear problem associated with the string equations of motion in AdS_3. It turns out that for the case $p_{\text{in}} = R$, the corresponding linear problem is equivalent to one that was recently studied [55] in the context of the Sinh-Gordon model. Remarkably, in [55] the exact wronskian (Q-function) of that linear problem was constructed. Starting from this Q-function we were able to construct the solution (2.58).

[16] Actually, $R \to 0$ will always correspond to the circle even if p_{in} has zeros inside the unit circle. For $R \to \infty$ the boundary curve will approach a $(4+2n)$-gon for p_{in} with n zeros inside of the unit disk.

[17] This is actually somewhat difficult due to the complicated analytic structure of the integrand and the difficulty of computing the Mathieu functions throughout the necessary range of variables. We will comment more on the former point in section 2.3.

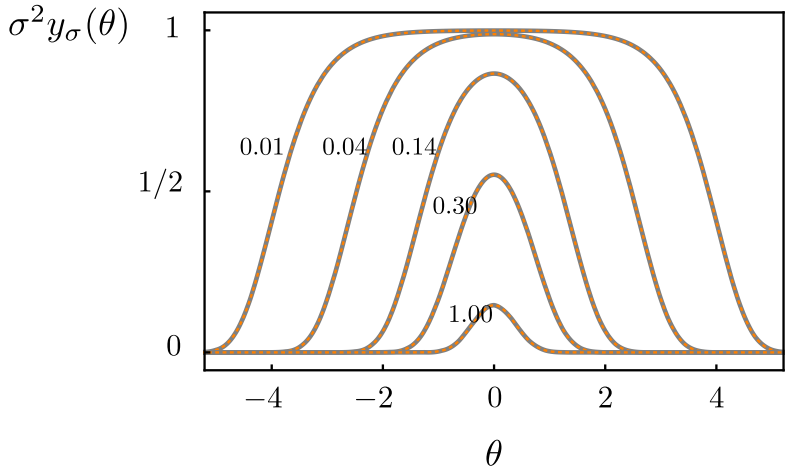


Figure 2.12: Comparison of the y-function constructed from (2.58) with the results of solving the CTBA with $z_\sigma = R(e^{i\sigma_2} - e^{i\sigma_1})$. Here we plot the results for $R = 1$. Due to the rotational symmetry of p_{in} the y-functions depend on σ_1 and σ_2 only through the combination $\sigma \equiv \sigma_2 - \sigma_1$. We plot $\sigma^2 y_\sigma(\theta)$ for $\sigma = 1.00, 0.30, 0.14, 0.04, 0.01$. The thick grey curves are the results of the numerics while the dashed orange curves are the exact result. Note the formation of plateaus for small values of σ and that the plateau height diverges as $1/\sigma^2$, all in accord with the discussion of section 2.2.3.

Here we only compute the area and the boundary contour, however perhaps it is possible to actually construct the full embedding surface. An extremely interesting generalization of this solution for the case when when $p_{\text{in}} = R\, z^n$ was recently proposed in [36].

2.3 CTBA numerics

In this section we turn to a numerical study of the CTBA (2.37). The main purpose of this section is to develop the proper numerical techniques to integrate these equations. This allows us to demonstrate that these equations are more than a formal curiosity, but are also a practical tool for computing minimal areas. The second goal of this section is to perform a final check of the CTBA as well as the numerical recipe that we present below. For this we directly numerically integrate the (Pohlmeyer reduced) string equations of motion. The results obtained from the two approaches agree well within the expected numerical error from each side.

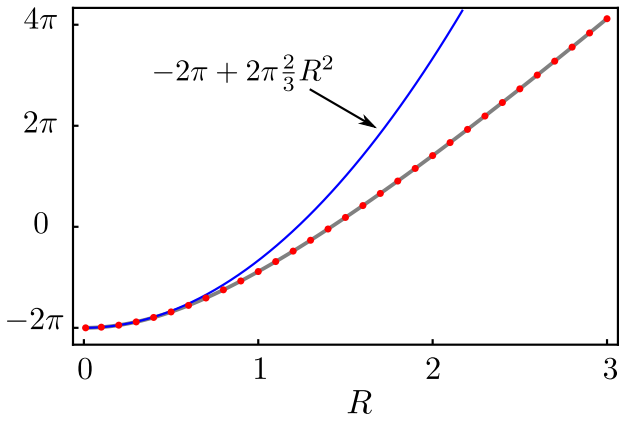


Figure 2.13: Area as a function of R. The gray curve is $A(R)$ computed from (2.59). The red points are the results of the relaxation numerics applied to the problem 2.56. Note that $A(R) \geq A^{\text{circ}} = -2\pi$ with the equality holding only at $R = 0$. The blue curve is the small R expansion of $A(R)$. One can check that this expansion is in precise agreement with the wavy line approximation [54]. The large R expansion (not shown here) is also consistent with the approach to the 4 cusp solution as it develops divergences in R which one can interpret as cusp divergences.

2.3.1 CTBA

For the purposes of this section it is useful to work in a new worldsheet coordinate $w(z)$ defined via

$$w(z) = \int_0^z dz \sqrt{p_{\text{in}}(z)}, \qquad p_{\text{in}}(z) = \left(\frac{dw(z)}{dz}\right)^2 \tag{2.60}$$

Because $p_{\text{in}}(z)$ has no zeros inside the unit disk this domain maps to a simply connected region of the w-plane with boundary $w_\sigma = w(e^{i\sigma})$. The primary utility of this set of coordinates is that the integrals $z_{\sigma_1\sigma_2} = w_{\sigma_2} - w_{\sigma_1}$ are just the (complex) "distances" between points on the curve w_σ. Thus, in this set of coordinates p_{in} is eliminated and the shape of the AdS boundary contour x_σ is encoded in the complex curve w_σ. In analogy with our previous notation we use the shorthand $w_{\boldsymbol{\sigma}} = w_{\sigma_2} - w_{\sigma_1}$.

The first order of business is to find a suitably regularized function for which to solve. There are two types of singularities in the region $\sigma_2 \sim \sigma_1$ which make $y_{\boldsymbol{\sigma}}(\theta)$ an unsuitable function to use for numerics. In the limit $\sigma_2 \to \sigma_1$ with fixed θ the y-function has the

expansion

$$y_{\boldsymbol{\sigma}} \sim \frac{1}{(\sigma_1-\sigma_2)^2} - \frac{1}{3}|\partial_{\sigma_1} w_{\sigma_1}|^2 e^{-2\theta} - \frac{1}{3}|\partial_{\sigma_1} w_{\sigma_1}|^2 e^{+2\theta} + \ldots \tag{2.61}$$

where the +... represents terms finite as $\sigma_2 \to \sigma_1$ and $\theta \to \pm\infty$. The first term presents one type of singularity which is simple to treat as it has only to do with short distances in σ. One can remove it by forming combinations like $y_{\boldsymbol{\sigma}} - y^{\rm circ}_{\boldsymbol{\sigma}}$. The second, more difficult, type of singularity is due to the θ-dependent terms in (2.61) which reflect a subtle order of limits that occurs at small separation in σ and large values of θ. From (2.61) we see that if we take $\sigma_2 \to \sigma_1$ and then take θ large then the y-function diverges exponentially in θ. On the other hand, if we first send $\theta \to \pm\infty$ and then take $\sigma_2 \to \sigma_1$ in the y-function, the result will be zero. To see this, first note that the expansion (2.61) is not valid in this limit. For finite separation in σ the dominant large θ behavior is given by dropping the kernel term in (2.37). From this it is clear that the y-function will go to zero double exponentially at large theta for any nonzero separation in σ. In other words, there is a sort of "boundary layer" at $\sigma_2 - \sigma_1 = 0$ whose height diverges exponentially in θ. This divergent boundary layer is quite toxic for the numerics, but fortunately it can be regulated easily. The key point is that the kink y-function captures this behavior exactly. Thus a fully regulated function is given by

$$y^{\rm reg} = y - y^{\rm kink} - y^{\rm akink} + y^{\rm circ} \tag{2.62}$$

where $y^{\rm akink}$ is given by

$$y^{\rm akink}(\theta) = y^{\rm kink}(-\theta)^* \tag{2.63}$$

Note that $y^{\rm kink}$ regulates the boundary layer at $\theta \to -\infty$, $y^{\rm akink}$ regulates the boundary layer at $\theta \to +\infty$ and the double poles in (2.61) at $\sigma_2 \sim \sigma_1$ cancel between the terms on the right hand side of (2.62). Thus $y^{\rm reg}$ is a good function for numerics.

The integral equation obeyed by $y^{\rm reg}$ can be obtained by recalling that $y^{\rm kink}$ obeys CTBA equation (2.37) but with $\cosh\theta \to 1/2e^{-\theta}$ as explained in appendix A. Similarly, $y^{\rm akink}$ obeys the CTBA but with $\cosh\theta \to 1/2e^{+\theta}$. Finally, $y^{\rm circ}$ obeys the CTBA with $|z_{\boldsymbol{\sigma}}| \to 0$ in the source. Putting all of this together yields the $y^{\rm reg}$ equation.

$$y^{\rm reg} = -y^{\rm kink} - y^{\rm akink} + y^{\rm circ} + \frac{y^{\rm kink} y^{\rm akink}}{y^{\rm circ}} e^{-K\star y^{\rm reg}} \tag{2.64}$$

where $K \star f$ represents the action of the CTBA kernel on the function f.

Now that we have the equation (2.64) suitable for numerics, we will describe our numerical method which is based on the usual iteration scheme used for TBA equations. That is, we start with some initial guess for the y-function ($y^{\rm reg}_{(0)} = 0$, for example), plug it into

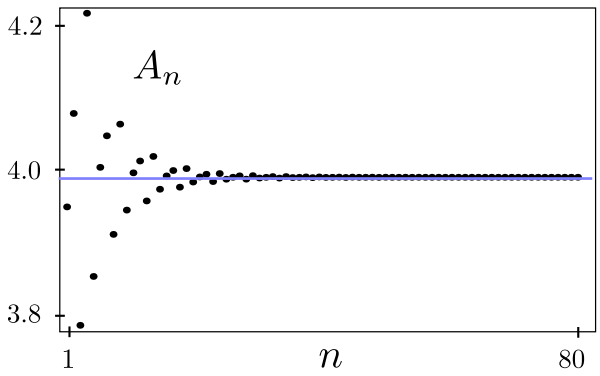


Figure 2.14: Convergence of CTBA. Area A_n computed from (2.55) from the nth iteration of the CTBA starting from the initial iterate $y^{\text{reg}}_{(0)} = 0$. We use the curve $w_\sigma = e^{i\sigma} + 1/5\, e^{2i\sigma}$. We used a 10×30 grid for (α_1, α_2). More points are required in the α_2 direction due to some remaining non-analytic behavior even after the regulation (2.62) (i.e. there is a remaining boundary layer of finite height due to the +... in (2.61)). The theta variable is cutoff at ± 10 and contains 32 grid points. The CTBA converges to $A^{\text{CTBA}} = 3.990....$ The blue line is at $A^{\text{EOM}} = 3.989...$ and indicates the value obtained by direct numerical integration of the Pohlmeyer-reduced string equations of motion which we explain in section 2.3.2.

the RHS of the CTBA, and integrate to produce an updated y-function $y^{\text{reg}}_{(1)}$ which we plug back into the CTBA to produce $y^{\text{reg}}_{(2)}$ and so on. We repeat the process until it converges. At first sight this seems painfully slow. At each iteration one must perform a triple integration for each point in a suitable $\{\theta, \sigma_1, \sigma_2\}$ grid. This grid typically contains around 10^4 points and direct numerical implementation of the integration (using `Mathematica`'s `NIntegrate`, for example) takes several seconds for each point in this grid. The end result is that each iteration takes a few hours. Given that a few hundred iterations are needed for good convergence, this approach is clearly too slow to be practical or useful. This difficulty can be circumvented with the use of Fourier methods which allow one to convert the integrations into matrix multiplication such that each iteration (i.e. evaluation of the entire $\{\theta, \sigma_1, \sigma_2\}$ grid) can be performed in under a second even in an un-parallelized code. We will now explain this in greater detail.

First consider the θ'-integration. Using convolution theorem we can write the integrations in (2.37) as

$$\int_{\sigma_2}^{\sigma_1+2\pi} d\tau_2 \int_{\sigma_1}^{\sigma_2} d\tau_1 \; \mathcal{F}_\theta^{-1}\left[e^{\delta_{\sigma\tau}\,\omega}\mathcal{F}_\omega\left[(\pi\cosh\theta')^{-1}\right]\mathcal{F}_\omega\left[y^{\text{reg}}_{\boldsymbol{\tau}}(\theta')\right]\right] \tag{2.65}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are forward and reverse Fourier transforms, ω is the Fourier variable conjugate to θ, and[18]

$$\delta_{\boldsymbol{\sigma\tau}} = (\pi/2 + \varphi_{\boldsymbol{\sigma}} - \varphi_{\boldsymbol{\tau}}) \tag{2.66}$$

Now consider the τ integration. The τ-variables are compact and thus we can expand $y^{\text{reg}}_{\boldsymbol{\tau}}(\theta')$ in modes as

$$\sum_{a_1=-\infty}^{\infty} \sum_{a_2=0}^{\infty} e^{i\pi a_1 \alpha_1} \cos(\pi a_2 \alpha_2)\, u_{a_1 a_2}(\theta') \tag{2.67}$$

where we have introduced the useful shorthand

$$\alpha_1 = \frac{-2\pi + \tau_2 + \tau_1}{2\pi}, \qquad \alpha_2 = \frac{\tau_2 - \tau_1}{2\pi} \tag{2.68}$$

and made use of the $\tau_2 \leftrightarrow \tau_1$ symmetry of y^{reg}. Plugging this expansion into (2.65) gives

$$\mathcal{F}_\theta^{-1}\left[\mathcal{C}^{a_1 a_2}_{\sigma_1\sigma_2}(\omega)\mathcal{F}_\omega\left[(\pi\cosh\theta')^{-1}\right]\mathcal{F}_\omega\left[u_{a_1 a_2}(\theta')\right]\right] \tag{2.69}$$

where

$$\mathcal{C}^{a_1 a_2}_{\sigma_1\sigma_2}(\omega) = \int_{\sigma_2}^{\sigma_1+2\pi} d\tau_2 \int_{\sigma_1}^{\sigma_2} d\tau_1\, e^{i\pi a_1\alpha_1} \cos(\pi a_2\alpha_2) e^{\delta_{\boldsymbol{\sigma\tau}}\omega} \tag{2.70}$$

The mode transfer matrix $\mathcal{C}$ is a fixed object: it is computed once and for all for a given w_σ and then is an input into the numerical algorithm. Once this transfer matrix is computed, the RHS of (2.69) gives an extremely numerically efficient representation of the integrations in (2.37). Even with a $\{\theta, \sigma_1, \sigma_2\}$ grid containing on the order of 10^4 points, each iteration of the CTBA can be performed in under one second on an ordinary modern computer with only a single core. With an appropriate damping scheme the method typically converges in around 100 or so iterations as shown in figure 2.14.

2.3.2 String Equations of Motion

Let us now turn to an alternative method for computing minimal areas, which is a brute force attack on the string equations of motion, or some reduced variant of them. This will allow us to perform a final check of the CTBA and the numerical recipe described above. Since we are only interested in the area (i.e. and not the full string embedding) it is useful to work only with $SO(2,2)$ scalars formed from the $\mathbb{R}^{2,2}$ embedding coordinates $X(z,\bar{z})$.

[18] Here one must take care to properly define the branches of the arg function appearing in $\varphi_{\boldsymbol{\sigma}} = \arg(w_{\boldsymbol{\sigma}})$. This subtlety does not arise in (2.37) due to the $2\pi i$ periodicity of the kernel.

In particular, it is useful to use work with the variable $\alpha(z,\bar{z})$ defined by (2.41) which appears in the on-shell string action (2.42). Working in terms of such scalar variables goes by the name of Pohlmeyer reduction [49–51], and is actually the starting point for the integrability-based method for computing minimal areas of surfaces with null polygonal boundaries [19, 46]. We will numerically integrate the sinh-Gordon equation (2.43) in the usual z-coordinate with $p \to p_{\text{in}} = dw(z)/dz$ and unit disk for the domain. This equation must be supplemented with boundary conditions that α approach the straight line solution α^{circ} given in (2.47) at the boundary.

For the purpose of numerics we must define a suitable function for which to solve. This is simple: a function which is regular everywhere in the unit disk and on its boundary is given by

$$\alpha^{\text{reg}} = \alpha - \alpha^{\text{circ}} \tag{2.71}$$

which obeys the boundary conditions $\alpha^{\text{reg}} \to 0$ at the boundary. Now we have a well defined numerical problem: solve (2.43) written in terms of α^{reg} subject to the boundary condition $\alpha^{\text{reg}} \to 0$ at the boundary of the unit disk and then evaluate (2.48)-(2.49). With (2.43) written in terms of the regular function α^{reg} we solve the resulting equation using a standard relaxation method which is suitable for the elliptic operator $\partial\bar{\partial}$. One could of course use faster integration schemes based on spectral methods, however we prefer relaxation for its simplicity and stability. The results of this procedure for an example case are shown in figure 2.15. In figure 2.14 we compare the results obtained from the numerical integration of the CTBA and the results obtained from integrating the Pohlmeyer reduced equations of motion. The numbers obtained from the two different methods agree within the expected error of the numerics on both sides.

As a final comment we note that, as in the case of the CTBA, the Pohlmeyer numerics is completely parameterized in terms of p_{in}. However, unlike the CTBA, we cannot directly recover the physical boundary curve $\mathbf{x}_\sigma$ after integration. In order to do that in this approach, one would need to further integrate the equations of motion for the embedding coordinates with (2.41) to obtain X near the boundary. Here we are only interested in checking the results of the CTBA and thus this inherent difficulty is of no consequence to us. Indeed, it clearly demonstrates the advantage of the integrability based approach over direct numerical methods.

2.4 Discussion

In this chapter we developed an integrability-based method for computing the area of minimal surfaces in AdS which end on smooth curves at the boundary. Our main result is

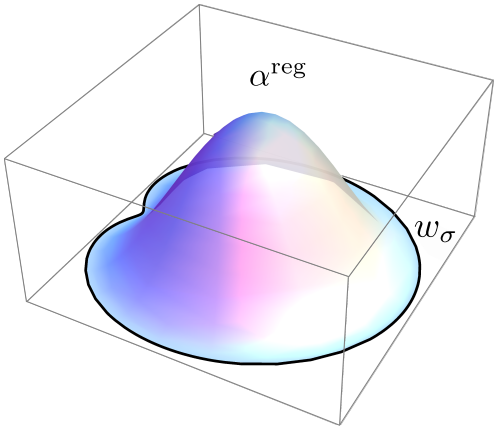


Figure 2.15: Here we plot $\alpha^{\text{reg}}(w, \bar{w})$ the regulated solution of (2.43) for the curve $w_\sigma = e^{i\sigma} + 2/5\, e^{2i\sigma}$ obtained from the relaxation numerics described in this section. In these coordinates the unit disk of the z-plane maps to a simply connected region bounded by the curve $w_\sigma = w(e^{i\sigma})$ where $w(z)$ is defined in (2.60).

the set of integral equations (2.37) and (2.55). These integral equations, which we dubbed the CTBA, provide a powerful tool for analytic study of minimal surfaces. The CTBA also provides a powerful tool for numerics and we developed an algorithm for numerically integrating these equations. It efficiently reproduces the results obtained from a brute force numerical integration of the string equations of motion as shown in figure 2.14.

Perhaps the most exciting aspect of this work is the myriad possibilities for future directions. First, it would be interesting to generalize the results here to minimal surfaces in the full AdS_5. It should be possible to do this by following the same steps used to derive the equations presented here, but starting from the AdS_5 version of the AMSV equations [19]. Another interesting direction would be to adapt the results described above for the case of *Euclidean* AdS. This would be of interest in the study of entaglement entropy and should also reveal a fascinating connection between (C)TBA and theta functions [33–37]. We have made much progress towards this end and will present the results elsewhere.

First and foremost, we see the results presented here as the first step in the study of smooth Wilson loops in $\mathcal{N} = 4$ at any value of the t'Hooft coupling. Although a daunting task, history has taught us that it is indeed possible as demonstrated by the exact solution of the spectrum problem as well as recent results from the OPE of null polygonal Wilson loops and 3-point functions. In these examples, the first step was to identify integrability in the extreme strong and weak coupling limits. In this paper we achieve the former. A natural next step is to study the weak coupling problem where there have been remarkable

advances in the study of null polygonal Wilson loops. It would be very interesting to study the continuum limit of these perturbative results and to try to identify some hidden integrability structure present at both weak and strong coupling.

Chapter 3

Correlators at strong coupling

We now turn our attention to the computation of correlators in $\mathcal{N} = 4$ at strong coupling. The computation of correlators is generally highly nontrivial and obtaining explicit expressions outside the perturbative regime is typically impossible. As in the previous chapter, our main tool will be the AdS/CFT correspondence [2–5] which due to (1.3) maps the problem at strong coupling to a problem of classical strings moving in a curved spacetime. In particular, the problem of computing the correlation function at strong coupling is that of finding the area of the minimal surface in $AdS_5 \times S^5$ that goes to the AdS boundary at the operator insertion points x_a.

In this chapter, we compute the AdS part of the correlation function for arbitrary heavy scalar operators inserted along a line. The method used here is inspired by the integrability techniques originally developed for the Null Polygonal Wilson loop problem [19, 46, 47] and later applied to the computation of three-point functions [38–40, 42–44]. As in these previous applications, integrability allows one to compute the minimal AdS action without knowing the explicit classical solution. For the four point correlation function the connection with Hitchin systems and the formalism developed in [52, 53] is used intensely. As in the previous chapter the starting point of the method is the map of the string equations of motion in AdS to a certain auxilliary linear problem by Pohlmeyer reduction. Ultimately the solution takes the form of a set of functional equations that we call a χ-*system*. These functional equations are similar in spirit to the Y-system that appeared in the solution of the null polygon problem [19] and which naturally arise in the solutions of integrable QFT's. While we have mostly focused on the 4-point computation, we note that the method used below could in principle be applied for the corresponding N-point computation.

For some specific BPS operators dual to strings spinning on the same great circle of S^5 the sphere contribution is well know. In this case we can construct the full strong-coupling correlation function. We emphasize that these 4-point functions are generically neither extremal nor protected. Complete, non-protected results for correlation functions of heavy operators at strong coupling are quite rare. For example, in [38], the AdS part of the three-point function is computed, but the only case for which the sphere is known (BPS operators) is protected.[1]

This chapter is based on the work [41]. The layout is as follows. In the section 3.1 we start by giving the general strategy of this work and discuss some physically relevant aspects of the problem. Then, in section 3.2 we write the AdS part of the correlation function in terms of objects which are naturally computed using the integrability of the string equations of motion. In section 3.3 we introduce a formalism that will lead to the χ-system which is the set of functional equations that allows one to compute the minimal AdS action. In section 3.4 we compute explicitly the AdS part of the correlation function using the χ-system and explain the mechanism by which the dependence on the spacetime points emerges. We also present some numerical tests of the results. In section 3.5.1, we compute the sphere part for the specific case of BPS operators with large charges. In section 3.5.2, we discuss the saddle points of the fourth insertion point of the correlation function. We then study the extremal limit of the correlation function which is an analytical test of the result. In section 3.6 we conclude and discuss some open problems.

3.1 Four point function generalities

For large 't Hooft coupling λ, the semi-classical computation of correlation functions corresponds to the evaluation of the AdS_5 and S^5 actions for classical solutions with the topology of a four punctured sphere. The boundary conditions are that the solution close to each puncture P_a, which is associated with the gauge theory operator $\mathcal{O}_a(x_a)$, approaches the AdS boundary at the point x_a in the same way as a 2-point function involving $\mathcal{O}_a(x_a)$ and some other heavy, scalar operator. In this paper, we study the simplest case where the operators are inserted on a line in $\mathbb{R}^4$. This implies that the string solution is contained in a Euclidean AdS_2 subspace of AdS_5. Moreover, there is only one independent cross-ratio. The conformal symmetry of $\mathcal{N}=4$ constrains the four-point correlation function to take

[1] Using the results of this paper it may be possible to extend the results of [**?**] to the complete N-point functions of GKP strings at strong coupling since the mathematical problem is similar to the one treated here.

the form

$$\langle \mathcal{O}_1(x_1)\mathcal{O}_2(x_2)\mathcal{O}_3(x_3)\mathcal{O}_4(x_4)\rangle = f(u)\prod_{a>b}^{4}(x_{ab})^{\Delta_{ab}}\,, \tag{3.1}$$

where $x_{ab} = x_a - x_b$, Δ_a is the dimension of operator $\mathcal{O}_a$, $\Delta_{ab} = \left(\sum_c \Delta_c\right)/3 - \Delta_a - \Delta_b$ and u is the conformal cross-ratio

$$u = \frac{x_{14}x_{23}}{x_{12}x_{34}} \tag{3.2}$$

Both the AdS and sphere contributions contain divergences as the string approaches the position of the operators at the boundary of AdS, which requires a cut-off $z = \mathcal{E}$. To describe the world-sheet we use complex variables $w, \bar{w}$. On the 4-punctured sphere, the physical cut-off $\mathcal{E}$ corresponds to cutting out small disks of radius ϵ_a around each puncture P_a at w_a. Ultimately, we will need to establish a precise relation between the cut-off's ϵ_a and $\mathcal{E}$. As we will review later, this is possible given the data accessible from integrability [56].

In this paper, we will consider operators with charges scaling as $\sqrt{\lambda}$, and without spin in AdS. Following the prescription developed in [38, 56], we account for the states in the sphere by introducing an extra contribution of wave-functions. Therefore, the semi-classical four-point function is given schematically by

$$\int dw_4\, e^{-\frac{\sqrt{\lambda}}{\pi}\int_{\Sigma\setminus\{\epsilon_a\}}\mathcal{L}_{AdS_2}} e^{-\frac{\sqrt{\lambda}}{\pi}\int_{\Sigma\setminus\{\epsilon_a\}}\mathcal{L}_{S^5}}\Psi_1\Psi_2\Psi_3\Psi_4 \tag{3.3}$$

where the actions are evaluated on a classical (Euclidean) string solution approaching the boundary of AdS at the positions of the insertion points x_a.

In principle, there is an integral over all four insertion-points on the worldsheet. In (3.3) we only integrate over the insertion w_4 since the position of the other punctures can be fixed by conformal transformations. Since we are considering the $\lambda \to \infty$ limit, one can evaluate the integral over w_4 by saddle point and the end result is the integrand of (3.3) evaluated at the dominant saddle point.

Let us consider the issue of the saddle point in some detail since it will be an interesting aspect of our computation. There are two issues here: the positions of the operators on the boundary and the positions of the insertion points on the sphere. We can use the target-space conformal symmetry to place three of the operators at $x_1 = 1$, $x_2 = \infty$, $x_3 = -1$ and the world-sheet conformal symmetry to fix $w_1 = 1$, $w_2 = \infty$, $w_3 = -1$.

The position x_4 is an input since we can put $\mathcal{O}_4$ anywhere along the line that contains $\mathcal{O}_{1,2,3}$. On the other hand, once we choose x_4 the position of the fourth puncture is fixed

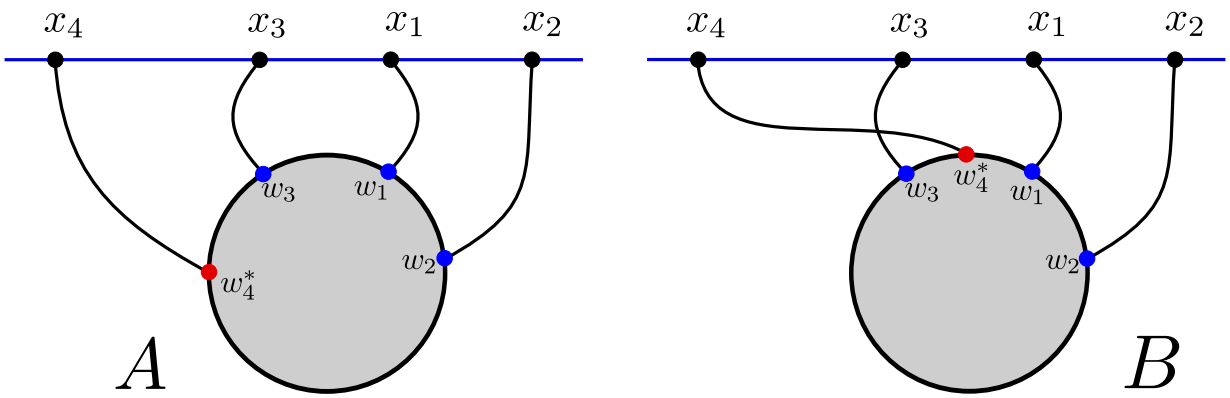


Figure 3.1: Insertions on the 4-punctured sphere. The gray ball represents the world-sheet (the complex plane plus the point at infinity, or simply 'the sphere') and the black boundary of the ball represents the equator of the sphere. The points w_a are the punctures on the sphere corresponding to the operators inserted at the positions x_a at the boundary of AdS_2, which is represented by the strait line. We fix the points w_1, w_2, w_3 and x_1, x_2, x_3 using the world-sheet and target-space conformal symmetry respectively. The position of the fourth insertion w_4 should be fixed at the dominant saddle point w_4^* of the integrand of (3.3). By symmetry we expect this saddle point to also be along the real axis, and thus we have a notion of an ordering of the 4 punctures. In particular, there is three distinct ranges for the location of w_4^*. Consider the ordering of the x_a shown in this figure. If the dominant saddle point is located between w_2 and w_3 (as in panel A) then the insertions will not cross and the string embedding will look schematically like the one shown in figure 3.2A. If the dominant saddle-point is located between w_3 and w_1 (as in panel B) or between w_1 and w_2 then the insertions cross each-other and we expect the string embedding to look like the one shown in figure 3.2B.

at $w_4 = w_4^*$ by the saddle-point condition. By symmetry we expect the dominant saddle-point to be located on the real axis and in this case we have a notion of an ordering of the punctures. In particular, there are three possible in-equivalent orderings depending on the position of w_4. Figure 3.1 shows two of these possibilities. If the ordering of the x_a is the same as the w_a then the insertions do not cross each other, as in figure 3.1A. If the ordering of the x_a is different from that of the w_a, then the insertions will cross as in figure 3.1B. These two possibilities lead to two types of string embeddings with distinctly different properties as is shown in figure 3.2. We will see that two types of solutions arise naturally in our construction. We are able to characterize the qualitative features of the

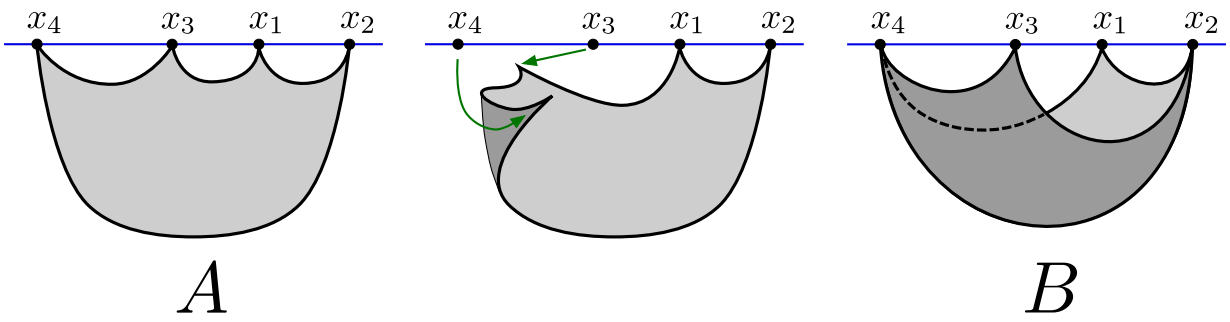


Figure 3.2: Two different possible string embeddings in AdS_2 which obey the required boundary conditions. These two solutions are shown in panels A and B. The center panel shows how to generate the configuration of panel B from that of panel A by interchanging the order in which the insertions on the sphere attach to the boundary; this interchange results in the characteristic folding shown in the embedding of panel B. These two types of solutions arise from the possibility that for a given choice of operator insertion points x_a the insertion point w_4^* (see figure 3.1) can be located in any of the three possible intervals (w_2, w_3), (w_3, w_1), (w_1, w_2).

spacetime embeddings and compute the minimal AdS action of both types of solutions. We will return to this topic below.

3.2 AdS_2 Pohlmeyer reduction

In this section we briefly review the Pohlmeyer-reduction process. We begin with a discussion of the string equations of motion and the stress-energy tensor, which is the starting point of the reduction. We then introduce the function γ in terms of which the AdS Lagrangian can be written. It turns out that γ satisfies a non-linear but *scalar* equation of motion that is a modified version of the well-know sinh-Gordon equation. Next we show how the different types of string embeddings discussed in section 3.1 are encoded though the boundary conditions imposed on γ. Finally we use the function γ to write the AdS action in a form where integrability is more readily applied.

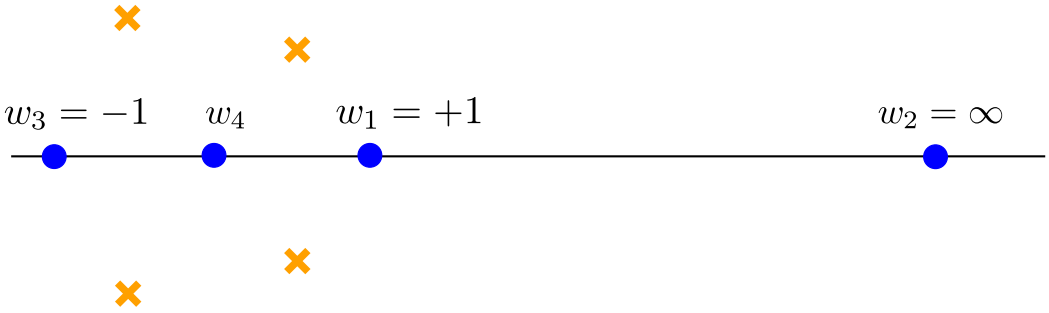


Figure 3.3: Schematic analytic structure of T. The blue dots represent the (double) poles of T at locations w_a and corresponding to the operator insertions $\mathcal{O}_a(x_a)$. The yellow crosses indicated zeros of T. We have fixed the positions of w_1, w_2 and w_3 using the world-sheet conformal symmetry. We have arbitrarily placed w_4 in the interval (w_3, w_1) although generically the saddle-point w_4^* can be located in any of the three intervals along the real axis.

3.2.1 Equations of motion and stress-energy tensor

Recall that we can consider (euclidean) AdS_2 as a surface in $R^{1,2}$ obeying the constraint

$$Y \cdot Y = (Y_1)^2 - (Y_2)^2 + (Y_3)^2 = -1. \tag{3.4}$$

We write the action for a string in AdS_2 as

$$S = \frac{1}{2} \int d^2\sigma \left[\partial_\alpha Y \cdot \partial^\alpha Y + \lambda \left(Y \cdot Y + 1 \right) \right] \tag{3.5}$$

and the resulting equations of motion as

$$\Box Y = \left(\partial Y \cdot \bar{\partial} Y \right) Y \tag{3.6}$$

The first term in the action is just the free string action in $R^{1,2}$; the second term is a Lagrange multiplier term that imposes (3.4).

The equations of motion (3.6) must be supplemented by the Virasoro constraints and boundary conditions. The Virasoro constraint requires $T_{AdS} + T_S = 0$. In particular, the AdS contribution to the stress-energy tensor does not vanish. Fortunately the boundary conditions allow us to completely fix the form of $T_{AdS} = -T_S$. Here we are interested in solutions with the topology of a four-punctured sphere where the punctures are at the position of the operator insertions and thus the boundary conditions give the behavior of

the string solutions near the insertion points. The correct prescription is to demand that the string goes to the boundary at the insertion points. Furthermore, it should approach the boundary in a specific way as dictated by the vertex operators. The behavior of the solution near the boundary will be dominated by the operator inserted there, independent of the properties or number of other operators inserted at different points. This means that the behavior near the insertion points can be determined from the 2-point function, where the string solution is know explicitly. From the explicit solution for the 2-point function one finds that the desired property of the solution near insertion point w_a is [38]

$$(\partial Y)^2 \equiv T(w) \sim \frac{\Delta_a^2}{4(w-w_a)^2} \qquad (w \to w_a) \tag{3.7}$$

where $T(w)$ is the holomorphic component of T_{AdS}. The corresponding property also is required for the anti-holomorphic component $\bar{T}(\bar{w})$. Thus we know that T should be an analytic function on the (4-punctured) Riemann sphere with double-pole singularities at the punctures. This fixes T to be a specific rational function.

First consider the denominator of the rational function T. The polynomial in the denominator is determined by the positions of the insertions. Three of the insertions can be fixed by conformal symmetry, leaving one final insertion. The integrand of (3.3) will be a function of this final insertion point. In the limit $\sqrt{\lambda} \to \infty$ the integral localizes at the saddle point $w_4 = w_4^*$, thus fixing completely the denominator of T.

Now consider the numerator of T. Without loss of generality we can consider the case where there is no insertion at infinity since we can perform a transformation that maps any arbitrary point to infinity. Then the polynomial in the numerator can be at most of degree 4 (otherwise T would not be regular at infinity) and therefore it is characterized by 5 parameters. Four of these parameters are fixed by the condition (3.7). The final unfixed parameter, which we will call U, parameterizes the single cross-ratio of the four operators (recall that four points in a line have only one independent cross-ratio). The precise map between the parameter U and the cross-ratio u is quite involved but fortunately we will not need it since the cross-ratio will be encoded in the χ-system in a simple way. The analytic structure of T is shown schematically in figure 3.3. We will use this sort of figure to represent T throughout this paper.

3.2.2 The function γ

Our objective is to evaluate the AdS part of the string action. In Poincaré coordinates the on-shell action becomes[2]

$$\partial Y \cdot \bar{\partial} Y = \frac{\partial x \bar{\partial} x + \partial z \bar{\partial} z}{z^2} = \sqrt{T\bar{T}} \cosh \gamma \tag{3.9}$$

where the above formula defines the function $\gamma(w, \bar{w})$. It follows from the equations of motion that γ satisfies the modified sinh-Gordon equation

$$\partial \bar{\partial} \gamma = \sqrt{T\bar{T}} \sinh \gamma. \tag{3.10}$$

It is well known that this equation is classically integrable, and in what follows we exploit this integrability to compute the AdS action.

Now let us determine what boundary conditions should be imposed on γ. For the 2-point function $\gamma = 0$. Recall that the string solution should approach that of the 2-point function as the string approaches the boundary at the operator insertion points x_a. Therefore we should require that $\gamma \to 0$ as $w \to w_a$ [38]. Furthermore, in order to have a non-singular world-sheet metric the right-hand side of (3.9) should never vanish. Thus when T has a zero γ must have a logarithmic singularity to cancel it. In summary, the boundary conditions on γ are

$$\gamma \to \pm \frac{1}{2} \log T\bar{T} \qquad (w \to z_a) \tag{3.11}$$

$$\gamma \to 0 \qquad (w \to w_a) \tag{3.12}$$

where z_a denotes a zero of T and w_a a pole of T. Notice that the regularity of the world-sheet metric does not fix the sign of the logarithmic 'spike' in (3.11) and, in principle, different choices are possible at each zero (recall that generically T will have 4 zeros for the 4-point function, as follows from the discussion of the previous section). These different choices correspond to different string solutions having differing properties, and generically different total action. We will refer to the spikes with the $+$ $(-)$ sign as u-spikes (d-spikes). We will now describe how the choice of these signs is related to the string embeddings shown in figure 3.2.

[2]The AdS_2 Poincaré coordinates are given by

$$Y^1 = -\frac{1}{2z}\left(1 - x^2 - z^2\right), \qquad Y^2 = \frac{1}{2z}\left(1 + x^2 + z^2\right), \qquad Y^3 = \frac{x}{z}. \tag{3.8}$$

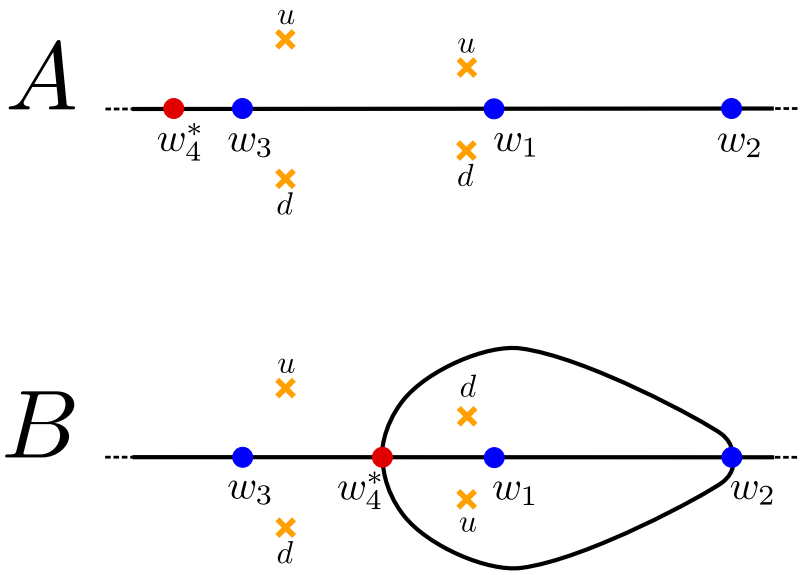


Figure 3.4: Contours where $\gamma = 0$ based on the choice of signs in equation (3.11). These contours are shown schematically by the black curves. The label u (d) at a zero indicates the choice of sign $+$ ($-$) in equation (3.11). We give a detailed discussion of why these are the only possible structures for these contours in appendix B.4. The key in relating these figures to the embeddings in figure 3.2 is that contours on the world-sheet where $\gamma = 0$ map onto folds of the embedding.

3.2.3 Spikes, fold-lines and string embeddings

As mentioned in the previous section there are 4 zeros of T and at each zero we have a $\mathbb{Z}_2$ ambiguity (see equation (3.11)) in the choice of spikes of γ. A priori there are 2^4 different choices for the spikes. However, it turns out that there are only 2 distinct choices that correspond to target-space solutions with the desired properties. These two possibilities are shown in figure 3.4. A discussion of why these are the only two possible choices is given in appendix B.4.[3] These two different possibilities correspond precisely to the two different possible string solutions shown in figure 3.2. The key ingredient in making this correspondence is the observation that contours on the world-sheet where $\gamma = 0$ correspond to fold-lines in the string embedding (see appendix B.4). The location of these contours is

[3]The main ideas are: first, configurations related by $\gamma \to -\gamma$ are not distinct since this is a symmetry of (3.10), and second, one should choose the spikes such that $\gamma \to -\gamma$ under reflection about the real axis. See appendix B.4.

directly connected with the choice of spikes. For example, between a u-spike and a d-spike we know that there must be at least one such contour. In figure 3.4 the $\gamma = 0$ contours are indicated by the black curves. In appendix B.4 we discuss in detail how the structure of these contours is inferred from the choice of spikes.

Let us describe in more detail how we relate the two spike configurations in figure 3.4 to the two string embeddings in figure 3.2. As mentioned above, the key ingredient is to study the fold lines in the two figures. First consider the target-space solution. In figure 3.2A it is clear that there is a single fold-line that runs through each of the punctures in sequence. That is, there is a fold-line directly connecting x_4 with x_3 then x_3 with x_1, etc. This is in agreement with the fold-structure implied by figure 3.4A since for this choice of spikes we can deduce that there is a single contour where $\gamma = 0$ running along the real axis connecting w_4 to w_3 then w_3 to w_1, etc. Thus the spike configuration of figure 3.4 describes a string embedding of the type shown in figure 3.2A.

Now consider the folds of the embedding shown in figure 3.2B. Insertions x_1 and x_3 are both connected by fold lines directly to the insertions x_2 and x_4. Furthermore, x_2 and x_4 are connected to each other by *two* fold-lines. This is because this configuration is double-folded along that line, as one can see from the construction shown in the center panel of figure 3.2. All of this is in perfect agreement with the fold-structure implied by figure 3.4B. In particular, for this choice of spikes both w_1 and w_3 are directly connected to w_2 and w_4^* by contours where $\gamma = 0$. Moreover, w_2 and w_4^* are connected by *two* contours where $\gamma = 0$, precisely corresponding to the double-fold line connecting x_2 and x_4 in figure 3.2B.

Let us comment on a subtle point regarding figure 3.4. Note that we have placed the saddle point w_4^* in different intervals in the two figures. On one hand, we should do this in order to be in agreement with figures 3.1 and 3.2. However, as we will see, given a cross-ratio u and the saddle point w_4^*, the fold structure is fixed. So, to compare two different fold structures for a given cross ratio we are forced to place the saddle point w_4^* in different intervals. This is in perfect agreement with the intuitive perspective of figures 3.1 and 3.2. We will return to this point in section 3.5.2.

3.2.4 The action as a wedge product

We will now return to the computation of the minimal AdS action (see equation (3.3)). Explicitly, the quantity we want to evaluate is

$$-\frac{\sqrt{\lambda}}{\pi}\int_{\Sigma\setminus\{\epsilon_a\}}\frac{\partial x\bar{\partial}x+\partial z\bar{\partial}z}{z^2} \tag{3.13}$$

where $\Sigma\backslash\{\epsilon_a\}$ denotes the sphere with small disks of radius ϵ_a cut out at each puncture. These cut-offs are not independent and are all fixed in terms of the single target-space cut-off $z = \mathcal{E}$; this is important in recovering the spacetime dependence of the correlation function and we will return to this point below [38]. It is convenient to separate the action into a piece that is independent of the cut-offs, and a piece where the dependence can be explicitly evaluated. This can be done because the solution near the punctures is know. In particular, we may write [19, 38]

$$\mathcal{A} = -\frac{\sqrt{\lambda}}{\pi}\int_{\Sigma}\sqrt{T\bar{T}}\,(\cosh\gamma - 1) - \frac{\sqrt{\lambda}}{\pi}\int_{\Sigma\backslash\{\epsilon_a\}}\sqrt{T\bar{T}} \tag{3.14}$$

To extend the integration to the full sphere in the first term we have used the fact that the action (3.9) goes like $\sqrt{T\bar{T}}$ near the punctures as follows from (3.12). We will refer to the first and second term in (3.14) as A_{fin} and A_{div} respectively. Since T is known A_{div} can be evaluated explicitly in terms of the ϵ_a, but to eliminate the ϵ_a in terms of $\mathcal{E}$ requires detailed information about the string solution itself. Fortunately, the tools necessary for computing A_{fin} will also provide the necessary information to complete the calculation of A_{div}. Thus, let us focus for the time being on the computation of A_{fin} and return to A_{div} afterwards.

We would like to write A_{fin} in a form where the integrability of (3.10) is more readily usable. Following [19, 38] we introduce the forms

$$\omega = \sqrt{T}dw \tag{3.15}$$

$$\eta = \frac{1}{2}\sqrt{\bar{T}}\,(\cosh\gamma - 1)\,d\bar{w} + \frac{1}{4}\frac{1}{\sqrt{T}}\,(\partial\gamma)^2\,dw \tag{3.16}$$

and then from a direct computation it follows that

$$A_{fin} = \frac{i}{2}\int_{\widetilde{\Sigma}}\omega\wedge\eta \tag{3.17}$$

where $\widetilde{\Sigma}$ denotes the double cover of the sphere defined by $y^2 = T(w)$. Extending the integration from Σ to $\widetilde{\Sigma}$ simply involves a factor of 2 since each form is odd under sheet-exchange. An important property of these forms is that they are both closed. The form ω is clearly closed since it is holomorphic, and the closure of η follows from the equations of motion for γ. Notice that (3.17) would be true for any choice of the dw component of η. The specific coefficient appearing in (3.16) is necessary for the closure of the form.

Now we would like to apply the Riemann bilinear identity (RBI) to reduce (3.17) to one-dimensional integrals over cycles on $\widetilde{\Sigma}$. There are two caveats in doing this – the

singularities in ω and the singularities in η. These issues where resolved in [38], and we follow the approach used there (see [38] for a more detailed treatment and also [39] for a different approach). The basic idea of the RBI is to write one of the forms of the wedge product as an exact form, $\omega = dF$ where $F = \int_{P_0}^{P} \omega$, which is always possible on a Riemann surface minus some contour, L. In the present case ω has single poles and thus F will have logarithmic cuts which need to be accounted for. A way to side-step this complication is to spread the single poles in ω into a small square-root cuts such that F has only square-root cuts and no singularities. The cost of doing this is that the genus of $\widetilde{\Sigma}$ increases, but the upside is that the application of the RBI is simplified. This takes care of the singularites in ω. Now consider the form η which behaves as $\eta \sim (w - z_a)^{-5/2}$ near the zeros of T. The prescription of [38] is to remove the points z_a from the domain by taking L to be the sum of the standard contour for a Riemann surface of genus g and small contours C_a encircling the points z_a. The integrand of (3.17) can then be written as $d(F\eta)$ (since $d\eta = 0$ on the domain) and then Stokes theorem can be used to reduce the surface integral to a line integral over the usual boundary of the genus g Riemann surface and the contours C_a. The end result is that each boundary C_a contributes a correction of $\pi/12$ to A_{fin} in (3.17) while the integral over the boundary of $\widetilde{\Sigma}$ gives the usual sum over cycles on $\widetilde{\Sigma}$ and thus we have [38]

$$A_{fin} = (\text{number of zeros})\frac{\pi}{12} - \frac{i}{2}\left(\oint_{\gamma_a}\omega\right) I^{-1}_{ab}\left(\oint_{\gamma_b}\eta\right) \tag{3.18}$$

where $\{\gamma_a\}$ is a complete basis of cycles on $\widetilde{\Sigma}$ and I_{ab} is their intersection matrix. For the four-point function there is generically 4 zeros and 4 poles. When we spread the four poles we introduce an additional 4 cuts and thus the surface is genus 5 and there will be 5 a-cycles and 5 b-cycles; that is $\{\gamma_a\} = \{\gamma_{a_1}, \gamma_{b_1}, \gamma_{a_2}, ..., \gamma_{a_5}, \gamma_{b_5}\}$. The main point is that we have reduced the computation of the surface integral (3.17) into a sum of 1-dimensional cycle integrals of a closed form. Such integrals are precisely what integrability is good at computing. In the following section we will see how to compute the cycles $\oint_{\gamma_a}\eta$ by exploiting the integrability of (3.10).

3.3 The linear problem

To compute the η-cycles appearing in (3.18) it is useful to consider the linear problem associated with equation (3.10). Consider a function ψ obeying

$$(\partial + J_w)\,\psi = 0, \qquad \left(\bar{\partial} + J_{\bar{w}}\right)\psi = 0 \tag{3.19}$$

where the components of the connection $J = J_w dw + J_{\bar{w}} d\bar{w}$ are given by

$$J_w = A_w + \frac{1}{\xi}\Phi_w, \qquad J_{\bar{w}} = A_{\bar{w}} + \xi\Phi_{\bar{w}} \tag{3.20}$$

where A and Φ are independent of the spectral parameter ξ and given in terms of γ and $T, \bar{T}$. We give the explicit forms of A and Φ in appendix B.1. Note that we will frequently write the spectral parameter as $\xi = e^{\theta}$.

Compatibility of equations (3.19) for all ξ is equivalent to the flatness of J, which is satisfied if γ obeys the equation of motion (3.10) and T ($\bar{T}$) is purely holomorphic (anti-holomorphic). In the following section we will discuss the relation between the solutions of the (3.19) and the η-cycles appearing in (3.18).

3.3.1 Basic properties

There are a few aspects of the linear problem which will be essential for the following analysis. Let us comment on each of them in turn.

- *Solutions near punctures.* Using (3.7) and (3.12) one can show that near the punctures P_a there are two linear-independent solutions of the form (see Appendix B.1)

$$\hat{\psi}^{\pm}(w) \equiv \left(T/\bar{T}\right)^{1/8} e^{\pm\frac{1}{2}\int^w \xi^{-1}\omega + \xi\bar{\omega}} |\pm\rangle \tag{3.21}$$

$$\sim (w - w_a)^{\pm\frac{1}{4}\Delta_a\xi^{-1} - \frac{1}{4}} (\bar{w} - \bar{w}_a)^{\pm\frac{1}{4}\bar{\Delta}_a\xi + \frac{1}{4}} |\pm\rangle \tag{3.22}$$

  where $|\pm\rangle$ are the eigenvectors of σ^3. Notice that there is a solution that is exponentially big and one that is exponentially small as one approaches the puncture P_a.[4]

- *'Small' solutions.* Demanding that a function is both a solution of the linear problem and also small at some puncture P uniquely defines that solution (up to overall normalization). Thus there is a family of 'small' solutions s_a each of which is small at puncture P_a. On the other hand, specifying that a solution has the big asymptotic near P does not uniquely determine the solution since one could create another solution obeying the same boundary conditions by adding an arbitrary multiple of s_P.

[4] In going from (3.21) and (3.22) we have been careless about the branches of ω. In particular, we may choose a particular branch at some P_a such that the near-puncture solutions take the form (3.22) but then if we smoothly continue $\sqrt{T}$ to some other puncture P_b it is possible that the small and big solutions correspond to the opposite components from the small and big solutions at P_a. This will be very important below, since it will usually be the case in the construction we will use.

- $\mathbb{Z}_2$ *symmetry and 'big' solutions.* Even though one cannot uniquely specify a solution by demanding that it has the big asymptotic near P, there is nevertheless a special solution big near P that is uniquely defined. This follows from the $\mathbb{Z}_2$ symmetry of the connection (3.20) which is given by

$$UJ(\xi)U^{-1} = J(e^{i\pi}\xi) \tag{3.23}$$

where $U = i\sigma^3$. This symmetry implies that if $s_P(\xi)$ (we are suppressing the w, $\bar{w}$ dependence) is the solution to (*B*.1) small at P then

$$\tilde{s}_P \equiv \sigma^3 s_P(e^{-i\pi}\xi) \tag{3.24}$$

is another solution of the linear problem. Moreover, from (3.21) it follows that $\tilde{s}_P$ is *big* at P. Thus we have a second uniquely defined family of solutions $\tilde{s}_a$, each of which is big at puncture P_a.

- *Products of solutions.* Given two solutions of the linear problem ψ_1 and ψ_2, there is a natural SL_2 invariant inner product

$$(\psi_1 \wedge \psi_2) \equiv \mathrm{Det}\left[\{\psi_1, \psi_2\}\right] \tag{3.25}$$

This inner product is equivalent to the Wronskian of the two solutions. Important properties of this Wronskian are that it is independent of w and $\bar{w}$, and thus only depends on the spectral parameter ξ. Further, the product will vanish if the two solutions are linearly *dependent*.

Now that we have introduced these basic facts of the linear problem, we can state what is perhaps the key ingredient in the whole computation.[5] We claim that the $\xi \to 0$ expansion of the inner product of two small solutions is the following [19, 38]

$$(s_a \wedge s_b) \sim \exp\left[\frac{1}{2}\xi^{-1}\varpi_{ab} + \frac{1}{2}\xi\bar{\varpi}_{ab} + \xi\int_a^b \eta + \mathcal{O}\left(\xi^2\right)\right] \tag{3.26}$$

where η is precisely the same form (3.16) that appears in the action formula (3.18) and ϖ_{ab}, $\bar{\varpi}_{ab}$ are explicitly known in terms of integrals of ω and $\bar{\omega}$.[6] A derivation of (3.26) is

[5] To our knowledge the following fact first appear in [19]. Later it was used in [38,39] for 3-point function computations. We give a derivation in appendix B.3; we thank Pedro Vieira and Amit Sever for explaining the basic components of the derivation used for [19]. A different derivation appears in [39].

[6] To be more precise, this expansion will be true for certain s_a and s_b depending on certain conditions stemming from the form of T and also depending on the value of $\mathrm{Arg}\,(\xi)$. Furthermore, the contour of integration will be precisely defined by these conditions. We will discuss these conditions in detail below.

given in appendix B.3. The point is that by computing the inner products $(s_a \wedge s_b)(\xi)$ we can extract the "puncture-puncture" integrals $\int_a^b \eta$ by extracting the $\mathcal{O}(\xi)$ coefficient of this inner product. All of the η-cycles appearing in (3.18) can be written in terms of linear combinations of these puncture-puncture integrals. Thus, we can compute area (3.18) by computing the inner products $(s_a \wedge s_b)$. The rest of this section is devoted to explaining how we compute such inner products using techniques from integrability.

3.3.2 Defining solutions globally

Let us now comment on how to globally define the small solutions. Suppose that we want to construct the small solution s_P away from puncture P, say at some generic point A. To do this we need to use the connection to transport the solution along some path from the neighborhood of P to the point A. However, it is clear from (3.21) that the solutions of the linear problem have non-trivial monodromies around the punctures and therefore homotopically different paths on the 4-punctured sphere will result in different values of the small solution at A. In other words, solutions of the linear problem live on a (generically infinite-sheeted) Riemann surface with branch points at the punctures. For the purposes of calculating it is convenient to fix some conventions for dealing with the multivaluedness of the solutions. We first define the sheets by cutting the Riemann surface as shown in figure 3.5. The cuts all join at a common point and the monodromy about that point is the identity since a path passing through all the cuts is contractable on the sphere. We then define the value of the small solution associated with puncture P at some point A as follows. Draw any curve from the neighborhood of P to A. In the neighborhood of P one starts with s_P. For every time the path crosses a cut emanating from some puncture Q in the *clockwise* (*counterclockwise*) sense attach a factor M_Q (M_Q^{-1}).[7] In this way, if we transport along a path that is homotopically equivalent to a path that does not cross any cuts then the small solution at A will be $s_P|_A$. If the path crosses the cut emanating from puncture Q once in the clockwise sense, then the value of the small solution at A will be $(M_Q s_P)|_A$, and so on (see figure 3.5). In the case when s_P is transported around the puncture P one can see from the explicit form (3.21) of s_P near P that the result will be multiplication by a constant. That is

$$M_P s_P = \mu_P s_P \tag{3.27}$$

$$M_P \tilde{s}_P = \tilde{\mu}_P \tilde{s}_P \tag{3.28}$$

[7]Note that the result of a monodromy can be expressed as the linear map M since both s and Ms are solutions of the linear problem. Therefore they can both be expanded in terms of two linearly independent solutions of the linear problem, and thus they are related to each other simply by a linear map, or in other words simply by multiplication by some matrix, M.

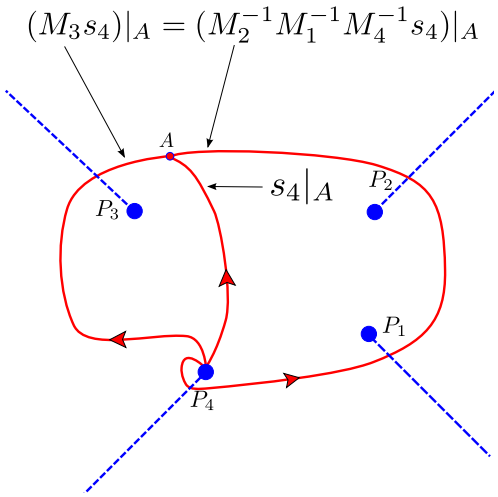


Figure 3.5: Our conventions for defining the solutions globally. The dashed blue lines emanating from the punctures indicate the conventions for 'cutting' the full Riemann surface, thus defining the sheets. The red lines indicate the parallel transport of a solution from P_4 to the point A along three paths. Two of the paths are homotopically equivalent due to the triviality of the total monodromy $M_4 M_3 M_2 M_1 = 1$ (which follows from the fact that any path encircling all the punctures with the same orientation is contractable on the sphere). The third path is homotopically distinct from the other two, and thus the value of the solution at A will differ by monodromy factors.

so that s_P and $\tilde{s}_P$ are eigenvectors of M_P with eigenvalues μ_P and $\tilde{\mu}_P = 1/\mu_P$ respectively. One cannot repeat such an analysis to evaluate $M_Q s_P$ since generically one does not know the explicit form of s_P in the neighborhood of Q.

3.3.3 WKB approximation and WKB Curves

As we will discuss shortly, it will be essential to have control over the $\xi \to 0, \infty$ asymptotics of the inner products $(s_P \wedge s_Q)(\xi)$. It is clear from (3.19) – (3.20) that these are both singular limits, and the basic idea of extracting this singularity – which is called the *WKB approximation* – is as follows.[8] As discussed above, we have good control over the solutions in the neighborhood of the punctures. Thus we would like to study, in the limits $\xi \to 0, \infty$, the transport of small s_P along a curve $w(t)$ which connects a neighborhood of a puncture

[8] See appendix B.3 or [52] for a more detailed treatment.

P with a neighborhood of another puncture Q. Let us consider the transport away from P (see figure 3.6). We will discuss the $\xi \to 0$ limit since the $\xi \to \infty$ limit is similar.

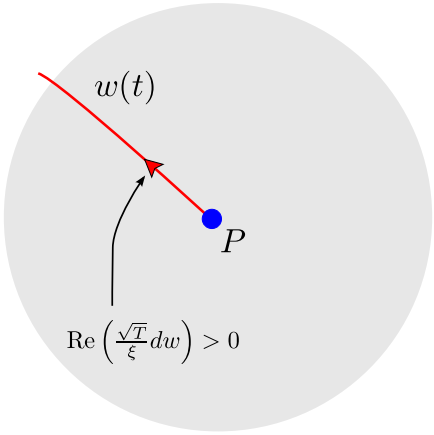


Figure 3.6: Transporting s_P away from P along $w(t)$. We have chosen the branch of Φ in (3.29) such that $s_P \propto |+\rangle$ near w_P. In other words, we have chosen the branch of Φ such that $\mathrm{Re}\left(\langle +| - \Phi_w/\xi dw|+\rangle\right) = \mathrm{Re}\left(dw\sqrt{T}/\xi\right) > 0$ for dw pointing along $w(t)$ *away* from P and thus $\exp\left(\int_{w'_P}^{w(t)} dw\sqrt{T}/\xi\right)$ is exponentially diverging as $\xi \to 0$.

At any point in Σ the matrix Φ has the two eigenvalues $\mp\omega/2 = \mp\sqrt{T}/2\, dw$ (which are single valued on the double cover $\widetilde{\Sigma}$), and thus we can choose a gauge along $w(t)$ where Φ is diagonal and given by

$$\Phi = \frac{1}{2}\begin{pmatrix} -\omega & 0 \\ 0 & \omega \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -\sqrt{T}dw & 0 \\ 0 & \sqrt{T}dw \end{pmatrix} \tag{3.29}$$

In the limit $\xi \to 0$ some component of Φ/ξ will dominate and thus the leading contribution to s_a at some point w along $w(t)$ will be given by

$$e^{-\int_{w'_a}^{w} \Phi/\xi}|\sigma\rangle \tag{3.30}$$

where the value of $\sigma = \pm$ depends on the branch of Φ we have chosen (recall that $|\pm\rangle$ are the eigenvectors of σ^3). This is the singular contribution in the limit $\xi \to 0$ for the same reason that it is the small solution – namely, because

$$\mathrm{Re}\left(\langle\sigma|\left(-\Phi/\xi\right)|\sigma\rangle\right) > 0 \tag{3.31}$$

along a path traveling *away* from P_a. The basic statement of the WKB approximation is that so long as we transport along paths such that (3.31) is true along the whole path then the leading contribution to s_P in the $\xi \to 0$ limit is indeed given by (3.30). In other words, as long as we transport along curves satisfying (3.31) everywhere, then we can reliably extract the singularity as $\xi \to 0$ as it is simply given by (3.30). The curves along which (3.31) is satisfied most strongly are those for which

$$\mathrm{Im}\left(\langle\sigma|\left(-\Phi/\xi\right)|\sigma\rangle\right) = 0 \tag{3.32}$$

Curves satisfying this condition are called *WKB curves*. If we transport along some curve satisfying (3.32) for $\mathrm{Arg}\,(\xi) = \phi$, then the condition (3.31) will be satisfied for $\mathrm{Arg}\,(\xi) \in (\phi - \pi/2, \phi + \pi/2)$. In fact, we will need to control the asymptotics of s_P in precisely such a wedge of the ξ-plane, and thus we should always transport along WKB lines. We will give the a very brief overview of the properties of these lines in the next subsection. For a detailed treatment see [52].

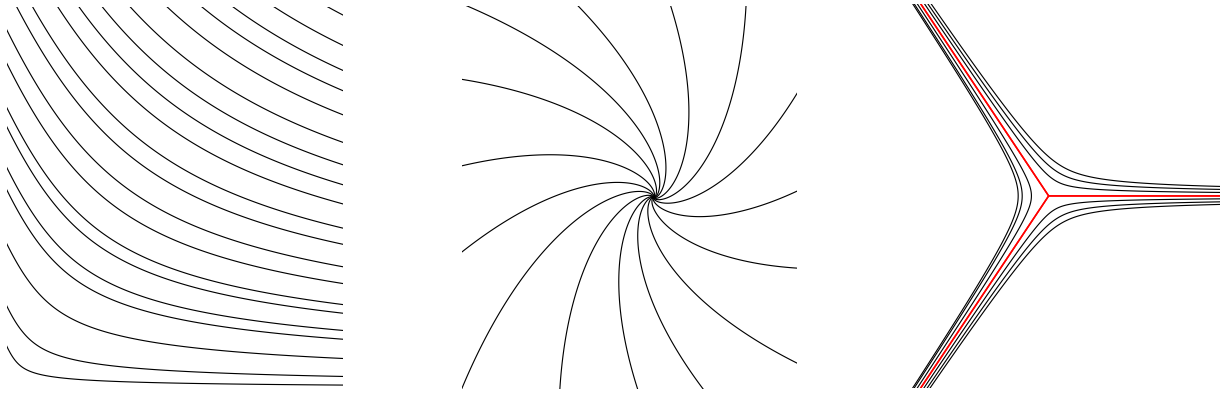
3.3.4 WKB triangulation

As we discussed in section 3.3.2 we define the solutions of (3.19) globally by transporting along specific paths. Transport of solutions along homotopically equivalent paths will lead to the same result, whereas transport along homotopically inequivalent paths generically will give different results. For this reason it is useful to set up a system of fiducial paths between the punctures which we will use to globally define the solutions. Because we will need to control the large/small ξ asymptotics of the Wronskians, it is best to choose these paths to be WKB curves – i.e. curves satisfying (3.32).

We will first consider the local structure of WKB curves. In the neighborhood of a generic point on the punctured sphere the WKB curves are smooth and non-intersecting (see figure 3.7A). In the neighborhood of a (double) pole of T the WKB curves follow logarithmic spirals that asymptote to the singular point (see figure 3.7B). All that will be important here is that the poles act as sources/sinks of WKB curves but the exact nature of these spirals will not be important. Finally, working in the neighborhood of a simple zero of T one can show that there are three special WKB curves that asymptote to the zero and which govern the WKB lines near the zero (see figure 3.7C).

Now consider the global structure of the WKB curves. All WKB curves fall into one of the following types [52]

- *Generic WKB curves* which are those that asymptote in both directions to a pole of T (potentially the same one);



A: generic point B: pole C: zero

Figure 3.7: Local structure of WKB lines in the neighborhood of, A: a generic point; B: a double pole of T; C: a simple zero of T. In the case of a generic point the WKB curves form continuous non-intersecting lines. In the case of a singular point they form logarithmic spirals for generic values of $\mathrm{Arg}(\xi)$. The exact nature of these spirals will not be important. What is important is that the singular points act as sources/sinks of WKB curves. In the case of a zero, there are three special WKB curves that asymptote to the zero which are the red curves in panel C. These special curves, called separating curves, determine the global structure of the WKB foliation.

- *Separating WKB curves* which asymptote to a zero of T in one direction and to a pole of T in the other;
- *Finite WKB curves* which are closed or asymptote in both directions to a zero of T (potentially the same one).

We will now describe how we use the WKB curves to set up a system of fiducial paths, or more specifically, a *triangulation*. By triangulation we mean a triangulation of the punctured sphere with all vertices at the punctures and at least one edge incident on each vertex. Consider fixed T and $\mathrm{Arg}(\xi)$ such that there are no finite WKB curves (this can always be done since such curves only appear at special, discrete values of $\mathrm{Arg}(\xi)$). First draw all of the separating WKB curves – there will be $3N_Z$ of these, where N_Z is the number of zeros of T (since for the moment we are not allowing finite WKB curves). These curves will divide the punctured sphere up into cells with each cell defining a family of homotopically equivalent *generic* WKB curves as shown in figure 3.8 for an example of the 4-punctured sphere. To construct the triangulation, choose a representative curve

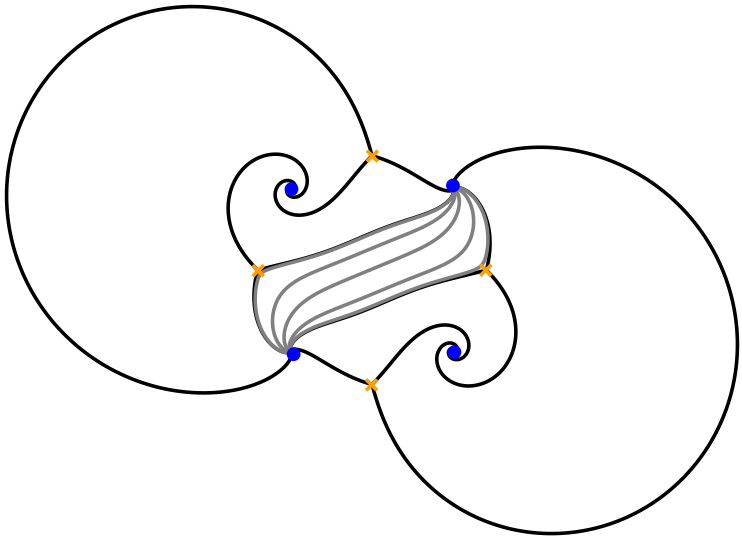


Figure 3.8: Global WKB structure for an example with 4 punctures. The separating curves are shown in black. In one cell we show several examples of homotopically equivalent curves (shown in gray) that sweep the cell. Each cell defined by the separating curves has a 1-parameter family of such curves. By choosing a representative curve from each family we obtain the triangulation shown in figure 3.9. Notice that near each puncture (the blue dots) we see the spiral structure shown in panel B of figure 3.7 and near each zero (yellow $\times$) we see the local structure shown in panel C of figure 3.7.

from each cell, e.g. any one of the silver curves shown in figure 3.8. The claim is that the collection of these representative curves, which we will call *edges*, gives the desired triangulation [52].[9] As a concrete example, the triangulation associated with the cell-construction of figure 3.8 is shown in figure 3.9. This same triangulation will play an important role in the 4-point function computation below.

We have now finished the discussion of how to construct the WKB triangulation for a given T and $\mathrm{Arg}\,(\xi)$. Before moving on the the next section let us discuss one final point. In the following it will be useful to lift edges of the triangulation to the double cover $\widetilde{\Sigma}$ and to endow the lifted edges with an orientation. Recall that $\omega = \sqrt{T}dw$ is a single valued form on $\widetilde{\Sigma}$. Let ∂_t be a tangent vector of the lifted edge E at a point on $\widetilde{\Sigma}$. There are of

[9] To see this in general consider a single zero of T as shown in figure 3.7. The zero is on the boundary of three cells. Choosing edges from the family of curves in each cell we see that they form a triangle. Thus the edges form a triangulation of the punctured sphere with each triangle containing a zero of T.

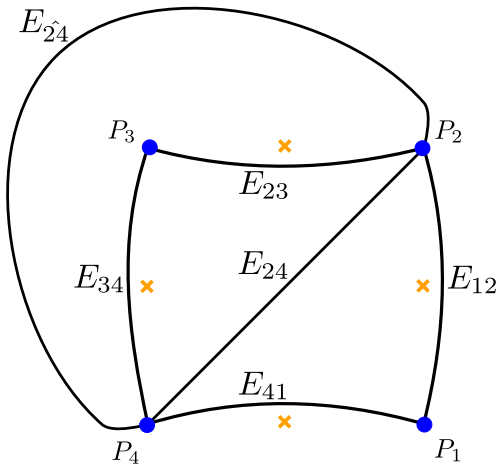


Figure 3.9: The WKB triangulation of the 4-punctured sphere following from the WKB foliation shown in figure 3.8. Each edge E_{ab} of the triangulation is a representative from one of the families of homotopotically equivalent lines in each cell of figure 3.8. This triangulation will be of central interest in the 4-point function computation.

course two possible orientations for ∂_t. Note that by virtue of (3.32) we have $e^{-i\phi}\omega\cdot\partial_t \in \mathbb{R}$. We define the orientation of the lifted edge E by the condition $e^{-i\phi}\omega\cdot\partial_t > 0$. Notice that each edge on the punctured sphere will lift to two edges – one on each sheet of $\widetilde{\Sigma}$ and that these two edges will have opposite relative orientation. Picking a particular orientation of some edge is equivalent to picking a particular sheet of $\widetilde{\Sigma}$.

3.3.5 Coordinates

From the WKB triangulation we will now construct the so-called Fock-Goncharov coordinates [52]. These are natural objects to work with because they are gauge invariant and independent of the normalization of the small solutions. From the coordinates we will be able to extract the η-cycles that we need to compute the action (3.18).

Consider some edge E of the triangulation. This edge is shared by precisely two triangles, and these triangles form the quadrilateral Q_E (see figure 3.10). Number the vertices of Q_E as shown in figure 3.10 with E going between P_1 and P_3. As we mentioned in section 3.3.1, associated with each puncture P_a there is a small solution s_a. The solutions cannot be made globally smooth and single valued on the punctured sphere due to the monodromy around each puncture. However, we can define them such that they are single valued and

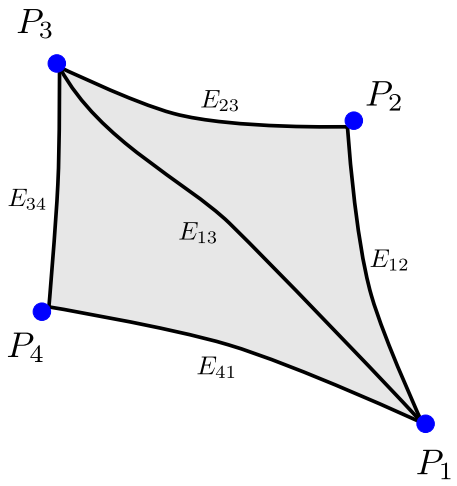


Figure 3.10: The two triangles sharing the edge E_{13}. These two triangles define the quadrilateral $Q_{E_{13}}$, which is shown in gray shading. Each blue dot represents a puncture, which are the verticies of the triangulation and each black line and is an edge.

smooth throughout Q_E.[10] We then define the Fock-Goncharov coordinate as [52]

$$\chi_E = (-1)\frac{(s_1 \wedge s_2)(s_3 \wedge s_4)}{(s_2 \wedge s_3)(s_4 \wedge s_1)} \tag{3.33}$$

where all the s_a are evaluated at a common point in Q_E.

As a concrete example consider the triangulation of the 4-punctured sphere shown in figure 3.9. In figure 3.11 we show how to apply the procedure just described to construct the coordinates corresponding to edges E_{24} and $E_{\hat{24}}$. Consider first the left panel of 3.11. We define each solution s_a throughout Q_{24} by parallel transporting from each P_a where the explicit form of the solutions is known – see (3.21). The red lines indicate the parallel transport of each s_a from P_a to a common point A; clearly we can define the small solutions at any point $A \in Q_{24}$ in this way. Further, if the paths never leave the quadrilateral (or at least is always homotopically equivalent to a paths that never leave the quadrilateral) then the solution defined in this way is guaranteed to be single-valued and smooth throughout the quadrilateral, as required. With the solutions defined at a common point in the quadrilateral we can construct the coordinate χ_{24}, which is independent of the choice of $A \in Q_{24}$. Now consider the right panel of figure 3.11 where the grey shading indicates the

[10] We will show this in some concrete examples momentarily.

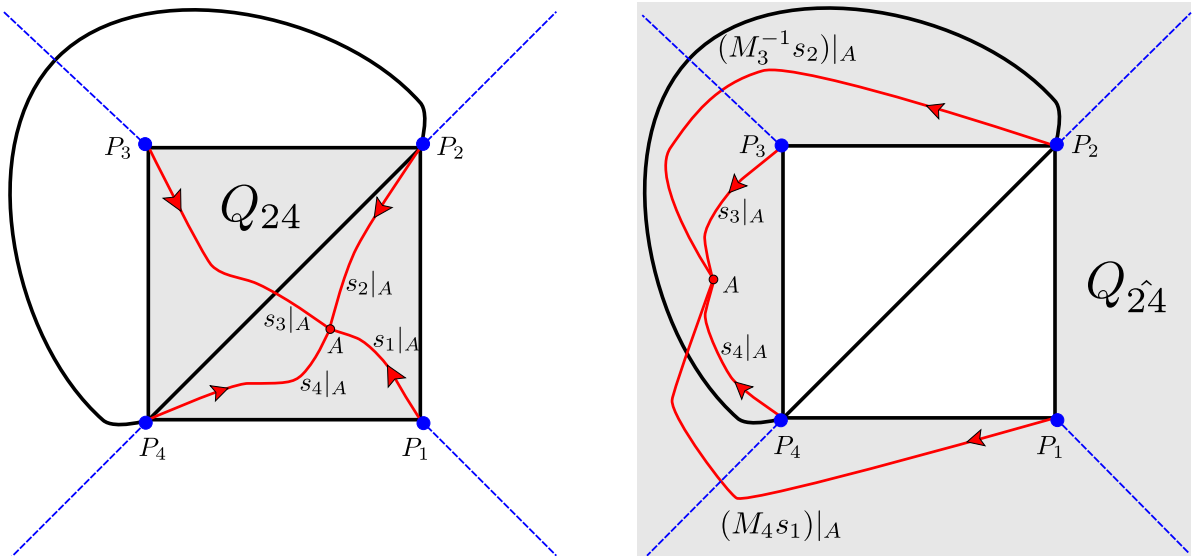


Figure 3.11: Here we show how to construct the coordinates χ_{24} (left panel) and $\chi_{\hat{2}\hat{4}}$ (right panel) of the triangulation of figure 3.9. The gray shaded regions represent Q_{24} and $Q_{\hat{2}\hat{4}}$ respectively. These figures should be pictured on the sphere. The dashed blue lines emanating from the punctures indicate our conventions for defining the sheets of the small solutions as explained in section 3.3.2. The red lines indicate how we globally define the solutions s_a by transporting away from P_a using the connection. We use paths that never leave the quadrilateral such that the solutions used to form the coordinates are guaranteed to be single-valued and smooth throughout the quadrilateral, as required.

quadrilateral associated to edge $Q_{\hat{2}\hat{4}}$. These figures should be imagined on the sphere. Now to transport the small solutions to a common point one cannot avoid passing under a cut onto new sheets of some of the small solutions. For example s_2 must pass onto a new sheet in order to be smoothly continued to the point A. This is because if we were to compare the s_2 of the left panel and the s_2 of the right panel (by moving each respective A to a common point A' along the edge E_{34}, for example) the two paths of continuation would differ by a holonomy around P_3, and thus the values at the point A' would not coincide but would differ by the action of $M_3^{\pm 1}$. Of course which solution we call s_2 and $M_3^{\pm 1}s_2$ is purely a matter of convention. Similarly, which solutions acquire factors of M_a depends on the choice of the point A. We stress that the coordinates are independent of all such ambiguities, as one can easily check using identities such as $(M_c s_a \wedge s_b) = (s_a \wedge M_c^{-1} s_b)$,

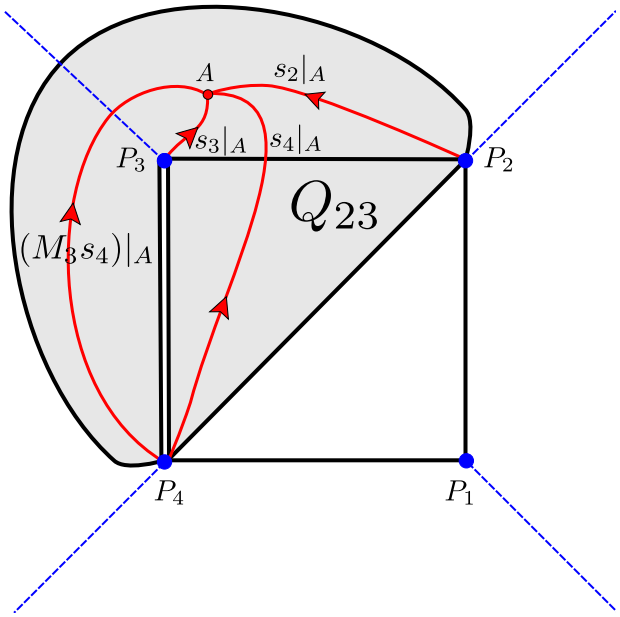


Figure 3.12: Here we describe the construction of the coordinate for the slightly degenerate case where the coordinate corresponds to an edge ending at a vertex that has only two incident edges (e.g. P_2 has only 2 incident edges: E_{12} and E_{23}). We construct the coordinate for edge E_{23} of the triangulation shown in figure 3.9. The quadrilateral prescription described above still applies, but one must take care to correctly define Q_E and the solutions inside Q_E. First of all, in order to have single-valued and smooth solutions throughout Q_{23} we must exclude a region between P_3 and P_4. Otherwise Q_{23} would contain P_3 and thus the solutions could not be single valued in Q_{23} (there would be a monodromy around P_3). Since the boundaries of the quadrilateral must be edges of the triangulation, the only choice is to remove a thin region running along edge E_{34} and then to treat the two 'sides' of E_{34} as different edges. In the figure we have represented this process by showing E_{34} as doubled and with the region between the new edges excluded from Q_{23}. We then define the solutions throughout Q_{23} in the same way as described in figure 3.11, by analytically continuing the solutions along paths from P_a to A that stay within Q_{23} which is represented as the shaded region. Once we have defined the solutions at a common point we form the coordinate χ_{23} given in equation (3.35).

etc. Then from figure 3.11 we read off

$$\chi_{24} = (-1)\frac{(s_2 \wedge s_3)(s_4 \wedge s_1)}{(s_3 \wedge s_4)(s_1 \wedge s_2)}, \qquad \chi_{\hat{24}} = (-1)\frac{(M_3^{-1}s_2 \wedge M_4 s_1)(s_4 \wedge s_3)}{(M_4 s_1 \wedge s_4)(s_3 \wedge M_3^{-1}s_2)}. \tag{3.34}$$

We will also need to construct coordinates in the slightly degenerate case where the coordinate corresponds to an edge ending at a vertex that has only two incident edges (including the edge under consideration) for example all edges in figure 3.9 except E_{24} and $E_{\hat{24}}$. We show how to construct this coordinate in figure 3.12. Using the procedure described there we find

$$\chi_{23} = (-1)\frac{(s_2 \wedge M_3 s_4)(s_3 \wedge s_4)}{(M_3 s_4 \wedge s_3)(s_4 \wedge s_2)}, \qquad \chi_{12} = (-1)\frac{(s_1 \wedge M_1^{-1} s_4)(s_2 \wedge s_4)}{(M_1^{-1} s_4 \wedge s_2)(s_4 \wedge s_1)} \tag{3.35}$$

The other two coordinates χ_{34} and χ_{14} are computed in a similar way.

We have now completed our discussion of how to construct the coordinates. Before we continue, let us comment on a useful property of these objects. Consider multiplying all of the coordinates associated with edges meeting a given puncture P. For example, the edges ending at P_2 in the triangulation of figure 3.9 are E_{12}, $E_{\hat{24}}$, E_{23} and E_{24}. Using (3.34)-(3.35) we have

$$\chi_{12}\chi_{\hat{24}}\chi_{23}\chi_{24} = \mu_2^2 \,. \tag{3.36}$$

This property is true in general since the inner-products in the coordinates telescopically cancel in the product and the only thing that remains is the effect of the monodromy around the puncture which produces a μ_P^2 factor. Thus we have the general rule [52]

$$\prod_{E \text{ meeting } P} \chi_E = \mu_P^2 \,. \tag{3.37}$$

3.3.6 WKB asymptotics of the coordinates

The advantage of using the WKB triangulation is that the $\xi \to 0, \infty$ asymptotics of the coordinates of the triangulation are easily extracted given the discussion of section 3.3.3. That is, because we have maximum control over the large/small ξ asymptotics of the small solutions when we transport along WKB curves. We give only the basic idea of the derivation of these asymptotics here and refer the reader to appendix B.3 and [52] for details.

To obtain the asymptotic of some χ_E one simply needs to use expression (3.30) for each inner-product of the coordinate, taking care to account for the direction of the WKB lines. Consider the coordinate associated with edge E_{24} in figure 3.13. The expression for this coordinate in terms of the small solutions is given in (3.34). We will now use formula (3.30) to compute the asymptotic of this coordinate in the $\xi \to 0$ limit. Let us take the directions of the WKB lines to be as given in figure 3.13. To evaluate the inner product $(s_2 \wedge s_3)$ we must transport the solutions to a common point. Since there is a WKB line

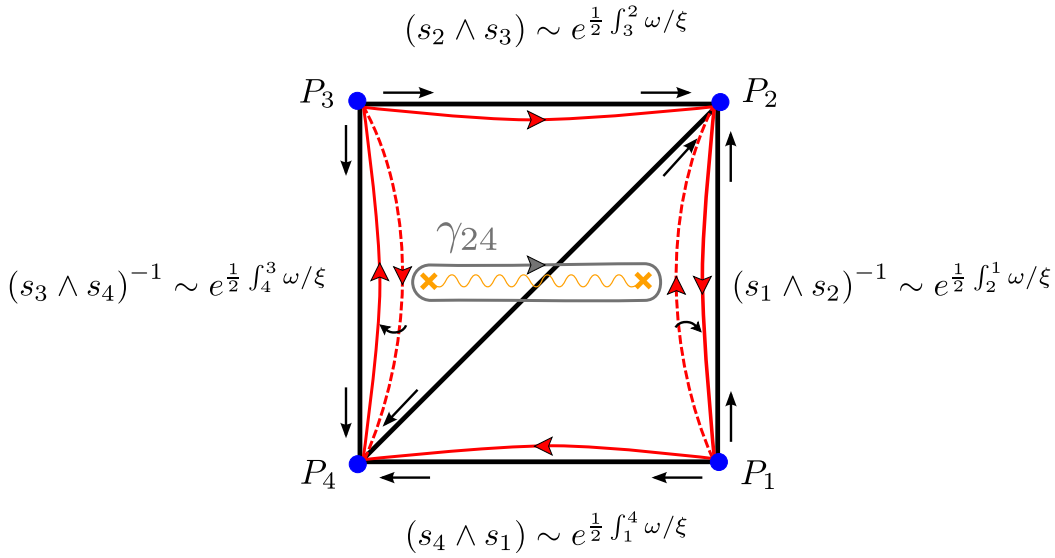


Figure 3.13: Computing the $\xi \to 0$ asymptotic of the coordinate χ_{24} for a typical WKB triangulation. The blue disks represent the punctures and the black lines represent edges of the triangulation. A yellow $\times$ represents a zero of ω and the wavy yellow line shows our convention for defining the sheets of $\widetilde{\Sigma}$. The black arrows running along the edges indicate the choice of the direction for the edges. Each red curve indicates the transport of a small solutions in the limit $\xi \to 0$. The dashed red lines correspond to the transport of a solution appearing in the *denominator* of the coordinate. The transports used to form the coordinate combine into the continuous integral of ω near the boundary of $\mathcal{Q}_{24}$, which can then be deformed into the cycle integral γ_{24} shown in gray.

flowing from P_3 to P_2 we can safely use (to leading order) expression (3.30) to transport s_3 to the neighborhood of P_2, giving a contribution of the form $(s_2 \wedge s_3) \sim e^{\frac{1}{2}\int_3^2 \omega/\xi}$. To evaluate $(s_3 \wedge s_4)$ we must transport s_3 to P_4 since that is the direction of the WKB line and thus we get the contribution $(s_3 \wedge s_4) \sim e^{\frac{1}{2}\int_3^4 \omega/\xi}$. We may then reverse the order of integration in $(s_3 \wedge s_4)$ and also move it to the numerator of the coordinate. Then the integrations from $(s_2 \wedge s_3)$ and $(s_3 \wedge s_4)$ combine nicely into a continuous integral running just inside the boundary of $\mathcal{Q}_{24}$ from the neighborhood of P_2 to P_3 to P_4. Repeating this analysis for the remaining two brackets one obtains a closed cycle integral passing along the boundary of $\mathcal{Q}_{24}$. Recall from the discussion of section 3.3.4 that each triangle in the WKB triangulation encloses one zero of ω. The integral of ω thus encloses two zeros and so it can be deformed to the cycle integral shown in figure 3.13. Thus the non-vanishing

contribution in the limit $\xi \to 0$ is given by

$$\chi_E \sim (-1) \exp\left(\frac{1}{2}\xi^{-1}\int_{\gamma_E}\omega + C_E^{(0)}\right) \tag{3.38}$$

The contour γ_E is the cycle encircling the two zeros contained in Q_E and its direction is the same as that of the WKB lines corresponding to the brackets in the numerator of the coordinate. The term $C_E^{(0)}$ is the $\mathcal{O}(\xi^0)$ contribution to the WKB expansion, which we will discuss momentarily. The overall (-1) prefactor in (3.38) is the same (-1) appearing in the definition of the coordinate (3.33).

To derive the subleading WKB corrections (in the $\xi \to 0$ limit, for example) is essentially a matter of perturbation theory once the singular contribution has been extracted. We give a detailed discussion of this in appendix B.3. Here we will simply focus on the result and its implications. We find the first subleading contribution is given by

$$C_E^{(0)} = \log(-1)^{u_E} \pm i\pi \tag{3.39}$$

where u_E is the number of u-spikes enclosed by γ_E.

Finally the $\xi \to \infty$ asymptotic follows in the same way as the $\xi \to 0$ and leads to a cycle integral around Q_E of $\xi\bar{w}$.

To summarize, the $\xi \to 0, \infty$ asymptotics for χ_E are given by

$$\chi_E \sim (-1)^{u_E} \exp\left[\frac{1}{2}\int_{\gamma_E}\left(\xi^{-1}\omega + \xi\bar{\omega}\right)\right] \tag{3.40}$$

where γ_E is the cycle encircling the two zeros contained in Q_E and its direction is the same as that of the WKB lines corresponding to the brackets in the numerator of the coordinate. Now it is clear how the choice of spikes (i.e. the choice of signs in (3.11)) is encoded into the coordinates – via the constant term in the WKB expansion which contributes the $(-1)^{u_E}$ factor in (3.40). Recall that u_E is the number of u-spikes encircled by γ_E.

3.3.7 Shift relation.

In section 3.3.1 we explained that there are two special solutions s_P, $\tilde{s}_P$ associated with each puncture P and that they are related to each other by a shift in the spectral parameter: $\tilde{s}_P(\xi) = \sigma^3 s_P(e^{-i\pi}\xi)$. Here we give an alternative relation between the small and big solutions that does not involve shifting the spectral parameter. The solutions s_P and $\tilde{s}_P$

are linearly independent and thus we can expand any solution s_Q in terms of them. In particular we have

$$s_Q = \left(\frac{\tilde{s}_P \wedge s_Q}{\tilde{s}_P \wedge s_P}\right) s_P + \left(\frac{s_P \wedge s_Q}{s_P \wedge \tilde{s}_P}\right) \tilde{s}_P \tag{3.41}$$

$$M_P s_Q = \left(\frac{\tilde{s}_P \wedge s_Q}{\tilde{s}_P \wedge s_P}\right) \mu_P s_P + \left(\frac{s_P \wedge s_Q}{s_P \wedge \tilde{s}_P}\right) \mu_P^{-1} \tilde{s}_P \tag{3.42}$$

For the second equality we have used (3.27)-(3.28). Combining these two equations it follows that

$$\left(\frac{M_P s_Q \wedge s_Q}{M_P s_Q \wedge s_P}\right) = (1-\mu_P^2) \left(\frac{\tilde{s}_P \wedge s_Q}{\tilde{s}_P \wedge s_P}\right) \tag{3.43}$$

The utility of this equation is that it allows us to replace certain wronskians involving big solutions (as on the RHS of (3.43)) in terms of small solutions with monodromies. This will play a key role in the derivation of the functional equations that we present in the following section.

3.3.8 χ-system.

We will now derive a set of functional equations for the coordinates which, together with certain analytic properties, allows us to determine the coordinates completely. Our inspiration comes from the solution of the bosonic Wilson-loop problem at strong coupling [19] where the solution involves a set of functional equations of the schematic form[11]

$$Y_a^+ Y_a^- = F_a(Y) \tag{3.44}$$

where $f^{n\times\pm} \equiv f(\theta \pm ni\pi/2)$. On the RHS of (3.44) the function F_a can depend on all of the Y_a, but with their arguments un-shifted. The only shifts in the spectral parameter occur on the LHS of (3.44). For the Wilson-loop problem the F_a are such that (3.44) takes the form of a so-called Y-system which commonly appear in the context of $1+1$ dimensional integrable QFT's. Here, using the general formalism of [52], we will arrive at a set of functional equations with the same schematic form as (3.44) but with the F_a of a different form than that occurring in the Wilson-loop problem. We will call this type of functional

[11] The linear problem associated with that problem is very similar to the one considered here and the Y_a are (up to shifts in the spectral parameter) the coordinates associated with that problem. We are referring here to the special case where the Wilson loop lives in an $\mathbb{R}_{1,1}$ subspace.

equation a χ-system.

To derive a relation of the form (3.44) we begin with the LHS. Using (3.24) we have

$$\chi_E \widetilde{\chi}_E = \chi_E \chi_E^{++} \tag{3.45}$$

where $\widetilde{\chi}_E$ is defined by taking χ_E and replacing each small solutions $s_a \to \tilde{s}_a$. To obtain the schematic form (3.44) we need to rewrite (3.45) in terms of only un-shifted small solutions. That is, we need to get rid of all the tildes without introducing any shifts in the spectral parameter. For this we can use (3.43) after applying the Schouten identity[12] to (3.45) to obtain

$$\chi_E \chi_E^{++} = \chi_E \widetilde{\chi}_E = \frac{(1+A_{ab})(1+A_{cd})}{(1+A_{bc})(1+A_{da})} \tag{3.46}$$

where we have defined the useful auxiliary quantity

$$A_{PQ} = (-1)\frac{(s_Q \wedge \tilde{s}_P)\,(s_P \wedge \tilde{s}_Q)}{(s_P \wedge \tilde{s}_P)\,(s_Q \wedge \tilde{s}_Q)} \tag{3.47}$$

$$= (-1)\left(1+\mu_P^2\right)^{-1}\left(1+\mu_Q^2\right)^{-1}\left(\frac{M_P s_Q \wedge s_Q}{M_P s_Q \wedge s_P}\right)\left(\frac{M_Q s_P \wedge s_P}{M_Q s_P \wedge s_Q}\right) \tag{3.48}$$

Here, the edge E is the edge ac in Q_E where the vertices are labeled $abcd$ in counter-clockwise order. To go from (3.47) to (3.48) we used the shift relation (3.43). The last step is to rewrite the wronskians appearing in (3.48) in terms of the coordinates. Once this is done, combining (3.45) – (3.48), we can assemble a functional equation of the form (3.44). To do this (following [52]) we introduce the quantity

$$\Sigma\left(P; Q \to Q\right) = 1 + \chi_{P,a}\left(1 + \chi_{P,a-1}\left(1 + \ldots \chi_{P,2}\left(1 + \chi_{P,1}\right)\right)\right) \tag{3.49}$$

The coordinates appearing in this object are shown in figure 3.14. By repeatedly applying the Schouten identity (see footnote 12) starting with $(1+\chi_{P,1})$ one can see that the Wronskians in (3.49) telescopically cancel so that[13]

$$\Sigma\left(P; Q \to Q\right) = \frac{(s_0 \wedge s_{a+1})(s_P \wedge s_a)}{(s_{a+1} \wedge s_a)(s_0 \wedge s_P)} = \frac{(s_P \wedge s_a)(M_P s_Q \wedge s_Q)}{(s_Q \wedge s_a)(M_P s_Q \wedge s_P)} \tag{3.50}$$

In going from the first equality to the second in (3.50) we have accounted for the monodromy acquired by the small solutions when they are analytically continued around P (see figure 3.14). Then, from (3.50) and (3.48) we have

$$(1+\mu_P)^2(1+\mu_Q)^2 A_{PQ} = \chi_{PQ}\Sigma\left(P; Q \to Q\right)\Sigma\left(Q; P \to P\right) \tag{3.51}$$

[12] $(s_a \wedge s_b)(s_c \wedge s_d) + (s_a \wedge s_c)(s_d \wedge s_b) + (s_a \wedge s_d)(s_b \wedge s_c) = 0$.

[13] An easy way to see this in general is to use induction [52]. The case $a = 1$ is simple to prove using Schouten identity. Then one can show (again using Schouten) that $\Sigma\left(P; Q_{a+2} \to Q_0\right) = 1 + \chi_{P,a+1}\Sigma\left(P; Q_{a+1} \to Q_0\right)$.

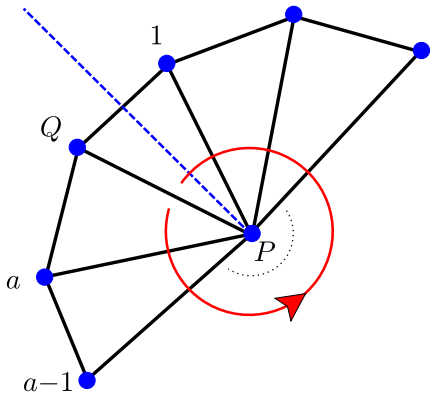


Figure 3.14: Graphical rules for constructing $\Sigma(P; Q \to Q)$. Start at edge E_{PQ} and continue in a *counterclockwise* fashion about P forming the nested product (3.49) by multiplying the coordinates for each edge encountered along the way (i.e. the coordinates associated with each edge intersected by the red line in the order indicated by the arrow). The dashed blue line indicates our convention for cutting the solutions to account for the monodromy around P. The small solutions used to form the coordinates are defined in the vicinity of P by analytically continuing them throughout the triangles along the direction indicated by the red arrow and thus if we use s_Q in $\chi_{P,a}$ then we must include a monodromy matrix when the solution is continued around P to form $\chi_{P,1}$.

Finally, using (3.51) in (3.46) and noting (3.49) we obtain a closed functional equation for χ_E of the form (3.44). Repeating this procedure for the coordinate associated to each edge in a given triangulation gives the desired set of functional equations. Note that this procedure can be applied to derive the χ-system for an arbitrary number of punctures. In section 3.4 we will apply this procedure to the triangulation (3.9), which is one of the triangulations of interest for the four-point function computation.

3.3.9 Inverting χ-systems

In the previous section we showed how to derive the χ-system associated with a given triangulation of the N-punctured sphere. In this section we will discuss how to use the χ-system along with certain analytic properties of the coordinates to obtain integral equations that determine the χ_E uniquely.

The basic idea behind the inversion of a χ-system is to Fourier transform (the log of) each equation since in Fourier space these nonlocal relations become local as the shifts in

the parameter θ can be undone in the usual way. For such a procedure to be successful one must have a certain amount of control of the analytic properties of the coordinates. Let us discuss this carefully. The equations that we want to Fourier transform have the form

$$\log \chi_E^- (\theta + i\phi) + \log \chi_E^+ (\theta + i\phi) = \log F_E \left(\chi^\pm (\theta + i\phi) \right) \tag{3.52}$$

where $F_E(\chi)$ is an explicit function of the coordinates which follows from the discussion of section 3.3.8. We have introduced the arbitrary shift ϕ for reasons that will be explained momentarily. Note that $\chi_E \chi_E^{++} = \chi_E \chi_E^{--}$ since the small solutions are $2\pi i$-periodic, which is why we can have either shift $F_E(\chi^\pm)$ on the RHS. The choice of this shift is arbitrary since the objects we will eventually compute (the η-cycles) are functionals of the coordinates only through A_{PQ} which is $i\pi$-periodic and thus does not care about the choice of shift. As a convention we choose the shift $-i\pi/2$.

To Fourier transform the relationship (3.52) one must be sure that the transform converges. Moreover, to undo the shifts on the LHS, one must account for the singularities (if any) of $\log \chi_E$ in the strip of width π centered along the line where the transform has been performed. We will now discuss each of these issues in turn.

The information from the WKB analysis will allow us to ensure the convergence of the Fourier transform, provided certain conditions are satisfied. First consider the LHS of (3.52). We need to ensure that the transform of each *individual* term converges. We can ensure this if we know the asymptotics of the coordinates in the full strip $\mathrm{Im}(\theta) \in (\phi - \pi/2, \phi + \pi/2)$. The coordinates should be derived from the triangulation that one has at $\mathrm{Im}(\theta) = \phi$. Then the WKB analysis guarantees that the asymptotics are given by (3.40) in a strip that includes the region $\mathrm{Im} \in (\phi - \pi/2, \phi + \pi/2)$. Each term on the LHS can be made safe to transform by making (on the LHS only) the replacement $\chi_E \to \chi_E / \chi_E^{(0)}$ where $\chi_E^{(0)}$ is the asymptotic (3.40). This replacement does not modify the equation since $\left(\chi_E^{(0)}\right)\left(\chi_E^{(0)}\right)^{++} = 1$.

Now consider the RHS of (3.52), which has the form (see equation (3.46))

$$\log F_E \left(\chi^\pm (\theta + i\phi) \right) = \log \left[\frac{(1 + A_{ab})(1 + A_{cd})}{(1 + A_{bc})(1 + A_{da})} (\theta \pm i\pi/2 + i\phi) \right] \tag{3.53}$$

Each A_{PQ} is computed by (3.51) and (3.49). For the RHS of (3.53) to be decaying it is sufficient for all of the A_{PQ} in (3.53) to be decaying. If all the χ-functions are decaying then from (3.51) and (3.49) it is clear that all of the A_{PQ} will decay; the μ-factors will decay by virtue of the rule (3.37). On the other hand, if all the χ-functions are growing the μ-factors in (3.51) will dominate the RHS of (3.51) so that A_{PQ} is still decaying; to see this one should re-express the μ-factors in terms of the coordinates using (3.37). Thus

the RHS of (3.53) will decay if all of the χ-functions are growing, or alternatively if they are all decaying. For generic ϕ it will generally not be true that the RHS of (3.53) is well behaved, and one must try to find a range of ϕ-values for which the χ_E are all decaying or are all growing. If a suitable ϕ can be found, then (3.52) can be directly solved by Fourier-transform. In all of the examples we have considered (in particular, those relevant for the 4-point function) it has been possible to find such a ϕ.

Concerning the issue of singularities within the strip of inversion, it follows from (3.19) that the Wronskians $(s_a \wedge s_b)(\theta)$ are (in an appropriate normalization) analytic away from $\theta = \pm\infty$. It is, however, possible for these objects to have *zeros* and in the following it is an assumption that there are no zeros in the strip where we do the inversion.[14] In section 3.4.1 we perform numerical tests that support this assumption.

Finally, we use the Fourier analysis to obtain

$$\log X_E(\theta) = \log X_E^{(0)}(\theta) - \int_{\mathbb{R}} \frac{d\theta'}{2\pi i} \frac{\log F_E(X(\theta'))}{\sinh(\theta' - \theta + i0)} \tag{3.54}$$

where $X_E(\theta) = \chi_E(\theta + i\phi - i\pi/2)$ and $X_E^{(0)}$ is the (shifted) asymptotic (3.40) and $F_E(X)$ is an explicit function of the coordinates which follows from the discussion of the previous section.

The equations (3.54) can easily be solved for the X_E by iterating them in a computer. In the next subsection we will show how to extract the η-cycles of formula (3.18) from the X_E which are computed using (3.54). We will then perform some numerical tests in section 3.4.1.

3.3.10 Extracting η-cycles

Once the coordinates are computed according to the prescription of the preceding section we extract the η-cycles as follows. What we need to compute are the individual Wronskians $(s_a \wedge s_b)$. For this, note that from (3.47) and footnote 12 we have

$$(1 + A_{ab}) \quad = \quad \frac{(s_a \wedge s_b)(\tilde{s}_a \wedge \tilde{s}_b)}{(s_a \wedge \tilde{s}_a)(s_b \wedge \tilde{s}_b)} \tag{3.55}$$

[14] In the limit where the WKB approximation holds, i.e. when $\theta \to \pm\infty$ or in the limit of large zero modes $|Z_E| \to \infty$ [52], it is clear that (in an appropriate normalization) the Wronskians will not have any zeros since (suppose we compute the Wronskian near P_b) then s_a will be the *big* solution near P_b and is thus linearly independent of s_b which is small at P_b. For finite values of θ (or alternately of $|Z_E|$) we have no concrete way of arguing that these zeros are not present.

We can choose a guage where $(s_P \wedge \tilde{s}_P) = 1$. The final result will be gauge independent. With this gauge choice we have

$$\log(s_a \wedge s_b)^- + \log(s_a \wedge s_b)^+ = \log\left(1 + A_{ab}^-\right) \tag{3.56}$$

Here we will use the notation $\theta \to \theta + i\phi$ where θ and ϕ are real. We then insert the zero-modes on the LHS in the same way as for the χ-system (see section 3.3.9). We are only interested in P_a and P_b that are connected by a WKB line when $\mathrm{Arg}(\xi) = \phi$, and thus we have good control over the asymptotics in the required strip. Performing the Fourier transforms we obtain

$$\log(s_a \wedge s_b)(\theta + i\phi) = \left(\frac{1}{2}e^{-\theta - i\phi}\varpi_{ab} + \frac{1}{2}e^{\theta + i\phi}\bar{\varpi}_{ab}\right) + \int_{\mathbb{R}} \frac{d\theta'}{2\pi} \frac{\log\left(1 + A_{ab}^-(\theta' + i\phi)\right)}{\cosh(\theta - \theta')} \tag{3.57}$$

where we have defined

$$\varpi_{ab} \equiv \lim_{w_a' \to w_a} \lim_{w_b' \to w_b} \left[\int_{E_{ab}} \sqrt{T} dw + \frac{\Delta_a}{2}\log(w_a - w_a') + \frac{\Delta_b}{2}\log(w_b - w_b')\right] \tag{3.58}$$

The integration in (3.58) is performed along edge E_{ab}. The direction of integration is the same as the direction of the edge E_{ab} (see appendix B.3). Note that the logarithmic terms precisely cancel the divergence from the endpoints of integration in (3.58) so that the ϖ_{ab} are finite. In going from (3.56) to (3.57) we have used the asymptotics for $(s_a \wedge s_b)$ derived in appendix B.3.

Expanding (3.57) around $\theta \to -\infty$, and comparing with (3.26) with $\xi = e^{\theta + i\phi}$ we read off

$$\int_{E_{ab}} \eta = \int_{\mathbb{R}} \frac{d\theta}{\pi} e^{-\theta - i\phi} \log\left(1 + A_{ab}^-(\theta + i\phi)\right) \tag{3.59}$$

The contour of integration in $\int_{E_{ab}} \eta$ is along the WKB line connecting P_a and P_b and the direction of integration is the same as the direction of the edge E_{ab}. This formula allows us to compute the η-cycles from the χ-functions since the A_{PQ} are explicit functions of the coordinates.

3.4 The AdS action

3.4.1 Regularized AdS action

Now that we have introduced the needed tools we are ready to calculate the action (3.18). We will demonstrate for the case of the 4-point function, but the method is general and

could be performed for any number of operators inserted along a line. The computation will be as follows. First we will introduce the relevant WKB triangulation which will be topologically equivalent to the triangulation shown in figure 3.9. Second, using the procedure of section 3.3.8 we will derive the χ-system satisfied by the coordinates of this triangulation. Supplementing these functional relations by the WKB asymptotics we will invert these functional relations using the technique of section 3.3.9 to obtain a set of integral equations that uniquely determine the coordinates. Finally, from coordinates we extract the η-cycles using the method of section 3.3.10. Once we have the η-cycles, we compute the action using (3.18).

Stress-energy tensor and WKB triangulation

For the purpose of the following computation, a useful parameterization of the stress energy tensor is

$$T(w) = \frac{1}{(w-w_4)^2}\left(c_\infty + \frac{c_0 + c_1 w + c_2 w^2 + U w^3}{(1+w)^2(1-w)^2}\right) \tag{3.60}$$

Here we have fixed three of the insertion points at $w_1 = +1$, $w_2 = \infty$, $w_3 = -1$ using the world-sheet conformal symmetry. The fourth insertion point is left at the position w_4 which should be fixed at the saddle point w_4^* once the full action is assembled. For the purpose of demonstration we will take w_4 to be between $w_3 = -1$ and $w_1 = +1$. When the dominant saddle point is located in one of the other intervals one can proceed by a similar procedure. The constants $c_a = c_a(w_4, \Delta)$ are functions of w_4 and dimensions of the operators and are fixed by the condition (3.7). Their explicit expressions are given in appendix B.6. The parameter U is unfixed by the condition (3.7) and implicitly parameterizes the cross ratio of the four operators (recall that they are inserted along a line in the boundary theory so that there is only one cross ratio). The analytic structure of T, the resulting WKB-structure and the WKB triangulation are shown in figure 3.15.

χ-system for the 4-point function

From equation (3.46) and figure 3.15 we have

$$\chi_{24}\chi_{24}^{++} = \left(\chi_{\hat{24}}\chi_{\hat{24}}^{++}\right)^{-1} = \frac{(1+A_{23})(1+A_{14})}{(1+A_{34})(1+A_{12})} \tag{3.61}$$

$$\chi_{12}\chi_{12}^{++} = \left(\chi_{14}\chi_{14}^{++}\right)^{-1} = \chi_{34}\chi_{34}^{++} = \left(\chi_{23}\chi_{23}^{++}\right)^{-1} = \frac{(1+A_{24})}{(1+A_{\hat{24}})} \tag{3.62}$$

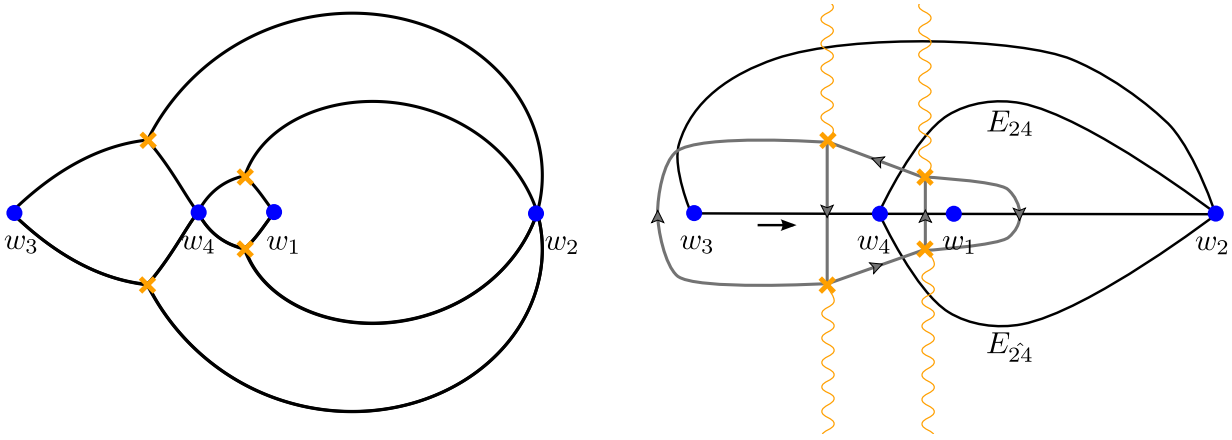


Figure 3.15: Constructing the triangulation for the 4-point function. In the left panel we show the WKB cells for $\mathrm{Arg}\,(\xi) = 0$. The cell walls are formed by the separating WKB curves as described in section 3.3.4; as described there, inside each cell there is a 1-parameter family of generic WKB curves and by taking a representative curve from each family we obtain the triangulation shown in the right panel. In the right panel the black lines are the edges of the WKB triangulation and the wavy yellow lines show our convention for defining the branches of ω. Notice that this triangulation is topologically equivalent to the one shown in figure 3.9. This means that we can borrow the results derived for that example. In particular, the coordinates can be carried over from that example by making the proper identifications. The cycles corresponding to each coordinate are represented by the gray curves – we show only the portion of each cycle on the sheet of ω where the edge E_{34} has orientation *towards* P_4 as indicated by the black arrow along edge E_{34}.

To compute each A_{PQ} we use formulas (3.51) and (3.49) along with the rules given in figure 3.14. In that way we find

$$A_{24} = \frac{\chi_{24}\,(1+\chi_{12}\,(1+\chi_{\hat{2}\hat{4}}\,(1+\chi_{23})))\,(1+\chi_{43}\,(1+\chi_{\hat{4}\hat{2}}\,(1+\chi_{41})))}{(1-\mu_2^2)\,(1-\mu_4^2)} \tag{3.63}$$

$$A_{23} = \frac{\chi_{23}\,(1+\chi_{34})\,(1+\chi_{24}\,(1+\chi_{12}\,(1+\chi_{\hat{2}\hat{4}})))}{(1-\mu_2^2)\,(1-\mu_3^2)} \tag{3.64}$$

with the rest of the A_{PQ} being related by relabeling (see appendix B.6 for the explicit formulas). These expressions and equations (3.61) – (3.62) provide a closed system of functional equations for the 6 coordinates associated with the triangulation shown in figure 3.15.

These functional equations can be converted into integral equations of the form (3.54) using the technique described in section 3.3.9. To apply the procedure of section 3.3.9 one must find a ϕ such that the RHS of (3.52) is decaying, and for this one should appeal to the WKB analysis. The WKB cycles which determine the asymptotics of the coordinates are shown in figure 3.15. When $\Delta_1 \sim \Delta_3$ and $U \sim 0, w_4 \sim 0$ the cycles shown in figure 3.15 all have $\text{Arg}(\oint_{\gamma_E} \omega) \sim \pi/2$.[15] In this case $\phi = 0$ is a suitable choice since then all χ_E^- will be growing and (3.53) will decay rapidly.[16] In summary, the integral equations in the region of present interest are given by equations (3.54) with F_E given by (3.61) – (3.64). These equations will remain valid for all values of the parameters Δ_a, U, and w_4 such that the triangulation is unchanged. If one deforms these parameters too much the triangulation will jump. One can then easily write the χ-system for the new triangulation and apply the same procedure to obtain the integral equations for that region of parameters.[17]

By numerically iterating these equations (using $\chi_E^{(0)}$ as the initial iterate for each χ_E) we obtain the χ-functions. The η-cycles are then extracted from the χ-functions using the procedure of section (3.3.10). In the following section we will write the regularized AdS action in terms of these η-cycles.

Finite part of AdS action

Now that we are able to compute the η-cycles (see previous subsection) we can use the formula

$$A_{fin} = \int_\Sigma \sqrt{T\bar{T}}\,(\cosh\gamma - 1) = \frac{\pi}{3} - \frac{i}{2}\left(\oint_{\gamma_a}\omega\right) I^{-1}_{ab}\left(\oint_{\gamma_b}\eta\right) . \tag{3.65}$$

(see section 3.2.2 and equation (3.18)) to compute the regularized part of the AdS action. To use (3.65) there are few steps. These steps are simple but tedious and we will only list them here (see appendix B.6 for a detailed implementation). As described in section 3.2.2 one should first modify T by spreading the double poles slightly such that $\omega = \sqrt{T}dw$ has an additional square-root cut at each of these points. Then one should choose a complete

[15] Interestingly, when $\Delta_1 = \Delta_3$ and $U = w_4 = 0$ there is a symmetry which causes the RHS of the χ-system to trivialize (i.e. to reduce to 1 for all χ_E) and the χ-functions can be computed explicitly (they are just equal to their zero-mode part). This is reminiscent of the case for the three-point function and, in fact, there is also a change of coordinates that maps the specific case $\Delta_1 = \Delta_3$ and $U = w_4 = 0$ to two copies of a three-point function.

[16] This will continue to be the case as long as the $\text{Arg}(\oint_{\gamma_E}\omega)$ remain in the upper-half plane. In other-words, the inversion procedure will be valid for all U and w_4 such that the triangulation is unchanged since the triangulation will jump precisely when one of the $\oint_{\gamma_E}\omega$ crosses the real-axis [52].

[17] Another (more elegant) approach would be to find a systematic way of analytically continuing the integral equations from one region of parameters to another as was done for the TBA equations of [19].

basis of a- and b-cycles (five of each is needed for the 4-point function). One can then apply formula (3.18) and then take the limit in which the small cuts close to form simple poles in ω. Once this is done the area will generically be expressed in terms of three different types of η-cycles: cycles connecting two punctures, cycles connecting a puncture with a zero and cycles connecting two zeros. The latter two can be expressed as linear combinations of the puncture-puncture cycles as described in appendix B.6. Once this is done, the final result takes the elegant form

$$A_{fin} = \frac{\pi}{3} - i \sum_{E \in \mathcal{T}} \omega_E \, \eta_E \tag{3.66}$$

where the sum runs over the edges in the triangulation (see figure 3.15), $\eta_{E_{ab}}$ is defined in (3.59) while $\omega_{E_{ab}}$ is the ω-cycles that intersects edge E_{ab} (i.e. the integral of ω that is associated with the coordinate χ_{ab}; these integrals are shown as the gray contours in figure 3.15).[18]

Formula (3.66) and the procedure of section 3.3 for computing the η-cycles solve the problem of computing the regularized AdS contribution to the 4-point function. In the next section we present some numerical tests of the procedure. Let us note that the procedure of section 3.3 is general and can be implemented for any number of punctures. Further, while we have only proved equation (3.66) for the case of the 4-point function, given its simplicity one might suspect that the formula holds in general (with $\pi/3 \to \pi/12 \times$ (#number of zeros of T), of course).[19] Even if the general result does not take the simple form (3.66), for a given T (i.e. for any number of punctures) the procedure described in section 3.2.2 is still valid and one can still write A_{fin} in terms of the η_E for the corresponding triangulation). In principle this solves the problem of computing the regularized AdS contribution to the N-point function. We have performed numerical tests only for the case of the 4-point function. We present these numerical results in the following section.

Numerical tests

We now present numerical tests of the method described above. We solved numerically the modified sinh-Gordon equation (3.10) for the function γ and then using this numerical

[18] Note that in formula (3.66) both integrals ω_E and η_E are the *segment* integrals between the appropriate limits. For example, the $\omega_E = \frac{1}{2} \oint_{\gamma_E} \omega$. In this sense we are abusive with the term 'cycle'.

[19] It would be a simple matter to check this, but we have not pursued this issue. We did check that the formula holds for the 3-point function (see appendix B.7).

U	Δ_3	Δ_4	Δ_1	Δ_2	Numerics	χ-system
1/5	1	2	1	2	0.84807	0.84812
1/2	1	2	1	2	0.82421	0.82423

Table 3.1: Comparison of the A_{fin} obtained by numerically integrating (3.10) and the area computed from the χ-system. The results are for the spike configuration of figure 3.4B.

solution to directly compute A_{fin} via

$$\int_\Sigma \sqrt{T\bar{T}}\,(\cosh\gamma - 1) \tag{3.67}$$

The general set-up of the numerical problem essentially follows that of [38] with some modifications. However, the numerical method that we use to solve the PDE (3.10) is quite different from that of [38].[20] We place the punctures at $w_3 = -1$, $w_4 = 0$, $w_1 = 1$, and $w_2 = \infty$. We then map the sphere to a square domain with the point at infinity mapping to the boundary of the square and the real axis mapping onto itself. Since γ must vanish at the punctures, we should impose $\gamma = 0$ along the boundary of the square domain since w_2 maps to the boundary of the square in the new coordinates. Further, since for either configuration of spikes (see section 3.2.3 and figure 3.4) there is a fold-line along the real axis, we know $\gamma\,(x, 0) = 0$ where we are using the coordinates $w = x + iy$ and writing $\gamma = \gamma(x, y)$. Thus we can solve the problem in half of the square with the Dirichlet boundary conditions $\gamma = 0$ on the boundaries. Lastly, we must remove the logarithmic singularities (3.11) in order to have a nice smooth function to solve for. A suitable function is

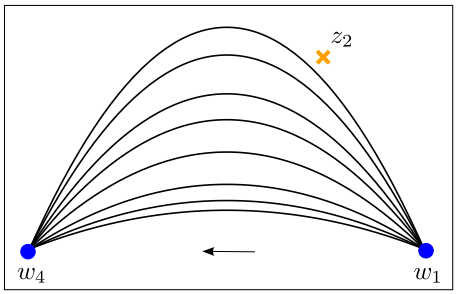
$$2\gamma_{reg} = \gamma + \frac{1}{2}\sum_a \sigma_a \log\left[\frac{(w - z_a)\,(\bar{w} - z_a)}{(1 + w\bar{w})}\right] \tag{3.68}$$

where we $\sigma_a = \pm 1$ is determined by $\gamma \sim -\sigma_a \frac{1}{2}\log T\bar{T}$ at z_a. The numerator of (3.68) removes the log divergences (3.11) in γ while the denominator is included to kill off these additional log terms at infinity. In the numerical implementation we fix the spike configuration we want to describe by choosing the set of $\{\sigma_a\}$. Finally, to numerically integrate the equation (3.10) (re-written in terms of γ_{reg}, of course) we use a standard relaxation method with an uniform grid.

In table 3.1 we compare the numerical results with the analytic results. The numerical results are obtained by the area computed using (3.67) with the numerical solution for γ.

[20] We are very grateful to Romuald Janik for providing us with a copy of the code used in [38] which was very useful in helping us to develop and test our own numerics.

$\eta_{14}[A]$	$\eta_{14}[B]$
−0.033166	−0.031624
−0.033166	−0.031624
−0.033167	−0.031626
−0.033168	−0.031628
−0.033169	−0.031629
−0.033168	−0.031629
−0.033167	−0.031627
−0.033166	−0.031626



$\eta_{34}[A]$	$\eta_{34}[B]$
−0.014496	−0.013581
−0.014497	−0.013582
−0.014499	−0.013585
−0.014501	−0.013587
−0.014503	−0.013589
−0.014503	−0.013589
−0.014502	−0.013589
−0.014502	−0.013588

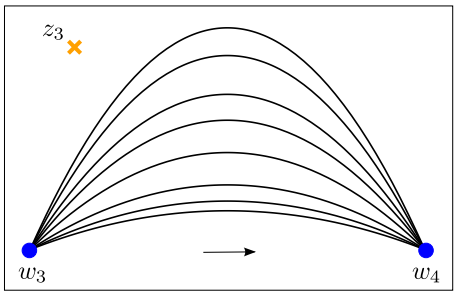


Figure 3.16: Here we show the values of η_{14} and η_{34} evaluated along several different contours. For example, the column labeled $\eta_{14}[A]([B])$ shows the values of η_{14} for the spike configuration of figure 3.4$A(B)$ for each of the contours shown to the right of the column. We use the parameter values $\Delta_3 = \Delta_1 = 1$, $\Delta_2 = \Delta_4 = 2$ and $U = 1/5$ for both spike configurations. There are five digits that we trust since they are unchanged for the different contours and they should be compared with our result from the functional equations that is $\eta_{14}^{\chi\text{–system}}[A] \approx -0.033169$, $\eta_{34}^{\chi\text{–system}}[A] \approx -0.014503$ and $\eta_{14}^{\chi\text{–system}}[B] \approx -0.031628$, $\eta_{34}^{\chi\text{–system}}[B] \approx -0.013588$. In the digits where the forms are closed there is perfect agreement with the analytic results.

The analytic result is obtained from (3.66) with the η-cycles computed using the χ-system procedure. These results show a good agreement of our formula with the numerics.

A sharper measure of the agreement between the analytics and numerics is to compare directly the η-cycles. In figure 3.16 we show the numerical results for η_{14} and η_{34} computed along several different contours. This allows us to test the closure of the numerical η which we obtain from the numerical γ via (3.16). Note that closure of η implies that γ must obey (3.10) and thus this is a good measure of the numerical error. Indeed, one can see in figure 3.16 that the numerical cycles agree with the analytical predictions in all digits in which they are closed. That is, the numerics is in agreement with the analytics in all of the digits for which the numerics can be trusted.

Finally, it would be interesting to perform numerical tests for a larger portion of the parameter space (i.e. more values of the Δ_a, U and w_4). To perform a systematic study will probably require an improvement of our numerical method as our current method, while extremely simple, has very slow convergence.

3.4.2 Divergent part

In section 3.4.1 we completed the task of computing the first term in formula (3.14). In this section we will discuss the second term

$$-\int_{\Sigma\setminus\{\epsilon_a\}} d^2w\sqrt{T\bar{T}} = -\frac{\pi}{2}\sum_a \Delta_a^2 \log\epsilon_a - A_{reg} \tag{3.69}$$

where A_{reg} is finite at $\epsilon_a \to 0$. The contribution A_{reg} can be computed by simple but tedious application of the Riemann bilinear identity and there are many ways to write the result. For example

$$A_{reg} = i\sum_{E\in\mathcal{T}} \varpi_E\omega_E - i\frac{1}{2}(\varpi_{24} - \varpi_{\hat{24}})(\omega_{24} - \omega_{\hat{24}}) \tag{3.70}$$

where the sum is over the triangulation shown in figure 3.15 and $\varpi_{E_{ab}} \equiv \varpi_{ab}$ is defined in (3.58). The ω_E are defined in the same way as in (3.66). One can check this formula by comparing with the direct 2D numerical integration of $\sqrt{T\bar{T}}$ with small circular disks cut out around the puncture (in Mathematica one can use NIntegrate along with the Boole command, for example).

We recall that (3.69) came from the regularization of the string action where we have added and subtracted $\sqrt{T\bar{T}}$ from the integrand of the AdS action. This integral depends explicitly on the cut-off ϵ_a around the punctures. It will be important to understand the connection with the physical cut-off $\mathcal{E}$ at the boundary of AdS. Fortunately we can extract the needed information from the linear problem since we have good analytic control over the solutions near the insertion points. To proceed by this route (which parallels the discussion of [38] for the 3-point function) we must first describe how the string embedding coordinates are recovered from the linear problem formalism, which is via the aptly-named *reconstruction formulas*. We will discuss this in the next subsection, 3.4.2. After that, in section 3.4.2 we will use the reconstruction formulas to eliminate the ϵ_a in favor of $\mathcal{E}$. From this procedure we will recover the standard spacetime dependence in (3.1) along with a contribution to the function $f(u,v)$. This will complete the computation of the semiclassical AdS contribution to (3.1).

Reconstruction formulas

The reconstruction formulas allow us to express the string embedding coordinates in terms of solutions of the linear problem. This point is crucial in our construction for the following reasons. First, we have introduced some regulators in the world-sheet, ϵ_a, that must be related to the physical cut-off in the boundary of AdS, $z = \mathcal{E}$. Second, by using them we will be able make the spatial dependence explicit in the final result, namely the insertion points x_a of the operators in the gauge theory.

Consider two solutions of the linear problem, ψ_A and ψ_B normalized as $(\psi_A \wedge \psi_B) = 1$, and construct a matrix Ψ as

$$\Psi = (\psi_A \ \psi_B)\,. \tag{3.71}$$

The matrix Ψ obeys the same equations of motion as $\psi_{A,B}$ (3.19), namely

$$(\partial + J_w)\Psi = 0, \quad (\bar{\partial} + J_{\bar{w}})\Psi = 0\,. \tag{3.72}$$

where J_w and $J_{\bar{w}}$ are defined in (3.19)-(3.20). One can verify using (3.9) that the quantity

$$y^I \equiv -\frac{1}{2}\,\mathrm{Tr}\left(\tilde{\sigma}^I \sigma^2 \Psi^T \sigma^1 \Psi\right)\big|_{\theta=0} \tag{3.73}$$

with $\tilde{\sigma}^1 = \sigma^1$, $\tilde{\sigma}^2 = -i\sigma^2$, $\tilde{\sigma}^3 = \sigma^3$, satisifes the same equations of motion as Y^I and also the constraint $y \cdot y = -1$ (with the AdS metric). In this way we establish a correspondence between target space coordinates and solutions of the linear problem,

$$\frac{1}{z} = Y^2 - Y^1 = 2i\,\Psi_{11}\Psi_{21}, \qquad \frac{x}{z} = Y^3 = i\,(\Psi_{11}\Psi_{22} + \Psi_{12}\Psi_{21}) \tag{3.74}$$

In order to relate the operator insertion points x_a and physical cut-off $\mathcal{E}$ with the linear problem data, it is convenient to express ψ_A and ψ_B in terms of the elementary solutions s_a and $\tilde{s}_a$ whose behavior close to the punctures is given by (3.21),

$$\psi_A = (\psi_A \wedge \tilde{s}_a)\,s_a + (s_a \wedge \psi_A)\,\tilde{s}_a, \qquad \psi_B = (\psi_B \wedge \tilde{s}_a)\,s_a + (s_a \wedge \psi_B)\,\tilde{s}_a \tag{3.75}$$

Close to the punctures the solution $\tilde{s}_a$ becomes dominant. Then, using (3.74) and the explicit form of $\tilde{s}_a$ close to the puncture P_a we get that

$$z = \frac{1}{i\,(s_a \wedge \psi_A)_0^2}|w - w_a|^{\Delta_a} \tag{3.76}$$

where the subscript 0 indicates that the solutions are evaluated at $\theta = 0$ (recall that this is the value where the physical problem is recovered – see equation (3.73)). Equation (3.76) is the relation needed to make the connection between the world-sheet and physical cut-off's

$$\Delta_a \log \epsilon_a = \log \mathcal{E} + \log |\,(s_a \wedge \psi_A)\,|_0^2 \tag{3.77}$$

Finally, using once again (3.74) we express the insertion points x_a of the operators in the gauge theory as

$$x_a = \frac{(s_a \wedge \psi_B)_0}{(s_a \wedge \psi_A)_0} \tag{3.78}$$

Physical regulator and spacetime dependence

We can now use (3.77) to eliminate the ϵ_a in (3.69) in favor of the physical cut-off at the boundary of AdS $z = \mathcal{E}$. We have

$$\sum_a \Delta_a^2 \log \epsilon_a = \left(\sum_a \Delta_a \log \mathcal{E} + \sum_a \Delta_a \log |\,(s_a \wedge \psi_A)\,|_0^2 \right) \tag{3.79}$$

where a and A refer respectively to the small solution s_a and one generic solution ψ_A appearing in the reconstruction formulas. Now we will eliminate the factors $|\,(s_a \wedge \psi_A)\,|_0$ in terms of objects that we can compute.

The terms $|\,(s_a \wedge \psi_A)\,|_0^2$ can be related to the insertion points x_a in target space and overlaps of the elementary solutions evaluated at $\theta = 0$ through expression (3.78). Using Schouten's identity one can verify that

$$|\,(s_a \wedge \psi_A)\,|_0^2 = \frac{x_{bc}}{x_{ba} x_{ca}} \frac{|\,(s_b \wedge s_a)\,|_0 |\,(s_c \wedge s_a)\,|_0}{|\,(s_c \wedge s_b)\,|_0} \tag{3.80}$$

for a, b, c distinct. This solution is unique up to different ways of rewriting the spatial dependence using the cross-ratio

$$u = \frac{x_{14} x_{23}}{x_{12} x_{34}} = \frac{(s_1 \wedge s_4)_0 \, (s_2 \wedge s_3)_0}{(s_1 \wedge s_2)_0 \, (s_3 \wedge s_4)_0} \tag{3.81}$$

where we have used (3.78). Note that we can compute the brackets appearing in (3.77)-(3.78) using (3.57). In particular we have

$$\log (s_a \wedge s_b)_0 = \left(\frac{1}{2} \varpi_{ab} + \frac{1}{2} \bar{\varpi}_{ab} \right) + \int_{\mathbb{R}} \frac{d\theta}{2\pi} \frac{\log \left(1 + A_{ab}^-\right)}{\cosh \theta} \tag{3.82}$$

This formula is valid when there is a WKB line connecting P_a and P_b. If a bracket appears for which we do not have a WKB line, we can simply use the cross ratio (3.81) to eliminate it in terms of brackets that can be computed using (3.82).

Finally, using (3.80) in (3.79) and massaging the resulting spacetime dependence by extracting multiples of u and $(1+u)$ we find

$$e^{2\times\frac{\sqrt{\lambda}}{2}\Delta_a^2\log\epsilon_a}=\prod_{a>b}^{4}\left(|s_a\wedge s_b|_0\right)^{-\sqrt{\lambda}\Delta_{ab}}\left(\frac{x_{ab}}{\mathcal{E}}\right)^{\sqrt{\lambda}\Delta_{ab}} \tag{3.83}$$

where $\Delta_{ab}=\left(\sum_c\Delta_c\right)/3-\Delta_a-\Delta_b$. The extra factor of 2 in the exponent on the left hand side of (3.83) anticipates the sphere regularization which turns out to be similar to the AdS part and will be treated in section 3.5.1.

We recognize in (3.83) the canonical spacetime dependence in the 4-point function of a conformal field theory (compare with equation (3.1)). The appearance of the cut-off in (3.83) is related to the renormalization of the operators. In fact, if we define $\tilde{\mathcal{O}}_{\Delta_a}\equiv\mathcal{E}^{\Delta_a}\mathcal{O}_{\Delta_a}$ this will cancel the $\mathcal{E}$ factors in (3.83). To be more precise, we should define a 4-point function that is independent of the operator renormalization. The standard procedure is to divide by the appropriate product of 2-point functions such that normalization factors cancel. The same factors of $\mathcal{E}$ will appear in these 2-point functions and will cancel with those in (3.83). We will thus drop the factors of $\mathcal{E}$ in the formulas below.

3.4.3 Summary of the AdS and divergent contributions

We have now computed all the parts of (3.14). In this section we summarize the full result. The semiclassical limit of the 4-point function (3.1) is given by

$$\left(f_{fin}^{AdS}f_{div}^{AdS\times S}f_{fin}^{S}\right)^{*}\prod_{a<b}^{4}(x_{ab})^{\Delta_{ab}}\,, \tag{3.84}$$

where the $*$ denotes evaluation at $w_4=w_4^*$ and we define

$$f_{fin}^{AdS}\left(w_4\right) = e^{-\frac{\sqrt{\lambda}}{\pi}A_{fin}} \tag{3.85}$$

$$f_{div}^{AdS\times S}\left(w_4\right) = e^{-2\frac{\sqrt{\lambda}}{\pi}A_{reg}}\prod_{a>b}^{4}\left(|s_a\wedge s_b|_0\right)^{-\sqrt{\lambda}\Delta_{ab}} \tag{3.86}$$

and f_{fin}^{S} will be defined momentarily. The contribution A_{fin} is given by (3.66), A_{reg} is given in (3.70), the brackets in $f_{div}^{AdS\times S}$ are given by (3.82).

The sphere part of the correlation function contains divergences of the same type as

AdS. We therefore regularize it also by subtracting $\sqrt{T\bar{T}}$. Such finite contribution is what we denote by f^S_{fin}

$$f^S_{fin} \equiv e^{-\frac{\sqrt{\lambda}}{\pi}\int_\Sigma \left(S^5 \text{ contribution} - \sqrt{T\bar{T}}\right)} . \tag{3.87}$$

where S^5 contribution stands for the S^5 Lagrangian and wavefunctions [38]. To compensate this subtraction, we include the factor of 2 in front of A_{reg} in expression (3.86). In general we cannot complete the construction of the 4-point function because we are unable to compute the contribution f^S_{fin}. Fortunately, for correlators involving only BPS operators of the same type (e.g. only Z and $\bar{Z}$) the sphere part is known and we can assemble the full result. This is the subject of the next section.

3.5 Full correlation function for BMN operators

In this section we compute the full correlation function for operators of the type $\mathrm{Tr} Z^\Delta$ when Δ scales as $\sqrt{\lambda}$. For these type of operators, the sphere part f^S_{fin} was already known [57] and therefore we can complete our computation. We stress that, unlike the three point function, this four point correlator is not protected. In section 3.5.2, we fix the location of the puncture w_4 by the saddle point method and discuss some issues on the multiple string embedding configurations. In section 3.5.3 we perform an analytical check of our procedure by studying the extremal limit where $\Delta_2 = \Delta_1 + \Delta_3 + \Delta_4$, which is known to be protected from quantum corrections.

3.5.1 Sphere part

The sphere part of the correlation function involves the classical wave-functions associated to the external states. We consider specifically the correlation function of four BMN operators[21]

$$\langle \mathrm{Tr} Z^{\hat{\Delta}_1}(x_1)\ \mathrm{Tr} Z^{\hat{\Delta}_2}(x_2)\ \mathrm{Tr}\bar{Z}^{\hat{\Delta}_3}(x_3)\ \mathrm{Tr}\bar{Z}^{\hat{\Delta}_4}(x_4)\ \rangle\,, \tag{3.88}$$

for which the wave-functions are known [58, 59]. The string dual of these operators corresponds geometrically to a string that is point-like in the sphere and rotates around an equator [60]. The surface developed by the worldsheet is not extended in the sphere.

Let X_i $(i = 1, ..., 6)$ be the coordinates in S^5. This particular string state can be expressed as

$$X_1 + iX_2 = e^{i\varphi} \qquad\qquad X_i = 0, \quad i = 3, \ldots, 6 \tag{3.89}$$

[21] We are using the following notation for the dimensions of the operators $\hat{\Delta}_a = \sqrt{\lambda}\Delta_a$.

where φ is an azimuthal angle of the sphere. The wave-functions for $\mathrm{Tr}Z^{\hat{\Delta}_a}$ and $\mathrm{Tr}\bar{Z}^{\hat{\Delta}_a}$ are given respectively by

$$\Psi_{\hat{\Delta}_a} = e^{i\hat{\Delta}_a\varphi(w_a,\bar{w}_a)}, \qquad \bar{\Psi}_{\hat{\Delta}_a} = e^{-i\hat{\Delta}_a\varphi(w_a,\bar{w}_a)} \tag{3.90}$$

where the field φ is evaluated at the puncture corresponding to the respective operator insertion.

As the wave-functions scale exponentially with $\sqrt{\lambda}$, they will act as sources for the equations of motion for φ. The total sphere contribution is then given by

$$\exp\left[-\frac{\sqrt{\lambda}}{\pi}\left(\int d^2w\,\partial\varphi\bar{\partial}\varphi + i\pi\left(\Delta_3\varphi_{w=-1} + \Delta_4\varphi_{w=w_4} - \Delta_1\varphi_{w=1} - \Delta_2\varphi_{w=\infty}\right)\right)\right]. \tag{3.91}$$

Considering both the contributions from the S^5 action and wave-functions as an effective action, we obtain the equations of motion for φ which are solved by

$$\varphi(w,\bar{w}) = i\left(\Delta_3\log|w+1| + \Delta_4\log|w-w_4| - \Delta_1\log|w-1|\right). \tag{3.92}$$

This solution has an additional singularity at infinity with charge $-\Delta_3 - \Delta_4 + \Delta_1 (\equiv -\Delta_2)$, corresponding to the wave-function inserted at infinity. This is consistent with R-charge conservation. We may now plug (3.92) into (3.91), introducing cut-off's around the punctures to regulate this contribution. This amounts to evaluate the solution at a distance ϵ away from the punctures. As in the case of the AdS action, the logarithmic divergences

$$\exp\left[\frac{\sqrt{\lambda}}{2}\sum_i \Delta_i^2\log\epsilon_i\right] \tag{3.93}$$

need to be regularized. We do this by subtracting $\sqrt{T\bar{T}}$ from the integrand. To compensate, we add a similar contribution to the divergent part, that was already treated in the previous section (indeed, this regularization procedure is responsible for the factor of 2 appearing in front of A_{reg} in expression for $f_{div}^{AdS\times S}$, see (3.86)). The dependence on the cut-off's then disappears yielding the following expression for the regularized sphere action and wave-functions

$$f_{fin}^S = \exp\left[\sqrt{\lambda}\left(A_{reg} - \log 2^{\Delta_3\Delta_1} - \log\frac{|w_4-1|^{\Delta_1\Delta_4}}{|w_4+1|^{\Delta_3\Delta_4}}\right)\right], \tag{3.94}$$

where f_{fin}^S was defined in (3.87)

3.5.2 Fixing the fourth insertion point

We have shown how to compute the quantities (3.85)-(3.86) as a general function of w_4. However, to compute (3.84) we must evaluate at the saddle point $w_4 = w_4^*$. Let us now describe how we determine the saddle point in practice. All the physical information that we input is contained in the stress-energy tensor. Besides the conformal dimensions of the operators and the position of the punctures, there is the extra parameter U that translates the additional degree of freedom of the cross ratio. The two are implicitly connected by the expression

$$\chi_{24}(\theta = 0; U) = \frac{x_{14}\, x_{23}}{x_{12}\, x_{34}} \equiv u \tag{3.95}$$

by formulas (3.81) and (3.34). If we use (3.95) to fix the parameter U than we have one unfixed parameter w_4. This remaining freedom allows us to impose the Virasoro constraint

$$T \equiv T_{AdS} = -T_S = -(\partial_w \varphi)^2 \tag{3.96}$$

which so far we have not imposed on T. Plugging in (3.92) and using R-charge conservation $\Delta_2 + \Delta_1 = \Delta_3 + \Delta_4$ we find the saddle point is given by

$$w_4 = \frac{\Delta_1(2-U) + \Delta_3(2+U) + \Delta_4 U}{2(\Delta_1 - \Delta_3)} \tag{3.97}$$

Depending on the cross-ratio and the conformal dimensions the saddle point can be located in any of the three intervals along the real axis. In section 3.1 and appendix B.4, we discussed the different AdS_2 string embedding geometries and its connection with the different boundary conditions (3.11) that one can impose on γ. At the level of the functional equations, we have seen that the different boundary conditions manifest in the different $\xi \to 0, \infty$ asymptotics of the coordinates. More precisely they will affect the constant $C_E^{(0)}$ in the expression (3.40). One may ask which of the configurations in figure 3.2 we should find given a cross ratio and a set of conformal dimensions. We now see that once we fix the external data (cross ratio and conformal dimensions) the saddle point is fixed by (3.97) and thus embedding is also fixed. This is in perfect agreement with the mapping between figure 3.1 and 3.2 and it is non-trivial that the integral equations encode this mapping.

3.5.3 Extremal Limit

In this section, we study the correlation function

$$\langle \mathrm{Tr}\bar{Z}^{\hat{\Delta}}(x_1)\ \mathrm{Tr}Z^{\hat{\Delta}_2}(x_2)\ \mathrm{Tr}\bar{Z}^{\hat{\Delta}}(x_3)\ \mathrm{Tr}\bar{Z}^{\hat{\Delta}_4}(x_4)\rangle \tag{3.98}$$

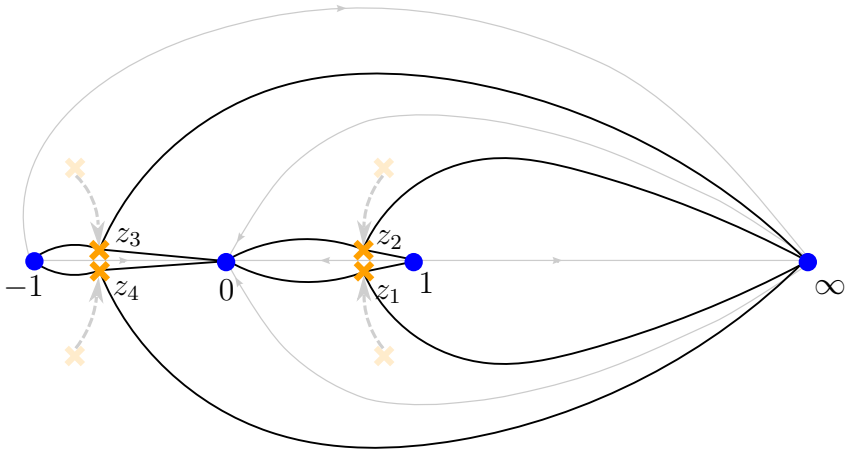


Figure 3.17: In the extremal limit, the main feature is that the zeros collide on the real axis. The black lines represent the WKB cells whereas the gray lines represent the WKB triangulation. At the exact extremal configuration, there are no WKB lines connecting 1 to 0 nor -1 to 0. We interpret this as a manifestation of the field theory fact that at tree level all operators are Wick contracted only to the fourth operator.

in the extremal limit when

$$\Delta_2 = 2\Delta + \Delta_4. \tag{3.99}$$

Such correlator is protected from quantum corrections as conjectured in [61] and later proved in [62]. Thus, we expect to obtain the tree level gauge theory result which in the planar limit is simply given by Wick contractions

$$\frac{1}{x_{12}^{2\hat{\Delta}}\, x_{23}^{2\hat{\Delta}}\, x_{24}^{2\hat{\Delta}_4}}. \tag{3.100}$$

The AdS part of our formula is universal in the sense that it only depends on the dimensions of the operators. On the other hand, the sphere part of the correlation function involves the precise details of the operators inserted. Compared to the previous sphere calculation (3.88), computing (3.98) just amounts to take the complex conjugate of the wave function located at x_1, due to the replacement of $Z \to \bar{Z}$.

Let us start by studying the case when the cross ratio is $u = 1$, where we know the saddle point is $w_4^* = 0$. From this we will be able to see the general mechanism that gives the expected simplification of our result. The first important observation is that in this limit the zeros of $T(w)$ collide on the real axis as depicted in figure 3.17. Let us start by analyzing what this implies at the level of the χ-system. As the integrals ω_{14} and ω_{34}

vanish, the χ's associated to these cycles, namely χ_{34} and χ_{14}, tend to -1. This observation has the remarkable consequence that the right hand side of all equations in the χ-system becomes trivially equal to 1 as one can easily verify [22]. As a result, all χ-functions are *exactly* given by leading term of the WKB expansion (3.40)[23]. For convenience, let us introduce an infinitesimal δ defined by the condition $\delta = 2\Delta + \Delta_4 - \Delta_2$. At the end of the day, we will take $\delta \to 0$. In this limit, the solutions of the χ-system are then given by

$$\chi_{23} = \chi_{12} = -e^{-\frac{\pi(4\Delta-\delta)}{2}\cosh\theta}, \quad \chi_{34} = \chi_{14} = -e^{-\frac{\pi\delta}{2}\cosh\theta}, \quad \chi_{24} = \chi_{\hat{24}} = e^{-\frac{\pi(2\Delta_z-\delta)}{2}\cosh\theta} \tag{3.101}$$

We may now plug this solution in the expression (3.51) and extract the cycles using as described in section 3.3.10. We find that all A's vanish in the limit $\delta \to 0$ except for A_{14} and A_{34}, which tend to -1 as δ goes to zero. This implies that all $\eta_{E_{ab}}$ vanish except for η_{14} and η_{34}, which diverge since the integrand of these cycles becomes singular in this limit. However, one must go back to the area formula (3.66) and realize that such cycles are multiplied by a vanishing quantity. Indeed, (3.66) simplifies to

$$\frac{1}{4}\pi\delta\eta_{14} + \frac{\pi}{3}\,. \tag{3.102}$$

In the limit $\delta \to 0$, the first term of this expression is explicitly given by

$$\delta \int_0^\infty d\theta\, \cosh\theta \log\left(1 - e^{-\frac{1}{2}\pi\delta\cosh\theta}\right) + \mathcal{O}(\delta) = -\frac{\pi}{3} + \mathcal{O}(\delta). \tag{3.103}$$

Hence, it turns out that the finite AdS contribution vanishes in the extremal limit. We believe this is the general mechanism for any value of the cross ratio.

The computation of the sphere contribution follows the same steps as before, with a slight change on one vertex operator (recall that to get the extremal case, we replaced the operator located at x_1 in (3.88) by $\mathrm{Tr}\bar{Z}^{\hat{\Delta}}$). The new solution for the equations of motion is

$$\varphi(w,\bar{w}) = i\left(\Delta\log|w+1| + \Delta_4\log|w| + \Delta\log|w-1|\right)\,. \tag{3.104}$$

[22] This trivialization of the χ-system is general and follows *just* from the fact that the two cycles ω_{14} and ω_{34} vanish which implies that the χ-functions χ_{34} and χ_{14} become -1. In the specific case of $U = w_4 = 0$ and $\Delta_1 = \Delta_3$, which turns out to correspond to cross ratio 1, the χ-system is already trivial because of the symmetry of the stress energy tensor in this particular point of the parameter space, see footnote 15). Nevertheless, we emphasize that the trivialization of the χ-system in general does not rely on this specific symmetry of the stress energy tensor.

[23] Indeed, when the right hand side of the χ-system is 1, the kernel term in equation (3.54) vanishes and we are left with leading WKB contribution.

Now when we compute the contribution of the sphere action and wavefunctions on this solution, we find that it *exactly* cancels the term $\sqrt{T\bar{T}}$ for Δ's satisfying (3.99). Consequently, the sphere part of the correlation function also vanishes in the extremal limit.

The divergent piece in the extremal becomes simply

$$e^{-\frac{\sqrt{\lambda}}{\pi}A_{reg}}\prod_{a>b}^{4}(|s_a\wedge s_b|_0)^{-\sqrt{\lambda}\Delta_{ab}}\to\delta\int_{-\infty}^{\infty}\frac{d\theta}{2\pi\cosh\theta}\log(1-e^{-\frac{1}{2}\pi\delta\cosh\theta})+\mathcal{O}(\delta)=\frac{\delta}{\pi}\log\frac{\pi\delta}{2}+\mathcal{O}(\delta) \tag{3.105}$$

which goes to zero as $\delta\to 0$. We are left with the spatial dependent part which, using that the cross-ratio is 1, can be written as

$$\frac{1}{\left(\frac{x_{12}}{\mathcal{E}}\right)^{2\hat{\Delta}}\left(\frac{x_{23}}{\mathcal{E}}\right)^{2\hat{\Delta}}\left(\frac{x_{24}}{\mathcal{E}}\right)^{2\hat{\Delta}_4}}. \tag{3.106}$$

This is nothing but the tree level result (3.100) of the gauge theory.

3.6 Discussion

In this chapter, we have computed the AdS part of the four point function for heavy scalar operators in $\mathcal{N}=4$ SYM in the classical limit. For the particular case of BPS operators on a line with a single scalar field, the sphere part is known and thus we can construct the full strong coupling four point function.

The main ingredient of our method is the integrability of the string equations of motion in AdS_2. Specifically, we use the method of Pohlmeyer reduction to map the problem to that of solving a certain modified Sinh-Gordon equation which is known to be integrable. We construct the linear problem associated with this equation, which has the form of an $SU(2)$ Hitchin system. This approach was used in the solution of the Null Polygonal Wilson-Loop problem at strong coupling [19] as well as in the study of three-point functions of heavy operators at strong coupling [38–40, 42–44].

Let us mention that while our approach was inspired by these previous works, to solve the $N>3$ point function problem required significant generalization of [19,38–40] as well as nontrivial new ingredients. For the case of the Null-Polygonal Wilson loop the world-sheet has the topology of a disk, whereas in our problem it is that of an N-punctured sphere and this changes the boundary conditions that one imposes. This issue was addressed in [38,39] for the case of the 3-punctured sphere, however in those works the total monodromy condition was enough to derive the functional equations that determine the necessary

objects. These functional equations are linear and can be easily inverted using standard techniques. For the case of 4 or more punctures the situation becomes significantly more complex. First of all, the total monodromy condition no longer provides enough information to fix the necessary objects. We have made heavy use of the formalism developed in [52,53] to derive the functional equations. Second, the inversion of these functional equations is more subtle due to their complexity. The result turns out to be some integral equations resembling the usual TBA equations.

There are a multitude of interesting applications and extensions of the results presented. Let us consider each of these in turn.

- *Multiple configurations and phase transitions.* An important physical outcome of this paper is the emergence of multiple string configurations in AdS_2. Each of these configurations is associated to the existence of several saddle points. A natural question is to figure out whether the dominant saddle point depends on the parameters of the theory. If so it would be interesting to study the phase diagram and the possible transitions.

- *OPE.* A natural question to ask given any 4-point function in a conformal field theory is what can be learned from its OPE decomposition. In particular, important information about the spectrum and structure constants of the theory can be extracted.

- *GKP string.* An interesting aspect of [38,39] is the similarity between the mathematical formalism employed despite the differences in the physical problem: [38] describes strings without spin in AdS_2 whereas [39] describes spinning strings in AdS_3. In the formalism of [39,40], one expresses the N-point function of GKP string in terms of a universal AdS contribution and a contribution from vertex operators, both of which can be computed for the case of the three-point function. It is possible that one could use the formalism developed in this work to calculate the AdS contribution to the N-point function of GKP strings.

- *N-point functions.* The formalism developed here does not depend in any special way on having only 4 punctures and in principle one could use the same methods to study the N-punctured sphere for any N. It would be interesting to understand how the functional equations generalize to higher N. Furthermore, since for the N-point function there will be $N-3$ unfixed insertion points, the moduli space of possible configurations should be quite interesting.

- *TBA equations.* We should note that the techniques developed in [52], in principle, allow one to write the functional equations derived in this paper in the usual form of

a Y-system. Typically this Y-system will involve an infinite number of Y-functions. This form of the equations could be useful for various applications including analytic continuation of parameters and generalization to N-point functions.

- *Generalizing out of the line and WL/CF duality* A natural step would be to generalize this work for operators not inserted on a line. In this case the string is embedded in a higher dimensional AdS space, which involves a more complicated Pohlmeyer reduction scheme. It would be interesting to study the question of whether the multiple string configurations/ saddle points we have found is special to AdS_2 case. Another promising application of such generalization would be the possibility of studying the OPE for the Null Polygonal Wilson Loop [9–17]. One could also investigate the duality between Null Polygonal Wilson Loops and Correlation functions of null separated local operators at strong coupling [63–68].

Many of these points present interesting opportunities to try to learn about finite coupling and weak/strong coupling interpolation and this is probably the most stimulating reason for pursuing them.

Part II

Conformal and S-matrix Bootstrap

Chapter 4

The boundary bootstrap

In this chapter we aim at carving out the space of massive QFTs based only on fundamental properties of any Lorentz invariant, unitary theory. As we discussed in the invitation 1.2 by considering QFT in AdS we can study the boundary correlators using the technology of the CFT bootstrap as summarized in the cartoon 1.1. We call this procedure the *boundary bootstrap*. The goal of this chapter is to give life to this cartoon. We start in section 4.1 by establishing some basic properties of QFT in AdS, especially the relation between the bulk and boundary theories and the justification for applying the conformal bootstrap to the boundary theory. We then work out a concrete dictionary for translating the boundary CFT data into flatspace QFT data. In section 4.2 we recall the basic setup of the conformal bootstrap with special attention to problem of bounding structure constants. In section 4.3 we momentarily return to a flatspace perspective in order to formulate interesting QFT questions that can be attacked with the tools of sections 4.1-4.2. Finally in section 4.4 we combine the results of all these previous sections in order to obtain bounds on couplings in $1+1$ dimensional QFT's thereby realizing a concrete implementation of the boundary bootstrap. The content of this chapter is based on the work [69].

4.1 QFT in hyperbolic space

In this section we shall consider QFT in AdS in greater detail. Our main objectives are to justify the boundary bootstrap and to explain the relation between the large Δ limit of the conformal theory data and the physical data of the flat-space QFT.

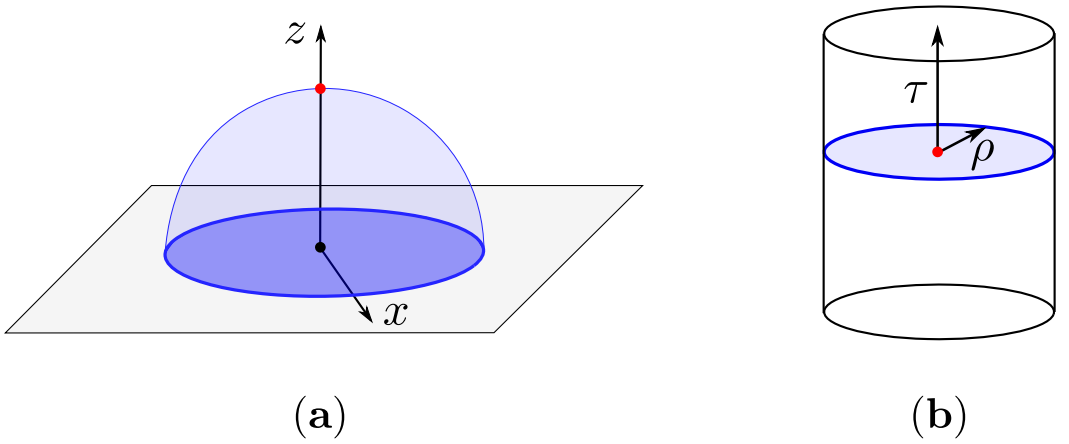


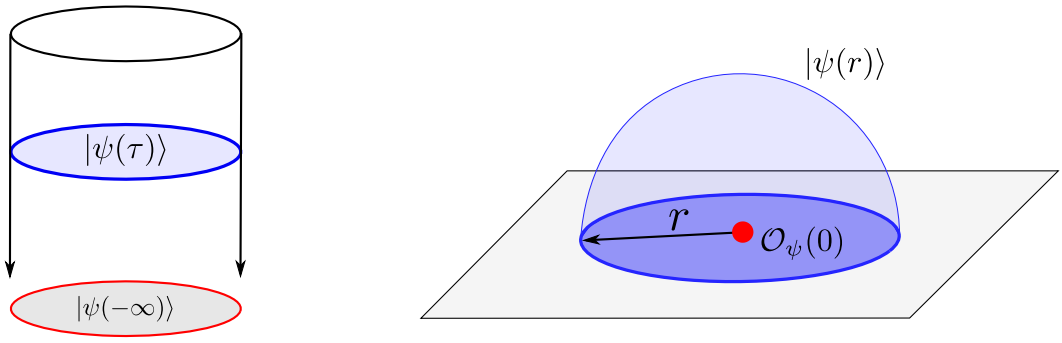
Figure 4.1: Euclidean AdS in Poincaré coordinates (**a**) and global coordinates (**b**). In Poincaré z is the bulk coordinate and x are the coordinates on the $\mathbb{R}^d$ conformal boundary at $z = 0$. A hemisphere of radius a centred at $z = |x| = 0$ maps to a slice of constant global time $\tau = \log a$ in global coordinates. In particular, the point $z = |x| = 0$ maps to $\tau = -\infty$.

Boundary operator - Bulk state correspondence and OPE

The first order of business is to establish the relationship between states in the bulk and operators in the boundary theory. For this it is useful to consider two different representations Euclidean AdS_{d+1}. First consider Poincaré coordinates where the metric takes the form

$$ds^2 = R^2 \frac{dz^2 + dr^2 + r^2 d\Omega_{d-1}^2}{z^2} \tag{4.1}$$

where $z > 0$ and r is a radial coordinate for $\mathbb{R}^d$. In these coordinates there is a flat conformal boundary at $z = 0$ as shown in figure 4.1a. The operators of the boundary theory are defined by pushing local bulk operators to this bounary. Correlation functions of boundary operators naturally inherit the $SO(d+1,1)$ isometry of AdS which is the conformal group in $\mathbb{R}^d$. They thus have the same global transformation properties as correlation functions of local operators in a d-dimensional CFT. For example, the 2- and



(a) **(b)**

Figure 4.2: In global coordinates (**a**) a state $|\psi(\tau)\rangle$ defined on a constant global time slice can be produced by time evolution from an appropriate state $|\psi(-\infty)\rangle$ in the infinite past. In Poincaré coordinates (**b**) this state lives on a hemisphere of radius $r = \log \tau$ and evolving to the infinite past shrinks this hemisphere to a point. We can thus create the state $|\psi(-\infty)\rangle$ by acting on the vacuum with a local operator $\mathcal{O}_\psi$ at $z = r = 0$.

3-point functions can be written as [70]

$$\begin{aligned} \langle \mathcal{O}(x_1)\widetilde{\mathcal{O}}(x_2)\rangle &= \frac{\delta_{\mathcal{O}\widetilde{\mathcal{O}}}}{|x_{12}|^{-2\Delta}} \\ \langle \mathcal{O}_a(x_1)\mathcal{O}_b(x_2)\mathcal{O}_c(x_3)\rangle &= \frac{C_{abc}}{x_{12}^{-\Delta_c+\Delta_a+\Delta_b} x_{13}^{-\Delta_b+\Delta_a+\Delta_c} x_{23}^{-\Delta_a+\Delta_b+\Delta_c}} \end{aligned} \tag{4.2}$$

where $x_{ab} = x_a - x_b$ are boundary coordinates and $\delta_{\mathcal{O}\widetilde{\mathcal{O}}} = 1$ if $\mathcal{O} = \widetilde{\mathcal{O}}$ and zero otherwise. For this reason we say that the bulk QFT "induces" a conformal theory at the boundary.[1]

Now consider changing to global coordinates via $z = e^\tau \cos\rho$, $r = e^\tau \sin\rho$ where the metric takes the form

$$ds^2 = R^2 \frac{d\tau^2 + d\rho^2 + \sin^2\rho d\Omega_{d-1}^2}{\cos^2\rho} \tag{4.3}$$

with $-\infty < \tau < \infty$ and $0 < \rho < \pi/2$ as shown in figure 4.1b. We can produce a bulk state $|\psi(\tau)\rangle$ defined on a slice of constant global time τ by starting with a specific state in

[1] We refrain from calling this theory a conformal *field* theory since it lacks a stress tensor. Instead we shall refer to it simply as a conformal theory.

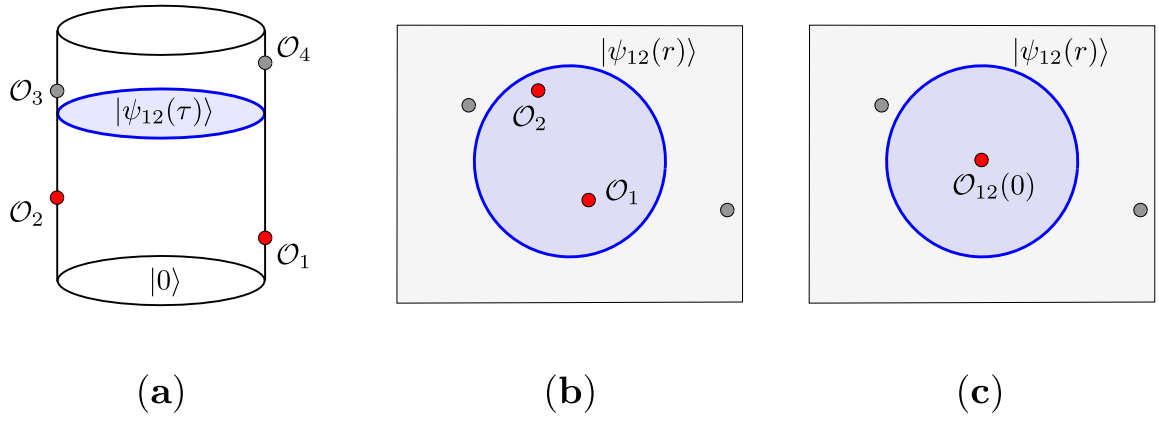


Figure 4.3: **(a)** Inserting the operators $\mathcal{O}_1$ and $\mathcal{O}_2$ produces the state $|\psi_{12}(\tau)\rangle$ at fixed global time τ. **(b)** Same picture in Poincaré coordinates. We show only the boundary. **(c)** According to the boundary operator - bulk state correspondence, the state $|\psi_{12}(r)\rangle$ can equally be produced by inserting the local operator O_{12} at the origin.

the infinite past $|\psi(-\infty)\rangle$ and evolving it using the generator of global time translations H as show in figure 4.2a. Now consider this same process in Poincaré coordinates. The state $|\psi(\tau)\rangle$ lives on the hemisphere defined by $\sqrt{r^2+\tau^2} = e^{\tau}$ as show in figure 4.2b. As we evolve to the infinite past this surface shrinks to an infinitesimal hemisphere about the point $r = z = 0$ and thus we can think of the state $|\psi(-\infty)\rangle$ as being created by the insertion of a local operator $\mathcal{O}_{\psi}(0)$ at this point. In this way we see that there is a one-to-one correspondence between boundary operators and bulk states. Specifically, since the hamiltonian H is just the operator which generates dilatations about the point $r = z = 0$ its eigenstates correspond to the primaries (and their descendants) of the conformal boundary theory. Thus, the states are organized according to representations of the d-dimensional conformal group which are labeled by the conformal dimension Δ and the $SO(d)$ irreducible representation of the corresponding primary operator.

OPE of boundary operators

Conformal symmetry of the boundary theory requires the OPE to take the form [70]

$$\mathcal{O}_a(x)\mathcal{O}_b(0) = \sum_c C_{abc}\,(x^2)^{\frac{1}{2}(\Delta_c-\Delta_b-\Delta_a)}\,[\mathcal{O}_c(0) + \text{decendants}] \tag{4.4}$$

where C_{abc} is the c-number called the structure constant which appears in 4.2. A important consequence of the correspondence between boundary operators and bulk states is that this OPE is actually convergent for finite x rather than just an asymptotic statement for small x [71]. We can see this as follows. Start with the vacuum state at $\tau = -\infty$ and insert two operators on the boundary before the time τ as shown in figure 4.3a. The effect of these two insertions is to create a certain bulk state at the time slice τ in global coordinates, or on a hemisphere of radius $r = \log \tau$ in Poincaré coordinates as shown in figure 4.3a and 4.3b respectively. However, by the operator-state correspondence this bulk state could equally well have been prepared by acting on the vacuum with a local operator at $\tau = -\infty$ as show in figure 4.3c so that we have

$$\mathcal{O}_1(x_1)\mathcal{O}_2(x_2)|0\rangle = \mathcal{O}_{12}(0)|0\rangle \tag{4.5}$$

The local operator $\mathcal{O}_{12}$ can be written as a sum over primary operators of definite dimension and spin, thus yielding the OPE (4.4). When derived in this way the convergence of the OPE follows from a basic theorem about Hilbert spaces: a scalar product of two states converges when one of the states is expanded in a complete basis of states. The argument above can break down when other operators are present. In particular, the argument will hold as long as we can find a time slice τ such that only $\mathcal{O}_1$ and $\mathcal{O}_2$ are inserted before τ. For example, in 4.3a as we move operator $\mathcal{O}_2$ upward the OPE will continue to converge until $\mathcal{O}_2$ passes $\mathcal{O}_3$. This translates to the general rule in Poincaré coordinates: inside correlation functions the OPE of two operators will converge as long as one can make a sphere around them that does not contain any other operators.

Flat space limit

We will now explain how flat-space data is extracted from the boundary theory [69]. As we have already seen a primary operator $\mathcal{O}_a$ corresponds to a massive bulk state where $\Delta_a \sim m_a R$ in the flat space limit. Thus the mass spectrum of the flat space QFT can be obtained from the limit

$$\frac{m_a}{m} = \lim_{\Delta\to\infty} \frac{\Delta_a}{\Delta} \tag{4.6}$$

where m is the mass of the lightest bulk particle dual to the operator of smallest dimension Δ.

Scattering amplitudes can be extracted from the flat space limit of the boundary correlators. This is essentially an AdS version of the LSZ prescription where one extracts the S-matrix by properly amputating the correlator. This can be done using the Mellin-space formalism, which leads to a general formula for scattering amplitudes in terms of

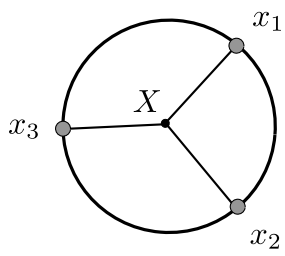


Figure 4.4: Witten diagram giving the leading contribution to the boundary three-point function.

the Mellin transform of the boundary correlator [69]. Here we shall only need the result for the simplest case of a three-point coupling, which can be extracted from the flat-space limit of the appropriate boundary 3-point function. The amputation procedure simply amounts to dividing off the contributions from propagating to the boundary of AdS. As a concrete example (which will be essential for our purposes) consider a bulk theory whose lightest particle is a scalar of mass m with a cubic coupling g. This particle is dual to the boundary operator $\mathcal{O}$ of lowest conformal weight Δ. To see explicitly the relation between the QFT coupling g and the structure constant $C_{\mathcal{O}\mathcal{O}\mathcal{O}}$ consider the leading contribution to the three-point function given by the Witten diagram in figure 4.4 which is

$$\langle \mathcal{O}(x_1)\mathcal{O}(x_2)\mathcal{O}(x_3)\rangle = gR^{\frac{5-d}{2}} \int_{AdS} d^{d+1}X \sqrt{-g_{AdS}}\, G_{B\partial}(x_1, X) G_{B\partial}(x_2, X) G_{B\partial}(x_3, X) \quad (4.7)$$

where the propagator from the bulk point $X = (z, x)$ to the boundary point $(0, x_i)$ is given by

$$G_{B\partial}(x_i, X) = \sqrt{\mathcal{C}_\Delta} \frac{z^\Delta}{[z^2 + (x - x_i)^2]^\Delta}, \qquad \mathcal{C}_\Delta = \frac{\Gamma(\Delta)}{d\pi^{\frac{d}{2}}\Gamma(\Delta - d/2 + 1)}\,. \quad (4.8)$$

The normalization has been chosen such that the two-point functions have unit norm as is our convention in (4.2) and the factors of R in (4.7) have been inserted to make the cubic coupling g dimensionless. At the same time, we know from (4.2) that

$$\langle \mathcal{O}(x_1)\mathcal{O}(x_2)\mathcal{O}(x_3)\rangle = \frac{C_{\mathcal{O}\mathcal{O}\mathcal{O}}}{(x_{12}^2 x_{23}^2 x_{13}^2)^{\frac{\Delta}{2}}}\,. \quad (4.9)$$

Computing the integral (4.7) and comparing with this expectation gives

$$C_{\mathcal{O}\mathcal{O}\mathcal{O}} = gR^{\frac{5-d}{2}} \frac{\pi^d \Gamma(\frac{3\Delta - d}{2})\Gamma^3(\frac{\Delta}{2})\mathcal{C}_\Delta^{3/2}}{2\Gamma^3(\Delta)}\,. \quad (4.10)$$

As we discussed above, the flat-space limit corresponds to the limit of large R with fixed m which means large Δ. We thus find

$$m^{\frac{d-5}{2}} g^{\text{flat space}} = \left(\frac{3}{2}\right)^{\frac{d}{2}} \sqrt{3}\pi^{\frac{d-2}{4}} \lim_{\Delta\to\infty} \Delta^{\frac{d-2}{4}} e^{\Delta\frac{3}{2}\log\frac{4}{3}} C_{\mathcal{O}\mathcal{O}\mathcal{O}} \,. \tag{4.11}$$

Although our argument was perturbative the relation (4.11) is exact. The explanation of this is quite simple. There are two types of diagrams that can appear at higher orders – those correcting the individual leg factors and those connecting various leg factors. As for the later type of diagram, exchanges that occur deep inside AdS are already accounted for in g since it is defined as the exact cubic coupling (reside of the pole of the S-matrix at $s = m^2$). Exchanges that do not occur near the centre of AdS are suppressed in the flat space limit. As for the corrections to the leg factors, we are already using propagators corresponding to the physical mass m and any wave function renormalization factors are already included in the exact g.

In summary, we now have the concrete formulas for relating the conformal data of the boundary theory to the flat space QFT data. In particular, we have learned how to extract the flat space mass spectrum from the spectrum of conformal dimensions (4.6) and the flat space couplings from the structure constants (4.11). Through these relations, we see that by using the conformal bootstrap to bound $C_{\mathcal{O}\mathcal{O}\mathcal{O}}$ in the d-dimensional conformal theory, we can bound $g^{\text{flat space}}$ of the $d+1$ dimensional massive QFT![2] We will refer to this procedure as the boundary bootstrap.

4.2 The conformal bootstrap program

In the previous section we saw that a massive QFT in AdS induces a conformal boundary theory and that its conformal data (dimensions and structure constants) encode the mass spectrum and couplings of the bulk theory. In this section we shall review the basic ingredients of the conformal bootstrap which places strong constraints on the conformal data of the boundary theory and thus on physical data of the bulk QFT.

Our main object of study will be the four-point function in a d-dimensional conformal theory. Consider a scalar primary operator ϕ with conformal dimension Δ_ϕ. It's four point function is constrained by conformal symmetry to take the form

$$\langle\phi(x_1)\phi(x_2)\phi(x_3)\phi(x_4)\rangle = \frac{g(u,v)}{|x_{12}|^{2\Delta_\phi}|x_{34}|^{2\Delta_\phi}} \tag{4.12}$$

[2]Of course it could happen that the bound becomes trivial (diverges) in the flat-space limit. We shall see that, at least for a $1+1$ dimensional bulk this does not occur.

where u and v are the cross-ratios

$$u = \frac{x_{12}^2 x_{34}^2}{x_{13}^2 x_{24}^2}, \qquad v = \frac{x_{14}^2 x_{23}^2}{x_{13}^2 x_{24}^2}\,. \tag{4.13}$$

The two essential ingredients in the CFT bootstrap are *crossing symmetry* and the *conformal block decomposition* which follows from the OPE. Let us now quickly recall these basic properties.

- Crossing symmetry: This is simply the statement that since the left side of (4.12) is invariant under the interchange of two x_a, the right side must also be invariant. This leads to a set of crossing symmetry constraints. Symmetry under $x_1 \leftrightarrow x_3$ gives[3]

$$v^{\Delta_\phi} g(u,v) = u^{\Delta_\phi} g(v,u) \tag{4.14}$$

- Operator product expansion: The OPE expresses the product of two primary operators at finite separation in terms of a sum of local primaries and can be written schematically as

$$\phi(x)\phi(0) = \sum_O C_O\, P(x,\partial_x)\, O(x) \tag{4.15}$$

  where the sum is over all primary operators and the coefficients P are fixed by conformal symmetry. The structure constant C_O is the constant appearing in the three point function $\langle\phi\phi O\rangle$. Note that we are suppressing all spin dependence for notational simplicity.

- Conformal block decomposition: This follows from applying the OPE (4.15) to the four-point function (4.12). Applying this relation to the products $\phi(x_1)\phi(x_2)$ and $\phi(x_3)\phi(x_4)$ gives the conformal block decomposition

$$g(u,v) = 1 + \sum_O \lambda_O^2\, g_O(u,v) \tag{4.16}$$

  where we have explicitly separated out the contribution of the identity operator. The functions g_O are called conformal blocks. They receive contributions from the two point function of the operator O as well as those of all its descendants. Closed form expressions for these objects in terms of hypergeometric functions exists for even d, while in odd dimensions there are power series representations and also recursion relations which allow them to be efficiently computed. In any dimension the conformal blocks for identical scalars depend only on the spin and dimension of the exchanged operator as well as the cross ratios $g_O = g_{\Delta,l}(u,v)$.

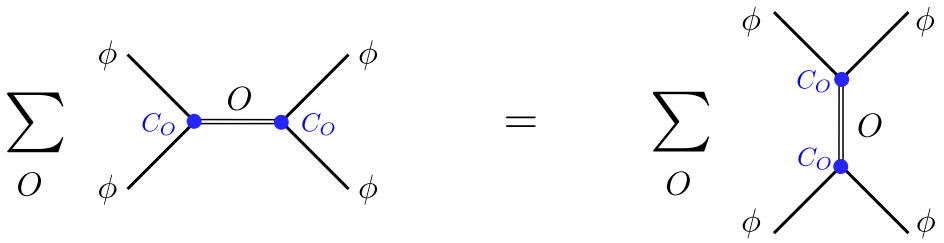


Figure 4.5: CFT bootstrap condition. Equating the conformal partial wave expansion (4.16) in two different channels leads to the sum rule (4.17).

The conformal block decomposition (4.16) and the crossing equations (4.14) are all the ingredients needed for the conformal bootstrap. The combination of the two turn out to be a rather strong constraint on the conformal data. Applying the crossing constraint to the conformal block decomposition. gives the so-called *sum rule*

$$u^{\Delta_\phi} - v^{\Delta_\phi} + \sum_O C_O^2 \left[v^{\Delta_\phi} g_O(v,u) - u^{\Delta_\phi} g_O(u,v) \right] = 0 \tag{4.17}$$

The derivation and content of this equation is summarized in figure 4.5. This equation must be valid for any four-point function of identical scalars in any conformal theory. It can be thought of as a continuous set of functional constraints on the (potentially infinite) set of structure constants C_O appearing in the $\phi \times \phi$ OPE.

We will now show how the sum rule can be used to derive rigorous constraints on the conformal data [21]. We write the sum rule as

$$1 + C_{O_1}^2 F_{O_1}(u,v) + \sum_{O_{a\neq 1}} C_{O_a}^2 F_{O_a}(u,v) = 0 \tag{4.18}$$

where we defined

$$F_O(u,v) = F_{\Delta,l}(u,v) = \frac{v^{\Delta_\phi} g_{\Delta,l}(v,u) - u^{\Delta_\phi} g_{\Delta,l}(u,v)}{u^{\Delta_\phi} - v^{\Delta_\phi}} \tag{4.19}$$

and we have explicitly separated out the contribution of the operator O_1 for reasons that will become clear momentarily. Now consider acting on (4.18) with a linear functional of

[3]There is an additional constraint from $x_1 \leftrightarrow x_2$ which gives $g(u,v) = g(u/v, 1/v)$. However, this will not play a role in our considerations.

the form

$$\Lambda[f] = \sum_{a,b=0}^{N} \lambda_{ab}\, \partial_u^{2a}\, \partial_v^{2b}\, f(u,v)\Big|_{u=v=1/4} \tag{4.20}$$

where λ_{ab} are arbitrary real numbers and N is a parameter which counts the number of derivatives in our functional. Now, given a tentative spectrum (dimensions and spins) of operators appearing in the $\phi \times \phi$ OPE we can evaluate all of the $F_{\Delta,l}(u,v)$ explicitly. Then, if one can find a Λ such that

$$\Lambda[1] > 0 \quad \text{and} \quad \Lambda[F_{\Delta,l}] > 0 \tag{4.21}$$

then this spectrum is excluded since acting on (4.18) with this Λ generates a contradiction. To obtain this contradiction it is essential that $C_O^2 \geq 0$ which is ensured by the reality of the structure constants [21].

With a slight variation of this argument, we can obtain bounds on the structure constants given a tentative spectrum of dimensions and spins. For this we seek a functional which satisfies

$$\Lambda[F_{O_{a\neq 1}}] > 0, \qquad \Lambda[F_{O_1}] = 1 \tag{4.22}$$

which gives the bound

$$C_{O_1}^2 \leq -\Lambda[1] \tag{4.23}$$

Note that if $\Lambda[1] > 0$ we have the situation in (4.21) and spectrum is excluded, but if $\Lambda[1] < 0$ we obtain a genuine bound on the structure constant C_{O_1} for any theory with the spectrum we started with. Note that increasing N (the number of derivatives in Λ) can only make a bound stronger. Any coupling or spectrum that was previously ruled out with smaller N will still be ruled out (but previously acceptable couplings and spectra may be excluded). Thus even for finite N the bounds are *exact* – they can go down but never up.

Typically, one makes assumptions about the lowest few dimensions in the spectrum but allows an arbitrary spectrum above some gap Δ_{gap}. For example, we could consider theories whose spectra consist of a scalar operator $\mathcal{O}$ of dimension Δ and then no other operators below some Δ_{gap}. Given a spectrum of this form, we can ask what is the maximum allowed value of $C_{\mathcal{O}\mathcal{O}\mathcal{O}}$. For this one must discretize the space of Δ's between Δ_{gap} and some cutoff Δ_{max}. Additionally one must truncate the sum over spins at some l_{max}. Since the conformal blocks converge rapidly for large dimension and spin, this is a good approximation. In this way, we can bound the structure constants appearing in *any* CFT whose lowest dimensions satisfy some desired properties.

So far we have avoided commenting on how one actually goes about searching the space of functionals (4.20). It turns out that this is a standard problem in linear algebra

which goes by the name of Linear Programming. There are efficient algorithms such as `JuliBootS` [72] and `SDPB` [73] designed precisely for such a task. With these algorithms, the ideas presented in this section become a practical and powerful tool for constraining the space of CFT's which has already been applied with great success to a number of important problems [74–83].

4.3 An interesting QFT question

In this section we devise a set of interesting QFT questions that have a natural formulation in terms of the boundary bootstrap.

QFT perspective

Consider a D-dimensional QFT whose lightest stable particle is a scalar of mass m and consider the elastic scattering of two such particles $m + m \to m + m$. In flat space this scattering process is described by the S-matrix element $S(s, t)$ where s and t are the usual mandelstam variables. We denote by $\mathbb{m} = \{m_1, m_2, \dots\}$ the mass spectrum of stable single-particle states that appear in the scattering event S. Such single particle processes generate poles in the S-matrix at the mass of the intermediate particles $s = m_a^2$ and the residue of each pole is precisely the square of the cubic coupling g_a^2 for the process $m + m \to m_a$. We take the list $\mathbb{m}$ to be ordered $m_1 < m_2 < \cdots < 2m$ so that g_1 is always the coupling to the lightest exchanged particle.[4]

Let us consider the simplest case $\mathbb{m} = \{m\}$ which corresponds to theories where the only intermediate single-particle state that can occur in $m + m$ scattering is m itself. It is well known that scalar exchange generates an attractive potential and thus we expect that the attractive force becomes stronger as g_1^2 increases such that eventually the formation of a bound state $m + m \to m_2$ cannot be avoided. Thus we are led to the very interesting question: what is the maximum value of g_1^2 consistent with the spectrum $\mathbb{m} = \{m\}$? More generally, we can ask:

- Question: given some spectrum $\mathbb{m}$ what is the maximum possible value of a given coupling g_a^2?

[4] In some cases the lightest exchanged particle is the external particle (which is the lightest in the theory) so that $m_1 = m$; we will still refer to the coupling as g_1.

In practice we shall focus on maximizing the coupling to the lightest particle $m_1 \in \mathbb{m}$. In chapter 5 we shall explore this question in $1+1$ dimensions in great detail. For now, for the sake of concreteness, let us specialize to two specific scenarios:

- Scenario I: Max g_1: $\mathbb{m} = \{m, m_2\}$.
- Scenario II: Max g_1: $\mathbb{m} = \{m_1\}$.

Of course in Scenario II we have $m_1 \geq m$ by assumption. As we shall now see, both scenarios have a natural formulation in terms of the boundary bootstrap.

Boundary bootstrap perspective

In the boundary theory a scalar bulk state with mass m_a corresponds to the scalar primary operator $\mathcal{O}_a$ with dimension $\Delta_a \sim m_a R$. Thus the analog of the S-matrix element S is the flat space limit of the four point function of the scalar operator $\mathcal{O}$ with the smallest dimension Δ. Further, the spectrum of intermediate states $m_a \in \mathbb{m}$ corresponds to a spectrum of scalar operators with dimensions $\Delta_a < 2\Delta$ appearing in the $\mathcal{O} \times \mathcal{O}$ OPE. Thus the two scenarios formulated above translate to the following restrictions on the form of the $\mathcal{O} \times \mathcal{O}$ OPE

- Scenario I: $\mathcal{O} \times \mathcal{O} = 1 + C_1\, \mathcal{O} + C_2\, \mathcal{O}_2 + (\text{operators with dimension} > 2\Delta)$
- Scenario II: $\mathcal{O} \times \mathcal{O} = 1 + C_1\, \mathcal{O}_1 + (\text{operators with dimension} > 2\Delta)$

Assuming these forms of the OPE we can run the machinery of the CFT bootstrap and ask for the largest possible value of the structure constant C_1 as outlined in section 4.2. Using (4.11) we then translate this to a bound on g_1. In the next sectiom we will do precisely this for the case of QFT in $D = 2$ dimensions.[5]

[5] In practice in Scenario I we only impose a gap up to Δ_2 and then allow for an arbitrary spectrum. This turns out to be equivalent to the statement of Scenario I that we have given since optimization sets the structure constants to zero between Δ_2 and $2\Delta_1$. This is somewhat mysterious from the boundary bootstrap perspective, but can be simply explained in the language of the flat-space S-matrix.

4.4 The 1D Boundary Bootstrap

In this section we shall describe the implementation and results of the boundary bootstrap for 2D QFT or a 1D boundary conformal theory. The restriction to a 1D boundary provides a significant simplification as there is no spin to sum over in the OPE. However, the problem is still far from easy. As we shall see the limit of large Δ significantly complicates the application of the conformal bootstrap techniques described in section 4.2.

Analysis of Scenario I

In scenario I we focus on a spectrum with lowest dimension $\Delta = \Delta_1$ and then a gap until $\Delta_2 < 2\Delta$. We assume a non-zero structure constant $C_1 = C_{\mathcal{O}\mathcal{O}\mathcal{O}}$ which we seek to maximize. The application of the technology of section 4.2 yields a bound which depends on three parameters. First we have the non-physical parameter N which counts the number of derivatives we use in the functional (4.20). As explained above, the bound becomes stronger for larger N and thus we want to take N as large as possible or at least large enough until convergence is observed. The second parameter is the dimension of the external operator Δ which must also be made large in order to achieve the flat space limit.[6] Finally, we have the relative gap Δ_2/Δ, which we would like to keep finite since we want to study the dependence of this bound on this parameter. Figure 4.6 shows the bound as a function of Δ for several values of N and a gap of $\Delta_2 = 1.825\Delta$. From (4.11) we see that in order to obtain a finite bound we must have

$$\log C_1^{\max} \sim -\left(\frac{3}{2}\log\frac{4}{3}\right)\Delta + \frac{1}{4}\log\Delta + \text{finite} \tag{4.24}$$

for large Δ. Thus, as a function of Δ we should observe and approximately linear curve with a precise slope of $-\frac{3}{2}\log\frac{4}{3}$. In plot 4.6 we observe that the bounds start out with a negative slope, but then turn upward and seem to diverge for large Δ. However, the curves for different N become more widely spaced for large Δ indicating that convergence has not been achieved there. Notably, we also observe that the bound monotonically decreases with increasing N. The value $N = 100$ is already quite computationally difficult, and thus we seek to access the large N limit through extrapolation. We fit the data for fixed Δ, Δ_2 to a large degree polynomial in $1/N$ and then evaluate the result at $N = \infty$.[7] The result of this extrapolation is thick red curve in 4.6. Remarkably, this curve has exactly

[6] Since Δ is the lowest dimension this implies all other dimensions become large in this limit.

[7] Adding data points of larger N does not significantly affect the results of the extrapolation and thus we have confidence that we are accurately extracting the large N limit through this procedure.

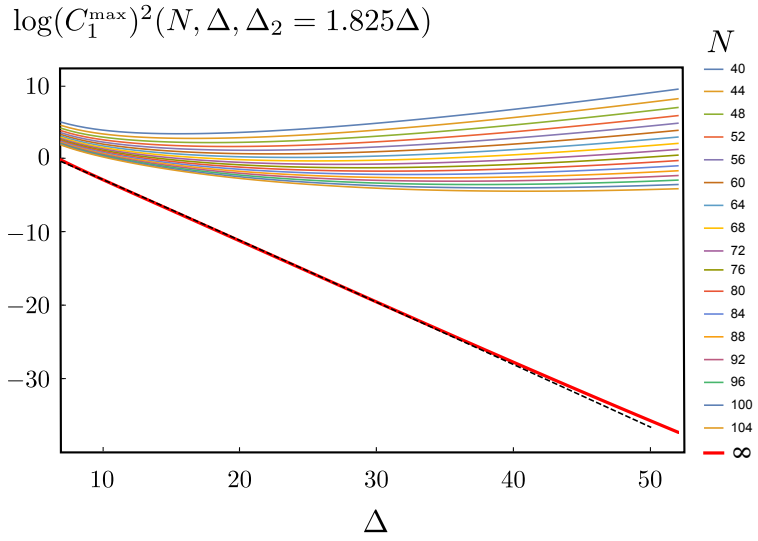


Figure 4.6: We bound the structure constant C_1 assuming the OPE in Scenario I with $\Delta_2 = 1.825\Delta$ using several values of N. We then extrapolate to $N = \infty$ as explained in the main text which gives the thick red curve. The dashed black line indicates the expected slope from the linear term in (4.24). Bounds obtained with `JuliaBot`.

the expected slope![8] We can thus use (4.11) to extract a finite result for the maximum flat space coupling $g_1^{\max}$. For this, we perform one more extrapolation in Δ – we fit to a low degree polynomial in $1/\Delta$ (typically no more than quadratic) and then evaluate at $\Delta = \infty$. The reason we must perform such a low degree fit in this case is that the data already contains a reasonable amount of error due to the extrapolation in N. Repeating this procedure for many values of the gap results in the plot shown in figure 4.7.

Figure 4.7 represents the culmination of all the ideas developed so far in this chapter. It is a concrete realization of the boundary bootstrap which gives life to the cartoon 1.1 in the Invitation 1.2. Each data point represents a bound on the coupling of any 2D QFT that falls within Scenario I. For example, no such QFT can exist with $(m_2/m)^2 \simeq 3.5$ and $\log(g_1)^2 \gtrsim 6$. Note that as per (4.11) the flat space coupling is given in units of the mass of the lightest particle. In the limit $m_2^2 \to 4m^2$ we recover the spectrum $\mathbb{m} = \{m\}$ discussed in 4.3 which inspired our investigation of Scenarios I and II. We find that such a system cannot exist in $1+1$ dimensions with $\log(g_1)^2 \gtrsim 3.03$; if the coupling is any larger, a bound state *must* form. Another interesting limit is $m_2^2 \to 3m^2$ where the bound diverges. From

[8] The small deviations for large delta are due to the subleading log term in (4.24).

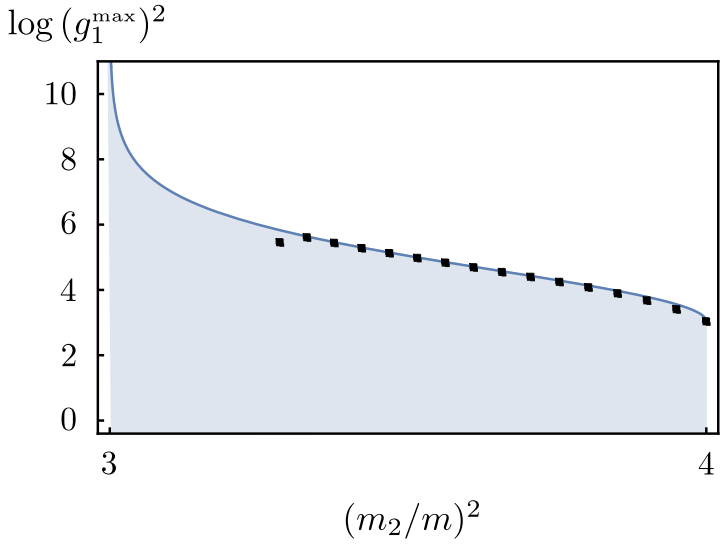


Figure 4.7: Bound on the flat space coupling g_1 for Scenario I. The black squares are the result of the boundary bootstrap numerics. All theories above these points are excluded. To generate each point (i.e. for fixed Δ_2) we perform the procedure to generate the thick red curve in 4.6 (result of extrapolating in N) and then we extrapolate this red curve to infinity (extrapolate to infinite Δ). In this way we measure the finite term in (4.24) which gives the flat space coupling according to (4.11). The solid curve is an analytic bound that we will derive in the next chapter 5 from the S-matrix bootstrap.

boundary bootstrap perspective the loss of the bound at this point is somewhat mysterious, although it may be possible to understand it from the analysis of [84]. However, as we shall see in the next chapter, from the perspective of the S-matrix bootstrap there is a very simple kinematical explanation for this phenomenon. Note that the bootstrap numerics become more difficult as we approach a point where the bound is lost which accounts for the poor agreement of the leftmost datapoint with the exact bound.[9]

[9] Although the conformal bootstrap produces exact bounds for finite N as explained in section 4.2, this statement is no longer true after the extrapolations in N and Δ. Thus it is no contradiction that some of the data points sit below the exact bound.

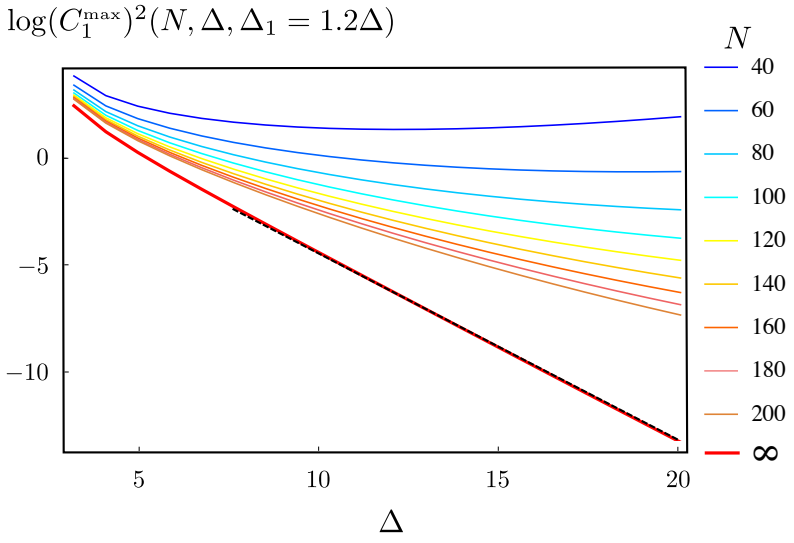


Figure 4.8: We bound the structure constant C_1 assuming the OPE in Scenario II with $\Delta_1 = 1.2\Delta$ using several values of N. We then extrapolate to $N = \infty$ which gives the thick red curve. The dashed black line indicates the expected slope. Bounds obtained using `SDPB`.

Analysis of Scenario II

In Scenario 2 we assume a single scalar operator with $\Delta_1 \geq \Delta$ and then a gap until 2Δ. With these restrictions on the spectrum we again maximize the first OPE coefficient C_1 in the self-OPE of the operator of lowest dimension. We again have three parameters N, Δ, Δ_1. The results for $\Delta_1 = 1.2\Delta$ are shown in figure 4.8 and several values of N. We see that again extrapolation to large N is necessary. Once this is done we beautifully recovered the expected slope required for a finite flat-space bound which is indicated by the black dashed line in figure 4.8 . We repeat this procedure for several values of Δ_1, in each case extrapolating to $\Delta = \infty$ as described in the analysis of Scenario I. The results are shown in figure 4.9.

There are a few interesting regimes of the bound for Scenario II. First, note that when $m_1 \to 1$ we again recover the case $\mathbb{m} = \{m\}$; we see that in this case the results agree with those of Scenario I. Second, as $m_1^2 \to 2m^2$ the bound diverges. Again it is not obvious why this occurs from the conformal bootstrap perspective, but we shall see in the next chapter that it has a simple kinematical explanation in the S-matrix bootstrap. Finally, as $m_1^2 \to 4m^2$ the coupling vanishes. This is intuitive since in this limit m_1 can be thought

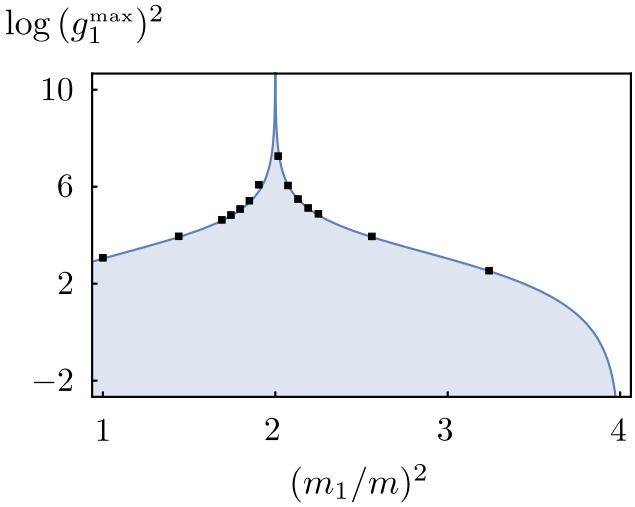


Figure 4.9: Bound on the flat space coupling g_1 for Scenario II. The black squares are the result of the boundary bootstrap numerics. We generate each point from an analysis like that of figure 4.8 using the same procedure used to generate the points in figure 4.7 (see corresponding caption). The solid curve is the analytic result that we will derive in the next chapter from the S-matrix bootstrap.

of as a bound pair of the lightest particle m with a very small binding energy; hence the attractive interaction should be correspondingly small.[10]

[10]This regime is also non-relativistic and can be studied using ordinary quantum mechanics.

Chapter 5

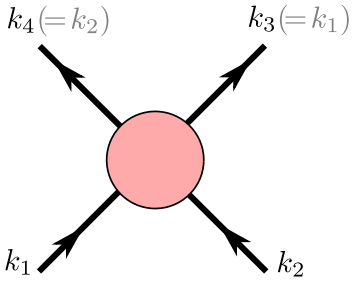
The 2D Amplitude Bootstrap

Inspired by the success of the boundary bootstrap of section 4.4, in this chapter we shall try to establish the bounds directly within the framework of flat space QFT using only the usual crossing symmetry and unitarity of the S-matrix. As mentioned in the invitation 1.2, it is intuitive that for a fixed spectrum the coupling g_1 cannot be made arbitrarily large. In this section we systematically study these bounds in two dimensions where everything is simpler in the S-matrix world (the kinematical space simplifies significantly and crossing symmetry can be taken care of very explicitly). Not only do we find the above mentioned bounds but also manage to identify known integrable theories which saturate the bounds at special points. We hope these results will constitute the first steps in a general program aimed at extending the successful CFT bootstrap to massive QFT's.

Our main object of study will be the $2 \to 2$ S-matrix elements of a relativistic two dimensional quantum field theory. We will further focus on the scattering process involving identical chargeless particles of mass m. For the most part, we shall take the external particles to be the lightest in the theory.[1]

Let us very briefly review a few important properties of this object, setting some notation along the way. A major kinematical simplification in $2 \to 2$ scattering in two dimensions is that there is only a single independent Mandelstam invariant. In particular, for scattering involving particles of identical masses there is zero momentum transfer as

[1] Strictly speaking, what we shall use is that any two particle cut in the S-matrix element opens up after the two particle cut of the external particles. The $2 \to 2$ S-matrix element of the lightest particles is also free of Coleman-Thun singularities [85](which render the analysis more involved and which will not be considered here). Sometimes, symmetry alone forbids such cuts or poles. In those case, the restriction to the lightest particle can be relaxed.



$$s \equiv (k_1 + k_2)^2$$
$$t \equiv (k_2 - k_3)^2 \ = 4m^2 - s$$
$$u \equiv (k_3 - k_1)^2 = 0$$

Figure 5.1: The $2 \to 2$ S-matrix element. Time runs vertically in this figure. In two dimensions energy-momentum conservation implies there is only one independent Mandelstam variable such that $S = S(s)$ with $\sqrt{s}$ the centre of mass energy.

depicted in figure 5.1. If all external particles are identical, crossing symmetry which flips t and s simply translates into

$$S(s) = S(4m^2 - s)\,, \tag{5.1}$$

while unitarity states that for physical momenta, i.e for centre of mass energy greater than $2m$, probability is conserved,

$$|S(s)|^2 \leq 1\,, \qquad s > 4m^2\,. \tag{5.2}$$

We shall come back to this relation in more detail below, in section 5.2.

Finally, we have the analytic properties of $S(s)$ depicted in figure 5.2. Of particular importance for us are the S-matrix poles located between the multi-particle cuts. Such poles are associated to single-particle asymptotic states. Note that there is no conceptual difference between fundamental particles or bound-states here. We refer to both as *particles* in what follows. The poles in S always come in pairs as

$$S \simeq -\mathcal{J}_j \frac{g_j^2}{s - m_j^2} \qquad \text{and} \qquad S \simeq -\mathcal{J}_j \frac{g_j^2}{4m^2 - s - m_j^2}\,, \qquad \left(\mathcal{J}_j = \frac{m^4}{2m_j\sqrt{4m^2 - m_j^2}}\right) \tag{5.3}$$

corresponding to an s- or t-channel pole respectively. Here we normalize g_j^2 to be the residue in the invariant matrix element T which differs from S by the subtraction of the identity plus some simple Jacobians related to the normalization of delta functions in the connected versus disconnected components. This justifies the prefactors $\mathcal{J}_j$ in (5.3).[2] Note

[2] We have $\mathbf{S} \equiv \mathbf{1} \times S(s) = \mathbf{1} + i(2\pi)^2\delta^{(2)}(P)\ T$. The contribution $\mathbf{1} = (2\pi)^{(2)}\sqrt{E_1E_2E_3E_4}(\delta(\vec{k}_1 -$

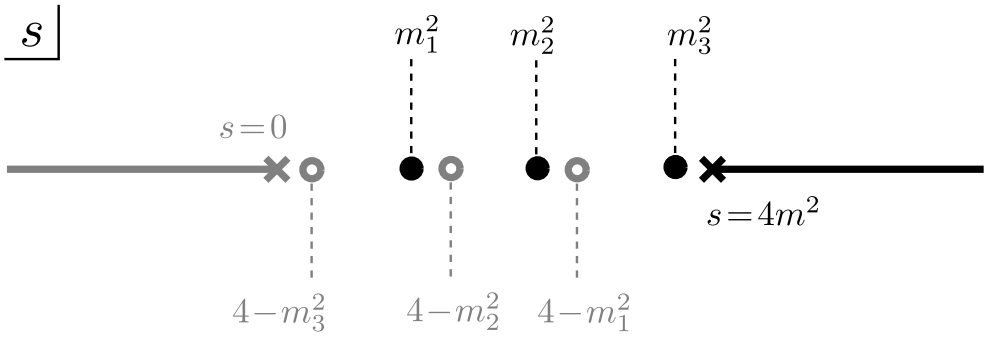


Figure 5.2: Analytic properties of the S-matrix element $S(s)$ for the scattering of the lightest particle of the theory. We have a cut starting at $s = 4m^2$ corresponding to multi-particle s-channel processes. As implied by (5.1), we have another cut starting at $t = 4m^2$ (or $s = 0$) describing multi-particle t-channel processes. The segment $s \in [0, 4m^2]$ can contain poles corresponding to intermediate propagation of single particle asymptotic states. We distinguish s and t channel poles (solid and empty circles respectively) by the sign of their residues. When the external particles are not the lightest in the theory, we sometimes have more singularities such as further two particle cuts and/or Coleman-Thun poles.

that we can always clearly tell the difference between an s- or a t-channel pole: since in a unitary theory g_j^2 is positive, an s-channel pole has a negative residue (in s) while a t-channel pole has a positive residue (in s).

This concludes the lightning review of two dimensional scattering. We now have all the ingredients necessary to state the problem considered in this paper. As input we have a fixed spectrum of stable particles of masses $m_1 < m_2 < \cdots < m_N$ which can show up as poles in $S(s)$. Note that by definition of stable asymptotic state (be it a bound-state or a fundamental particle) we have $m_j < 2m$. Note also that m_1 might be equal to the mass m of the external particle itself if the cubic coupling $m + m \to m$ is non-vanishing

$\vec{k}_3)\delta(\vec{k}_2 - \vec{k}_4) + (\vec{k}_1 \leftrightarrow \vec{k}_2))$ represents the (disconnected) contribution of the free propagation while T accounts for the connected contribution. Here we are we denoting the momentum component as $\vec{k}$ even though it is just a number just to distinguish it from the 2-momentum k. Now, the delta function multiplying T is the energy-momentum conservation delta function $\delta^{(2)}(P) = \delta^{(d)}(k_1+k_2-k_3-k_4)$. On the support of the solution $\vec{k}_1 = \vec{k}_3$, $\vec{k}_2 = \vec{k}_4$ we have $E_1E_2\,\delta^{(1)}(\vec{k}_1-\vec{k}_3)\delta^{(1)}(\vec{k}_2-\vec{k}_4) = (2s\sqrt{4m^2-s})^{-1}\delta^{(2)}(P)$. This Jacobian relating the δ-functions results in the denominator in the definition of $\mathcal{J}_j$ in (5.3). The m^4 numerator is just dimensional analysis. It is there so that g_1 is dimensionless. In other words, as defined, g_1 is the coupling measured in units of the external mass.

(e.g. if such a cubic coupling is not forbidden by a $\mathbb{Z}_2$ symmetry for instance). Similarly, there might be other stable massive particles which are not produced in scattering of the lightest particle and thus do not appear as poles of $S(s)$ – e.g. for symmetry reasons. The question we ask is then what is the maximum possible value of the coupling to the lightest exchanged particle (i.e g_1) compatible with such a spectrum,

$$g_1^{\max} \equiv \max_{\text{fixed } m_j} g_1 = ? \tag{5.4}$$

Physically, we expect the right hand side to be less than infinity. After all, as we increase the coupling to m_1 we expect this to generate an attractive force mediated by the particle m_1 between the two external masses. At some point, this force is such that new bound states are bound to show up thus invalidating the spectrum we took as input. This should then set a bound on g_1. This question bears strong resemblance with very similar questions recently posed in the conformal bootstrap approach mentioned above. There also we can put upper bounds on the OPE structure constants given a fixed spectra of dimensions [76].

We will approach this simple problem from two complementary angles. First in section 5.1 we will combine numerics with dispersion relation arguments to find a numerical answer. In section 5.2 we present an analytic derivation of this bound exploring the power of analyticity for two dimensional kinematics. The content of this chapter is based on the work [86].

5.1 Dispersion Relations and the Numerical Bootstrap

On the physical sheet the S-matrix has singularities corresponding to physical processes but is otherwise an analytic function. Analyticity places strong constraints on $S(s)$ which can be summarized in a so-called dispersion relation which relates the S-matrix at any complex s to its values at the cuts and poles, see e.g. [22, 23]. To set the notation and to specialize to two dimensions, we briefly recall the argument here. We start with the identity

$$S(s) - S_\infty = \oint_\gamma \frac{dx}{2\pi i} \frac{S(x) - S_\infty}{x - s} \tag{5.5}$$

where γ is a small counterclockwise contour around the point s away from any pole or cut. Now consider blowing the contour outward. For simplicity we assume that $S(s)$ approaches a constant $S_\infty \in [-1, 1]$ as $s \to \infty$ although this restriction can easily be lifted by means

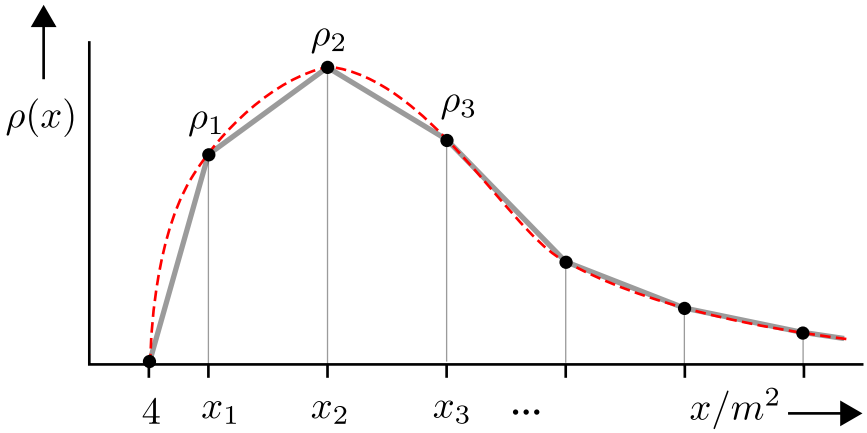


Figure 5.3: Approximation of an arbitrary density with a linear spline. The red dashed line represents some unknown $\rho(x)$ which we approximate with the grey spline passing through the points (ρ_n, x_n). Explicitly we have $\rho(x) \approx \rho_n \frac{(x-x_{n+1})}{(x_n-x_{n+1})} + \rho_{n+1} \frac{(x-x_n)}{(x_{n+1}-x_n)}$ for $x \in [x_n, x_{n+1}]$. We use this approximation up to some cutoff x_M after which we assume the density decays as $\rho(x) \sim 1/x$. That is, we have $\rho(x) \approx \rho_M \, x_M/x$ for $x \geq x_M$ which allows us to explicitly integrate the tail from x_M to ∞.

of so-called subtractions.[3] In this case we can drop the integration over the arcs at infinity so that we have only the integration around the poles and cuts giving

$$S(s) = S_\infty - \sum_j \mathcal{J}_j \left(\frac{g_j^2}{s - m_j^2} + \frac{g_j^2}{4m^2 - s - m_j^2} \right) + \int_{4m^2}^{\infty} dx \, \rho(x) \left(\frac{1}{x - s} + \frac{1}{x - 4m^2 + s} \right) \tag{5.6}$$

where we have defined the discontinuity $2\pi i \, \rho(s) \equiv S(s + i0) - S(s - i0)$ and we have further used the crossing equation (5.1) to replace the discontinuity across to the t-channel cut in terms of the s-channel discontinuity.

Equation (5.6) is the sought after dispersion relation: it simultaneously encodes the analyticity constraints as well as the the crossing condition and thus provides a concrete

[3]The basic idea of the subtraction procedure is to start with an identify of the form $S(s) = \oint \frac{dx}{2\pi i} \frac{S(x)}{x-s} \prod_{a=1}^{n} \frac{s-x_a}{x-x_a}$ where $n = 1, 2, \ldots$ is the number of subtractions. As we blow the contour, the integrand in the new identity is now more suppressed at large s such that dropping the arc at infinity is often safer. In the end, this leads to similar albeit a bit more involved dispersion relations as compared to (5.6) below. We checked on a few examples that the numerics described below yield similar results with a few subtractions. More generally we expect never to need more than $n = 1$ in two dimensions.

framework for addressing the question (5.4). In this form, the question becomes: what is the largest value of g_1 for which one can find $g_2, ..., g_N$ and $\rho(x)$ such that (5.2) is satisfied?

Let us describe a concrete numerical solution to this question. Denote by ρ_n the value $\rho(x_n)$ where $x_n \in [4m^2, \infty)$. We can choose a set of x_n and approximate $\rho(x)$ by a linear spline connecting the points (x_n, ρ_n) as shown in figure 5.3. We can then analytically perform the integral in (5.6) to obtain

$$S(s) \approx S_\infty - \sum_j \mathcal{J}_j \left(\frac{g_j^2}{s - m_j^2} + \frac{g_j^2}{4m^2 - s - m_j^2} \right) + \sum_{n=1}^{M} \rho_n K_n(s) \tag{5.7}$$

where $K_a(s)$ are explicit functions of s given in appendix C.1. Evaluating this expression at some value $s_0 > 4m^2$ and plugging it into equation (5.2) gives us a quadratic constraint in the space of variables g_j^2, ρ_n and S_∞. The space of solutions of the constraints is then the intersection of all these regions for all values of $s_0 > 4m^2$. (We can visualize it as the intersection of a bunch of cylinders in a very high dimensional space.) It now suffices to start inside this region and move in the direction of increasing g_1^2 until we hit the boundary of the region and can move no more.

In practice, these numerics are simple enough that they can be performed in a few seconds in `Mathematica` using the built-in function `FindMaximum` which allows one to search for the maximum value of a function inside of some constraint region. For more details see appendix C.1.

To illustrate, let us consider scenario II from section 4.3. This corresponds to the case in which only a particle of mass m_1 couples to the external particle of mass m. In other words, we consider an S-matrix with a single s-channel pole whose residue we are trying to maximize. We can then follow the procedure outlined in this section to find the maximum value of the coupling $g_1^{\max}$ for each value of m_1/m. The results are depicted in figures 5.4 and 5.5. The maximum coupling plotted in 5.4 is the same result plotted in figure 4.9 where one can see that it agrees perfectly with the results from the boundary bootstrap!

The numerical results depicted in these plots reveal various interesting features. First, we have the spike in figure 5.4. It has a simple kinematical explanation. As $m_1 \to \sqrt{2}m$ the s- and t-channel poles in (5.6) collide and thus annihilate each other. As such we can no longer bound the residue at this point. The symmetry $g_1^{\max}(m_1^2) = g_1^{\max}(4m^2 - m_1^2)$ observed in the numerics is equally simple to understand. Each solution to the problem with $m_1 > \sqrt{2}m$ can be turned into a solution to the problem with $m_1 < \sqrt{2}m$ provided we re-interpret who is the s- and who is the t- channel pole which we can easily do if we must multiply the full S-matrix by -1. The plots in figures 5.5 corroborate this viewpoint.

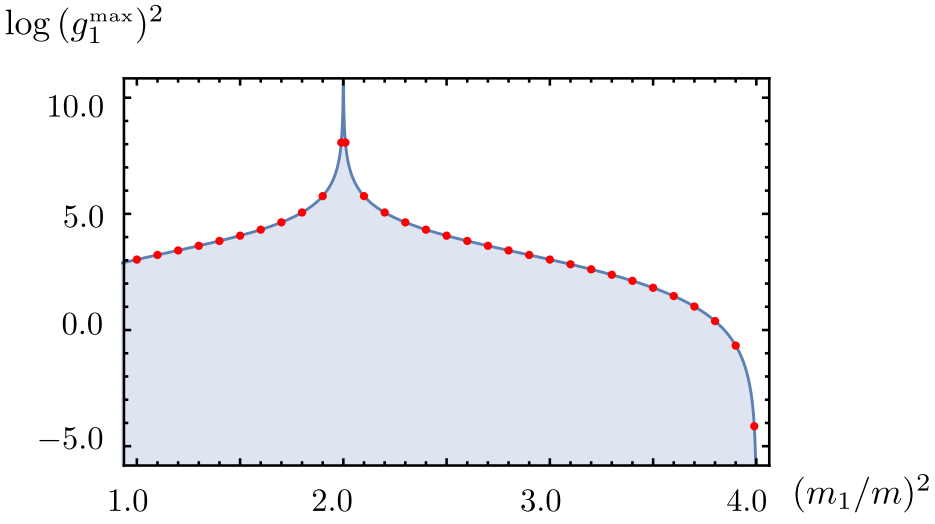


Figure 5.4: Maximum cubic coupling $g_1^{\max}$ between the two external particles of mass m and the exchanged particle of mass m_1. Here we consider the simplest possible spectrum where only a single particle of mass m_1 shows up in $S(s)$. The red dots are the numerical results obtained from the discretized dispersion relation. The solid line is an analytic curved guessed below (i.e. the coupling extracted from (5.9)) and derived in the next section. It is the same curve plotted in figure 4.9. The blue (white) region corresponds to allowed (excluded) QFT's for this simplest spectrum.

Another interesting regime is that where the exchanged particle is a weakly coupled bound-state of the external particles, that is $m_1 \simeq 2m$. As $m_1 \to 2m$ we see in the numerics that the maximum coupling vanishes. This is an intuitive result: only a small coupling can be compatible with this spectrum as a larger coupling would decrease the mass of the bound state. Note that this corner of our bounds can be tested in perturbation theory in various theories.

Finally, and most importantly, we observe in the plots in figure 5.5 that the numerical solutions for the S-matrices with the maximal residues actually saturate unitarity at *all* values of $s > 4m^2$. This observation has immediate implications. It implies the absence of $2 \to n$ particle production for any $n > 3$. After all,

$$|S_{2\to 2}(s)|^2 = 1 - \sum_{\text{other stuff X}} |S_{2\to X}(s)|^2 \,, \qquad s > 4m^2 \,. \tag{5.8}$$

Absence of particle production is the landmark of integrable models. S-matrices which

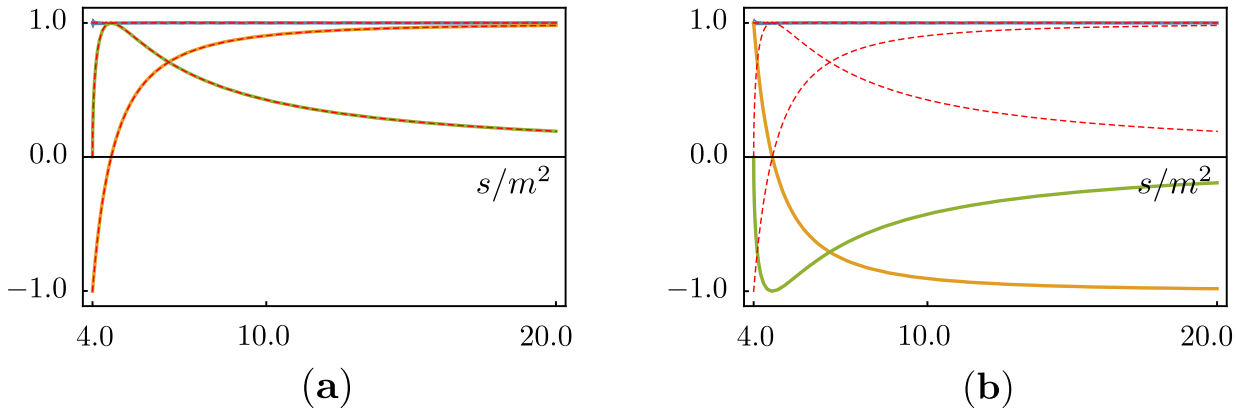


Figure 5.5: Result of numerics for (a) $m_1 = \sqrt{3}$ and (b) $m_1 = 1$. In both figures the green, orange and blue curves are $\mathrm{Im}(S)$, $\mathrm{Re}(S)$, $|S|$ respectively. Note that the blue curve is flat and equal to 1. In other words, the S-matrix that maximizes g_1 saturates unitarity at all values of $s > 4m^2$. The red dashed lines are real part, imaginary part and magnitude of the sine-Gordon S-matrix (5.9). In figure (a) the numerical results match perfectly with (5.9), while in figure (b) the numerics give precisely (-1) times the sine-Gordon S-matrix as explained in the text.

saturate unitarity often show up in the integrable bootstrap and can usually be determined analytically. When $m_1 > \sqrt{2}m$, for instance, there is a well known S-matrix obeying $|S(s)|^2 = 1$ for $s > 4m^2$ and with a single bound-state s-channel pole at $s = m_1^2$. It is the Sin-Gordon S-matrix describing the scattering of the lightest breathers in this theory; and the bound-state is the next-to-lightest breather. Explicitly, it reads [87]

$$S_{SG}(s) = \frac{\sqrt{s}\sqrt{4m^2 - s} + m_1\sqrt{4m^2 - m_1^2}}{\sqrt{s}\sqrt{4m^2 - s} - m_1\sqrt{4m^2 - m_1^2}}\,. \tag{5.9}$$

The dashed lines in figure 5.5a correspond to the values of the real and imaginary parts of this S-matrix element. Clearly, it agrees perfectly with the numerics. Our claim is that there is no unitary relativistic quantum field theory in two dimensions whose S-matrix element for identical particles has a single bound-state pole at $s = m_1^2 > 2m^2$ and a bigger residue than that of the Sine-Gordon breather S-matrix.

Also, according to what we discussed above, we conclude (and cross-check in figure 5.5b) that the S-matrix with the maximum coupling $g_1^{\max}$ and with a bound-state $m_1 < \sqrt{2}m$ is given by an S-matrix which differs from the Sine-Gorgon S-matrix by a mere minus sign,

$S(s) = -S_{SG}(s)$. We do not know of any theory with this S-matrix.[4]

In the next section we will explain that the phenomenon we encountered empirically here – i.e. saturation of unitarity – is actually generic and not merely a peculiarity of this simplest example with a single exchanged particle. This will open the door toward an analytic derivation of $g_1^{\max}$ for any bound-state mass spectrum of $\{m_1/m, m_2/m, \dots\}$.

5.2 Castillejo-Dalitz-Dyson factors and the Analytic Bootstrap

An important hint arose from the numerics of the last section: For the simplest possible mass spectrum (with a single s-channel pole), we found that the optimal S-matrix – leading to a maximum coupling $g_1^{\max}$ – *saturates unitarity* at any $s > 4m^2$ (see the blue curves in figure 5.5). This simple example suggests that one should be able to borrow standard machinery from the integrable bootstrap literature to tackle this problem analytically. This is what we pursue in this section. Ultimately, this will lead to an analytic prediction for $g_1^{\max}(m_1/m, \dots)$ for an arbitrary spectrum of masses. Actually, our analysis will determine the full S-matrix element corresponding to this maximal coupling.

To proceed, it is convenient to change variables from s to the usual hyperbolic rapidity θ with $s = 4m^2 \cosh^2(\theta/2)$. The mapping from s to θ is shown in figure 5.6. The strip $\mathrm{Im}(\theta) \in [0, \pi]$ covers the full physical s-plane of figure 5.2 and is thus called the physical strip. We recall in appendix C.2 a few useful properties of this parametrization. In terms of θ we write crossing and unitarity as

$$S(\theta) = S(i\pi - \theta)\,, \qquad S(\theta + i0)S(-\theta + i0) = f(\theta)\,, \tag{5.10}$$

Where f is the right hand side of (5.8) which we do not know. We do know that, by definition, this absorption factor takes values in $f \in [0, 1]$ for physical momenta, that is for $\theta \in \mathbb{R}$. Now, a solution to (5.10) can always be written as

$$S(\theta) = S_{\mathrm{CDD}}(\theta) \exp\left(-\int \frac{d\theta'}{2\pi i} \frac{\log f(\theta')}{\sinh(\theta - \theta' + i0)}\right) \tag{5.11}$$

[4]It is also conceivable that such theory does not exist at all; the bound for $m_1/m > \sqrt{2}$ must be optimal since Sine-Gordon theory exists but the left region of the plot in figure 5.4 for $m_1/m < \sqrt{2}$ might still move down as we include into the game further constraints such as those coming from S-matrix elements involving other particles in the theory as external states. Similar things happened in the conformal bootstrap story.

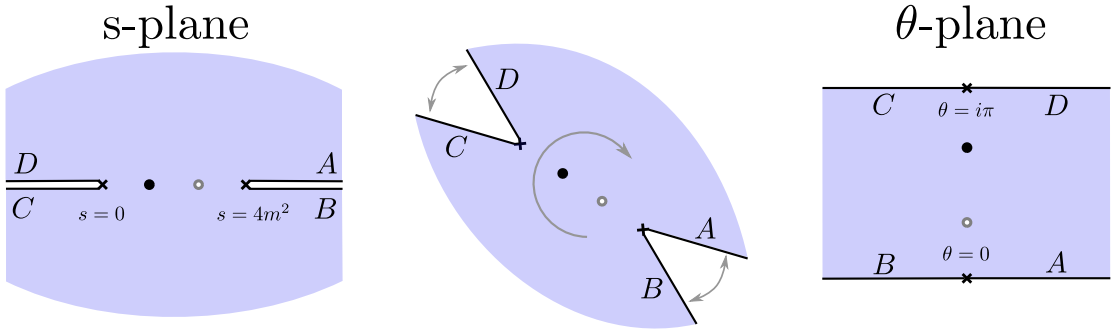


Figure 5.6: Mapping from s to θ. The map "opens" the cuts and rotates clockwise by $\pi/2$. The physical sheet of the s-plane is mapped to the strip $\mathrm{Im}(\theta) \in [0, \pi]$ with $s = 0$ ($s = 4m^2$) mapping to $\theta = i\pi$ ($\theta = 0$).

where the exponential factor is a particular solution to (5.10) – known as the minimal solution – while $S_{\mathrm{CDD}}(\theta)$ is a solution to (5.10) with $f = 1$. Note that the minimal solution has no poles (or zeros) in the physical strip; any poles (or zeros) are taken into account by S_{CDD}.

It is now rather straightforward to understand why the process of maximizing the coupling to the lightest exchanged particle leads to S-matrices which saturate unitarity, i.e. for which $f = 1$. Indeed, using the fact that f is an even function, we can symmetrize the integral in the minimal solution to get

$$S(it) = S_{\mathrm{CDD}}(it) \times \exp\left(\int_{-\infty}^{+\infty} \frac{d\theta'}{2\pi} \underbrace{\frac{\sin(t)\cosh(\theta')}{|\sinh(it-\theta')|^2}}_{\text{positive for } t \in [0,\pi]} \times \underbrace{\log f(\theta')}_{\text{negative}} \right). \tag{5.12}$$

in the segment $t \in [0, \pi]$ corresponding to $s \in [0, 4m^2]$ where the potential poles of the S-matrix lie. We see that the minimal solution always decreases the magnitude of the S-matrix in this segment unless $f = 1$. Therefore, if we are to maximize some residue in this region it is always optimal to set $f = 1$. This simple observation explains and the saturation of unitarity observed experimentally in the last section and establishes it for any spectrum of poles.

Next we have the Castillejo-Dalitz-Dyson (CDD) factor which solves the homogenous problem

$$S_{\mathrm{CDD}}(\theta) = S_{\mathrm{CDD}}(i\pi - \theta)\,, \qquad S_{\mathrm{CDD}}(\theta) S_{\mathrm{CDD}}(-\theta) = 1\,. \tag{5.13}$$

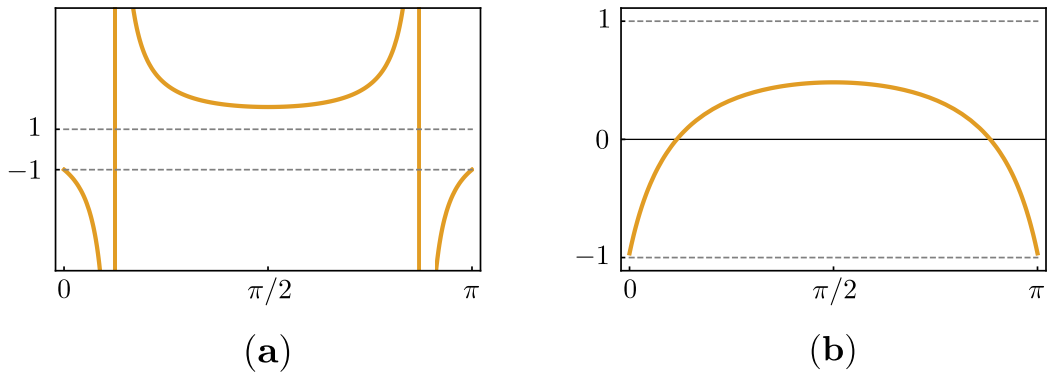


Figure 5.7: Panel (a) shows a CDD pole $[\pi/8]$ for θ purely imaginary between 0 and $i\pi$. Note that the magnitude of this factor is always greater or equal 1. Also note that it is positive between its s- and t- channel poles, while the tails of the function are negative. Panel (b) shows a CDD zero $[-\pi/8]$ in the same interval. The magnitude of this function in this interval is always less than or equal to 1.

There are infinitely many solutions to this homogenous problem which we can construct by multiplying any number of so-called CDD factors,

$$S_{\text{CDD}}(\theta) = \pm \prod_j [\alpha_j] \,, \qquad [\alpha] \equiv \frac{\sinh(\theta) + i \sin(\alpha)}{\sinh(\theta) - i \sin(\alpha)} \,. \tag{5.14}$$

Without loss of generality, we take α to be in the strip $\text{Re}(\alpha) \in [-\pi, \pi]$. Still, depending on its value these CDD factors $[\alpha]$ can represent very different physics. There are basically three different instances to consider:

Consider first the case when α is in the right half of the above mentioned strip, i.e. $\text{Re}(\alpha) \in [0, \pi]$. In this case the corresponding CDD factor will have a pole at $\theta = i\alpha$ in the physical strip. Because of locality such poles should always be located in the segment $s \in [0, 4m^2]$ corresponding to θ purely imaginary between 0 and $i\pi$. Therefore if α is in the right half of its strip, it ought to be purely real with $\alpha \in [0, \pi]$. In this case, the CDD factor $[\alpha]$ is referred to as a **CDD-pole**; an example is plotted in figure 5.7a. Clearly, these factors are very important. They are the only factors which give rise to poles in the S-matrix corresponding to stable asymptotic particles.

When α is in the left half of the above mentioned strip there are less physical constraints on its admissible values. The reason is that in this case the corresponding factor induces a pole at $\theta = i\alpha$ which is now no longer in the physical strip. In terms of s it would

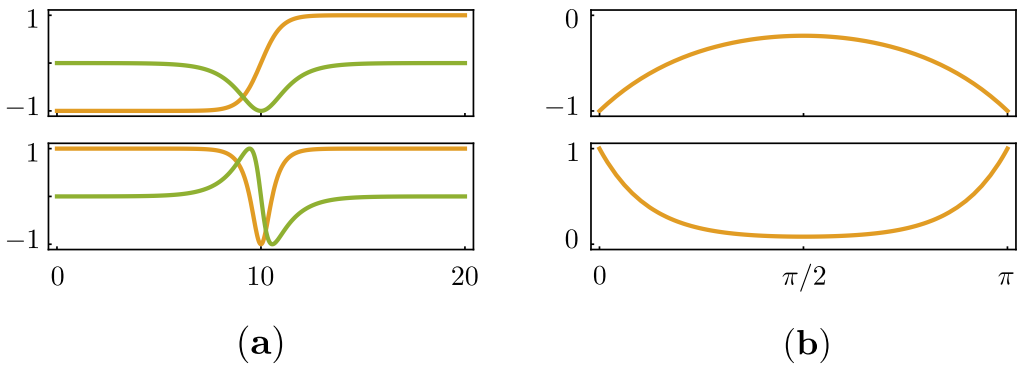


Figure 5.8: Panel (a) shows the behaviour of two types of CDD resonances for real θ. The upper and lower plots show $[-\pi/2 + 10i]$ and $[-\pi/5 - 10i][-\pi/5 + 10i]$ respectively. The orange curve is the real part while the green curve is the imaginary part. Resonances can be added at very little cost. If some parameters are large, for example, their effect only shows up at very high energies nearly not affecting low energy physics as expected. Panel (b) shows the behaviour of two resonance factors for θ purely imaginary between 0 and $i\pi$. The upper and lower panel show $[-\pi/2 + i]$ and $[-\pi/3 - i][-\pi/3 + i]$ respectively. In the former case the resonance factor is purely real in this interval while in the later case the product is real although the individual factors are not. Note that in this interval CDD resonances always have magnitude less than 1 and that each individual CDD resonances never changes sign.

be on another sheet after crossing some of the cuts in figure 5.2. A priori, there is not much we can say about possible positions of poles which leave the physical strip. It is still convenient to separately consider two possible cases. If α is purely real in the left strip – that is if $\alpha \in [-\pi, 0]$ – we say $[\alpha]$ is a **CDD-zero**. The reason is clear: such factor has a zero at $\theta = -i\alpha$ inside the physical strip and along the very same segment where possible poles will be. An example of a CDD zero is plotted in figure 5.7b. We can also have complex values of α provided they are carefully chosen not to spoil real-analyticity of S-matrix which requires that $S(\theta)$ should be real in the segment between 0 and $i\pi$. One possibility, for example would be to have $\alpha = -\pi/2 + i\beta$ where β is purely real. Another option would be to have a pair of complex conjugate α's such that their product would lead to a real contribution in the above mentioned segment. Such CDD contributions also lead to zeros in the physical strips, this time at complex values of θ. We refer to such factors as **CDD-resonances**. Examples of CDD resonances are plotted in figure 5.8.

Let us now discuss some general features of these three CDD factors which are relevant

for our present purposes. We see in figure 5.7a that a CDD-pole factor has magnitude greater than one at any point in the segment $\theta = [0, i\pi]$. On the other hand from figure 5.7b and 5.8 we see that CDD-zeros and CDD-resonances have magnitude always smaller or equal to one in this segment. As such, one may (incorrectly) conclude that the S-matrix which maximizes g_1 and is compatible with a given spectrum of asymptotic stable particles $\{m_1/m, m_2/m, \dots\}$ is simply given by a product of CDD-poles, one for each stable particle.

This is too hasty for the simple reason that such a naive product of CDD-poles will generically have wrong signs for the corresponding residues contradicting (5.3).[5] Hence, a more thoughtful conclusion is that while we can indeed discard any CDD-resonances, CDD-zeros are sometimes a necessary evil. In contradistinction with the CDD-ressonances and also with the minimal solution discussed above, CDD-zeros change sign in the segment $\theta = [0, i\pi]$ so we can – and must – use them to flip the wrong signs of any residues. The correct prescription is therefore to dress the product of CDD-poles by a potential overall sign plus a minimal amount of CDD-zeros such that the signs of all the residues come out right. The position of the CDD-zeros is then fixed such that g_1 is maximal. Appendix C.4 contains the final outcome of this maximization problem for the most general mass spectrum. Here we find it instructive to proceed in some detail with some simple examples.

Let us begin with the simplest case in which there is a single particle with $m_1 < 2m$. We wish to maximize the coupling for the process $m + m \to m_1$. This was the case considered in section 5.1 and for which the results of the numerics are given in figures 5.5 and 5.4. Since there is only a single bound state, we require only one pole and thus the solution is given by $S = \pm[\alpha_1]$ where α_1 is fixed by the condition

$$m_j^2 = 4\cosh^2(i\alpha_j/2) \tag{5.15}$$

and the $\pm$ is fixed such that the residue of the s-channel pole is positive. This leads to $S = [\alpha_1]$ for $m_1 > \sqrt{2}$ and $S = -[\alpha_1]$ for $m_1 < \sqrt{2}$.

Now suppose we have two particles such that $m_1 < m_2 < 2m$ and again we wish to maximize the coupling for the process $m + m \to m_1$. Clearly we should start with at least two CDD factors to accomodate bound-state poles at $s = m_1^2$ and $s = m_2^2$. However, the analysis is complicated by the requirement that the residues of these poles be positive since each individual CDD factor changes sign at its poles (see figure 5.7a). We must consider the four distinct configurations of s- and t-channel poles shown in figure 5.9a. First consider

[5] Translating (5.3) to θ-space we have that a proper s-channel pole corresponding to a mass $m_j^2 = 4\cosh^2(\theta_j/2)$ should behave as $S \simeq i\Gamma_j^2/(\theta - \theta_j)$ with Γ_j^2 positive and related to g_j^2 by some trivial Jacobians. Correspondingly, the associated t-channel pole will be located at $\theta = i\pi - \theta_j$ and will have a negative residue $S \simeq -i\Gamma_j^2/(\theta - i\pi + \theta_j)$.

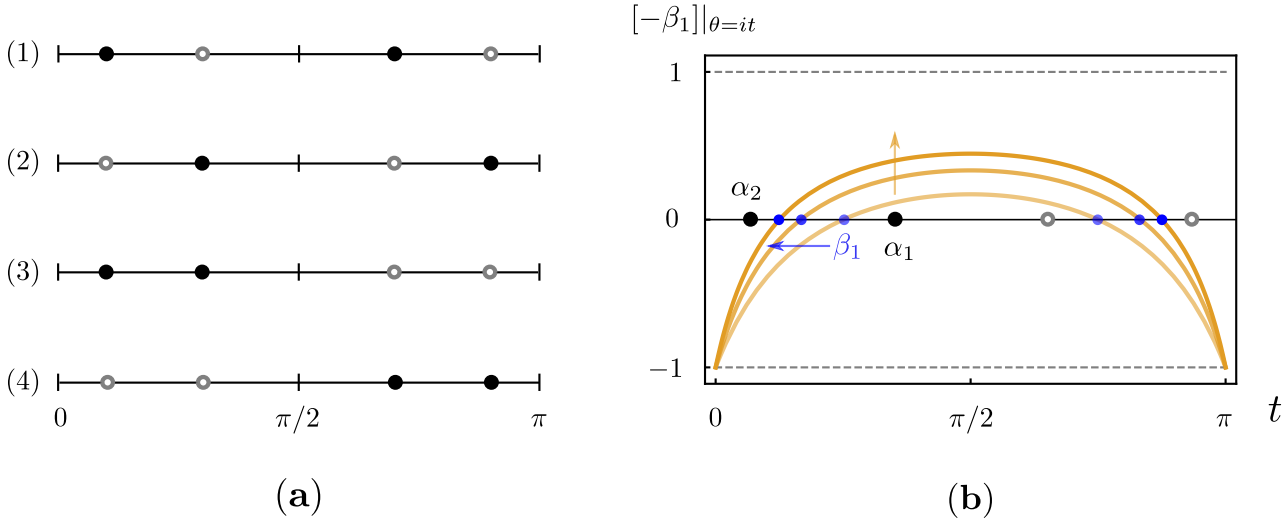


Figure 5.9: Panel (**a**) shows the four possible configurations of poles for a spectrum $m_1 < m_2 < 2m$ and no cubic coupling. Cases (1) and (2) correspond to $m_1 < \sqrt{2} < m_2$ the former with $m_1^2 > 4 - m_2^2$ and the latter with $m_1^2 < 4 - m_2^2$. Cases (3) and (4) correspond to $\sqrt{2} < m_1 < m_2$ and $m_1 < m_2 < \sqrt{2}$ respectively. The residues of a product of CDD factors alternate between positive and negative since a CDD factor changes sign at each of its poles and nowhere else. Thus in case (1) and (2) we can arrange for (5.3) to be satisfied simply by fixing the overall sign of the S-matrix. Cases (3) and (4) cannot be repaired in this way. Instead we must multiply by a CDD zero in order to fix the signs. Panel (**b**) shows a CDD zero factor $[-\beta_1]$ with $\alpha_2 < \beta_1 < \alpha_1$ such that is changes sign between the two s-channel poles and also between the two t-channel poles. In this way the product $\pm[\alpha_1][-\beta_1][\alpha_2]$ will have the correct residues (the overall sign which can be then fixed as in cases (1) and (2)). The precise value of β_1 must then be fixed to maximize g_1^2 which is the residue at α_1. We see that $[-\beta_1]$ grows monotonically as we shift the zero to the left toward α_2. Optimizing then implies that we must collide this zero with the pole at α_2.

cases (1) and (2) which correspond to $m_1 < \sqrt{2} < m_2$. Here the solution is simply given by $S = \pm\,[\alpha_1]\,[\alpha_2]$. Once the correct overall sign is selected, the residues of the poles work out since the poles alternate between s and t channel. We fix the overall sign as follows. Notice from figure 5.7a that an individual CDD factor is positive between its poles and negative before and after – i.e. the tails of the CDD factors are always negative. Further, the pole of an individual CDD factor closest to $i\pi$ has the form $i\,(-1) \times$ (positive). Thus, for a general product of such factors the sign of the residue closest to $i\pi$ has the form

$i\,(-1)^N \times$ (positive). If $m_1^2 > 4 - m_2^2$ this pole will be t-channel as in case (1) of figure 5.9a and since $N = 2$ we should choose the overall sign (-1). On the other hand when $m_1^2 < 4 - m_2^2$ the first pole is s-channel is in case (2) and thus we should choose the overall sign $(+1)$.[6]

Now consider the case (3) in figure 5.9a which corresponds to $\sqrt{2} < m_1 < m_2$. Now a simple product of two CDD poles cannot have the correct signs for its residues. The signs alternate at each pole but we have two consecutive s-channel poles with no t-channel pole in between. To correct for this, we are forced to insert a CDD zero $[-\beta_1]$ between the two s-channel poles $\alpha_2 < \beta_1 < \alpha_1$. Such a factor also has a zero between the two t-channel poles since it is crossing symmetric. The precise position of this zero is then fixed by the condition that g_1^2 be maximized – i.e we want to maximize the value of the CDD zero at the position α_1. From figure 5.9b we see that this means we should move the zero as far away from α_1 as possible. In particular, it implies that we should collide the zero with the pole at α_2, thus decoupling that state from the scattering of the lightest particle. In other words, the the optimal S-matrix is given by $S = [\alpha_1]$ for $\sqrt{2} < m_1 < m_2$. Note that this does not contradict our assumption that there is a particle m_2 in the spectrum. Rather, it simply implies that the S-matrix that maximizes g_1 has no coupling to this asymptotic state (i.e. $g_2 = 0$). Lastly, case (4) is related to (3) by reflection about $\pi/2$ so in that case we have $S = -[\alpha_1]$. The final result of all this analysis is summarized in figure 5.10.

Note that this example contains scenario I from section 4.3 as a special case. In particular, it corresponds to the slice of this plot along the plane $m_1/m = 1$ which is the upper-left boundary of the plotted region. A range of this slice is precisely what is plotted in figure 4.7 where one can see that it agrees beautifully with the data from the boundary bootstrap in the region where numerical results from the later are available.

The case $N = 2$ that we have just discussed demonstrates all the salient features of the general case. In particular for a set of masses $m_1 < m_2 < ... < m_N < 2m$ corresponding to $\{\alpha_1, ..., \alpha_N\}$ the optimizing $S(s)$ will always be given by (5.14) where the product runs over a *subset* of the masses. The product is only over a subset because the collision of zeros and poles we observed in the $N = 2$ case is a feature present in the general solution. That is, whenever the poles do not alternate between s- and t-channel, we are forced to insert CDD zeros so that the residues obey (5.3). Maximizing with respect to the position of these zeros always forces them to collide with a pole, thus decoupling that state from the scattering process. Precisely which poles get canceled is explained in appendix C.4. Finally, the overall sign in (5.14) is fixed by considering whether the pole closest to $i\pi$ is

[6] In general, configurations of poles which are related by reflection about $\pi/2$ have an S-matrix related by an overall sign. This follows from the dispersion relation (5.6) as explained in section 5.1.

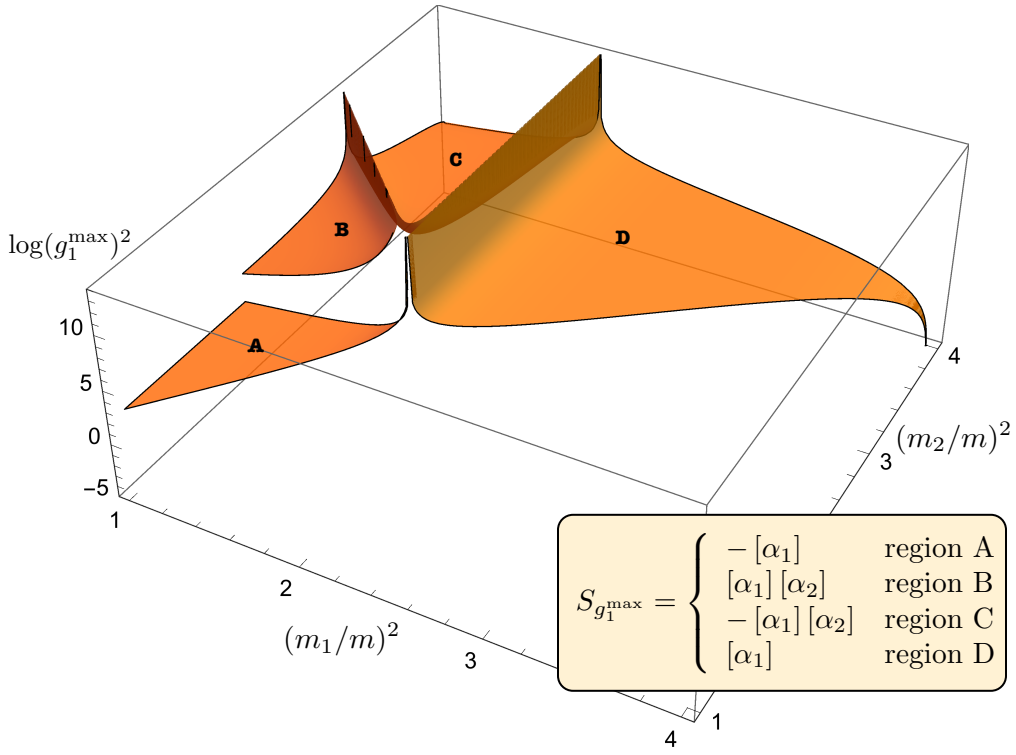


Figure 5.10: Maximal coupling $g_1^{\max}(m1/m, m2/m)$ for the spectrum $m_1 < m_2 < 2m$ and no cubic coupling. Each region corresponds to one of the four configurations of poles shown in figure 5.9a with regions A,B,C,D corresponding to (4),(2),(1),(3), respectively.

s- or t-channel. The end result of this analysis is formula (C.7) given in appendix C.4. As an application of this formula which will be relevant in the next section, in figure 5.11 we present the maximal coupling for the case $m_1 = m$ (i.e. a cubic coupling $m + m \to m$) and generic $m_1 < m_2 < m_3 < 2m$. Finally, we have verified in all these cases that performing the numerics of section 5.1 for the various configuration of poles confirms the CDD solutions given above.

We will now conclude with some comments regarding the CDD solution (5.14). First we note this solution (5.14) does not cover the full space of solutions of (5.13). We can

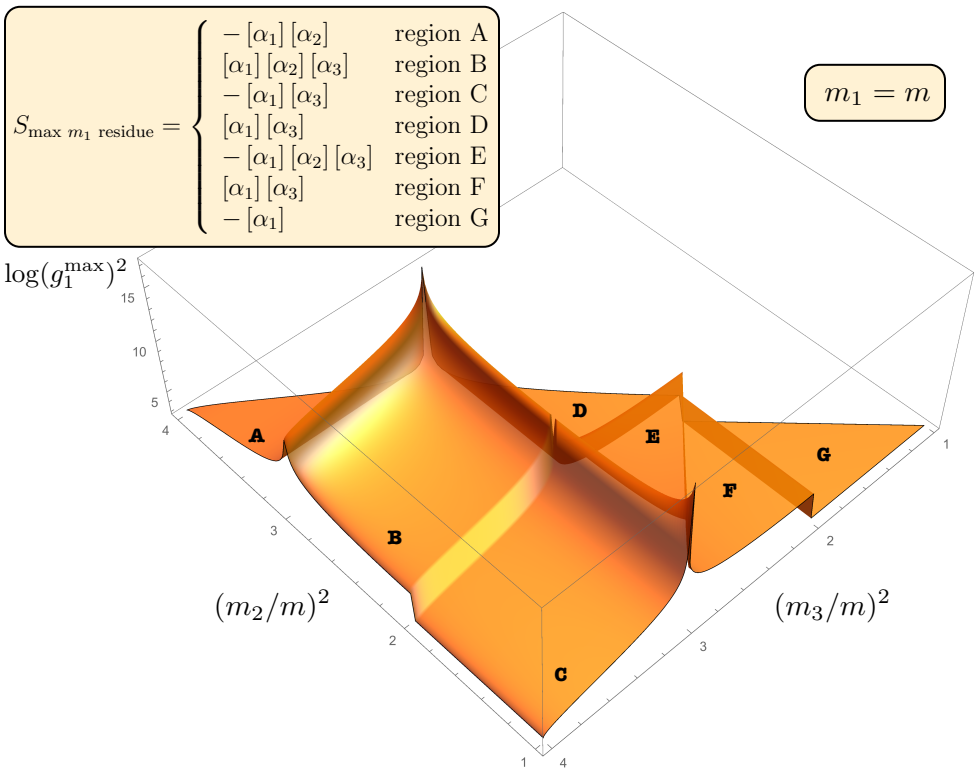


Figure 5.11: Maximal coupling $g_1^{\max}(m_2/m, m_3/m)$ for the spectrum $m_1 = m$ (i.e. a cubic coupling $m + m \to m$) and generic $m < m_2 < m_3 < 2m$.

always multiply (5.14) by any other solution of (5.13) such as the factor

$$S_{\text{grav}}(s) = e^{il_s^2\sqrt{s(s-4)}} = e^{2il_s^2 m^2 \sinh\theta} \tag{5.16}$$

with an arbitrary parameter l_s^2. This solution, called a "gravitational dressing factor" was recently introduced in [88]. For our purposes we can rule out the possibility of such a factor since $S_{\text{grav}} \in [0,1]$ in the segment $\theta = [0, i\pi]$ and thus will always decrease the value of g_1. We are not aware of any other solutions of (5.13) and the agreement of our results with numerics indicates that we have identified the correct class of solutions for our considerations.

Second, note that the general CDD solution (5.14) saturates unitarity ($|S_{\text{CDD}}| = 1$ for θ real) which implies the absence of particle production in the scattering $m + m$. As we

have already mentioned in section 5.1 absence of particle production is an indication of integrability. Thus one may wonder if each point on the surfaces of figures 5.10 and 5.11 correspond to some integrable model. As we shall see in the next section, generic points in these plots cannot correspond to integrable models without the addition of new particles into the spectrum. As such, for a fixed spectrum only very special points correspond to integrable theories.

5.3 The Ising Model with Magnetic Field

Figures 5.10 and 5.11 are examples of bounds on couplings of a quantum field theory given some mass spectrum. An obvious question is whether there are interesting field theories saturating these bounds. Also, when the answer is *no* what can we do to lower the bounds further until the answer is *yes*?

In some regions of these plots we already know the answer to these questions. Take for example the $m_2 = 2$ section of figure 5.10. As $m_2 \to 2$ this particle enters the two-particle continuum thus disappearing from the spectrum. We are thus left with a single exchanged particle m_1. This was precisely the case discussed in the simple numerics example and depicted in figure 5.4. For any $m_1 > \sqrt{2}m$ we do know of a theory which saturates this bound: it is the Sine-Gordon integrable theory when we identify m as the first breather and m_1 as the second breather.

What about the more general bounds in figures 5.10 and 5.11? All the optimal S-matrices there which maximize g_1 saturate unitarity and thus admit no particle production. Do they correspond to proper S-matrices of good integrable quantum field theories with their respective mass spectra? We will now argue that the answer to this question is *no*.

As an example we will focus on region B in figure 5.11. That is we will focus on the space of theories where there are three stable particles: the lightest particle itself with $m_1 = m$ and two other heavier particles with

$$\sqrt{2}m_1 < m_2 < \sqrt{3}m_1 < m_3 < 2m_1 \,. \tag{5.17}$$

In this region the S-matrix which maximizes g_1 is a simple product of three CDD factors,

$$S(\theta) = \frac{\sinh(\theta) + i\sin(2\pi/3)}{\sinh(\theta) - i\sin(2\pi/3)} \times \frac{\sinh(\theta) + i\sin(\alpha_2)}{\sinh(\theta) - i\sin(\alpha_2)} \times \frac{\sinh(\theta) + i\sin(\alpha_3)}{\sinh(\theta) - i\sin(\alpha_3)} \,, \qquad m_j = 2\cos(\alpha_j) \,. \tag{5.18}$$

We will now argue that in the region (5.17) of parameter space our bound should *not* be the strongest possible bound except at a single isolated point which we will identify with

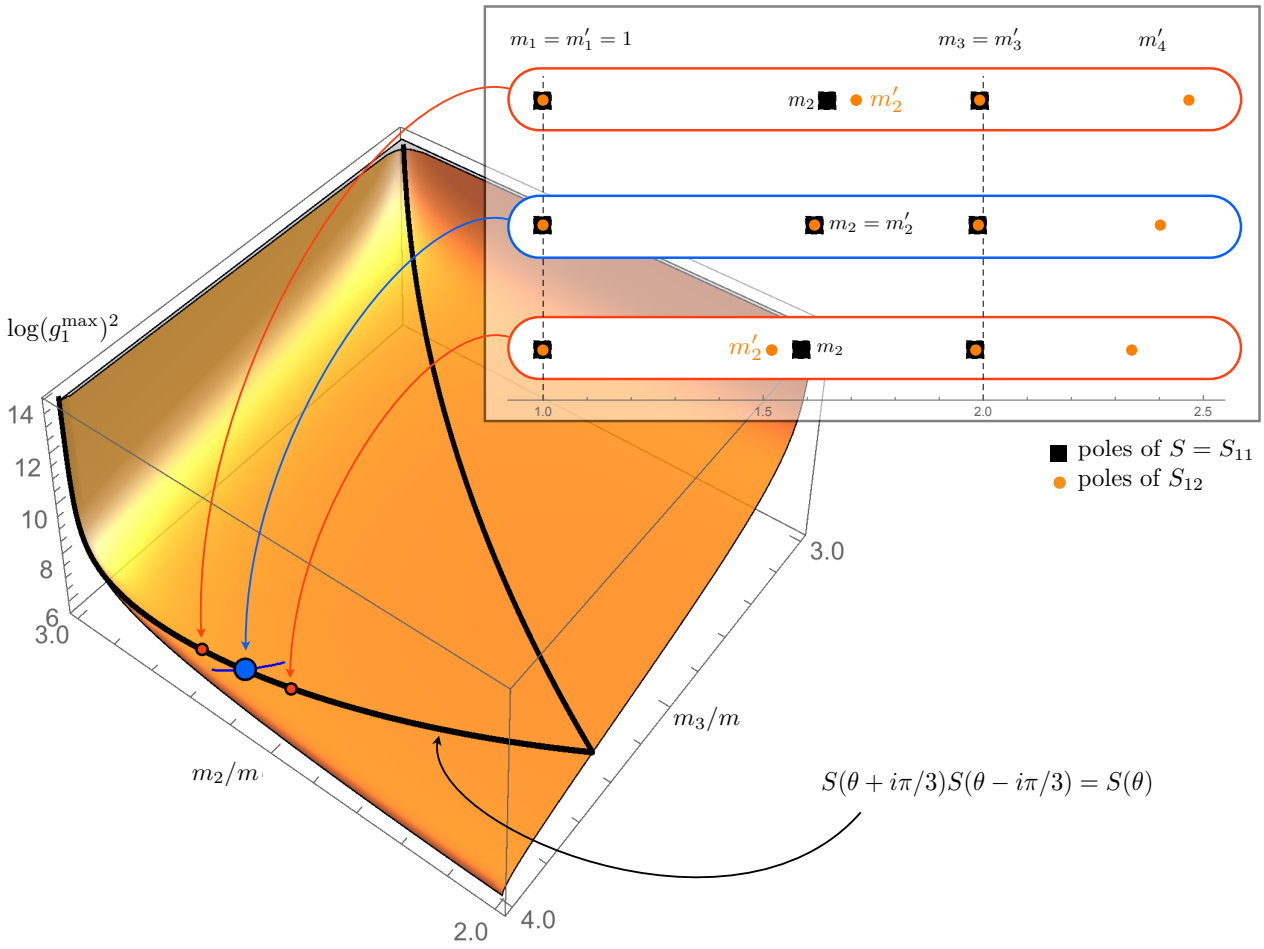


Figure 5.12: Blow up of region B from figure 5.11. The thick black line is where the cubic fusion property (5.20) holds (i.e. assumption (1) in the discussion of section 5.3). In the upper right corner we plot the s-channel poles of S_{12} versus those of S. We see that, following the thick black line, only at the blue dot does S_{12} have poles at the same locations as S indicating that assumption (3) from section 5.3 also holds.

a well known and very interesting field theory.[7] We will do this by observing some simple pathologies with (5.18) which are resolved once α_2 and α_3 take some particular values

[7]The reader fond of section titles probably guessed which one.

which we identify below.

To proceed we need to make three natural assumptions about a putative theory living in the boundary of our bounds for a fixed mass spectrum $\mathbb{m}$:

A1 The theory is integrable.[8]

A2 The exchanged particle with mass $m_1 = m$ is the same as the external particle and not just another particle in the theory with the same mass as the external particle.

A3 There are no other stable particles below the two particle threshold $2m_1$ other than those in $\mathbb{m}$.

In an integrable theory we can construct bound-state S-matrix elements from the fundamental S-matrix by so called *fusion*. If the stable particle shows up as a pole at $\theta = i\alpha_j$ in $S(\theta)$ then the S-matrix of this bound-state with the fundamental particle of mass m can be built by scattering both its constituents,

$$S_{\text{j}^{\text{th}}\text{ bs, fund}}(\theta) = S(\theta + i\alpha_j/2)S(\theta - i\alpha_j/2)\,. \tag{5.19}$$

This relation can be easily established starting with the $3 \to 3$ S-matrix which is factorized as a product of three two-body S-matrices. We can then take two of the three particles in the initial state and form a bound-state. This will then describe a scattering of that bound-state with the remaining fundamental particle. (Because the theory is integrable, the individual momenta in the out state are the same as in the in-state so automatically we will be fusing into another bound-state in the future.) In this fusion process one of the three S-matrices (the one involving the particles being fused into a bound-state) simplifies (it yields a single pole of which we extract the residue) leaving us with two S-matrices which are nothing but the right hand side of (5.19). We can also justify (5.19) in a more physical way as depicted in figure 5.13.

With the fusion property (5.19) following from assumotion A1 we will now show that powerful constraints on the spectrum follow from assumptions A2 and A3.

If a theory has a cubic coupling and $m_1 = m$ shows up as a pole in the S-matrix then it can itself be thought of as a bound-state. That is, under the assumptions (1) and (2) above we conclude that we must have

$$S(\theta) = S(\theta + i\pi/3)S(\theta - i\pi/3)\,. \tag{5.20}$$

[8]This is of course very natural since the S-matrices we found saturate unitarity and thus admit no particle production. Absence of particle production is of course a necessary condition for integrability. In most cases it is also a sufficient condition, see e.g. [89].

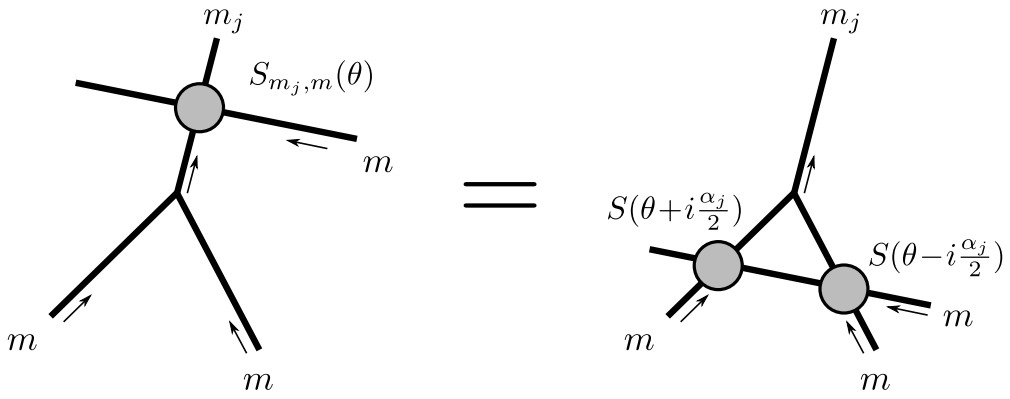


Figure 5.13: Suppose we take two (to be) constituents of a bound-state and throw them very slowly at each other so that they travel (almost) parallel to each other in space-time until they are close enough to feel each other and thus form the bound-state. Now suppose we want to scatter a fundamental particle with this bound-state as indicated on the left in this figure. This is how we would compute the left hand side of (5.19). In an integrable theory we can shift at will the position of the wave packet of this fundamental particle. So we can shift it far into the past such that it scatters instead with the constituents of the bound-state well before they were bound together as represented on the right. This leads to the right hand side of (5.19).

This is an important self-consistency constraint. We can now plug the solution (5.18) in this relation. We observe that it is generically *not* satisfied. However, there is a line $\alpha_3(\alpha_2)$ or equivalently $m_3(m_2)$ where it holds. This is the thick black line in figure 5.12. Away from this black line we can already conclude that our bound is either not the optimal bound *or* some of the assumptions A1 or A2 (or both) should not hold.

Sticking to the black line and continuing with assumption (3) we can do even better. We can now construct the S-matrix element $S_{12}(\theta) = S(\theta + i\alpha_2/2)S(\theta - i\alpha_2/2)$ for the scattering $m_1 + m_2 \to m_1 + m_2$ involving the lightest and the next-to-lightest particles. We can then look at the poles of this S-matrix which will correspond to asymptotic particles of the theory. There is a point in the black line, marked with the blue dot in figure 5.12 where these poles correspond perfectly to the spectrum $\mathbb{m} = \{m_1(= m), m_2, m_3\}$. Namely we find precisely three s-channel poles at $s = m_1^2, m_2^2, m_3^2 < (2m_1)^2$ which are the very same locations in the fundamental S-matrix $S(\theta)$. However, as we move away from this blue point something bad happens. We see that the poles at $s = m_1^2$ and $s = m_3^2$ are as expected however the pole at m_2^2 shifts to a nearby position $m_2'^2$. This would indicate

the presence of a new particle not in $\mathbb{m}$ with a mass close to that of m_2. This violates assumption A3.

Ultimately, only the blue dot in figure 5.12 which is located at

$$m_2 = 2\cos(\pi/5)m_1\,, \qquad m_3 = 2\cos(\pi/3)m_1, \tag{5.21}$$

survives! We conclude that under the assumptions A1–A3 the maximal coupling in region B of plot 5.11 (which corresponds to masses satisfying (5.17)) should be lower than the one we found everywhere except perhaps at the blue point.[9]

What about this blue dot? Is there a special integrable theory with these masses and an S-matrix given by (5.18)? *Yes*, it is the Scaling Ising model field theory with magnetic field [90]. Its is a very interesting strongly coupled integrable theory with E8 symmetry which describes the massive flow away from the critical Ising model when perturbed by magnetic field (holding the temperature fixed at its critical value).[10] Thus the CDD solution provides a sharp (i.e. as strong as possible) upper bound on g_1 for this value of the masses. In what follows we shall refer to the blue dot in figure 5.12 as the *magnetic* point.

The thin blue line in figure 5.12 represent the variation of the masses of the stable particles m_2 and m_3 of the scaled Ising model as we move away from the magnetic point by shifting the Ising model temperature away from its critical value. The slope $\delta m_2/\delta m_3$ defining this line can be computed using so-called form factor perturbation theory as recalled in appendix C.3. As we turn on the temperature the corresponding field theory is no longer integrable (see [91] for a review of the scaling Ising model with temperature and magnetic field turned on). Particle creation shows up to linear order in the thermal deformation but since this same particle production only shows up quadratically on the right hand side of (5.8), its effect of the elastic component S should be subleading. As such we expect that our bound for g_1 also captures the residue of the Scaled Ising model in the vicinity of the magnetic point. This is what we check in detail in appendix C.3.

A conclusion of the discussion above is that for finite deformations away from the magnetic point, the bound in figure 5.12 is not optimal. The obvious question is then how to improve it? One strategy would be to include other S-matrix elements into our

[9]Note that we can not exclude having other integrable theories living in the black line provided we accept more stable particles below threshold showing up in other S-matrix elements. We could also drop assumption A2 and conceive integrable theories where m_1 is not the same particle as the external one (despite having the same mass). If we keep assumption A3, the conclusion leading to the blue dot as a special isolated theory still holds.

[10]This is perhaps not that surprising. After all, many of the conditions we just imposed are simple recast of standard integrable bootstrap logic as used, for instance, in [90].

analysis. In particular, it would be very interesting to include into the game the simplest absorptive components which are the inelastic $2 \to 2$ processes $m + m \to m + m_2$ and $m + m \to m_2 + m_2$. Their existence, away from the integrable magnetic point, will forbid us to saturate unitarity for $S(\theta)$ since they will show up in the right hand side of (5.8). By taking them into account we expect therefore to be able to improve our bound. As we add these components to our analysis, it would be formidable if a ridge-like feature passing the magnetic point represented by the blue dot would develop in figure 5.12. By moving along this ridge we would hopefully be moving along the non-integrable thermal deformation thus accessing the full Scaling Ising model with temperature *and* magnetic field. We are currently studying this problem and hope to report on progress in this direction in the near future. In the CFT bootstrap, adding further components to the analysis proved to be a very powerful idea [82]. Hopefully the same will be true here. It would also be very interesting to consider multi-particle scattering such as $2 \to 3$ processes but these are kinematically more complicated and we did not dare explore them yet.

5.4 Discussion

Armed with the insights of the remarkable recent progress in the *conformal bootstrap* and with the well ironed technology of the *integrable bootstrap*, we revisited here the *S-matrix bootstrap* program. We found bounds on the maximal couplings in massive two dimensional quantum field theories with a given mass spectrum. We obtained these bounds numerically (see section 5.1) and analytically (see section 5.2) with perfect agreement between the two methods, see e.g. figure 5.4. These bounds also made contact with well known integrable theories. We found, for example, that there is no unitary relativistic quantum field theory in two dimensions whose S-matrix element for identical particles has a single bound-state pole at $s = m_1^2 > 2m^2$ and a bigger residue than that of the Sine-Gordon breather S-matrix.

In the section 4.4 we attacked this same bound problem from yet a different perspective. There, we considered a Gedankenexperiment where we put massive (d-dimensional) quantum field theories into a (Anti de Sitter fixed background) box. We can then study their landscape by analyzing the conformal theories they induce at the ($d-1$ dimensional) boundary of this space-time. This allows us to make use of well developed numerical methods of the conformal bootstrap for putting bounds on conformal theory data which then translate into bounds on the flatspace QFT data. An important difference with respect to previous works on conformal bootstrap is that the setup used here requires all conformal dimensions involved in the bootstrap to be very large. This is how we make sure the AdS box is large and the physics therein is equivalent to that in flat space. This poses

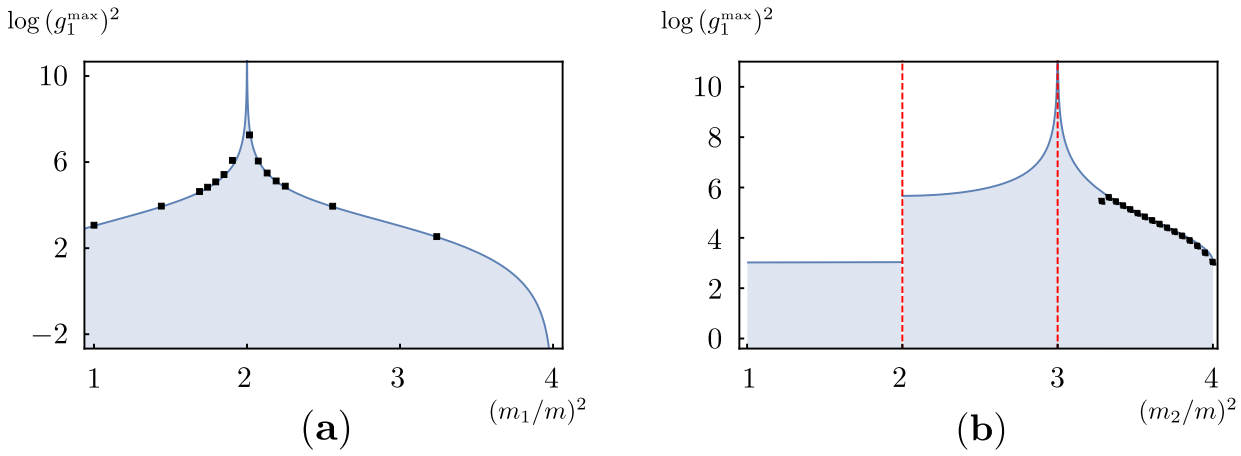


Figure 5.14: Maximal coupling $g_1^{\max}$ for (**a**) a single exchanged particle of mass m_1 and (**b**) a particle of mass $m_1 = m$ plus an heavier particle of mass m_2. The solid blue lines are the analytic results of the two dimensional S-matrix bootstrap. These are nothing but the top right and top left slices of the more general figure 5.10. The black squares are the outcome of the one dimensional conformal bootstrap numerics from [69]. These numerics are obtained using `SDPB` in (**a**) and `JuliaBot` in (**b**). In either case, within the precision of the numerics, the agreement with the analytic result is striking. It is worth emphasizing that the solid curves are very non-trivial functions. The right-most branch of (**b**), for instance, corresponds to the analytic result $(g_1^{\max})^2 = 12(x(6\sqrt{4-x^2} - \sqrt{3}x(x^2-4)) + 3\sqrt{3})/(x^4 - 4x^2 + 3)$ with $x = m_2/m$.

formidable technical challenges as discussed in detail in [69]. This method of extract QFT bounds is very onerous and requires several hours of computer time whereas the numerical method described in section 5 takes a few seconds. Beautifully, in the end, we reproduce precisely the same bounds as encountered in this paper as illustrated in figure 5.14.

We find the agreement between the conformal bootstrap and the S-matrix bootstrap to be conceptually very interesting. (At least in the case at hand corresponding to $d = 2$) we observe that the $d-1$ dimensional conformal bootstrap *knows* about the d dimensional massive S-matrix bootstrap. From an AdS/CFT-like intuition this is perhaps to be expected since we can always put whatever we want into boxes. On the other hand, we still find it comforting albeit counterintuitive that we can learn about massive quantum field theories from massless conformal theories in lower dimensions.

There are two natural follow up directions to this work. One is to explore further the

two dimensional world by including into the analysis S-matrix elements involving heavier particles. When these other components do not vanish, unitarity is not saturated and therefore we expect in this way to make contact with interesting non-integrable theories. One may learn, for example, about the full scaling ising model with magnetic and temperature deformations as discussed at the end of section 5.3. The second promising direction would be to stick with the simplest S-matrix element involving identical lightest particles but move to higher dimensions. In both cases we no longer expect the luxury of analytic results as obtained here. The hope, however, is that proper generalizations of the numerical methods – both the S-matrix and the conformal bootstrap one – will survive.

From the conformal bootstrap point of view, either direction seems straightforward although technically challenging. The technology for dealing with multiple correlators exists [82] even though it is well known that the numerics are much slower in this case. Going to higher dimensions also seems very doable. Indeed, almost all of the recent progress in the conformal bootstrap program has been focused on higher dimensions. A complication as compared to the one dimensional bootstrap is that we will also be forced to sum over spins in the OPE. Hopefully this is only a technical hurdle. For some preliminary results in this direction see [69]. In short, the conformal bootstrap provides a conceptually clear path along which we may proceed for both multiple correlators and higher dimensions. The price for this clarity is the difficult nature of the conformal bootstrap numerics and the feeling that one is dealing with a bit of a black box.

From the perspective of the numerical S-matrix bootstrap it seems simple to include amplitudes involving heavier particles. We no longer expect to find analytic bounds and we expect the conformal bootstrap numerics to provide a valuable way of testing these generalizations. Going to higher dimensions, however, seems much more difficult and it is less clear how to proceed. We now have to deal with two mandelstam variables and with angular momentum. Moreover, there are many representations of S-matrix elements which converge in different regions of physical parameters and which make different properties manifest. It is not yet clear which provides the best starting point. In particular, it is not clear how to efficiently impose unitarity and crossing at the same time. This is in contrast to 2d where crossing can be made manifest and unitarity can still be easily imposed.

In any case, it seems very fruitful to pursue the conformal and S-matrix bootstrap hand-in-hand. Both for the multiple correlator story as well as for higher dimensions, having a conformal bootstrap bound, even if it is numerically hard to get, would serve as an invaluable hint. Such lamp posts are extremely valuable and may provide key insights to the S-matrix bootstrap which were missing in the 60's.

Chapter 6

Concluding remarks

At the start of this thesis we asked the two basic QFT questions Q1 and Q2: How do we describe strongly interacting QFT and how do we describe QFT in terms of physical parameters? We attacked these questions from two complementary angles. In Part I we gain insight by studying the very special theory $\mathcal{N} = 4$ SYM while in part II we study the space of QFTs obeying a minimal set of fundamental properties. To conclude this thesis we return to these overarching questions and examine what progress we have made toward their resolution.

Let us first consider Part I. We are of course not the first to note that $\mathcal{N} = 4$ suggests a very natural answer to Q1 which is that strongly coupled gauge theories are naturally described in terms of strings. Although $\mathcal{N} = 4$ is a very special theory, this sort of duality is expected to be a general feature of gauge theories as was already recognized in the pioneering works of Wilson [92] and t' Hooft [1]. In particular, in the seminal paper from which the term "Wilson-loop" was born [92], Wilson observed that the strong-coupling limit of such loops are given as a sum over surfaces weighted by their area very much like a string partition function. A concrete realization of this idea was not borne out until many decades later with the advent of the gauge/gravity duality and in particular the paper [6] which gave the precise map (1.4) and (1.3). Wilson loops are arguable still the most natural variables for studying the transformation of a gauge theory into a string theory since they are hybrid objects that have a simple meaning on both sides of the duality as made explicit in the case of $\mathcal{N} = 4$ by (1.4). Moreover, $\mathcal{N} = 4$ is the perfect laboratory for calculating wilson loops. Its beta function is identically zero so the coupling does not run and much of the problems associated with Q2 are avoided. Furthermore, it is an integrable theory in the sense that the string dual has integrable world-sheet dynamics. This gives us hope that we may be able to use the tools of integrability to explicitly compute Wilson loops

at any finite coupling. Much progress has been made in this direction for null polygonal Wilson loops (which are dual to scattering amplitudes in N=4) where now an all-loop representation is available [9–17]. In this thesis we made the first step towards a finite coupling description of *smooth* Wilson loops, which was to formulate the minimal surface problem that arises at strong coupling in terms of integrability. Our hope is that an explicit solution for any coupling will give some insight into how the gauge theory transforms into a theory of strings as the coupling is increased. For example, we see very concretely in the case of $\mathcal{N} = 4$ that the dual string theory lives in a larger space than the gauge theory having and additional holographic direction as well as a compact manifold. It is suspected that the QCD string also requires such a holographic direction [2]. What general lessons can we learn from $\mathcal{N} = 4$? Perhaps we can say that $\mathcal{N} = 4$ is to gauge/string dualities as the $2D$ ising model is to renormalization group. In this sense it is a theory where we should learn/compute as much as possible.

Let us now turn to Part II. There we revisited the old idea of the S-matrix bootstrap. We returned to this idea armed with the recent revelations of the CFT bootstrap as well as modern ideas of holography as our inspiration. Our perspective in chapter 4 is to consider placing a massive quantum field theory in AdS. The massive bulk theory induces a conformal theory on the AdS boundary that can be studied using the conformal bootstrap. By studying the flat space limit we can try to constrain the space of possible theories in any dimension. We showed that this can be borne out explicitly in the case of a $1 + 1$ dimensional QFT in the bulk where the boundary bootstrap yielded numerical bounds on the space of couplings of any gapped quantum field theory that is unitary and lorentz invariant. Then in chapter 5 we derived analytical bounds on the couplings in $1+1$ directly from unitarity, crossing and analyticity of the S-matrix. These analytical bounds are in stunning agreement with the bounds obtained from the boundary bootstrap. Even in $1+1$ this is an extremely non-trivial result since there are many gapped, strongly interacting theories in two dimensions for which we have no analytical means of calculating mass spectra and couplings. Nevertheless we have shown that it is possible to place analytical bounds on these quantities. The incarnation of the S-matrix bootstrap that we have used provides a possible solution to both Q1 and Q2 – it is manifestly non-perturbative and the S-matrix is manifestly parameterized in terms of the physical data of the theory (e.g. the mass spectrum and couplings). There are two questions at this point. First and foremost, can the method be applied in higher dimensions? The boundary bootstrap employed in chapter 4 is applicable in any dimension although if the boundary dimension is greater than one there are spins in the OPE and this greatly complicates the numerics. This does not a priori guarantee bounds, since the bounds obtained from the conformal bootstrap can be lost in the flat-space limit. As for the direct S-matrix bootstrap in flat space, it seems

much more difficult to proceed in higher dimensions. The two-to-two S-matrix is now a function of two complex variables s and t. Furthermore one needs to find a representation of the S-matrix where crossing symmetry and unitarity can be efficiently implemented. A second important question is how to strengthen the bounds by bootstrapping S-matrix elements involving other particles of the theory. Even in two dimensions this is still an open and interesting question. The hope is that as the bounds are strengthened surfaces such as those shown in figure 5.12 will start to develop features which correspond to physical theories. The basis of this hope is that exactly such a phenomena occurs in the conformal bootstrap where the bounds often exhibit knee-like features at the location of important physical theories, the 3D ising model being the most famous example. By measuring the location of these features one is able to measure the physical data of a *specific* theory by studying general bounds of the space of theories. At an intuitive level, it is natural that important physical theories sit at distinguished points in such a space. In our incarnation of the S-matrix bootstrap, such features would allow us to directly measure the mass spectrum and couplings from the location of the feature. Will we some day be able to use our S-matrix bootstrap to measure the spectrum of important physical theories such as pure glue and QCD?

Finally, let us conclude this thesis by commenting on interrelation between Parts I and II. In Part II we have seen that integrable theories play an important role in the $1+1$ dimensional S-matrix bootstrap. They appear as theories saturating the coupling bounds and provide a means by which we may understand the mechanism for the existence of a bound as well as how/when the bound can be improved. Of course our ultimate goal is to perform the S-matrix bootstrap in higher dimensions, especially $3+1$. The closest thing we have to integrable theories in higher dimensions is theories with string duals whose world-sheet dynamics posses the usual two dimensional integrability. In fact, one may take this as a working definition of "higher dimensional integrability" which manages to evade the Coleman-Mandula theorem. Will such theories appear in the bootstrap in an analogous way as the integrable theories of two dimensions appeared in the two dimensional bootstrap? Of course $\mathcal{N}=4$ is a massless theory so it can not play an analogous role to the massive integrable theories of the two dimensional bootstrap. However, it provides the canonical example of higher dimensional integrability and thus the best laboratory for understanding this phenomenon. For this reason we believe the study of such special theories is an important ingredient in the success of a higher dimensional S-matrix bootstrap program and that these two avenues should be pursued in parallel.

References

[1] Gerard 't Hooft. A Planar Diagram Theory for Strong Interactions. *Nucl. Phys.*, B72:461, 1974.

[2] Alexander M. Polyakov. String theory and quark confinement. *Nucl. Phys. Proc. Suppl.*, 68:1–8, 1998.

[3] Juan Martin Maldacena. The Large N limit of superconformal field theories and supergravity. *Int. J. Theor. Phys.*, 38:1113–1133, 1999. [Adv. Theor. Math. Phys.2,231(1998)].

[4] S. S. Gubser, Igor R. Klebanov, and Alexander M. Polyakov. Gauge theory correlators from noncritical string theory. *Phys. Lett.*, B428:105–114, 1998.

[5] Edward Witten. Anti-de Sitter space and holography. *Adv. Theor. Math. Phys.*, 2:253–291, 1998.

[6] Juan Martin Maldacena. Wilson loops in large N field theories. *Phys. Rev. Lett.*, 80:4859–4862, 1998.

[7] Iosif Bena, Joseph Polchinski, and Radu Roiban. Hidden symmetries of the AdS(5) x S**5 superstring. *Phys. Rev.*, D69:046002, 2004.

[8] V. A. Kazakov, A. Marshakov, J. A. Minahan, and K. Zarembo. Classical/quantum integrability in AdS/CFT. *JHEP*, 05:024, 2004.

[9] Luis F. Alday, Davide Gaiotto, Juan Maldacena, Amit Sever, and Pedro Vieira. An Operator Product Expansion for Polygonal null Wilson Loops. *JHEP*, 04:088, 2011.

[10] Benjamin Basso, Amit Sever, and Pedro Vieira. Spacetime and Flux Tube S-Matrices at Finite Coupling for N=4 Supersymmetric Yang-Mills Theory. *Phys. Rev. Lett.*, 111(9):091602, 2013.

[11] Benjamin Basso, Amit Sever, and Pedro Vieira. Space-time S-matrix and Flux tube S-matrix II. Extracting and Matching Data. *JHEP*, 01:008, 2014.

[12] Benjamin Basso, Amit Sever, and Pedro Vieira. Space-time S-matrix and Flux-tube S-matrix III. The two-particle contributions. *JHEP*, 08:085, 2014.

[13] Benjamin Basso, Amit Sever, and Pedro Vieira. Collinear Limit of Scattering Amplitudes at Strong Coupling. *Phys. Rev. Lett.*, 113(26):261604, 2014.

[14] Benjamin Basso, Amit Sever, and Pedro Vieira. Space-time S-matrix and Flux-tube S-matrix IV. Gluons and Fusion. *JHEP*, 09:149, 2014.

[15] Benjamin Basso, Joao Caetano, Lucia Cordova, Amit Sever, and Pedro Vieira. OPE for all Helicity Amplitudes. *JHEP*, 08:018, 2015.

[16] Benjamin Basso, Joao Caetano, Lucia Cordova, Amit Sever, and Pedro Vieira. OPE for all Helicity Amplitudes II. Form Factors and Data analysis. *JHEP*, 12:088, 2015.

[17] Benjamin Basso, Amit Sever, and Pedro Vieira. Hexagonal Wilson Loops in Planar $\mathcal{N} = 4$ SYM Theory at Finite Coupling. 2015.

[18] Nikolay Gromov, Vladimir Kazakov, and Pedro Vieira. Exact Spectrum of Anomalous Dimensions of Planar N=4 Supersymmetric Yang-Mills Theory. *Phys. Rev. Lett.*, 103:131601, 2009.

[19] Luis F. Alday, Juan Maldacena, Amit Sever, and Pedro Vieira. Y-system for Scattering Amplitudes. *J. Phys.*, A43:485401, 2010.

[20] A. A. Belavin, Alexander M. Polyakov, and A. B. Zamolodchikov. Infinite Conformal Symmetry in Two-Dimensional Quantum Field Theory. *Nucl. Phys.*, B241:333–380, 1984.

[21] Riccardo Rattazzi, Vyacheslav S. Rychkov, Erik Tonni, and Alessandro Vichi. Bounding scalar operator dimensions in 4D CFT. *JHEP*, 12:031, 2008.

[22] R.J. Eden, R.J. Eden, P.V. Landshoff, D.I. Olive, and J.C. Polkinghorne. *The Analytic S-Matrix*. Cambridge University Press, 2002.

[23] R.J. Eden. *High energy collisions of elementary particles*. Cambridge U.P., 1967.

[24] Marcel Froissart. Asymptotic behavior and subtractions in the Mandelstam representation. *Phys. Rev.*, 123:1053–1057, 1961.

[25] Alexander B. Zamolodchikov and Alexei B. Zamolodchikov. Relativistic Factorized S Matrix in Two-Dimensions Having O(N) Isotopic Symmetry. *Nucl. Phys.*, B133:525, 1978. [JETP Lett.26,457(1977)].

[26] Curtis G. Callan, Jr. and Frank Wilczek. INFRARED BEHAVIOR AT NEGATIVE CURVATURE. *Nucl. Phys.*, B340:366–386, 1990.

[27] J Lagrange. Essai d'une nouvelle méthode pour determiner les maxima et les minima des formules integrales indefinies. *Miscellanea Taurinensia*, pages 173–195, 1760.

[28] J Pérez W Meeks. The classical theory of minimal surfaces. *Bull. Amer. Math. Soc. 48 (3)*, pages 325–407, 2011.

[29] David J. Gross and Hirosi Ooguri. Aspects of large N gauge theory dynamics as seen by string theory. *Phys. Rev.*, D58:106002, 1998.

[30] Nadav Drukker, Simone Giombi, Riccardo Ricci, and Diego Trancanelli. Supersymmetric Wilson loops on S**3. *JHEP*, 05:017, 2008.

[31] Luis F. Alday and Juan Martin Maldacena. Gluon scattering amplitudes at strong coupling. *JHEP*, 06:064, 2007.

[32] Luis F. Alday and Juan Maldacena. Comments on gluon scattering amplitudes via AdS/CFT. *JHEP*, 11:068, 2007.

[33] Riei Ishizeki, Martin Kruczenski, and Sannah Ziama. Minimal area surfaces, Euclidean Wilson loops and Riemann Theta functions. *J. Phys. Conf. Ser.*, 411:012017, 2013.

[34] Martin Kruczenski and Sannah Ziama. Wilson loops and Riemann theta functions II. *JHEP*, 05:037, 2014.

[35] Martin Kruczenski. Wilson loops and minimal area surfaces in hyperbolic space. *JHEP*, 11:065, 2014.

[36] Andrew Irrgang and Martin Kruczenski. Euclidean Wilson loops and minimal area surfaces in lorentzian AdS_3. *JHEP*, 12:083, 2015.

[37] Changyu Huang, Yifei He, and Martin Kruczenski. Minimal area surfaces dual to Wilson loops and the Mathieu equation. 2016.

[38] Romuald A. Janik and Andrzej Wereszczynski. Correlation functions of three heavy operators: The AdS contribution. *JHEP*, 12:095, 2011.

[39] Yoichi Kazama and Shota Komatsu. On holographic three point functions for GKP strings from integrability. *JHEP*, 01:110, 2012. [Erratum: JHEP06,150(2012)].

[40] Yoichi Kazama and Shota Komatsu. Wave functions and correlation functions for GKP strings from integrability. *JHEP*, 09:022, 2012.

[41] J. Caetano and J. Toledo. chi-Systems for Correlation Functions. 2012.

[42] Daigo Honda and Shota Komatsu. Classical Liouville Three-point Functions from Riemann-Hilbert Analysis. *JHEP*, 03:038, 2014.

[43] Yoichi Kazama and Shota Komatsu. Three-point functions in the SU(2) sector at strong coupling. *JHEP*, 03:052, 2014.

[44] Y. Kazama, S. Komatsu, and T. Nishimura. Classical Integrability for Three-point Functions: Cognate Structure at Weak and Strong Couplings. 2016.

[45] Luis F. Alday and Juan Maldacena. Minimal surfaces in AdS and the eight-gluon scattering amplitude at strong coupling. 2009.

[46] Luis F. Alday and Juan Maldacena. Null polygonal Wilson loops and minimal surfaces in Anti-de-Sitter space. *JHEP*, 11:082, 2009.

[47] Luis F. Alday, Davide Gaiotto, and Juan Maldacena. Thermodynamic Bubble Ansatz. *JHEP*, 09:032, 2011.

[48] Jonathan C Toledo. Smooth Wilson loops from the continuum limit of null polygons. 2014.

[49] K. Pohlmeyer. Integrable Hamiltonian Systems and Interactions Through Quadratic Constraints. *Commun. Math. Phys.*, 46:207–221, 1976.

[50] H. J. De Vega and Norma G. Sanchez. Exact integrability of strings in D-Dimensional De Sitter space-time. *Phys. Rev.*, D47:3394–3405, 1993.

[51] Antal Jevicki, Kewang Jin, Chrysostomos Kalousios, and Anastasia Volovich. Generating AdS String Solutions. *JHEP*, 03:032, 2008.

[52] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Wall-crossing, Hitchin Systems, and the WKB Approximation. 2009.

[53] Davide Gaiotto, Gregory W. Moore, and Andrew Neitzke. Four-dimensional wall-crossing via three-dimensional field theory. *Commun. Math. Phys.*, 299:163–224, 2010.

[54] Gordon W. Semenoff and Donovan Young. Wavy Wilson line and AdS / CFT. *Int. J. Mod. Phys.*, A20:2833–2846, 2005.

[55] Sergei L. Lukyanov. Critical values of the Yang-Yang functional in the quantum sine-Gordon model. *Nucl. Phys.*, B853:475–507, 2011.

[56] Romuald A. Janik, Piotr Surowka, and Andrzej Wereszczynski. On correlation functions of operators dual to classical spinning string states. *JHEP*, 05:030, 2010.

[57] E. I. Buchbinder and A. A. Tseytlin. On semiclassical approximation for correlators of closed string vertex operators in AdS/CFT. *JHEP*, 08:057, 2010.

[58] Alexander M. Polyakov. Gauge fields and space-time. *Int. J. Mod. Phys.*, A17S1:119–136, 2002.

[59] Arkady A. Tseytlin. On semiclassical approximation and spinning string vertex operators in AdS(5) x S**5. *Nucl. Phys.*, B664:247–275, 2003.

[60] S. S. Gubser, I. R. Klebanov, and Alexander M. Polyakov. A Semiclassical limit of the gauge / string correspondence. *Nucl. Phys.*, B636:99–114, 2002.

[61] Eric D'Hoker, Daniel Z. Freedman, Samir D. Mathur, Alec Matusis, and Leonardo Rastelli. Extremal correlators in the AdS / CFT correspondence. 1999.

[62] B. Eden, Paul S. Howe, C. Schubert, E. Sokatchev, and Peter C. West. Extremal correlators in four-dimensional SCFT. *Phys. Lett.*, B472:323–331, 2000.

[63] Luis F. Alday, Burkhard Eden, Gregory P. Korchemsky, Juan Maldacena, and Emery Sokatchev. From correlation functions to Wilson loops. *JHEP*, 09:123, 2011.

[64] Burkhard Eden, Gregory P. Korchemsky, and Emery Sokatchev. From correlation functions to scattering amplitudes. *JHEP*, 12:002, 2011.

[65] Burkhard Eden, Gregory P. Korchemsky, and Emery Sokatchev. More on the duality correlators/amplitudes. *Phys. Lett.*, B709:247–253, 2012.

[66] A. V. Belitsky, G. P. Korchemsky, and E. Sokatchev. Are scattering amplitudes dual to super Wilson loops? *Nucl. Phys.*, B855:333–360, 2012.

[67] Burkhard Eden, Paul Heslop, Gregory P. Korchemsky, and Emery Sokatchev. The super-correlator/super-amplitude duality: Part I. *Nucl. Phys.*, B869:329–377, 2013.

[68] Burkhard Eden, Paul Heslop, Gregory P. Korchemsky, and Emery Sokatchev. The super-correlator/super-amplitude duality: Part II. *Nucl. Phys.*, B869:378–416, 2013.

[69] Miguel Paulos, Joao Penedones, Jonathan Toledo, Balt van Rees, and Pedro Vieira. Conformal and S-matrix Bootstrap I: QFT in hyperbolic space. *In preparation.*

[70] P. Di Francesco, P. Mathieu, and D. Senechal. *Conformal Field Theory.* Graduate Texts in Contemporary Physics. Springer-Verlag, New York, 1997.

[71] Duccio Pappadopulo, Slava Rychkov, Johnny Espin, and Riccardo Rattazzi. OPE Convergence in Conformal Field Theory. *Phys. Rev.*, D86:105043, 2012.

[72] Miguel F. Paulos. JuliBootS: a hands-on guide to the conformal bootstrap. 2014.

[73] David Simmons-Duffin. A Semidefinite Program Solver for the Conformal Bootstrap. *JHEP*, 06:174, 2015.

[74] Vyacheslav S. Rychkov and Alessandro Vichi. Universal Constraints on Conformal Operator Dimensions. *Phys. Rev.*, D80:045006, 2009.

[75] Francesco Caracciolo and Vyacheslav S. Rychkov. Rigorous Limits on the Interaction Strength in Quantum Field Theory. *Phys. Rev.*, D81:085037, 2010.

[76] David Poland and David Simmons-Duffin. Bounds on 4D Conformal and Superconformal Field Theories. *JHEP*, 05:017, 2011.

[77] Riccardo Rattazzi, Slava Rychkov, and Alessandro Vichi. Central Charge Bounds in 4D Conformal Field Theory. *Phys. Rev.*, D83:046011, 2011.

[78] Riccardo Rattazzi, Slava Rychkov, and Alessandro Vichi. Bounds in 4D Conformal Field Theories with Global Symmetry. *J. Phys.*, A44:035402, 2011.

[79] Alessandro Vichi. Improved bounds for CFT's with global symmetries. *JHEP*, 01:162, 2012.

[80] David Poland, David Simmons-Duffin, and Alessandro Vichi. Carving Out the Space of 4D CFTs. *JHEP*, 05:110, 2012.

[81] Sheer El-Showk, Miguel F. Paulos, David Poland, Slava Rychkov, David Simmons-Duffin, and Alessandro Vichi. Solving the 3D Ising Model with the Conformal Bootstrap. *Phys. Rev.*, D86:025022, 2012.

[82] Sheer El-Showk, Miguel F. Paulos, David Poland, Slava Rychkov, David Simmons-Duffin, and Alessandro Vichi. Solving the 3d Ising Model with the Conformal Bootstrap II. c-Minimization and Precise Critical Exponents. *J. Stat. Phys.*, 157:869, 2014.

[83] Idse Heemskerk, Joao Penedones, Joseph Polchinski, and James Sully. Holography from Conformal Field Theory. *JHEP*, 10:079, 2009.

[84] Hyungrok Kim, Petr Kravchuk, and Hirosi Ooguri. Reflections on Conformal Spectra. *JHEP*, 04:184, 2016.

[85] Sidney R. Coleman and H. J. Thun. On the Prosaic Origin of the Double Poles in the Sine-Gordon S Matrix. *Commun. Math. Phys.*, 61:31, 1978.

[86] Miguel Paulos, Joao Penedones, Jonathan Toledo, Balt van Rees, and Pedro Vieira. Conformal and S-matrix Bootstrap II: The 2D Amplitude Bootstrap. *In preparation.*

[87] A. B. Zamolodchikov. Exact s Matrix of Quantum Sine-Gordon Solitons. *JETP Lett.*, 25:468, 1977.

[88] Sergei Dubovsky, Victor Gorbenko, and Mehrdad Mirbabayi. Natural Tuning: Towards A Proof of Concept. *JHEP*, 09:045, 2013.

[89] D. Iagolnitzer. Factorization of the Multiparticle s Matrix in Two-Dimensional Space-Time Models. *Phys. Rev.*, D18:1275, 1978.

[90] A. B. Zamolodchikov. Integrals of Motion and S Matrix of the (Scaled) T=T(c) Ising Model with Magnetic Field. *Int. J. Mod. Phys.*, A4:4235, 1989.

[91] Gesualdo Delfino. Integrable field theory and critical phenomena: The Ising model in a magnetic field. *J. Phys.*, A37:R45, 2004.

[92] Kenneth G. Wilson. Confinement of Quarks. *Phys. Rev.*, D10:2445–2459, 1974. [,45(1974)].

[93] Alexander Zamolodchikov. Ising Spectroscopy II: Particles and poles at $T > T_c$. 2013.

[94] Pedro Fonseca and Alexander Zamolodchikov. Ising spectroscopy. I. Mesons at T ¡ T(c). 2006.

[95] P. Fonseca and A. Zamolodchikov. Ising field theory in a magnetic field: Analytic properties of the free energy. 2001.

[96] Gesualdo Delfino, Paolo Grinza, and Giuseppe Mussardo. Decay of particles above threshold in the Ising field theory with magnetic field. *Nucl. Phys.*, B737:291–303, 2006.

[97] Gleb Arutyunov, Sergey Frolov, and Matthias Staudacher. Bethe ansatz for quantum strings. *JHEP*, 10:016, 2004.

[98] Benjamin Basso, Shota Komatsu, and Pedro Vieira. Structure Constants and Integrable Bootstrap in Planar N=4 SYM Theory. 2015.

[99] Benjamin Basso, Vasco Goncalves, Shota Komatsu, and Pedro Vieira. Gluing Hexagons at Three Loops. 2015.

APPENDICES

Appendix A

Regulation of continuum area

Consider replacing $p_{\text{in}} \to \delta^2 p_{\text{in}}$ so that $|z_{\boldsymbol{\sigma}}| \to \delta |z_{\boldsymbol{\sigma}}|$ in (2.37) and (2.52). Then we have

$$4 \int_D dz d\bar{z}\, e^{2\alpha^{\text{circ}}} = \lim_{\epsilon \to 0} \lim_{\delta \to 0} \frac{1}{2\pi} \int_0^{2\pi} d\sigma_2 \int_{0+\epsilon}^{\sigma_2 - \epsilon} d\sigma_1 \int_{-\infty}^{+\infty} d\theta\; \delta |z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, y^{\delta}_{\boldsymbol{\sigma}}(\theta) \tag{A.1}$$

where $y^{\delta}_{\boldsymbol{\sigma}}(\theta)$ is defined as the solution of the modified CTBA equation

$$\log \frac{y^{\delta}_{\boldsymbol{\sigma}}(\theta)}{y^{\text{circ}}_{\boldsymbol{\sigma}}} = -4\delta |z_{\boldsymbol{\sigma}}| \cosh\theta - \int_{\sigma_2}^{\sigma_1 + 2\pi} d\tau_2 \int_{\sigma_1}^{\sigma_2} d\tau_1 \int_{-\infty}^{+\infty} d\theta' \, \frac{y^{\delta}_{\boldsymbol{\tau}}(\theta') - y^{\text{circ}}_{\boldsymbol{\tau}}}{i\pi \sinh(\theta - \theta' + i\varphi_{\boldsymbol{\sigma}\boldsymbol{\tau}})} \tag{A.2}$$

As we have already mentioned in section 2.2.2 when $\delta = 0$ this equation admits θ-independent solutions of the form (2.36). Here is is important to consider how the solutions approach these values. As $\delta \to 0$ each $y^{\delta}_{\boldsymbol{\sigma}}$ form broad plateaus in θ of width $\sim 2\log(\delta |z_{\boldsymbol{\sigma}}|)$; at the edge of this plateau there is a "kink" where the solution decays rapidly to zero. In the limit $\delta \to 0$ the plateaus become infinite in width and their height is given by constants of the form (2.36).[1] As we will see shortly, the important contribution to the area comes from these kinks in the y-function. The shape of the kinks become independent of δ in the limit – they simply move further and further apart. We can write an equation for the shape of the left-kink by changing variables $\theta \to \theta + \log\delta$ and then taking the limit $\delta \to 0$ which gives

$$\log \frac{y^{\text{kink}}_{\boldsymbol{\sigma}}(\theta)}{y^{\text{circ}}_{\boldsymbol{\sigma}}} = -2|z_{\boldsymbol{\sigma}}| e^{-\theta} - \int_{\sigma_2}^{\sigma_1 + 2\pi} d\tau_2 \int_{\sigma_1}^{\sigma_2} d\tau_1 \int_{-\infty}^{+\infty} d\theta' \, \frac{y^{\text{kink}}_{\boldsymbol{\tau}}(\theta') - y^{\text{circ}}_{\boldsymbol{\tau}}}{i\pi \sinh(\theta - \theta' + i\varphi_{\boldsymbol{\sigma}\boldsymbol{\tau}})} \tag{A.3}$$

[1] As we have already mentioned, any function of the form (2.36) will solve (A.2) with $\delta = 0$. Precisely what values one obtains for these constants depends on how we take p_{in} to zero. Below we will find the explicit values of the constants when the limit is taken by sending $p_{\text{in}} \to \delta^2 p_{\text{in}}$ and then $\delta \to 0$ in (A.2).

where we have defined

$$y^{\text{kink}}_{\boldsymbol{\sigma}}(\theta) = \lim_{\delta\to 0} y^{\delta}_{\boldsymbol{\sigma}}(\theta + \log\delta) \tag{A.4}$$

With a bit of guess-work we find, quite remarkably, that equation (A.3) can be solved for arbitrary p_{in} with the solution given by[2]

$$y^{\text{kink}}_{\boldsymbol{\sigma}}(\theta) = z^2_{\boldsymbol{\sigma}}(\partial_{\sigma_1}\partial_{\sigma_2}\log z_{\boldsymbol{\sigma}})\, e^{-2(\theta+i\varphi_{\boldsymbol{\sigma}})}\text{csch}^{-2}(e^{-\theta}|z_{\boldsymbol{\sigma}}|) \tag{A.5}$$

With this solution in hand, we can now take $\delta \to 0$ in (A). We have

$$\lim_{\delta\to 0}\int_{-\infty}^{+\infty} d\theta\ \delta|z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, y^{\delta}_{\boldsymbol{\sigma}}(\theta) = \int_{-\infty}^{+\infty} d\theta\ |z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, y^{\text{kink}}_{\boldsymbol{\sigma}}(\theta) \tag{A.6}$$

where we made the change of variables $\theta \to \theta + \log\delta$ and then take the limit $\delta \to 0$.[3] Plugging this result into gives

$$4\int_D dz d\bar{z}\, e^{2\alpha^{\text{circ}}} = \lim_{\epsilon\to 0}\frac{1}{2\pi}\int_0^{2\pi} d\sigma_2 \int_{0+\epsilon}^{\sigma_2-\epsilon} d\sigma_1 \int_{-\infty}^{+\infty} d\theta\ |z_{\boldsymbol{\sigma}}|\, e^{-\theta}\, y^{\text{kink}}_{\boldsymbol{\sigma}}(\theta) \tag{A.7}$$

which is the result used in the derivation of the continuum area formula (2.54).

[2]This can be most easily checked by substitution into equation (A.3) and evaluating the integrals numerically for some p_{in} and various values of θ and $\boldsymbol{\sigma}$. In later versions of `Mathematica` it is useful to use `NIntegrate` with the option `Method -> {"GlobalAdaptive", Method -> "GaussKronrodRule", "SingularityDepth" -> Infinity}`.

[3]In (A.6) there is a competition between the vanishing of the explicit factor of $\delta|z_{\boldsymbol{\sigma}}|$ and the divergence of the θ-integrals as the plateaus become infinite. The region of integration that contributes in this limit comes from the edge of the plateau. One can see this, for example, by adding a derivative to the exponential $e^{\theta} = \partial_\theta e^{\theta}$ and then moving the derivative to y via integration by parts. The result of this competition is the finite integral on the right hand side of (A.6).

Appendix B

Four-point function appendices

B.1 The linear problem

B.2 Summary of the linear problem

The linear problem associated with (3.10) is given by

$$(\partial + J_w)\,\psi = 0, \qquad \left(\bar{\partial} + J_{\bar{w}}\right)\psi = 0 \tag{B.1}$$

where the connection has the form

$$J_w = \frac{1}{\xi}\Phi_w + A_w, \qquad J_{\bar{w}} = \xi\Phi_{\bar{w}} + A_{\bar{w}} \tag{B.2}$$

$$\Phi_w = \begin{pmatrix} 0 & -\frac{1}{2}e^{\widetilde{\gamma}} \\ -\frac{1}{2}Te^{-\widetilde{\gamma}} & 0 \end{pmatrix} \tag{B.3}$$

$$\Phi_{\bar{w}} = \begin{pmatrix} 0 & -\frac{1}{2}\bar{T}e^{-\widetilde{\gamma}} \\ -\frac{1}{2}e^{\widetilde{\gamma}} & 0 \end{pmatrix} \tag{B.4}$$

$$A_w = \partial_w \begin{pmatrix} \frac{1}{2}\widetilde{\gamma} & 0 \\ 0 & -\frac{1}{2}\widetilde{\gamma} \end{pmatrix} \tag{B.5}$$

$$A_{\bar{w}} = \partial_{\bar{w}} \begin{pmatrix} -\frac{1}{2}\widetilde{\gamma} & 0 \\ 0 & \frac{1}{2}\widetilde{\gamma} \end{pmatrix} \tag{B.6}$$

For compactness we have introduced the combination $\widetilde{\gamma} = 1/2(\gamma + \log\sqrt{T\bar{T}})$. The function γ is defined as the solution of the following problem

$$\begin{aligned} \partial\bar{\partial}\gamma &= \sqrt{T\bar{T}}\sinh\gamma \\ \gamma &\to \pm\frac{1}{2}\log T\bar{T} \qquad (w \to z_a) \\ \gamma &\to 0 \qquad\qquad\quad (w \to w_a) \end{aligned} \tag{B.7}$$

where z_a and w_a are the zeros and the poles of T, respectively.

For the near-puncture analysis as well as the WKB analysis it is useful to make the field redefinition $\psi \to \hat{\psi} = \hat{G}\psi$ where

$$\hat{G} = \frac{1}{2}\begin{pmatrix} +e^{-\gamma/2}T^{1/4}\bar{T}^{-1/4} & 1 \\ -e^{-\gamma/2}T^{1/4}\bar{T}^{-1/4} & 1 \end{pmatrix} \tag{B.8}$$

This is usually refered to as 'diagonal gauge' in the literature. In diagonal gauge we have

$$\hat{\Phi}_w = \frac{1}{2}\sqrt{T}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \tag{B.9}$$

$$\hat{\Phi}_{\bar{w}} = \frac{1}{2}\sqrt{\bar{T}}\begin{pmatrix} -\cosh\gamma & \sinh\gamma \\ -\sinh\gamma & \cosh\gamma \end{pmatrix} \tag{B.10}$$

$$\hat{A}_w = \partial_w\begin{pmatrix} \frac{1}{4}\gamma - \frac{1}{8}\log(T\bar{T}) & -\frac{1}{2}\gamma \\ -\frac{1}{2}\gamma & \frac{1}{4}\gamma - \frac{1}{8}\log(T\bar{T}) \end{pmatrix} \tag{B.11}$$

$$\hat{A}_{\bar{w}} = \partial_{\bar{w}}\begin{pmatrix} \frac{1}{4}\gamma + \frac{1}{8}\log(T\bar{T}) & 0 \\ 0 & \frac{1}{4}\gamma + \frac{1}{8}\log(T\bar{T}) \end{pmatrix} \tag{B.12}$$

We are now ready to consider the behavior of the solutions near the points w_a and z_a.

B.2.1 Solutions near w_a

Let us first consider the solutions of the linear problem in the neighborhood of one of the punctures. From (B.7) and the explicit expressions for $\hat{\Phi}$ and $\hat{A}$ for $w \to w_a$ we have

$$\hat{\Phi}_w \to \frac{1}{2}\sqrt{T}\begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix}, \qquad \hat{\Phi}_{\bar{w}} \to \frac{1}{2}\sqrt{\bar{T}}\begin{pmatrix} -1 & 0 \\ 0 & +1 \end{pmatrix} \tag{B.13}$$

$$\hat{A}_w \to \partial_w\left(-\frac{1}{8}\log T\bar{T}\right)\begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix}, \qquad \hat{A}_{\bar{w}} \to \partial_{\bar{w}}\left(\frac{1}{8}\log T\bar{T}\right)\begin{pmatrix} +1 & 0 \\ 0 & +1 \end{pmatrix} \tag{B.14}$$

Then the solution in the vicinity of puncture P_a is given by:

$$\hat{\psi}^{\pm}(w) \equiv \left(T/\bar{T}\right)^{1/8} e^{\pm\frac{1}{2}\int^w \xi^{-1}\omega + \xi\bar{\omega}}|\pm\rangle \sim (w - w_a)^{\pm\frac{1}{4}\Delta_a\xi^{-1} - \frac{1}{4}}\,(\bar{w} - \bar{w}_a)^{\pm\frac{1}{4}\bar{\Delta}_a\xi + \frac{1}{4}}\,|\pm\rangle \tag{B.15}$$

where $|\pm\rangle$ are the eigenvectors of the Pauli matrix σ^3. Note the characteristic monodromy of the solutions about w_a.

B.2.2 Solutions near z_a

Now we will consider the behavior of the solutions near the zeros z_a of T. Notice from (B.3) - (B.6) and (B.7) that the connection is regular or singular at z_a depending on the direction of the spike in γ at z_a. More specifically, the connection is regular if $\gamma \sim -\log|w-z_a|$ and thus the solution will be regular in the vicinity of a d-spike. However, the connection has a singularity if $\gamma \sim +\log|w-z_a|$ and at the u-spikes one can check that in gauge (B.3) - (B.6) there are two linearly independent solutions behaving as

$$\Psi \sim \Psi_{z_a} \equiv \begin{pmatrix} (w-z_a)^{-1/4}(\bar{w}-\bar{z}_a)^{+1/4} & 0 \\ 0 & (w-z_a)^{+1/4}(\bar{w}-\bar{z}_a)^{-1/4} \end{pmatrix} \tag{B.16}$$

where we have written the two solutions in matrix form as in (3.71). Notice that Ψ has square-root type singularity at z_a since it has a monodromy of $\Psi \to (-1)\,\Psi$ about z_a. The solutions associated with the punctures $\{s_P\}$ and $\{\tilde{s}_P\}$ inherit this square-root singularity as one can see by expanding them in the basis (B.16) near z_a.

In our analysis it is crucial to account for the additional monodromies originating from u-spikes. Let us explain our conventions for doing this. If there is a u-spike at z_a, one can always make the gauge-transformation $\Psi \to \Psi_{z_a}^{-1}\Psi$ that removes the square root singularity (Ψ_{z_a} is given in (B.16)). Of course this gauge transformation contains the same multivaluedness and one must still account for it at the end of the day. In the main text we use the point of view that this gauge transformation has been performed for each u-spike. The connection in this gauge will only have singularities at the punctures and the solutions in this gauge will only have non-trivial monodromies around the punctures. In this way we can define small solutions that are single valued throughout some Q_E, as is the prescription of [52]. We must then be sure to account for the multivaluedness of these gauge transformations whenever we have a holonomy that encloses an odd number of u-spikes. Such holonomies arise in the WKB expansion of the coordinates and we will return to this issue below.

B.3 WKB analysis

B.3.1 Statement of the WKB approximation

As we have discussed above, it is essential to have control over the $\xi \to 0, \infty$ asymptotics of the inner products. It is clear from ($B.1 - B.2$) that these are both singular limits, and the basic idea of extracting this singularity is as follows. As discussed above, we have good

control over the solutions in the neighborhood of the punctures. Thus what we would like to study is the transport

$$\mathrm{Pexp}\left[-\int_{\mathcal{C}(w_a'\to w)}\frac{1}{\xi}\Phi + A + \xi\bar{\Phi}\right]\psi\left(w_a'\right) \tag{B.17}$$

where $\mathcal{C}\left(w_a'\to w\right)$ is a curve starting at w_a', a point in the neighborhood of w_a, and terminating at a generic point w. Note that at any point on the punctured sphere C the Higgs field Φ has the two eigenvalues $\mp\omega/2 = \mp\sqrt{T}/2\,dw$ (which are single valued on the double cover $\widetilde{\Sigma}$), and thus we can choose a gauge along $\mathcal{C}$ where Φ is diagonal and given by

$$\Phi = \frac{1}{2}\begin{pmatrix}-\omega & 0\\ 0 & \omega\end{pmatrix} \tag{B.18}$$

Now consider the $\xi\to 0$ limit. First consider an infinitesimal segment of $\mathcal{C}$ in the neighborhood of P_a. In the neighborhood of P_a the connection (in diagonal gauge) becomes diagonal (see (B.13)-(B.14)) and thus one can break apart the path-ordered exponential. In particular, one can isolate the singular part $e^{-\int\Phi/\xi}|\pm\rangle$ which will have one component growing exponentially and one component decaying. Let us choose the branch of Φ such that the $|+\rangle$ component is the one that is growing as we transport along $\mathcal{C}$ *away* from P_a (although for the moment we are still working in a neighborhood of P_a). This will correspond to the small solution at P_a since it is exponentially decaying as it is transported *toward* from P_a. The *WKB approximation* is the statement that the exponentially growing part of the solution as $\xi\to 0$ will continue to be given by $e^{-\int\Phi/\xi}|+\rangle$ as we transport away from the neighborhood of P_a (now leaving the neighborhood of P_a) as long as we follow a curve such that at every point we have

$$\mathrm{Re}\left(\omega/\xi\right) > 0 \tag{B.19}$$

This condition is satisfied most strongly along a curve such that

$$\mathrm{Im}\left(\omega/\xi\right) = 0 \tag{B.20}$$

Condition (B.19) is called the *WKB condition* and curves satisfying (B.20) are called *WKB curves* [52]. Along a WKB curve defined for $\mathrm{Arg}\left(\xi\right)=\theta$ the WKB condition is satisfied for $\mathrm{Arg}\left(\xi\right)\in(\theta-\pi/2,\theta+\pi/2)$ and the WKB approximation is guaranteed to hold in this range. For example, suppose there is a WKB line connecting P_a to P_b for $\theta\in(\theta_-,\theta_+)$ but not outside that range. Then the WKB approximation will reliably give the $\xi\to 0,\infty$ asymptotic for $\theta\in(\theta_--\pi/2,\theta_++\pi/2)$. These statements are proven in [52] and we refer the reader there for a more detailed discussion.

B.3.2 Subleading WKB

We will now consider the $\xi \to 0$ limit of the inner products (or Wronskians) $(s_b \wedge s_a)(\xi)$. We consider the case when P_a and P_b are connected by a WKB line which will be an edge E_{ab} in the WKB triangulation. From the analysis of B.2.1 we know s_a and s_b in the neighborhood of P_a and P_b respectively. In order to evaluate the Wronskian we need to know the solutions at a common point. The approach here is to use the connection to transport the solution s_a along E_{ab} to a point w_b' in the neighborhood of P_b and then to evaluate the Wronskian at w_b'. That is, we want to study the $\xi \to 0$ behavior of

$$\langle s_b | \text{Pexp}\left[-\int_0^1 dt\ \frac{1}{\xi} H_0 + V\right] | s_a \rangle \tag{B.21}$$

where we defined

$$H_0 = \dot{w}\hat{\Phi}_w, \qquad V = \dot{w}\hat{A}_w + \dot{\bar{w}}\hat{A}_{\bar{w}} + \xi\dot{\bar{w}}\hat{\Phi}_{\bar{w}} \tag{B.22}$$

The contour of integration in (B.21) is the edge E_{ab} and the components of (B.22) are defined in appendix B.1. The basic idea of the computation is to expand in a perturbative series where $\xi^{-1}H_0$ acts as the free Hamiltonian. Such a procedure will be valid so long as the free part of the Hamiltonian is sufficiently larger than V for all points along the curve, which will be true along the edges of the WKB triangulation. Then we can expand (B.21) in the Born series

$$\begin{aligned}(-1)\hat{\psi}_b^- \hat{\psi}_a^+ \Bigg(\langle + | e^{-\int_0^1 H_0/\xi} | + \rangle \ - \ & \int_0^1 dt_1 \langle + | e^{-\int_{t_1}^1 H_0/\xi} V(t_1) e^{-\int_0^{t_1} H_0/\xi} | + \rangle \\ & + \int_0^1 dt_2 \int_0^{t_2} dt_1 \langle + | e^{-\int_1^{t_2} H_0/\xi} V(t_2) e^{-\int_{t_1}^{t_2} H_0/\xi} V(t_1) e^{-\int_0^{t_1} H_0/\xi} | + \rangle \Bigg)\end{aligned} \tag{B.23}$$

Let us explain a subtle point regarding the 'external states' in the above expression. We start with the small solution at P_a which we take to be ψ_a^+. We then transport it to P_b and then extract the coefficient of the exponentially growing part – that is, we take the inner product with the small part of this transported solution. Since $\psi_a^+ \sim |+\rangle$ grows as we transport it along a WKB curve (i.e. it decays as one follows the curve into P_a and thus grows as we transport it away from P_a) and H_0 is diagonal, we infer that the small part of the solution at P_b is the solution proportional to $|-\rangle$. Thus we take the out-state to be $\langle - | \psi_b^-$. Finally, since the inner product is the antisymmetric the $\langle - |$ gets flipped to a $\langle + |$.

Using the fact that $|\pm\rangle$ are eigenstates of the free Hamiltonian we can easily evaluate the order $\mathcal{O}(V^0)$ and $\mathcal{O}(V^1)$ terms in (B.23). For the $\mathcal{O}(V^2)$ term, we insert the identity $|+\rangle\langle+| + |-\rangle\langle-|$ between the two insertions of V. We find

$$(-1)\hat{\psi}_b^- \hat{\psi}_a^+ \, e^{+\frac{1}{2}\int_0^1 \omega/\xi} \left(1 - \int_0^1 dt_1 \langle +|V(t_1)|+\rangle + \frac{1}{2}\left[\int_0^1 dt_1 \langle +|V(t_1)|+\rangle\right]^2 + \int_0^1 dt_2 \int_0^{t_2} dt_1 e^{-\int_{t_1}^{t_2}\omega/\xi} \langle +|V(t_2)|-\rangle\langle -|V(t_1)|+\rangle \right) \tag{B.24}$$

Now concentrate on the second term on the $\mathcal{O}(V^2)$ contribution. As $\xi \to 0$ the factor $\exp\left(-\int_{t_1}^{t_2}\omega/\xi\right)$ will suppress the integrand except for the small range $t_2 = t_1 + \mathcal{O}(\xi)$ and thus the result of the first integration will already be $\mathcal{O}(\xi)$. So to order ξ we can take ω to be constant and $V(t_1) \to V(t_2)$. We then find for the second term in the $\mathcal{O}(V^2)$ contribution

$$e^{\frac{1}{2}\int_0^1 \omega/\xi} \int_0^1 dt_2 \xi \frac{|\langle +|V(t_2)|-\rangle|^2}{\dot{w}\sqrt{T}} \tag{B.25}$$

Putting everything together, we see that the result re-exponentiates and we find

$$(-1)\hat{\psi}_b^- \hat{\psi}_a^+ \exp\left[+\frac{1}{\xi}\int_0^1 dt \frac{1}{2}\sqrt{T} - \int_0^1 dt \langle +|V(t)|+\rangle + \xi \int_0^1 dt \frac{|\langle +|V(t)|-\rangle|^2}{\dot{w}\sqrt{T}} \right] \tag{B.26}$$

Grouping each term based on its order in ξ (including the prefactors $\hat{\psi}_b^- \hat{\psi}_a^+$ whose explicit expression are given in (B.15)) we find

$$(s_b \wedge s_a)(\xi) \sim \exp\left[+\frac{1}{2}\xi^{-1}\varpi_{ab} + \alpha_{ab} + \frac{1}{2}\xi\overline{\varpi}_{ab} + \xi\eta_{ab} \right] \tag{B.27}$$

where

$$\varpi_{ab} = \lim_{w_a' \to w_a} \lim_{w_b' \to w_b} \left[\int_{E_{ab}} \sqrt{T} dw + \frac{\Delta_a}{2}\log(w_a - w_a') + \frac{\Delta_b}{2}\log(w_b - w_b') \right] \tag{B.28}$$

$$\alpha_{ab} = -\int_{E_{ab}} \left(\frac{1}{4}\partial_w \left(\gamma - \log\sqrt{T\bar{T}} \right) dw + \frac{1}{4}\partial_{\bar{w}} \left(\gamma + \log\sqrt{T\bar{T}} \right) d\bar{w} \right) \tag{B.29}$$

$$\eta_{ab} = \int_{E_{ab}} \left(\frac{1}{2}\sqrt{\bar{T}} \left(\cosh\gamma - 1 \right) \, d\bar{w} + \frac{1}{4\sqrt{T}} \left(\partial\gamma \right)^2 \, dw \right) \tag{B.30}$$

This completes the derivation of formula (3.26) used in the main text. The integral ϖ_{ab} is defined as in (3.58). Note that the logarithmic terms in ϖ_{ab} in (B.27) are due to the prefactor $\hat{\psi}_b^- \hat{\psi}_a^+$. These terms precisely cancel the singularities at the endpoints of the integral $\int_a^b \omega$ so that ϖ_{ab} is finite as we continue the limits of integration all the way up the to punctures at w_a and w_b [38].

B.3.3 WKB expansion of the coordinates

In the previous section we derived the $\xi \to 0$ WKB expansion of $(s_a \wedge s_b)$ up to order $\mathcal{O}(\xi)$. To compute the WKB expansion of the coordinate χ_E we simply combine the expansions for each edge of the quadrilateral Q_E, taking care to account for the directions of the WKB lines as discussed in section 3.3.6. When this is done each of the integrals (B.28) - (B.30) become closed integrals along the cycle γ_E. The asymptotics of the χ-functions are needed for the inversion of the χ-system described in section 3.3.9. For that purpose only the non-vanishing contributions are needed in the $\xi \to 0, \infty$ limits.

There is one very important subtlety that must be addressed here, which is that of the monodromy around u-spikes discussed in appendix B.2.2. We take the point of view that we have made the (multi-valued) gauge-transformation (B.16) that removes the monodromy about each u-spike. The small solutions in this gauge are single valued throughout Q_E, but we must account for the monodromy of the gauge transformation about Q_E. This monodromy is simply $(-1)^{u_E}$ where u_E is the number of u-spikes in Q_E.

Combining the above discussion with (B.29), the constant term in the WKB expansion of χ_E is given by

$$C_E^{(0)} = \log(-1)^{u_E} - \frac{1}{4}\int_{\gamma_E} \left(d\gamma + *d\log\sqrt{T\bar{T}} \right) = \log(-1)^{u_E} \pm i\pi \tag{B.31}$$

To arrive at the last equality (B.31) we have used the fact that γ is single-valued on the 4-punctured sphere so that the integral of $d\gamma$ on any closed contour is zero. The integral of $*d\log\sqrt{T\bar{T}}$ is simple to do explicitly and gives the $\pm i\pi$ factor.[1]

The discussion of the $\xi \to \infty$ limit follows along the same lines as the $\xi \to 0$ limit. The singular term is given by $e^{\xi \int_{\gamma_E} \bar{\omega}/2}$. The constant term is the same. Thus the full non-vanishing WKB asymptotic is given by

$$\chi_E \sim (-1)^{u_E} \exp\left[\frac{1}{2} \int_{\gamma_E} \left(\xi^{-1}\omega + \xi\bar{\omega} \right) \right] \tag{B.32}$$

[1]The $\pm$ depends on the orientation of γ_E but both signs have the same overall effect so that the $\pm$ is irrelevant.

where we recall that u_E is the number of u-spikes enclosed in γ_E. This is the expression (3.40) used in the main text.

B.4 Fold lines and Properties of γ

In this appendix we discuss some properties of the function γ and how they are related to geometric features of the string embedding. In appendix B.5 we show that the world-sheet contours where $\gamma = 0$ map to the fold-lines of the target space solution; in appendix B.5.1 we discuss how the geometry of the string embedding near the boundary is deduced from the structure of these $\gamma = 0$ contours near the points w_a; finally, in section B.5.2 we show how the global structure of the $\gamma = 0$ contours is deduced from the choice of spikes in γ. The point of this appendix is to give the background details that were omitted in the discussion of section 3.2.3.

B.5 Fold lines

In this section we show that the contours on the worldsheet where $\gamma = 0$ map to the fold lines of the string embedding. This was pointed out in [38]. Recall the relation between γ and the world-sheet metric

$$\sqrt{T\bar{T}}\cosh\gamma = \frac{\partial x\bar{\partial}x + \partial z\bar{\partial}z}{z^2} \tag{B.33}$$

Furthermore, we have

$$T(w) = \frac{(\partial x)^2 + (\partial z)^2}{z^2}, \qquad \bar{T}(\bar{w}) = \frac{\left(\bar{\partial}x\right)^2 + \left(\bar{\partial}z\right)^2}{z^2} \tag{B.34}$$

Now, suppose that $\mathcal{C}$ is a curve on the worldsheet that maps to a fold-line of the string and consider a point $\mathcal{P}$ in that curve. We can choose local coordinates at $\mathcal{P}$ so that the derivative takes the form $\partial_w \to e^{i\phi}\left(\partial_t + i\partial_n\right)$ where the direction ∂_n is chosen such that $\partial_n z = 0$. The prefactor $e^{i\phi}$ is the Jacobian of the coordinate transformation (just a translation and rotation). The defining property of the fold-line is then that the x-coordinate reaches an local extrema and thus we also have $\partial_n x = 0$ as we cross the fold. Therefore along the fold-line we have (with $\partial_t x = \dot{x}$)

$$\sqrt{T\bar{T}}\cosh\gamma \to \frac{\dot{x}^2 + \dot{z}^2}{z^2}, \qquad T(w) \to e^{2i\phi}\frac{\dot{x}^2 + \dot{z}^2}{z^2}, \qquad \bar{T}(\bar{w}) \to e^{-2i\phi}\frac{\dot{x}^2 + \dot{z}^2}{z^2} \tag{B.35}$$

Using the last two equations to solve for $\sqrt{T\bar{T}}$ we see that they are consistent with the first equation only if $\gamma = 0$. Therefore, the worldsheet contours where $\gamma = 0$ map to the fold-lines of the string-embedding. For this reason, we frequently refer to the contours where $\gamma = 0$ as fold-lines.

B.5.1 Structure of γ near w_a

To gain some intuition about the structure of the contours where $\gamma = 0$ it is useful to study the behavior of γ near the points w_a. Recall that γ is defined as the solution to the PDE:

$$\partial\bar{\partial}\gamma = \sqrt{T\bar{T}}\sinh\gamma \tag{B.36}$$

subject to the boundary conditions

$$\gamma \rightarrow \pm\frac{1}{2}\log T\bar{T} \qquad (w \rightarrow z_a) \tag{B.37}$$

$$\gamma \rightarrow 0 \qquad (w \rightarrow P_a) \tag{B.38}$$

The boundary condition (B.38) simply imposes that γ is non-singular at the singularities of T and this condition is automatically imposed if we demand the solution be regular away from the zeros of T.

Since we know that γ must vanish at singularities of T, it's natural to study the function in the neighborhood of these points. Let us consider some P_a and use polar coordinates (r, ϕ) in which the origin is at w_a. Since γ is vanishing, we can linearize the RHS of (B.36). Further, we can take $\sqrt{T\bar{T}} \sim |\Delta|^2/(4r^2)$. The PDE becomes linear and separable and using standard techniques one finds the series solution

$$\gamma \sim g_0\, r^{\frac{1}{2}\Delta} + \sum_{m=1}^{\infty} g_m \sin\left(m\phi + \delta_m\right) r^{\frac{1}{2}\sqrt{\Delta^2+4m^2}} \tag{B.39}$$

Now consider a small circle centered at $r = 0$. As $r \rightarrow 0$ the series (B.39) is dominated by the lowest mode in the expansion. Thus along an infinitesimal circle centered at $r = 0$ the series (B.39) will vanish $2m^*$ times, where g_{m^*} is the smallest non-zero coefficient g_m, $m = 0, 1, 2, ...$ in the series. Thus, if g_0 is the smallest non-zero coefficient then the series will vanish only at the point w_a which will be a local extrema. If $m^* = 1$ then the series will vanish along a single curve passing through P_a; if $m^* = 2$ then γ will vanish along two curves that intersect at P_a, and so on.

The fact that the contours where $\gamma = 0$ map to fold-lines of the target-space solution

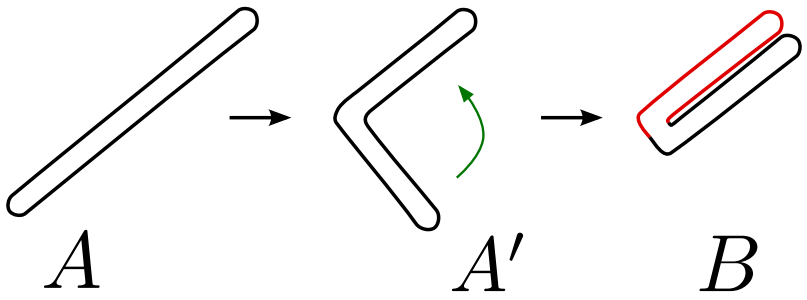


Figure B.1: Single-folded and double-folded string in panels A and B respectively.

gives a clear geometric meaning to each possible behavior $m^* = 1, 2, ...$ near an insertion point. For $m^* = 1$ we will cross two fold-lines as the world-sheet coordinate traverses a small loop around the point w_a. This means that near the insertion point the string is single-folded as shown in figure B.1A. For $m^* = 2$ we will cross two fold-lines as the world-sheet coordinate traverses a small loop around the point w_a. This means that near the insertion point the string is double-folded as shown in figure B.1B, for example. In general for $n > 0$ the case $m^* = n$ should correspond to an n-folded string. The only subtle case seems to be $m^* = 0$. Apparently if $m^* = 0$, as we traverse a closed loop around w_a the contour swept out in the target space does not close since there is no point at which the coordinates (x, z) can 'turn around'. In this paper we are only interested in solutions that are closed (i.e. the embedding coordinates have trivial monodromies around operator the insertion points x_a) and thus we will only study cases for which $m^* > 0$ at all w_a. This is further discussed in appendix B.5.2.

It is important to keep in mind that (as we mentioned above) the behavior of γ at P_a is not our choice, and is determined by regularity and the conditions ($B.37$). In other words, for fixed T the only remaining conditions one can specify are the choice of signs in ($B.37$). For each choice of signs there will be a unique m^* for each P_a. In the next section we demonstrate how this works using the T of the 4-point function discussed in the main text.

B.5.2 Structure of contours where $\gamma = 0$

In this section we describe why the spike configurations of figure 3.4 are the only two physically relevant configurations. Furthermore, we deduce the structure of the contours where $\gamma = 0$ for each of these spike configurations.

Consider T fixed to be that of the 4-point function discussed in the main text (see equation (3.60) and figure 3.15). There are 4 zeros and therefore 2^4 ways to choose the signs in (B.37). Because of the symmetry of (B.36) under $\gamma \to -\gamma$, without loss of generality we can fix one of the spikes to be up which leaves 2^3 choices. Now, because the string is embedded in AdS_2 we know that it must be folded. Moreover we know that the operator insertions x_a will sit along the fold-lines of the target-space solution. In the world-sheet coordinates this translates to the statement that we should require $m^* > 0$ at each w_a. That is, there should be at least one contour where $\gamma = 0$ running through each insertion point w_a. For the 4-point function T (see equation (3.60) and figure 3.15) the only obvious way to accomplish this in general is to choose the spikes such that $\gamma \to -\gamma$ under reflection about the real axis. This leaves only the spike configurations of figure B.2*A*,*B*, which are those of figure 3.4 used in the main text. We will now discuss the global structure of the $\gamma = 0$ contours for these two choices of spikes.

In figure B.2 we show the fold-structure for three different spike configurations. The black lines schematically represent the contours where $\gamma = 0$ and one can read off the m^* associated with each puncture. The structure of these contours is determined purely by the choice of the directions of the spikes of γ. We refer to these contours as 'fold lines' since they map onto the fold-lines of the target-space embedding (see appendix B.5). We guess the structure of the fold lines for each choice of the spikes as follows: u spikes must be separated from d spikes by at least one fold line; we use the minimum number of fold lines needed to accomplish this for all spikes. Note that fold lines must encircle at least one zero of T.[2] This restriction is useful because, for example, it allows one to rule-out the possibility of fold-lines corresponding to the gray contours in figure B.2C. This is important because if it was possible for the gray contours to be fold-lines then it might be possible to have a solutions with all $m^* > 0$ for configuration C. Configurations A and B are the physical configurations that we study in this paper and we have checked the fold structures of figure B.2*A*,*B* numerically. Configuration C is an example of a spike-configuration that does not correspond to a target-space solution with the desired properties; the corresponding fold structure is only our best guess but we have not checked it numerically.

To summarize this appendix, in appendix B.5 we showed that the world-sheet contours where $\gamma = 0$ map to the fold-lines of the target space solution; in appendix B.5.1 we

[2] Consider a closed contour along which $\gamma = 0$ and suppose (for a contradiction) that it does not enclose any zeros of T. Let $\mathcal{D}$ be the region enclosed by the contour. This contour must separate positive values from negative values (i.e. it cannot sit at the bottom of a 'valley' since this locally violates the equation (B.36)). Suppose for simplicity that $\gamma < 0$ in $\mathcal{D}$. Since γ is regular away from the zeros of T, there must be at least one local minimum inside $\mathcal{D}$, and therefore at least one point where $\left(\partial_x^2 + \partial_y^2\right)\gamma \geq 0$. Thus at such a point the LHS of (B.36) is positive or zero, but the RHS is strictly less than zero by assumption, which is the desired contradiction.

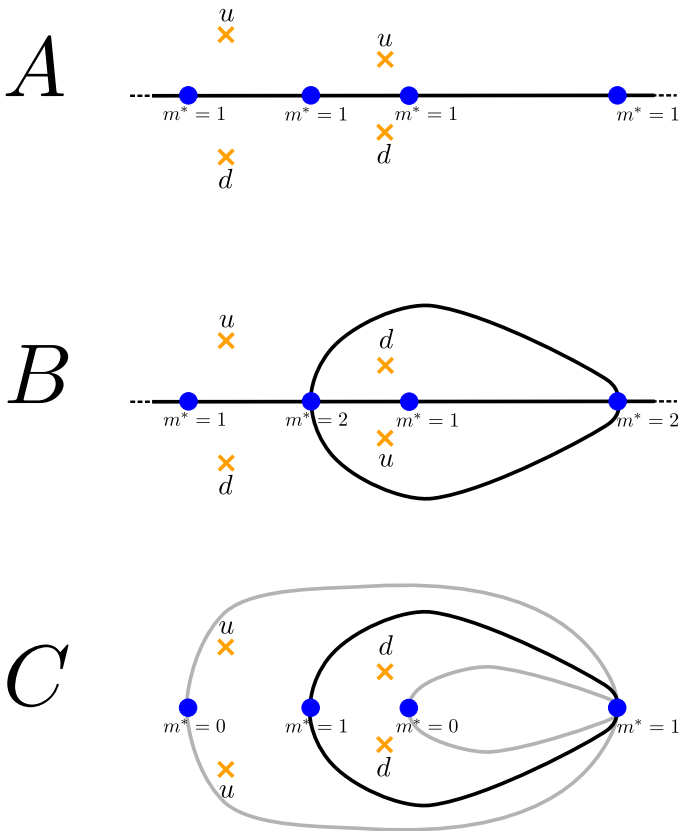


Figure B.2: Three different spike configurations and the corresponding structure of the $\gamma = 0$ contours. The black lines schematically represent the contours where $\gamma = 0$ and one can read off the m^* associated with each puncture. Panels A and B show the physically relevant configurations studied in the main text. Panel C shows a third spike configuration which is not physical due to the presence of $m^* = 0$ behavior at two of the insertion points. The gray contours in panel C indicate contours that cannot cannot correspond to fold-lines due to the restriction that $\gamma = 0$ contours must encircle at least one zero of T (see footnote 2). In this figure we are not indicating the location saddle-point w_4^* because it is not relevant for the present discussion (so long as it is located somewhere on the real axis).

discussed how the geometry of the string embedding near the boundary is deduced from the structure of these $\gamma = 0$ contours near the points w_a; finally, in section B.5.2 we discussed how the global structure of the $\gamma = 0$ contours is deduced from the choice of spikes in γ. From all of this one can deduce some qualitative global features of the string

embedding, which is discussed in detail in section 3.2.3.

B.6 Details of the 4-point function computation

B.6.1 Explicit expression for stress-energy tensor coefficients

For completeness we present the coefficients c_a of the stress-energy tensor in formula (3.60),

$$\begin{aligned} c_\infty &= \frac{\Delta_\infty^2}{4} \\ c_0 &= \frac{1}{4}\left[4\,U w_4 + 2w_4\,(1+w_4)\,\Delta_3^2 + (-1+w_4)\left(2w_4\Delta_1^2 + (1+w_4)\left(\Delta_2^2 - \Delta_4^2\right)\right)\right] \\ c_1 &= \frac{1}{2}\left[-2\,U + (-1+w_4)^{\,2}\Delta_1^2 - (1+w_4)^{\,2}\Delta_3^2\right] \\ c_2 &= \frac{1}{4}\left[-4\,U w_4 + 2\,(1+w_4)\,\Delta_3^2 + (-1+w_4)\left(-2\Delta_1^2 + (1+w_4)\left(-\Delta_2^2 + \Delta_4^2\right)\right)\right] \end{aligned} \tag{B.40}$$

B.6.2 Explicit expressions for χ-functions and A_{PQ}

For reference, we include here the explicit expressions for the χ-functions for the triangulation of figure 3.15. They are given by

$$\chi_{12} = (-1)\frac{(s_1 \wedge M_1^{-1} s_4)(s_2 \wedge s_4)}{(M_1^{-1} s_4 \wedge s_2)(s_4 \wedge s_1)} \tag{B.41}$$

$$\chi_{23} = (-1)\frac{(s_2 \wedge M_3 s_4)(s_3 \wedge s_4)}{(M_3 s_4 \wedge s_3)(s_4 \wedge s_2)} \tag{B.42}$$

$$\chi_{34} = (-1)\frac{(s_4 \wedge s_2)\left(s_3 \wedge M_3^{-1} s_2\right)}{(s_2 \wedge s_3)\left(M_3^{-1} s_2 \wedge s_4\right)} \tag{B.43}$$

$$\chi_{14} = (-1)\frac{(s_4 \wedge M_1 s_2)\,(s_1 \wedge s_2)}{(M_1 s_2 \wedge s_1)\,(s_2 \wedge s_4)} \tag{B.44}$$

$$\chi_{24} = (-1)\frac{(s_2 \wedge s_3)(s_4 \wedge s_1)}{(s_3 \wedge s_4)(s_1 \wedge s_2)} \tag{B.45}$$

$$\chi_{\hat{24}} = (-1)\frac{(M_3^{-1} s_2 \wedge M_4 s_1)(s_4 \wedge s_3)}{(M_4 s_1 \wedge s_4)(s_3 \wedge M_3^{-1} s_2)} \tag{B.46}$$

One can check that these coordinates satisfy the rule (3.37) at each puncture. The χ-system obeyed by these coordinates is given by

$$\chi_{24}\chi_{24}^{++} = \left(\chi_{\hat{24}}\chi_{\hat{24}}^{++}\right)^{-1} = \frac{(1+A_{23})\,(1+A_{14})}{(1+A_{34})\,(1+A_{12})} \tag{B.47}$$

$$\chi_{12}\chi_{12}^{++} = \left(\chi_{14}\chi_{14}^{++}\right)^{-1} = \chi_{34}\chi_{34}^{++} = \left(\chi_{23}\chi_{23}^{++}\right)^{-1} = \frac{(1+A_{24})}{(1+A_{\hat{24}})} \tag{B.48}$$

where the A_{PQ} are given by

$$A_{12} = \frac{\chi_{12}\,(1+\chi_{14})\,(1+\hat{\chi}_{24}\,(1+\chi_{23}\,(1+\chi_{24})))}{(1-\mu_1^2)\,(1-\mu_2^2)} \tag{B.49}$$

$$A_{23} = \frac{\chi_{23}\,(1+\chi_{34})\,(1+\chi_{24}\,(1+\chi_{12}\,(1+\chi_{\hat{24}})))}{(1-\mu_2^2)\,(1-\mu_3^2)} \tag{B.50}$$

$$A_{34} = \frac{\chi_{34}\,(1+\chi_{23})\,(1+\hat{\chi}_{24}\,(1+\chi_{14}\,(1+\chi_{24})))}{(1-\mu_3^2)\,(1-\mu_4^2)} \tag{B.51}$$

$$A_{14} = \frac{\chi_{14}\,(1+\chi_{12})\,(1+\chi_{24}\,(1+\chi_{34}\,(1+\hat{\chi}_{24})))}{(1-\mu_1^2)\,(1-\mu_4^2)} \tag{B.52}$$

$$A_{24} = \frac{\chi_{24}\,(1+\chi_{12}\,(1+\chi_{\hat{24}}\,(1+\chi_{23})))\,(1+\chi_{43}\,(1+\chi_{\hat{42}}\,(1+\chi_{41})))}{(1-\mu_2^2)\,(1-\mu_4^2)} \tag{B.53}$$

$$A_{\hat{24}} = \frac{\hat{\chi}_{24}\,(1+\chi_{23}\,(1+\chi_{24}\,(1+\chi_{12})))\,(1+\chi_{41}\,(1+\chi_{42}\,(1+\chi_{43})))}{(1-\mu_2^2)\,(1-\mu_4^2)} \tag{B.54}$$

Using the explicit expressions for the coordinates (B.41)-(B.46), schouten identity and the shift relation (3.43) one can directly verify the functional equations (B.47)-(B.48).

B.6.3 Finite part of AdS

In this section, we present some intermediate steps in the derivation of our formula (3.66) for the finite part of the AdS contribution. We want to compute

$$A_{fin} = \frac{\pi}{3} - \frac{i}{2}\left(\oint_{\gamma_a}\omega\right) I_{ab}^{-1}\left(\oint_{\gamma_b}\eta\right) . \tag{B.55}$$

according to the steps outline in section 3.4.1. The complete basis of five a-cycles and five b-cycles that we chose is depicted in figure B.3. From this figure we also read-off the

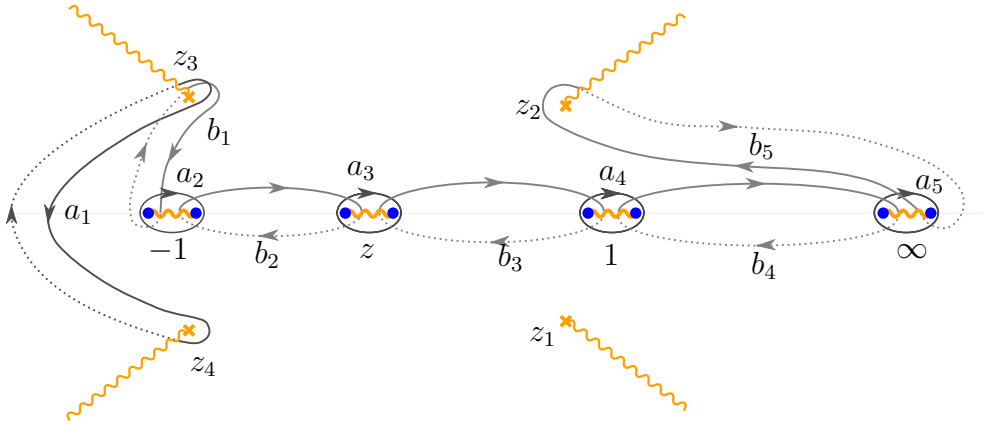


Figure B.3: The cycles for Riemann bilinear identity. The dashed line represents a contour in a different Riemann sheet. The wavy lines represent a choice of branch cuts. From this picture we also read the intersection matrix I_{ab} of the cycles. For each pair of cycles, say γ_a and γ_b, intersecting at a point with tangent vectors ∂_a and ∂_b respectively, we assign $I_{ab} = +1$ (-1) if $\det[\{\partial_a, \partial_b\}] > 0\,(< 0)$.

intersection matrix $I_{ab} = (\delta_{a+1,b} - \delta_{a-1,b})$ using the conventions described in the caption. The only other ingredient we need is

$$\int_{a_i} \eta = 0, \qquad i = 2, \ldots, 5 \tag{B.56}$$

which follows from the regularity of η at the poles of T. Plugging into (B.55) and computing we find

$$A_{fin} = \frac{\pi}{3} + i\left(\omega_{a_1}\eta_{z_3,z_2} + \omega_{a_2}\eta_{-1,z_2} + \omega_{a_3}\eta_{z,z_2} + \omega_{a_4}\eta_{1,z_2} + \omega_{a_5}\eta_{\infty,z_2}\right) - i\left(\sum_{i=1}^{5}\omega_{b_i}\right)\eta_{z_3,z_4} \tag{B.57}$$

where we are using the notation $\eta_{ab} = \int_a^b \eta$ and $\omega_{\mathcal{C}} = \int_{\mathcal{C}} \omega$ and the contours are defined in figures B.3 and B.4.

Each of these η_{ab} can be written as a linear combination of the $\eta_{E_{ab}} = \int_{E_{ab}} \eta$ where the integral is taken along the WKB-line from P_a to P_b and the direction of the contour is the same as that of the WKB line. The idea is to combine the $\eta_{E_{ab}}$ to form the contour that we want. Let us exemplify with η_{1,z_2}. From the WKB configuration, see figure B.4, we see that the large θ expansion of the ratio $\frac{(s_1 \wedge s_2)(s_1 \wedge s_4)}{(s_2 \wedge s_4)}$ involves a cycle that can be continuously deformed into *twice* the line integral connecting the puncture at $w = 1$ and

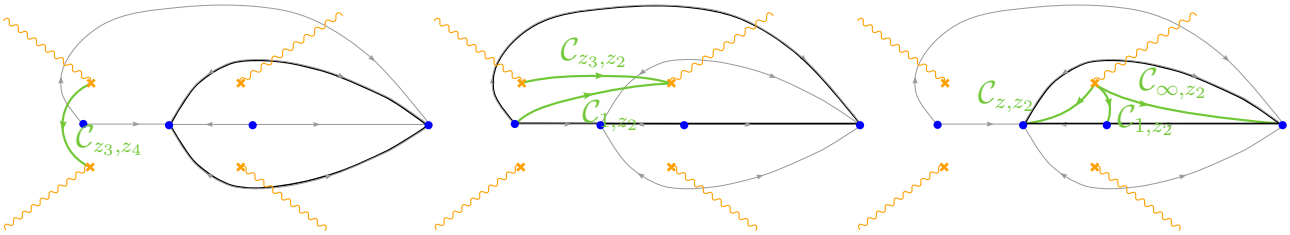


Figure B.4: To extract line integrals connecting a zero to puncture or connecting two zeros we combine products of elementary solutions that have WKB expansions involving integrals over the paths indicated by the black lines. The resulting closed contours can be continuously deformed into the contour that we want, indicated by the green lines. The precise way of combining these products is dictated by the direction of the WKB lines indicated by the gray arrows.

the zero at $w = z_2$. Therefore we have

$$\eta_{1,z_2} = \frac{1}{2}\int_{-\infty}^{\infty}\frac{d\theta}{\pi}e^{-\theta}\log\left[\frac{\left(1+A_{12}^{-}\right)\left(1+A_{14}^{-}\right)}{\left(1+A_{24}^{-}\right)}\right] = \frac{1}{2}\left(\eta_{E_{12}}+\eta_{E_{14}}-\eta_{E_{24}}\right) \tag{B.58}$$

In the same way we obtain

$$\eta_{z_3,z_4} = \frac{1}{2}\left(\eta_{24}-\eta_{\hat{24}}\right) \tag{B.59}$$

$$\eta_{-1,z_2} = \frac{1}{2}\left(2\eta_{34}+\eta_{12}-\eta_{14}-\eta_{24}\right) \tag{B.60}$$

$$\eta_{z,z_2} = \frac{1}{2}\left(\eta_{12}-\eta_{14}-\eta_{24}\right) \tag{B.61}$$

$$\eta_{\infty,z_2} = \frac{1}{2}\left(\eta_{14}-\eta_{12}-\eta_{24}\right) \tag{B.62}$$

$$\eta_{z_3,z_2} = \frac{1}{2}\left(\eta_{12}+\eta_{34}-\eta_{14}-\eta_{23}\right) \tag{B.63}$$

where the notation is the natural simplification of that used in (B.58). Plugging these expressions into (*B*.57) and re-collecting each η_E, one finds that the coefficient of η_E is simply ω_E where ω_E is the ω-cycle the intersects edge E, *not* the integral of ω along edge E (which would be divergent). That is, it's (1/2 of) the ω-cycle associated with the coordinate χ_E which are shown in figure 3.15. Thus we have

$$A_{fin} = \frac{\pi}{3} - i\sum_{E\in\mathcal{T}}\omega_E\,\eta_E \tag{B.64}$$

which is formula (3.66) as desired.

Equation (*B*.64) is perhaps the simplest possible result one could write from the triangulation data. Given this simplicity, it is probably possible to derive the result in a much more elegant way and perhaps even for any number of punctures. We have not pursued this issue but feel that it merits further exploration.

B.7 Three-point function in GMN language

In this section we apply the method developed in section 3.3 to the three point correlation function studied in [38]. We use the setup of [38], namely the same stress-energy tensor. We aim at deriving a set of functional equations to extract the cycles used there.

As a starting point, we introduce the WKB triangulation for this configuration from which we define the coordinates, see figure B.5. From this figure, we easily derive the χ-system. Since the quadrilateral is very degenerate it follows from (3.46) that the right hand side of the χ-system is equal to 1. The reason is that the same auxiliary A_{PQ}'s appear both in numerator and denominator canceling each other. Hence, the solution of the functional equations is simply given by the WKB asymptotics. More explicitly, the χ functions take the form

$$\chi_{ac} = (-1)\exp\left(\frac{1}{2}e^{\theta}\int_{\gamma_{ac}}\omega + \frac{1}{2}e^{-\theta}\int_{\gamma_{ac}}\bar{\omega}\right) = -\frac{\mu_a\mu_c}{\mu_b} \tag{B.65}$$

where a, b and c are distinct.[3] The spikes must be in pointing in opposite direction as follows from the discussion of appendix B.4. This is the the origin of the (-1) prefactor in (B.65). The cycles of ω are given in terms of the dimensions of the operators,

$$\int_{\gamma_{ac}}\omega = i\pi(-\Delta_a - \Delta_c + \Delta_b) \tag{B.66}$$

Having the solutions of the functional equations, we can easily find the auxiliary quantities A_{PQ} using the rules of section 3.3.8. The determination of the η-cycles is also straightforward. To compare with the result in [38] let us set $\Delta_1 = \Delta_2 = \Delta$ and $\Delta_3 = \Delta_\infty$. We use

[3] This result also follows directly from the definition of the coordinates in terms of the small solutions, $\chi_{ac} = -\frac{(s_c\wedge s_b)(s_a\wedge M_a^{-1}s_b)}{(M_a^{-1}s_b\wedge s_c)(s_b\wedge s_a)}$ for distinct a, b and c; all the inner-products cancel and one is left with only the monodromy factors in (B.65).

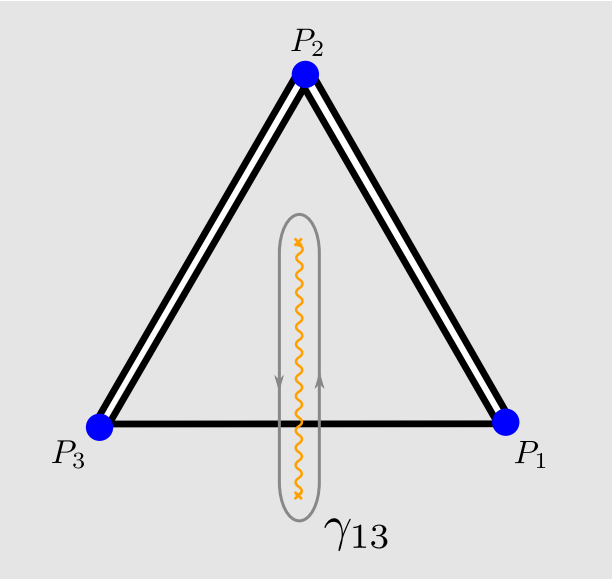


Figure B.5: The WKB triangulation for the 3-point function is composed of 3 edges forming two triangles on the sphere. Here we show the construction of the coordinate χ_{13}. We are using the edge-splitting procedure discussed in section 3.3.5 (in particular, see figure 3.12). The gray contour shows the cycle associated with the coordinate χ_{13}.

expression 3.57 to compute the cycles, and we get

$$\int_{-1}^{1}\eta = \int_{\mathbb{R}}\frac{d\theta}{\pi}e^{-\theta}\log(1+A_{12}^{-}) = h(2\Delta-\Delta_\infty)+h(2\Delta+\Delta_\infty)-2h(2\Delta) \tag{B.67}$$

$$\int_{1}^{\infty}\eta = \int_{\mathbb{R}}\frac{d\theta}{\pi}e^{-\theta}\log(1+A_{23}^{-}) = h(\Delta_\infty)+h(2\Delta+\Delta_\infty)-h(2\Delta)-h(2\Delta_\infty) \tag{B.68}$$

where we define

$$h(x) = \int_{\mathbb{R}}\frac{d\theta}{\pi}\cosh\theta\log\left(1-e^{-x\pi\cosh\theta}\right)\,. \tag{B.69}$$

This is precisely the result obtained in [38]. A last comment about the expression for the area in the three point function. It is easy to show using the same type of manipulation of the four point function case that the area can be expressed in terms of elements of the WKB triangulation as

$$A_{fin} = \frac{\pi}{6} - i\sum_{E\in\mathcal{T}}\omega_E\eta_E \tag{B.70}$$

where the sum is over the edges of the triangulation of figure B.5. As in the case of the four point function, we define $\eta_{E_{ab}}$ as the η-cycle that passes along edge E_{ab} from P_a to P_b and $\omega_{E_{ab}}$ as the ω-cycle that intersects edge E_{ab}.

Appendix C

Bootstrap appendices

C.1 S-matrix bootstrap numerics

In this appendix we give more details on the numerics described in section 5.1. We consider a grid $\{x_0, x_1 \dots x_M\}$ and measure everything in units of m so that $x_0 = 4$. Denote by ρ_n the value $\rho(x_n)$ and approximate $\rho(x)$ by a linear spline connecting the points (x_n, ρ_n) as shown in figure 5.3. We can then perform the integrals in (5.6) analytically giving (5.7) with

$$K_a(s) = \frac{(x_{a-1}-s)\log(x_{a-1}-s)}{x_{a-1}-x_a} + \frac{(x_{a+1}-s)\log(x_{a+1}-s)}{x_a - x_{a+1}} - \frac{(x_{a-1}-x_{a+1})(x_a-s)\log(x_a-s)}{(x_{a-1}-x_a)(x_a-x_{a+1})} + (s \to 4-s)\,,$$

with $a = 1, 2, \dots, M-1$ while for the last point of the grid

$$K_M(s) = \frac{(x_{M-1}-s)\log(x_{M-1}-s)}{x_{M-1}-x_M} - \frac{x_M \log(x_M)}{s} + \frac{(x_{M-1}-x_M-s)(x_M-s)\log(x_M-s)}{s(x_{M-1}-x_M)} - 1 + (s \to 4-s)\,.$$

Note that for $x > x_M$ we assume a tail of the form $\rho(x) \sim \rho_M x_M/x$ which leads to the above result for K_M.

We can now evaluate the approximate S-matrix (5.7) at a bunch of points with $s \geq 4$. It is convenient to evaluate on the gridpoints x_a themselves (although not necessary of course) so that we can make use of the identity

$$\operatorname{Im} S(x_a + i0) = \pi \rho_a. \tag{C.1}$$

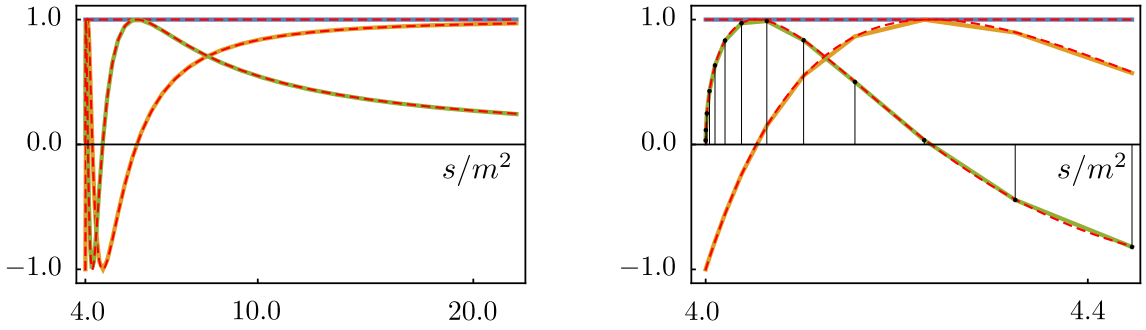


Figure C.1: Result of numerics for the spectrum $\mathbb{m} = \{1, 1.6, 1.8\}$ compared with the expectation (5.18) and its near-threshold close-up (on the right). The green, orange and blue curves are $\mathrm{Im}(S_{\mathrm{num}})$, $\mathrm{Re}(S_{\mathrm{num}})$, $|S_{\mathrm{num}}|$ where S_{num} is (5.7) evaluated on the result of the numerics given in (C.2). The black dots indicate the points (x_n, ρ_n); note that we use a grid which clusters points near threshold. The dashed red curves are the corresponding parts of the exact solution (5.18).

This gives set of M constraints[1]

$$\left[S_\infty - \sum_j \mathcal{J}_j \left(\frac{g_j^2}{x_a - m_j^2} + \frac{g_j^2}{4m^2 - x_a - m_j^2}\right) + \sum_{n=1}^{M} \mathrm{Re}\left[K_n(x_a)\right] \rho_a\right]^2 + (\pi\rho_a)^2 \leq 1$$

for $a = 1, ..., M$. The goal is, for a given set of masses m_j, to find the point in the space $\{S_\infty, g_1, g_2, ..., g_N, \rho_1, ..., \rho_M\}$ such that g_1 is as big as possible and the constraints (C.2) are satisfied. This amounts to a standard problem in quadratic optimization and the `Mathematica` program `FindMaximum` is conveniently designed to carry out such a task. The attached notebook contains our implementation of this problem in Mathematica. There we implement a function `MaxCoupling[m_]` which takes a spectrum $\mathbb{m}$ as input and returns the maximum value of g_1 along with the corresponding values of the variables $g_{j>1}$, ρ_n and S_∞. To illustrate with a typical example, the output of `MaxCoupling` for

[1]Note that $\mathrm{Re}[K_n(x_a)]$ can be computed simply by replacing $\log(\dots) \to \log(\mathrm{abs}(\dots))$ in the expressions for $K_n(s)$.

$\mathbb{m} = \{1, 16/10, 18/10\}$ (in units of m) is

```
{g[1] → 90.1938, g[2] → 98.5414, g[3] → 49.9688, S∞ → 1., ρ[1] → -0.0106059, ρ[2] → -0.0365801,
 ρ[3] → -0.0792579, ρ[4] → -0.136035, ρ[5] → -0.201675, ρ[6] → -0.264856, ρ[7] → -0.309091,
 ρ[8] → -0.314682, ρ[9] → -0.265642, ρ[10] → -0.159441, ρ[11] → -0.0111107, ρ[12] → 0.14126,
 ρ[13] → 0.261091, ρ[14] → 0.316077, ρ[15] → 0.298462, ρ[16] → 0.220909, ρ[17] → 0.108434,
 ρ[18] → -0.0121623, ρ[19] → -0.121235, ρ[20] → -0.207034, ρ[21] → -0.266571,
 ρ[22] → -0.301614, ρ[23] → -0.316666, ρ[24] → -0.316865, ρ[25] → -0.306922, ρ[26] → -0.290668,
 ρ[27] → -0.270967, ρ[28] → -0.249832, ρ[29] → -0.228599, ρ[30] → -0.2081, ρ[31] → -0.188812,
 ρ[32] → -0.170974, ρ[33] → -0.154669, ρ[34] → -0.139878, ρ[35] → -0.126544, ρ[36] → -0.114527,
 ρ[37] → -0.103793, ρ[38] → -0.0940492, ρ[39] → -0.0854934, ρ[40] → -0.0773852}
```

(C.2)

Note that this is within the parameter range (5.17) which is the region plotted in figure 5.12 and also region B in plot 5.11. Thus we expect the S-matrix to be given by (5.18) with the values of α_j chosen according to $\mathbb{m}$. We can see in figure C.1 that our numerics agree perfectly with expectation.

C.2 Hyperbolic Rapidity

In two dimensions, hyperbolic rapidities are a very useful parametrization of energy and momenta of relativistic particles. For two particles, for instance, we would write

$$p_1^\mu = (E_1, \vec{p}_1) = (m_1 \cosh(\theta_1), m_1 \sinh(\theta_1)) \,, \qquad p_2^\mu = (E_2, \vec{p}_2) = (m_2 \cosh(\theta_2), m_2 \sinh(\theta_2)) \,.$$

In this parametrization consider the elastic scattering of these two particles. In the final state, conservation of energy and momentum imply that the final individual momenta are the same as the initial one, that is $p_3 = p_1$ and $p_4 = p_2$ so that there is no momentum exchange $u = (p_3 - p_1)^2 = 0$. As for the other Mandelstam invariants we have

$$s = (p_1 + p_2)^2 = m_1^2 + m_2^2 + 2m_1 m_2 \cosh(\theta) \,, \qquad t = (p_2 - p_3)^2 = m_1^2 + m_2^2 - 2m_1 m_2 \cosh(\theta) \,,$$

where $\theta = \theta_1 - \theta_2$ is the difference of hyperbolic rapidities.

Note that these relativistic invariants are invariant under shifts of both rapidities. Indeed, boosts act as shifts of θ_1 and θ_2 such that θ is invariant.

Note also that $\theta \leftrightarrow i\pi - \theta$ is a crossing transformation which exchanges s and t. This is also nicely seen directly in terms of the two vectors above. For instance, if we keep θ_2 fixed and send $\theta_1 \to i\pi - \theta_1$ we get that $p_1^\mu \to (-E_1, p_1)$ as expected for a particle/anti-particle transformation. This sends p_1 to the future (and p_3 to the past) as expected for a crossing transformation.

The hyperbolic parametrization is also convenient when dealing with bound-states. Suppose for instance we form a bound-state out of two constituent particles with rapidities $\theta \pm i\eta$ and mass m. Then the total two-momenta of the bound-state would be

$$\begin{aligned} p_{\text{bound-state}} &= (m\cosh(\theta+i\eta)+m\cosh(\theta-i\eta), m\sinh(\theta+i\eta)+m\sinh(\theta-i\eta)) \\ &= (m_{\text{bound-state}}\cosh(\theta), m_{\text{bound-state}}\sinh(\theta))\,, \end{aligned} \tag{C.3}$$

where the bound-state mass

$$m_{\text{bound-state}} = 2m\cos(\eta)\,. \tag{C.4}$$

A necessary (but not sufficient) condition for such bound-states to form is the existence of a pole at $\theta = 2i\eta$ in the S-matrix element $S(\theta)$ describing the elastic process $m+m \to m+m$.

Some theories have a cubic coupling and the particle of mass m can also be though of a bound-state of two particles of mass m. In these cases $\eta = \pi/3$. The Ising field theory with magnetic field discussed in the main text is one such example.

C.3 Form Factor Expansion

The so called Scaled Ising Field Theory is a remarkable field theory, see [91] for a beautiful review. This theory describes the flow from the critical Ising model as we turn on magnetic field and temperature (measured as a deviation from its Curie value). When we turn on temperature only (without magnetic field) or magnetic field alone (without temperature) we end up with integrable theories. The first is that of free fermions while the second is Zamolodchikovs E8 theory [90]. We rediscovered this second special point in section 5.3 as the integrable theory with three stable particles of masses in the range (5.17) and whose cublic coupling to the lightest particle is maximal.

Away from these two Integrable points, the Scaled Ising Field Theory can be studied numerically, either from the lattice or using the so-called Truncated Conformal Space Approach [93–95]. We can also use Integrable Form Factor perturbation [96] theory to study small deformations away from the integrable points. Let us discuss briefly how our general bounds in figure 5.12 compare with this second analysis.

As we deform away from the E8 theory by turning on the temperature slightly the masses of the stable particles move. More precisely, we chose to measure everything in terms of the mass of $m_1 = 1$ which is thus kept fixed but m_2 and m_3 will move. This is a slightly different point of view compared to what is typically taken in the literature – see e.g. [96] – where masses are measured in unit of magnetic field. In this convention

all masses move as we deform away from the integrable point. The results (in this second notation) are given in equations (11) and (64) of [96]. Converting to our conventions we get therefore

$$\left.\frac{\delta m_2}{\delta m_3}\right|_{\text{here}} = \left.\frac{\delta(m_2/m_1)}{\delta(m_3/m_1)}\right|_{\text{there}} = \left.\frac{\delta m_2 - m_2/m_1\,\delta m_1}{\delta m_3 - m_3/m_1\,\delta m_1}\right|_{\text{there}} \simeq 1.57322\,. \tag{C.5}$$

In the small thin blue line in figure 5.12 we marked this slope. We can now compute the slope of our bound for $g_1^{\max}$ along this blue line. We find

$$\log(g_1) = 6.585891698 - 8.683281573\,\delta m_2 + O\left((\delta m_2)^2\right)\,. \tag{C.6}$$

This value must coincide with the variation of the coupling of the Scaled Ising model as we move away from this point or else we will violate our bound as we slightly increase or decrease the temperature. This is what we verified. It is a somehow involved check since extracting this residue from the form factor expansion is considerably harder than correcting the masses. Fortunately, attached to [96] is a long notebook with the four-particle form factor for the energy density operator. Using it we can construct the correction $\delta S(\theta)$ to the S-matrix. From it, we can read off the correction δg_1 to the cubic coupling to the lightest particle. In this way we obtain exactly the slope (C.6) (within the eleven digits of numerical precision of the notebook in [96]).

C.4 Most General Optimal CDD solution

A given mass spectrum $\{m_1/m, \ldots, m_N/m\}$ leads to $2N$ poles between $\theta = 0$ and $\theta = i\pi$. They come in pairs (for the s-channel and the t-channel contribution) related by $\theta \leftrightarrow i\pi - \theta$ and thus it is enough to focus on the segment $[0, i\pi/2]$. We order the poles in this segment and denote them as $\theta_j = i\gamma_j$ with $0 < \gamma_1 < \gamma_2 < \cdots < \gamma_N < \pi/2$. (Needless to say, this ordering is *not* the same as the order $m_1 < m_2 < \cdots < m_N$.) To each pole γ_j we associate a sign $\text{sgn}(j) = +1$ if this is an s-channel pole or $\text{sgn}(j) = -1$ for a t-channel pole. In this way, the set $\{(\gamma_1, \text{sgn}(1)), \ldots, (\gamma_N, \text{sgn}(N))\}$ encodes all the information about the mass spectrum. In terms of this useful notation, the optimal solution is simply

$$S_{g_1^{\max}}(\theta) = \text{sgn}(1)(-1)^{N-1} \times \prod_{j=1}^{J-1} [\gamma_j]^{\frac{1-\text{sgn}(j)\text{sgn}(j+1)}{2}} \times [\gamma_J] \times \prod_{j=J+1}^{N} [\gamma_j]^{\frac{1-\text{sgn}(j-1)\text{sgn}(j)}{2}} \tag{C.7}$$

where γ_J is the pole associated to the lightest exchanged particle, that is $m_1^2 = 2m^2(1 + \text{sgn}(J)\cos(\gamma_J))$ or $m_1^2(4m^2 - m_1^2) = 4m^2\sin^2(\gamma_J)$. In words, the optimal solution (C.7)

carefully removes CDD-poles whenever the alternating pattern between s- and t-channel poles is not observed. This immediately guarantees that all signs come out right. The optimal residue $g_1^{\max}$ can now be straightforwardly read from (5.3) and (C.7) as

$$(g_1^{\max})^2 = 16 \sin^2 \gamma_J \times (\Gamma_1^{\max})^2 \tag{C.8}$$

with

$$(\Gamma_1^{\max})^2 = \sigma(1)(-1)^{N-1}\, 2 \tan(\gamma_J) \prod_{j=1}^{J-1} \left(\frac{\sin(\gamma_J) + \sin(\gamma_j)}{\sin(\gamma_J) - \sin(\gamma_j)} \right)^{\frac{1-\sigma(j)\sigma(j+1)}{2}} \prod_{j=J+1}^{N} \left(\frac{\sin(\gamma_J) + \sin(\gamma_j)}{\sin(\gamma_J) - \sin(\gamma_j)} \right)^{\frac{1-\sigma(j)\sigma(j-1)}{2}} \tag{C.9}$$

where we are using the shorthand $\sigma(j) \equiv \text{sign}(j)$.