

This is the accepted manuscript made available via CHORUS. The article has been published as:

Evolution of segmented strings

Steven S. Gubser

Phys. Rev. D **94**, 106007 — Published 14 November 2016

DOI: [10.1103/PhysRevD.94.106007](https://doi.org/10.1103/PhysRevD.94.106007)

Evolution of segmented strings

Steven S. Gubser

Joseph Henry Laboratories, Princeton University, Princeton, NJ 08544, USA

Abstract

I explain how to evolve segmented strings in de Sitter and anti-de Sitter spaces of any dimension in terms of forward-directed null displacements. The evolution is described entirely in terms of discrete hops which do not require a continuum spacetime. Moreover, the evolution rule is purely algebraic, so it can be defined not only on ordinary real de Sitter and anti-de Sitter, but also on the rational points of the quadratic equations that define these spaces. For three-dimensional anti-de Sitter space, a simpler evolution rule is possible that descends from the Wess-Zumino-Witten equations of motion. In this case, one may replace three-dimensional anti-de Sitter space by a non-compact discrete subgroup of $SL(2, \mathbf{R})$ whose structure is related to the Pell equation. A discrete version of the BTZ black hole can be constructed as a quotient of this subgroup. This discrete black hole avoids the firewall paradox by a curious mechanism: even for large black holes, there are no points inside the horizon until one reaches the singularity.

1 Introduction

In [1, 2, 3], analyses of classical string trajectories were made with the starting point assumption that the string configurations were segmented. In flat space, this simply means that at a given moment in time, the string is piecewise linear, so the worldsheet is composed of facets each of which is a region in some affine image of $\mathbf{R}^{1,1}$. In AdS_D , the worldsheet is composed of facets each of which is a region in some AdS_2 space of AdS_D . An appealing feature of the analysis in both cases was that the evolution equations are reduced entirely to algebraic operations, first to track the movement of kinks where two string segments meet, and second to determine the outcome of a collision of two kinks.

The purpose of the present work is to simplify the formulation of segmented string motions so that it can easily be applied in AdS_D and dS_D for any D as well as flat space. I also explore the extent to which we can disregard continuum notions of spacetime and view strings as propagating on spacetimes defined in purely algebraic terms. The key is to think of the string worldsheet as a collection of spacetime events, namely the collisions of kinks. On one hand, such an approach may seem antithetical to the common understanding of why string theory works as a quantum theory of gravity—namely that the spatial extent of the string naturally avoids ultraviolet divergences. On the other hand, finite entropy across unit spacetime area suggests some sort of discreteness in spacetime, and we should therefore be interested in ways in which we can retreat from reliance on a continuous description in formulating string dynamics.

The organization of the paper is as follows. In section 2 I introduce the main methods in the simplified context of flat space, where classical string evolution is trivial because one can just add together a left-moving solution and a right-moving solution to the worldsheet wave-equation. I introduce in section 2.3 a notion of causal structure motivated by the observation that most of what one needs to know about spacetime in order to talk about segmented strings is the collection of forward-directed null displacements. Brief discussions of T-duality and worldsheet fermions follow. In section 3 I explain the evolution of segmented strings in AdS_D and dS_D , as well as in some algebraic generalizations of these spacetimes. In section 3.4 I explain how fundamental strings moving in AdS_3 supported by Neveu-Schwarz three-form flux have a simpler rule for their segmented evolution, relying on the group multiplication law of $SL(2, \mathbf{R})$. Because of this reliance, we may consistently demand that all kink collisions occur in some discrete subgroup of $SL(2, \mathbf{R})$, such as $SL(2, \mathbf{Z})$. Taking a stronger view, we can discard $SL(2, \mathbf{R})$ altogether as a real manifold and think of segmented strings propagating on a discrete spacetime. Pursuing this line of thought, I develop in

section 3.6 a discrete analog of the BTZ black hole. Concluding remarks can be found in section 4.

2 Flat spacetime

All classical results obtained in flat space must be trivial in the sense that they can be recovered from the standard treatment based on left-movers and right-movers. The putative advantage of the formalism developed here is that it generalizes to at least some curved geometries. The left-movers / right-movers approach is based on the solution to the string equations of motion:

$$X^\mu(\tau, \sigma) = Y_L^\mu(\sigma^+) + Y_R^\mu(\sigma^-) \quad \text{where} \quad \sigma^\pm = \tau \pm \sigma, \quad (1)$$

where $\sigma^\pm = \tau \pm \sigma$. The Virasoro constraints are satisfied provided the tangent vectors to $Y_L^\mu(\sigma^+)$ and $Y_R^\mu(\sigma^-)$ (considered as functions of a single variable, σ^+ or σ^-) are null. To obtain a closed string, we need X^μ to be periodic in the σ direction, say with period Σ . The periodicity conditions on Y_L^μ and Y_R^μ are

$$Y_L^\mu(\xi + \Sigma) = Y_L^\mu(\xi) + \pi\alpha' P^\mu \quad Y_R^\mu(\xi - \Sigma) = Y_R^\mu(\xi) - \pi\alpha' P^\mu, \quad (2)$$

and because Y_L^μ and Y_R^μ have null tangent vectors, P^μ must be timelike or null. Piecewise linear Y_L^μ and Y_R^μ lead to segmented string solutions. P^μ is the total momentum of the string. An elementary account of the string equations of motion and the Virasoro constraints, as well as the periodicity conditions (2), can be found in [4].

All the information about piecewise linear Y_L^μ and Y_R^μ is contained in the locations of their turning points: Let's say that within one period of the spatial components, these turning points occur in Y_L^μ at σ_i^+ for $i = 1, 2, \dots, N_L$, and in Y_R^μ at σ_i^- for $i = 1, 2, \dots, N_R$. We can extend the list of σ_i^\pm so that the index i runs across all the integers, and the sequences σ_i^\pm stretch all the way into the far past and the far future. Then these values are the entire list of times when Y_L^μ or Y_R^μ has a kink. Correspondingly, there is a grid of points on the worldsheet, conveniently labeled by i and j , where two kinks collide; and this grid is mapped by (1) to a set of points in spacetime, X_{ij}^μ . From the form of (1), we know that $X_{i+1,j}^\mu - X_{ij}^\mu$ and $X_{i,j+1}^\mu - X_{ij}^\mu$ are forward-directed null displacements. The periodicity conditions (2) imply

$$X_{i+N_L,j}^\mu = X_{ij}^\mu + \pi\alpha' P^\mu \quad X_{i,j+N_R}^\mu = X_{ij}^\mu - \pi\alpha' P^\mu. \quad (3)$$

2.1 Evolution of serrated slices

We would like to take a Hamiltonian perspective, where we consider an initial state of the string and then evolve it forward in time. To this end, it is easiest to start by considering maps from $\mathbf{Z} \times \mathbf{Z}$ to the target spacetime, $(i, j) \rightarrow X_{ij}$, without any notions of periodicity required for now. (The index μ has been suppressed for brevity.) We are still thinking of i as parametrizing the left-moving light-like direction on the worldsheet, while j parametrizes the right-moving light-like direction. Now consider traversing $\mathbf{Z} \times \mathbf{Z}$ diagonally, where each step in our traverse increments i by 1 or decrements j by 1 (but not both). Let's denote by S a definite choice of how to make this traverse: S then is an ordered subset of $\mathbf{Z} \times \mathbf{Z}$ that goes “all the way across,” from left to right, in the way we described. We will make one further stipulation: the range of values of i and j encountered within S must be unbounded in both directions. In other words, if we project S onto one of the two lightlike directions, it covers all of \mathbf{Z} . Each segment of S (say from ij to $i+1, j$ or from ij to $i, j-1$) should be thought of as null. But because we go all the way across the lattice without doubling back, S is effectively a spatial slice of the string worldsheet. We will describe it as a serrated slice.

Suppose we specify X_{ij} for all $ij \in S$, abiding by the requirement that $X_{i+1,j} - X_{ij}$ is forward-directed null whenever ij and $i+1, j$ are in S , and likewise all differences $X_{i,j+1} - X_{ij}$ in S must be forward-directed null. Then we would like to have a rule for figuring out what all other values of X_{ij} must be. More particularly, we want a rule that allows us to calculate X_{ij} at a lattice point which is, in an appropriate sense, one step forward from S . To make this precise, consider a point ij on S whose neighboring points are $i+1, j$ and $i, j+1$. We want to have a rule that specifies $X_{i+1,j+1}$ given X_{ij} , $X_{i+1,j}$, and $X_{i,j+1}$. In flat space, the obvious evolution rule is

$$X_{i+1,j+1} - X_{ij} = (X_{i+1,j} - X_{ij}) + (X_{i,j+1} - X_{ij}) \quad (4)$$

A little thought leads to the conclusion that repetitive use of (4) allows us to deduce all X_{ij} which can be reached by going forward from S : that is, incrementing i and/or j . We could run time evolution the other way by stipulating that if ij , $i-1, j$, and $i, j-1$ are all in S , then

$$X_{i-1,j-1} - X_{ij} = (X_{i-1,j} - X_{ij}) + (X_{i,j-1} - X_{ij}). \quad (5)$$

It is also easy to see that (4)-(5) together with the assumption that S goes all the way across $\mathbf{Z} \times \mathbf{Z}$ imply

$$X_{i+1,j} - X_{ij} \equiv \Delta_{Li} \quad X_{i,j+1} - X_{ij} \equiv \Delta_{Rj}, \quad (6)$$

where the j -independent quantities Δ_{Li} are null vectors, as are the i -independent quantities Δ_{Rj} . If we now build sequences Y_{Li} and Y_{Ri} so that

$$Y_{L,i+1} - Y_{Li} = \Delta_{Li} \quad Y_{R,i+1} - Y_{Ri} = \Delta_{Ri}, \quad (7)$$

with the added assumption that $X_{00} = Y_{L0} + Y_{R0}$, then we find immediately that

$$X_{ij} = Y_{Li} + Y_{Rj}. \quad (8)$$

Now let's return to periodicity requirements. The straightforward notion of periodicity on a serrated slice is to say that S is invariant under the shift $ij \rightarrow i - N_L, j + N_R$, where N_L and N_R are positive integers. We can think of a closed string as a map from a periodic S to spacetime satisfying

$$X_{i-N_L, j+N_R} = X_{ij} \quad \text{for all } i, j. \quad (9)$$

Clearly, (9) is implied by (3), but not vice versa unless we specify time evolution appropriately. If we do specify time evolution by (4)-(5), then we are led to the form (8), which if plugged into (9) can be used to conclude that

$$Y_{R, j+N_R} - Y_{Rj} = Y_{Li} - Y_{L, i-N_L} \quad \text{for all } i, j. \quad (10)$$

Denoting the common value of all possible $Y_{R, j+N_R} - Y_{Rj}$ and $Y_{Li} - Y_{L, i-N_L}$ by $\pi\alpha'P$, we immediately recover (3).

2.2 Examples

Let's look at some simple examples based on taking S to be the set of all integer pairs ij such that $i + j$ is either 0 or 1. Choose some null vector u , and set $X = 0$ for $i + j = 0$ and $X = u$ for $i + j = 1$. This assignment respects the periodicity property (9) with $N_L = N_R = 1$. Then the evolution equations imply for all ij that $X_{ij} = (i + j)u$, and we see that the classical trajectory is collapsed on a null trajectory in the direction of u . The spacetime momentum is $P = u/\pi\alpha'$.

An obvious variant of the previous construction is the "starburst string," where we set $X_{ij} = 0$ for $i + j = 0$ as before, and $X_{ij} = u_i$ for $i + j = 1$ where the u_i are an arbitrary collection of forward-directed null vectors. Let's focus on the case where the sequence of u_i repeats after N entries. Then we respect the periodicity property (9) with $N_L = N_R = N$.

The solution strategy leading to (8) can be applied, with the result that $Y_{Ri} = -Y_{L,-i}$ and $Y_{L,i+1} - Y_{Li} = u_{i+1}$. The overall momentum is $P = \frac{1}{\pi\alpha'} \sum_{i=1}^N u_i$, which typically is timelike but can be null if the u_i are all parallel.

A special case of the starburst for $N = 2$ is the familiar yo-yo string of [4]. Most simply, we can work in two dimensions and assume that $u_1 = (\ell, \ell)$ while $u_2 = (\ell, -\ell)$. Then the string expands from a point at time $t = 0$ to a string whose maximal extent at time $t = \ell$ is from $x = -\ell$ to $x = +\ell$. The string then contracts back to a point at time $t = 2\ell$. A useful check is that the string's energy, $E = 2\ell/\pi\alpha'$, may be calculated either from $P = (u_1 + u_2)/\pi\alpha'$, or by noting that the string's energy comes from its total length of 4ℓ at full extension. (Recall that the string doubles back on itself.)

So far we have had in mind that $X_{ij} \in \mathbf{R}^{d-1,1}$ for some d , but one can make different starting assumptions. For example, we could study strings on a lattice discretization of spacetime by replacing \mathbf{R} by $\mathbf{Z}a$ where a is an arbitrary length. If all the X_{ij} on an initial serrated slice of the discretized string are on the lattice, then all future and past X_{ij} will be too. Another interesting point is that if we replace \mathbf{R} by $\mathbf{Q}a$, where \mathbf{Q} denotes the rational numbers, then any *periodic* string motion (in particular any motion of a finite closed string) that starts out with $\mathbf{Q}a$ -valued coordinates will not only remain $\mathbf{Q}a$ -valued; it will in fact propagate entirely on some integer sublattice. That is, all coordinate entries will remain in $\mathbf{Z}\tilde{a}$ for some \tilde{a} commensurate with a . On the other hand, the full set of $\mathbf{Q}a$ -valued segmented string motions is dense in the set of all possible classical string motions on $\mathbf{R}^{d-1,1}$. So, in a sense, we can build an understanding all of classical string theory on flat space starting with integer-valued segmented string motions.

2.3 Causal structure

To generalize further, it is useful to start with some notion of forward-directed null displacements. Let M be a set of spacetime points, and for each $p \in M$ we associate a subset $N^+(p) \subset M$ comprising all the points of M that can be reached from p by a forward-directed null displacement, and also a subset $T^+(p) \subset M$ of all points that can be reached by a forward-directed timelike displacement. The minimal requirements that should be obeyed by these subsets are:

1. $N^+(p)$ and $T^+(p)$ are disjoint for any choice of $p \in M$.
2. $N^+(N^+(p)) \subset N^+(p) \cup T^+(p)$ for any $p \in M$.

By $N^+(Q)$ for any set $Q \subset M$, we mean the union of $N^+(q)$ for $q \in Q$. We will refer to (M, N^+, T^+) satisfying the above properties as a weak causal structure on M . Note that we can start from a weak causal structure (M, N^+, T^+) and immediately define a new one (M, N^-, T^-) by saying

$$N^-(q) = \{p: q \in N^+(p)\} \quad T^-(q) = \{p: q \in T^+(p)\}. \quad (11)$$

There are some additional properties one might desire, and which are obviously satisfied by null and timelike displacements in $\mathbf{R}^{d-1,1}$:

3. $p \in N^+(p)$ for any $p \in M$.
4. If $q \in N^+(p)$ and $q \neq p$, then $p \notin N^+(q)$.
5. $N^+(T^+(p))$, $T^+(N^+(p))$, and $T^+(T^+(p))$ are all subsets of $T^+(p)$ for any $p \in M$.

With these assumptions in place, one cannot have closed null loops or closed timelike loops, and it follows in particular that $N^+(p) \cap N^-(p) = \{p\}$ and $T^+(p) \cap T^-(p) = \emptyset$. If a weak causal structure satisfies the additional requirements 3 through 5, then we characterize it as a strong causal structure. We could also require:

6. $T^+(p) \subset \bigcup_{n=1}^{\infty} N^{+n}(p)$ for any $p \in M$, where by N^{+n} we mean the n -th iteration $N^+ \circ N^+ \circ \dots \circ N^+$.
7. If by $N_{\text{all}}(p)$ we mean $N^+(p) \cup N^-(p)$, then $M = \bigcup_{n=1}^{\infty} N_{\text{all}}^n(p)$ for any $p \in M$.

(The awkward notation $N_{\text{all}}(p)$ is due to the fact that we will later want to use $N(p)$ to mean something different in a particular example.) If a weak causal structure satisfies the additional requirements 6 and 7, then we characterize it as a complete causal structure.

With a weak causal structure in place, we can try to generalize the evolution scheme (4). It's best to start by defining “forward null triples” and “backward null triples” of spacetime points. More precisely: We define subsets F and B of the triple product $M \times M \times M$ as follows:

$$\begin{aligned} F &= \{(X_{00}, X_{10}, X_{01}): X_{00} \in N^-(X_{10}) \cap N^-(X_{01})\} \\ B &= \{(X_{11}, X_{10}, X_{01}): X_{11} \in N^+(X_{10}) \cap N^+(X_{01})\} \end{aligned} \quad (12)$$

Intuitively, an element of F —which we refer to as a forward null triple—is the smallest piece of a serrated slice which has enough information to be propagated forward one step, while an element of B (a backward null triple) has just enough information to be propagated

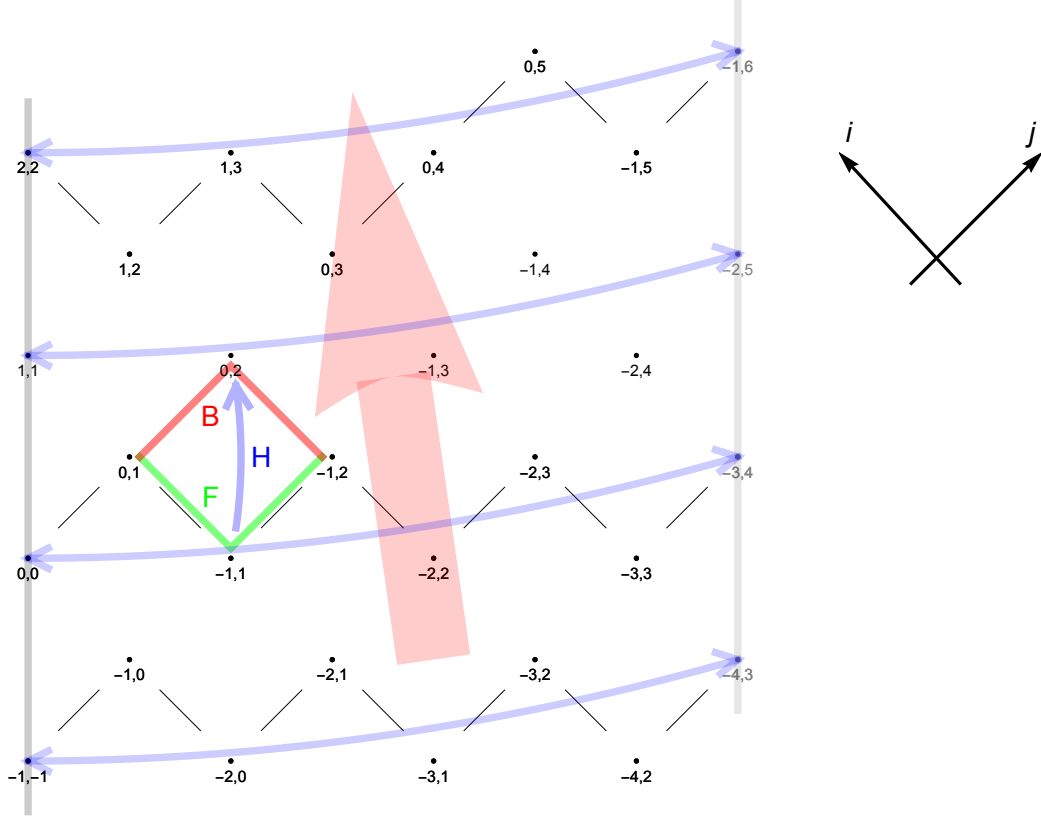


Figure 1: An example of evolution of serrated slices. Three serrated slices on the discretized worldsheet are indicated using black line segments from one worldsheet point to the next along the slice. An initial serrated slice comprises the worldsheet points $(0, 0)$, $(0, 1)$, $(-1, 1)$, $(-1, 2)$, $(-2, 2)$, $(-2, 3)$, $(-3, 3)$, and $(-3, 4)$ which is identified with $(0, 0)$. Thus $N_L = 3$ and $N_R = 4$. Similar identifications are made between $(-4, 3)$ and $(-1, -1)$, between $(-2, 5)$ and $(1, 1)$, and between $(-1, 6)$ and $(2, 2)$, as indicated by the light blue arrows. The overall direction of forward time evolution is indicated by the wide, light red arrow. The slice starting at $(-1, -1)$ is the result of seven backward evolution steps from the initial slice. The slice starting at $(2, 2)$ is the result of 12 forward evolution steps from the initial slice. The worldsheet representation of a forward null triple is shown in green, and the backward null triple that it gets mapped to by H is shown in red. More properly, H acts in spacetime to determine the spacetime image of $(0, 2)$ once the images of $(-1, 1)$, $(0, 1)$, and $(-1, 2)$ are known.

backward one step. For elements $f \in F$ and $b \in B$, we write $f \leftrightarrow b$ to mean that the X_{10} and X_{01} components of f and b are the same. Now we define an evolution scheme as a map H from a subset of F to a subset of B , such that $f \leftrightarrow H(f)$ for all f in the domain of H . For brevity, we could write

$$X_{11} = H(X_{00}, X_{10}, X_{01}) \quad (13)$$

on the understanding that what we really mean is that H maps $(X_{00}, X_{10}, X_{01}) \in F$ to $(X_{11}, X_{10}, X_{01}) \in B$. See figure 1 for an example of the action of H .

Initial data can then be specified in the form of a map from a serrated slice S to M such that $X_{i+1,j} \in N^+(X_{ij})$ whenever $i+1, j$ and ij are in S , and likewise $X_{i,j+1} \in N^+(X_{ij})$ whenever $i, j+1$ and ij are in S . As before, we evolve forward in time by identifying points $ij \in S$ such that $i+1, j$ and $i, j+1$ are also in S , and then H is applied to the triple $(X_{ij}, X_{i+1,j}, X_{i,j+1})$ to obtain $X_{i+1,j+1}$, as illustrated in figure 1. A caveat is that $(X_{ij}, X_{i+1,j}, X_{i,j+1})$ might not be in the domain of H ; in such a case, the simplest approach is to declare that the site ij simply doesn't evolve forward. By construction, $X_{i+1,j+1}$ is in the future of X_{ij} , that is $X_{i+1,j+1} \in N^+(X_{ij}) \cup T^+(X_{ij})$. If the causal structure is strong, then $X_{i'j'}$ is in the future of X_{ij} whenever $i' \geq i$ and $j' \geq j$. In other words, the evolution of the string is in the future time direction with respect to the causal structure on M .

We prefer to consider reversible evolution, namely a function H which is one-to-one, so that an inverse H^{-1} exists from the image of H to its domain. Then starting with a map from a serrated slice to M (the initial state of the string), one can develop through forward and backward time evolution a unique map from some region in $\mathbf{Z} \times \mathbf{Z}$ to M . The region can be thought of as the entire discrete string worldsheet, and it is either the entirety of $\mathbf{Z} \times \mathbf{Z}$, or it is bounded in the forward time direction by sites which cannot be evolved further forward, and in the backward time direction by sites which cannot be evolved further backward.

It is straightforward to spell out how one goes from knowledge of an evolution function H to the evolution scheme for serrated slices. Recall that a serrated slice S includes points ij for all integer values of i and all integer values of j . As a result, S includes infinitely many forward null triples and infinitely many backward null triples. An easy case to visualize is periodic S , with ij identified with $i - N_L, j + N_R$, where N_L and N_R are both necessarily positive. Then there must be at least one forward null triple and at least one backward null triple in every period. It is possible but not necessary that S is entirely comprised of overlapping forward and backward null triples. More likely, there are some forward and backward null triples separated by runs of points all along one lightlike direction on the worldsheet. In any case, a forward evolution step of a serrated slice consists of choosing

one of the forward null triples and mapping it through H to a backward null triple, as illustrated in figure 1. (In the case of a periodic serrated slice, of course we mean to map all periodic images of a given forward null triple through H at the same time.) If there are several forward null triples, it doesn't matter which one is evolved forward first. Backward evolution proceeds similarly.

2.4 T-duality

T-duality is a key feature of the theory of free strings; an authoritative account for bosonic strings is [5]. For simplicity, let's restrict to compactification on a circle of radius R . In the standard approach, spacetime is $\mathbf{R}^{d-1,1}$ modulo the relation $x \rightarrow x + 2\pi R$ where x is one of the spatial coordinates. Let's replace \mathbf{R} by $\mathbf{Z}a$ where a is a length. Then we need $2\pi R = pa$ where p is some positive integer (not necessarily a prime). Let us now set $a = 1$ for brevity; all the following formulas can be made dimensionally correct by restoring obvious powers of a .

We can present T -duality in the framework of discrete spacetime, though with a significant subtlety: We have to work in the “upstairs” picture where the compactified extra dimension is regarded as periodic rather than truly compact. The reason for this is that we will need to distinguish between a purely timelike displacement $(t, x) \rightarrow (t + 2\pi R, x)$ and a null displacement $(t, x) \rightarrow (t + 2\pi R, x + 2\pi R)$ that goes once around the circle. When spacetime is a manifold, it's OK to really identify x and $x + 2\pi R$ because one can tell whether a continuous trajectory is winding around the compact dimension. Once we have passed to a discrete subset of spacetime points, this distinction is at least much trickier—at least in the discrete topology, it is impossible.

Instead of the periodicity condition (9), we must now require only that the string comes back to itself up to some multiple of the $2\pi R$ displacement in the x direction. That is,

$$\sum_{\alpha \in S} s_{\alpha} \Delta_{\alpha}^x = pw, \quad (14)$$

where w is the winding number. In (14) and below, we use α rather than ij to label the null links in a serrated slice S , and we set $s_{\alpha} = 1$ for a link which goes from ij to $i + 1, j$ and $s_{\alpha} = -1$ for a link from ij to $i, j - 1$. By Δ_{α} we mean the forward-directed null vector that describes how link α is embedded in spacetime, and by Δ_{α}^x we mean its x component. Thinking in the upstairs picture means that Δ_{α}^x is an integer rather than an integer mod p , so that we can distinguish between $\Delta_{\alpha}^x = p$, meaning a displacement once around the circle,

and $\Delta_\alpha^x = 0$, which means no displacement in the x direction. The natural null condition on the Δ_α is the one that is obvious in the upstairs picture, i.e. Δ_α is null in $\mathbf{Z}^{d-1,1}$.

The spacetime momentum of each link is

$$P_\alpha = \frac{1}{2\pi\alpha'} \Delta_\alpha, \quad (15)$$

and so the total momentum in the x direction is

$$P^x = \frac{1}{2\pi\alpha'} \sum_\alpha \Delta_\alpha^x = \frac{n}{R} = \frac{2\pi n}{p}, \quad (16)$$

where in the second to last equality we imposed the standard quantum mechanical restriction on possible momenta in terms of an integer n . For this restriction to make sense, we must have

$$4\pi^2\alpha' = pq, \quad (17)$$

where q is also a positive integer. Then (16) can be rewritten simply as

$$\sum_{\alpha \in S} \Delta_\alpha^x = qn. \quad (18)$$

The form of (18) is almost identical to (14). To perform T-duality, we map

$$\Delta_\alpha^x \leftrightarrow s_\alpha \Delta_\alpha^x, \quad p \leftrightarrow q \quad w \leftrightarrow n. \quad (19)$$

It is easily checked that $p \leftrightarrow q$ is the usual map $R \rightarrow \alpha'/R$. From a lattice point of view, having exchanged the winding condition with the momentum condition means that the periodicity in the x direction is now $x \sim x + q$ rather than $x \sim x + p$. Switching the sign on precisely those Δ_α^x that correspond to links from ij to $i, j-1$ means switching the sign on Δ_{Rj}^x but not Δ_{Lj}^x . In other words, we are sending $Y_R \rightarrow -Y_R$ while leaving Y_L alone, which is the usual notion of T-duality. Without referring to Y_L and Y_R explicitly, we could simply note that performing T-duality on a serrated slice commutes with the standard evolution (4).

We can in principle work in the downstairs picture, where spacetime is $\mathbf{Z}^{d-2,1} \times \mathbf{Z}_p^1$ if we declare a displacement Δ to be null precisely if its projection onto the $\mathbf{Z}^{d-2,1}$ directions is null. Development along these lines seems formal in that we are no longer trying to track winding around the \mathbf{Z}_p direction; however, it is still consistent with the rules of section 2.3

¹We use \mathbf{Z}_p to mean the integers mod p .

to use the evolution scheme (4). A large part of what makes T-duality work in the form described in (19) is that the map $\Delta^x \rightarrow -\Delta^x$ is an automorphism of the additive structure on \mathbf{Z} . We can inquire whether in this downstairs setup some more general automorphism might be used. An example arises when coordinates are integer-valued and $p = q$, meaning that we are at the self-dual radius; also we require p to be odd. In the downstairs picture, (14) and (18) read

$$\sum_{\alpha \in S} s_\alpha \Delta_\alpha^x = 0 \pmod{p} \qquad \sum_{\alpha \in S} \Delta_\alpha^x = 0 \pmod{p}. \quad (20)$$

Because p is odd, these equations can be recast as

$$\sum_{\alpha \in S} \frac{1 - s_\alpha}{2} \Delta_\alpha^x = 0 \pmod{p} \qquad \sum_{\alpha \in S} \frac{1 + s_\alpha}{2} \Delta_\alpha^x = 0 \pmod{p}. \quad (21)$$

The map $\Delta_\alpha^x \rightarrow s_\alpha \Delta_\alpha^x$ means switching the sign of every Δ_α^x entering into the first equation in (21) but not the ones entering into the second equation. We could instead map all the right-moving displacements through an automorphism Σ of the additive structure on \mathbf{Z}_p : $\Delta_\alpha^x \rightarrow \Sigma(\Delta_\alpha^x)$ for those α with $s_\alpha = -1$. An example of such an automorphism is $\Delta_\alpha^x \rightarrow r \Delta_\alpha^x$ for r coprime to p , and this automorphism provides a candidate generalization of T-duality. The automorphism is usually not an involution; however, according to Carmichael's theorem, one recovers the identity map after $\lambda(p)$ iterations of Σ , where $\lambda(p)$ is the so-called reduced totient function.

2.5 Adding worldsheet fermions

The Ramond-Neveu-Schwarz formalism starts with the following equations of motion for the superstring:

$$D_{\theta^+} D_{\theta^-} \mathbf{X}^\mu = 0, \quad (22)$$

where

$$D_{\theta^\pm} = \partial_{\theta^\pm} + \theta^\pm \partial_{\sigma^\pm}. \quad (23)$$

The bosonic variables $\sigma^\pm = \tau \pm \sigma$ as before, while θ^\pm are Grassmann variables, and each superfield \mathbf{X}^μ is, *a priori*, a general function of σ^\pm and θ^\pm (assumed to be smooth). Without much effort one can show that the general solution of (22) is

$$\mathbf{X} = \mathbf{Y}_L(\sigma^+, \theta^+) + \mathbf{Y}_R(\sigma^-, \theta^-), \quad (24)$$

and the super-Virasoro constraints are

$$\partial_{\sigma^+} \mathbf{Y}_L \cdot D_{\theta^+} \mathbf{Y}_L = 0 = \partial_{\sigma^-} \mathbf{Y}_R \cdot D_{\theta^-} \mathbf{Y}_R. \quad (25)$$

If we expand

$$\mathbf{Y}_L = Y_L(\sigma^+) + i\theta^+ \psi_L(\sigma^+) \quad \mathbf{Y}_R = Y_R(\sigma^-) + i\theta^- \psi_R(\sigma^-), \quad (26)$$

then in place of the null-tangent condition on $Y_L(\sigma^+)$ and $Y_R(\sigma^-)$ that we had in the context of the bosonic string, we have from (25) the more complicated conditions

$$\partial_{\sigma^+} Y_L \cdot \partial_{\sigma^+} Y_L + \psi_L \cdot \partial_{\sigma^+} \psi_L = 0 \quad \psi_L \cdot \partial_{\sigma^+} Y_L = 0, \quad (27)$$

and similarly for the right-moving sector.

We would like to distinguish a set of simple solutions (24) to the superstring equations of motion which is analogous to the piecewise linear solutions to the bosonic string equations of motion. Let's examine the problem from the point of view of specifying a certain class of functions $\mathbf{Y}_L(\sigma^+, \theta^+)$. Analogous considerations will apply to \mathbf{Y}_R . An obvious choice is piecewise linear $Y_L(\sigma^+)$ and piecewise constant $\psi_L(\sigma^+)$, where the derivative of Y_L and the value of ψ_L change at the same values σ_i^+ of the null worldsheet coordinate. Following our earlier treatment, let's define $Y_{Li} = Y_L(\sigma_i^+)$; also, let's denote by $\psi_{L,i+1/2}$ the constant value of $\psi(\sigma^+)$ for $\sigma_i^+ < \sigma^+ < \sigma_{i+1}^+$. This notation makes clear the point of view that the bosonic data resides at the turning points while the fermionic data resides on the edges connecting them.

Plugging the ansatz just described into (27), we see from evaluating the first equation away from the special values σ_i^+ that each linear segment of Y_L must in fact be null: that is,

$$(Y_{L,i+1} - Y_{Li}) \cdot (Y_{L,i+1} - Y_{Li}) = 0. \quad (28)$$

The condition (28) plus the condition of piecewise linearity implies that the first term in the first equation of (27) vanishes everywhere.² Now consider the second term. In a neighborhood of σ_i^+ , we see that $\partial_{\sigma^+} \psi_L = (\psi_{L,i+1/2} - \psi_{L,i-1/2})\delta(\sigma^+ - \sigma_i^+)$ is proportional to a delta-function. In order for that delta function to drop out of the first equation in (27), we should require

²Precisely at $\sigma^+ = \sigma_i^+$, $\partial_{\sigma^+} Y_L$ changes its value. However, if we use for $\partial_{\sigma^+} Y_L$ its limiting value as $\sigma^+ \rightarrow \sigma_i^+$ from below, then $\partial_{\sigma^+} Y_L \cdot \partial_{\sigma^+} Y_L = 0$ at $\sigma^+ = \sigma_i^+$; likewise if we use the limiting value as $\sigma^+ \rightarrow \sigma_i^+$ from above. It is in this sense that we can say that the first term in the first equation of (27) vanishes everywhere.

that

$$\psi_{L,i-1/2} \cdot \psi_{L,i+1/2} = 0. \quad (29)$$

Note that we are effectively satisfying a stronger constraint than the first equation in (27), namely $\partial_{\sigma^+} Y_L \cdot \partial_{\sigma^+} Y_L = 0 = \psi_L \cdot \partial_{\sigma^+} \psi_L$. Passing on to the second equation in (27), we find by evaluating away from the special points σ_i^+ that

$$\psi_{L,i+1/2} \cdot (Y_{L,i+1} - Y_{Li}) = 0. \quad (30)$$

We can proceed to a full discretized worldsheet and serrated slices of it by keeping the discussion of the bosonic data exactly the same as it was in the bosonic string case, and adding the fermionic quantities to the links. To be precise, fermionic variables $\psi_{L,i+1/2,j}$ are located on the links between sites ij and $i+1, j$ for all i and j . We require

$$\psi_{L,i-1/2,j} \cdot \psi_{L,i+1/2,j} = 0 \quad (31)$$

if sites $i-1, j$, ij , and $i+1, j$ are all present on the serrated slice. Additionally, we require

$$\psi_{L,i+1/2,j} \cdot (X_{i+1,j} - X_{ij}) = 0 \quad (32)$$

if sites ij and $i+1, j$ are present on the serrated slice. One proceeds analogously with ψ_R . Then the bosonic evolution equation (4) can be supplemented with the fermionic evolution equation

$$\psi_{L,i+1/2,j+1} = \psi_{L,i+1/2,j} \quad \psi_{R,i+1,j+1/2} = \psi_{R,i,j+1/2}, \quad (33)$$

and it is possible to show that the constraints (31)-(32) are preserved by the evolution. It is clear from (33) that $\psi_{L,i+1/2,j} = \psi_{L,i+1/2}$, independent of j , and $\psi_{R,i,j+1/2} = \psi_{R,j+1/2}$, independent of i . Technically one might include in the discussion also an auxiliary field F , coming from the $\theta^+ \theta^-$ term in the expansion of \mathbf{X} and defined on the faces of the grid. This might be of use in less trivial spacetimes; but in flat space, F is immediately set to 0 through the equations of motion (22).

3 Anti-de Sitter and de Sitter spacetimes

It is natural to start the study of segmented strings in curved spacetime with an examination of AdS_2 . In section 3.1, starting from AdS_2 , we will find a simpler expression of the evolution

law studied in [1, 2, 3]. As we will explain, this evolution law can be applied naturally to anti-de Sitter and de Sitter space in any dimension, and to situations where real coordinates are replaced by coordinates valued in some other field, for instance the rationals. In section 3.4 we will explain that the evolution law found in section 3.1 is not the only one we could consider: At least for AdS_3 , there is another one naturally springing from the fact that AdS_3 is the group manifold of $SL(2, \mathbf{R})$. This group theoretic evolution law can be extended in an obvious way literally to any group; the main challenge is to decide in a sensible way what group elements generate forward-directed null displacements.

3.1 AdS_2

In [1, 2, 3] it was argued that segmented motions in AdS_3 could be composed of worldsheets with facets which are regions of AdS_2 subspaces of AdS_3 . In this section we ignore higher-dimensional considerations and focus simply on strings on the hyperboloid

$$-u^2 - v^2 + x^2 = -1 \tag{34}$$

in $\mathbf{R}^{2,1}$. Defining the dot product

$$A \cdot B = \eta_{\mu\nu} A^\mu B^\nu, \tag{35}$$

where $\eta_{\mu\nu} = \text{diag}\{-1, -1, 1\}$, we can rewrite (34) as $X \cdot X = -1$. To define time ordering on the hyperboloid, first define $z = u + iv$. Then for any two points X_1 and X_2 on the hyperboloid, form the corresponding z_1 and z_2 , and let $\phi = \arg z_2/z_1 \in (-\pi, \pi]$. We say that X_2 is at a later time than X_1 iff $\phi \in (0, \pi)$, and at an earlier time iff $\phi \in (-\pi, 0)$. Evidently, transitivity cannot be relied upon: If X_2 is later than X_1 and X_3 is later than X_2 , then it may happen that X_3 is earlier than X_1 ! This is the familiar problem of closed timelike curves on the hyperboloid, which is solved by passing to the covering space. The approach here will be simply to work on the hyperboloid itself. The presence of closed timelike curves will prevent the causal structure we define from satisfying the fifth condition in section 2.3, but we can think of eliminating them in the end by passing to the global cover. The point of treating the problem directly on the hyperboloid is that we do not actually need to forbid closed timelike curves at the level of the analysis of this section; it is enough that the hyperboloid has a weak causal structure in the sense of section 2.3.

To complete the definition of the causal structure, we first note that all the points on the hyperbola which are null separated from a given starting point X have the form $X + \Delta$

where $\Delta \cdot \Delta = 0$ and $X \cdot \Delta = 0$. Let's define $N^+(X)$ as the set of all $X + \Delta$ which are at a later time than X . Also define $T^+(X) = N^+(N^+(X)) - N^+(X)$, where minus in this context means complement. It is easily checked that (M, N^+, T^+) indeed defines a weak but complete causal structure, and that this causal structure is unchanged by the $SO(2, 1)$ isometries of the hyperboloid.

Let's now ask, what is the most natural evolution function H ? By use of an $SO(2, 1)$ isometry, we choose the position of X_{00} to be anything we please. Thus we can assume

$$X_{00} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad X_{10} = \begin{pmatrix} 1 \\ a \\ \pm a \end{pmatrix}, \quad \text{and} \quad X_{01} = \begin{pmatrix} 1 \\ b \\ \pm b \end{pmatrix}, \quad (36)$$

where a and b are positive and the two signs can be chosen independently. Suppose first we choose $-a$ and $+b$. Then $N^+(X_{10}) \cap N^+(X_{01})$ contains only a single point, namely

$$X_{11} = \frac{1}{1+ab} \begin{pmatrix} 1-ab \\ a+b \\ -a+b \end{pmatrix}. \quad (37)$$

It is not hard to check that the generalization of (36)-(37) which is invariant under the action of $SO(2, 1)$ is

$$X_{11} = X_{00} + \frac{\Delta_L + \Delta_R + (\Delta_L \cdot \Delta_R)X_{00}}{1 - \Delta_L \cdot \Delta_R/2} \quad (38)$$

$$\text{where } \Delta_L = X_{10} - X_{00} \quad \text{and} \quad \Delta_R = X_{01} - X_{00}.$$

To deduce (38), we had to assume that the two signs in (36) are opposite one another. If the two signs are the same, then Δ_L and Δ_R are parallel, so $\Delta_L \cdot \Delta_R = 0$, and (38) reduces to the abelian evolution law $X_{11} = X_{10} - X_{00} + X_{01}$. We could in principle consider variants of the evolution law, for example $X_{11} = X_{00} + \lambda(\Delta_L + \Delta_R)$, when Δ_L and Δ_R are parallel, but $\lambda \neq 1$ gives rise to problems with energy conservation, and we will not consider it further. In short, (38) defines the only reasonable evolution law for the hyperboloid. When formalized into a function $H: (X_{00}, X_{10}, X_{01}) \rightarrow (X_{11}, X_{10}, X_{01})$, one can easily demonstrate that H is a bijection from all of F to all of B . The key step in this demonstration is to show that $\Delta_L \cdot \Delta_R \leq 0$, which is a consequence of Δ_L and Δ_R both being forward-directed null displacements.

Without changing any formulas, we can consider another case, namely dS_2 , which is the same as the hyperboloid (34) but with a reversal of the identification of time and space,

so that the x coordinate is timelike. $N^+(X)$ should now be the set of all $X + \Delta$ with $\Delta \cdot \Delta = X \cdot \Delta = 0$ where the x -component $\Delta^x \geq 0$. (This is under the convention that increasing x is the future.) $T^+(X) = N^+(N^+(X)) - N^+(X)$ as before. (38) can stand as written, but to respect the causal structure of de Sitter, we must require that Δ_L and Δ_R should have positive x component. Now $\Delta_L \cdot \Delta_R \geq 0$, and we are in some danger of making $1 - \Delta_L \cdot \Delta_R/2$ zero or negative. Zero is obviously bad because $1 - \Delta_L \cdot \Delta_R/2$ occurs in the denominator of (38), but negative is bad too because X_{11} will not be in $T^+(X_{00})$. All this is in fact a feature not a bug. Physically, what's going on is that the expansion of de Sitter space drags the string apart, so that only finitely many kink collisions can occur before the string reaches the boundary in the infinite future. In other words, if $\Delta_L \cdot \Delta_R \geq 2$, then there is no X_{11} because the preceding kinks X_{10} and X_{01} are in different causal patches of de Sitter space, such that $N^+(X_{10}) \cap N^+(X_{01}) = \emptyset$.

3.2 Algebraic generalizations

With the formula (38) in hand, it is possible to generalize to higher-dimensional spacetimes, or algebraic generalizations of spacetimes. In place of $\mathbf{R}^{2,1}$ and η , consider a vector space V over a field K equipped with a quadratic form η that we use as in (35) to define a dot product. Then the equation $X \cdot X = -1$ defines the algebraic generalization of AdS_D or dS_D that we are interested in, call it M . Assume that the equations $X \cdot X = -1$ and $\Delta \cdot \Delta = 0$ have enough solutions so that we can set up interesting initial conditions on a serrated slice S on the discretized string worldsheet. Then (38) may be used unchanged. To go further, let's assume that $K \subset \mathbf{R}$ so that there is a natural notion of positive and negative, and that η has a complete set of eigenvectors over K . If two of the eigenvalues are negative and the rest are positive, then we have a generalized anti-de Sitter spacetime, while if only one is positive and the rest are negative, then we have a generalized de Sitter spacetime.

Causal structure can be defined in analogy to the the approaches for the hyperboloid and AdS_2 in section 3.1. For generalized anti-de Sitter spacetimes, we write a general vector $X \in V$ as

$$X = u\hat{u} + v\hat{v} + \sum_i x_i \hat{x}_i, \quad (39)$$

where \hat{u} and \hat{v} are the eigenvectors of η with negative eigenvalue while \hat{x}_i are the eigenvectors with positive eigenvalue. Defining $z = u + iv$, one obtains a non-transitive notion of time-ordering—which, as before, can be made transitive by passing to an appropriate covering

space. For generalized de Sitter spacetimes, we write instead

$$X = t\hat{t} + \sum_i x_i \hat{x}_i \quad (40)$$

where \hat{t} is the unique eigenvector with positive eigenvalue and \hat{x}_i are the other eigenvectors. A transitive time ordering is then available by ordering with respect to t . In both the anti-de Sitter and de Sitter cases, null displacements on the generalized hyperboloid $X \cdot X = -1$ correspond to null vectors Δ in V with X , and null displacements are forward directed if $X + \Delta$ is later than X . As in the two-dimensional case, we can use $T^+(X) = N^+(N + (X)) - N^+(X)$. By design, the causal structure is consistent with the evolution rule (38). In the anti-de Sitter case, one can again argue that (38) defines a bijection $H: F \rightarrow B$, while in the de Sitter case it defines a one-to-one map from part of F to part of B , as appropriate to the tendency of an inflating spacetime to drag a string apart into separate causal patches.

3.3 Consistency with equations of motion

If $K = \mathbf{R}$ and $\eta = \text{diag}\{-1, -1, 1, \dots, 1\}$, then the construction of the previous section gives ordinary AdS_D (up to passing to the global cover, which we will ignore in this section). If $K = \mathbf{R}$ and $\eta = \text{diag}\{1, -1, -1, \dots, -1\}$ then we have ordinary dS_D . The purpose of this section is to argue that the evolution rule (38) then defines a string motion that follows from the Nambu-Goto action.

The main part of the argument is as follows. Consider (X_{00}, X_{10}, X_{01}) in F such that the image (X_{11}, X_{10}, X_{01}) under H exists, and assume that that $X_{10} - X_{00}$ and $X_{01} - X_{00}$ are not parallel. Then X_{00} , X_{10} , and X_{01} together define a plane in \mathbf{V} that also includes the origin, and intersecting this plane with the hyperboloid $X \cdot X = -1$ gives a copy of AdS_2 or dS_2 . The point X_{11} as defined by (38) is easily seen to lie in the same plane, so there is a diamond shaped region of AdS_2 or dS_2 with corners X_{00} in the past, X_{10} and X_{01} on the sides, and X_{11} in the future. It is consistent with the local equations of motion following from the Nambu-Goto action for the string worldsheet to cover this diamond shaped region. To see this for the AdS_D case, we need only note that through an $SO(2, D-1)$ transformation we can map the AdS_2 subspace to a standard one at $x_2 = x_3 = \dots = x_{D-1} = 0$. A string stretched across this subspace is certainly a solution of the Nambu-Goto equations of motion, so at least locally within each diamond shaped region, we are solving the same equations of motion with $SO(2, D-1)$ images of the string stretched across the standard AdS_2 . A similar argument can be made in the dS_D case.

From the previous paragraph, the picture we have of segmented strings in AdS_D or dS_D (considered as real manifolds) is that they are assemblies of string segments, each of which is locally static with respect to a suitably defined global time whose choice varies segment by segment. Within each segment, the string equations of motion are satisfied by the argument of the previous paragraph. The case where $X_{10} - X_{00}$ and $X_{01} - X_{00}$ are parallel is a degeneration of the non-parallel case, where the AdS_2 or dS_2 region collapses to a null line along which there can be localized momentum and energy. It is possible to follow [6] and explicitly modify the Nambu-Goto action by adding terms that support localized null energy-momentum; or one can take the view that motions with localized energy-momentum are best treated as special limits of motions without. Taking this latter approach, we have only to ask whether the equations of motion are satisfied at each edge, or kink, where one diamond shaped region joins on to another, and whether the equations of motion are satisfied at the corners where kinks collide.

To argue that the equations of motion are satisfied at each edge, it helps to recall that these equations take the form

$$\partial_a P_\mu^a - \Gamma_{\mu\lambda}^\kappa \partial_a X^\lambda P_\kappa^a = 0 \quad \text{where} \quad P_\mu^a = -\frac{1}{2\pi\alpha'} \sqrt{-h} h^{ab} G_{\mu\nu} \partial_b X^\mu. \quad (41)$$

One matching condition at each edge is that X^μ is continuous. Usually at kinks one expects one more matching condition in the form of a first derivative condition, obtained by integrating the equation of motion over a small interval including the kink that joins the two solutions. Equivalently, we want there to be no delta function contribution, localized at the edge, coming from $\partial_a P_\mu^a$. The key point is that when this edge is lightlike on the worldsheet, there is no such delta function contribution. If present, it could only come from a second derivative of X^ν in P_μ^a . The second derivative terms, however, are proportional to $\partial_+ \partial_- X^\nu$, where the derivatives are with respect to lightlike worldsheet coordinates σ^\pm . If a delta function contribution arose in the equations of motion across an edge at $\sigma^+ = 0$, its coefficient would include a factor

$$\partial_- \left[X^\nu \Big|_{\sigma^+=\epsilon} - X^\nu \Big|_{\sigma^+=-\epsilon} \right], \quad (42)$$

where ϵ is a very small positive number. The quantity in square brackets vanishes as $\epsilon \rightarrow 0$ on account of the continuity condition, and the piecewise linearity of X^ν as a function of σ^\pm then implies that the whole expression (42) also vanishes. So indeed we see that no additional condition is imposed at the first derivative level.

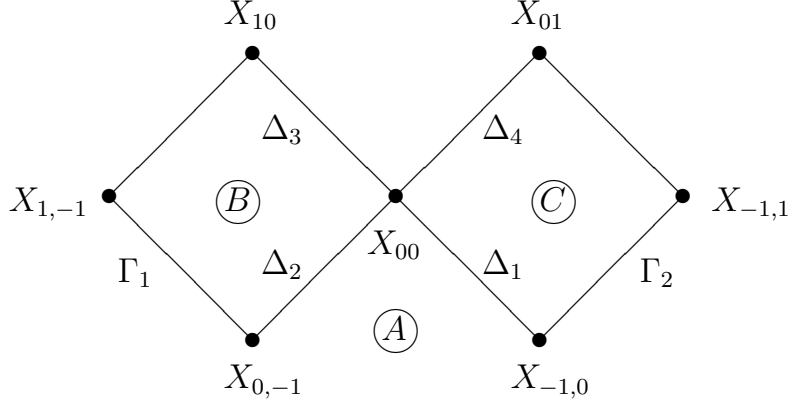


Figure 2: String evolution close to X_{00} .

Likewise we claim that there are no additional conditions imposed at a corner where two kinks collide. The quickest way to argue this is to consider a local description where we focus in on a small neighborhood around a particular corner, call it the 00 corner. Then the displacements $\Delta_1 = X_{00} - X_{-1,0}$, $\Delta_2 = X_{00} - X_{0,-1}$, $\Delta_3 = X_{10} - X_{00}$, and $\Delta_4 = X_{01} - X_{00}$ can be viewed as vectors in the tangent space. Locally the evolution is just like strings in flat space, so if there is a constraint on the four Δ_i at the corner, it should be present in flat space as well. But in flat space, these are manifestly independent quantities: In the notation of section 2.1, $\Delta_1 = \Delta_{L,-1}$, $\Delta_2 = \Delta_{R,-1}$, $\Delta_3 = \Delta_{L0}$, and $\Delta_4 = \Delta_{R0}$, and we recall that the Δ_{Li} and Δ_{Rj} are all independent of one another.

To readers familiar with [1, 2], it may seem odd that no conditions are imposed at corners where kinks collide, since in those papers much of the information about string propagation was in fact encoded at the corners. To see that there is no contradiction, let Δ_1 and Δ_2 be specified, and consider how we gather enough information to determine Δ_3 and Δ_4 . In the approach of the current paper, we must know the triples $(X_{0,-1}, X_{1,-1}, X_{00})$ and $(X_{-1,0}, X_{00}, X_{-1,1})$, and from them we deduce X_{10} and X_{01} using (38). In other words, we must know Γ_1 and Γ_2 as shown in figure 2 in addition to Δ_1 and Δ_2 ; then we can get X_{10} and X_{01} from (38) and from them Δ_3 and Δ_4 . In the approach of [2] (and [1] is similar), we require enough local information around the null segment from $X_{-1,0}$ to X_{00} in order to determine the AdS_2 regions labeled A and C in figure 2, and likewise we require enough information around the null segment from $X_{1,-1}$ to X_{00} in order to determine regions B and C . Then we have enough information to determine the directions of Δ_3 and Δ_4 , though not the magnitudes; but that is OK since information from edges further to the left and right will combine with the directions of Δ_3 and Δ_4 to determine X_{10} and X_{01} .

3.4 AdS_3 and a discretized WZW model

When we replace \mathbf{R} by a field $K \subset \mathbf{R}$, say the rationals, we are giving up most of the points in AdS_D but retaining a dense subset. An obvious goal is to go further and pass to some discrete set of points distributed more or less uniformly across it, but nowhere densely, similar to the lattice discretizations of flat spacetime. There is an obvious obstacle to such a step. Suppose on some AdS_2 subspace we are allowed to include initial null displacements $(1, 0, 0) \rightarrow (1, na, -na)$ and/or $(1, 0, 0) \rightarrow (1, nb, nb)$ where n is any positive integer and a and b are fixed and positive. If we replace $a \rightarrow na$ and $b \rightarrow nb$ in (36)-(37), we see that for n large, there is an accumulation point for X_{11} at $(-1, 0, 0)$. This has a perfectly physical interpretation, most easily stated when $a = b$: If we travel a long way toward the boundary of AdS_2 as measured by the affine parameter along a null direction, and then turn around and come back the same distance, we will wind up at the same point in space, but approximately half way around the timelike direction on the AdS_2 hyperboloid. This problem has nothing to do with the closed timelike curves on the hyperboloid (34); it is instead related to a feature of anti-de Sitter space, namely that there is a definite light-crossing time. Mathematically, the problem we are seeing has to do with the presence of $1 - \Delta_L \cdot \Delta_R/2$ in the denominator of (38). When division is part of the evolution law, it is hard to lay down a lattice structure and stay within it starting from generic initial conditions. So, while it's easy to restrict values of coordinates from \mathbf{R} down to a field $K \subset \mathbf{R}$, it is much harder to restrict from \mathbf{R} to a ring some of whose non-zero elements do not admit multiplicative inverses.

On the other hand, AdS_3 is the $SL(2, \mathbf{R})$ group manifold, and one might naturally expect that segmented strings can propagate so that their kink collisions lie in a discrete subgroup like $SL(2, \mathbf{Z})$. This is indeed possible, but not with the evolution law (38). A different algebraic evolution law is possible on AdS_3 ; in fact it follows from changing the action from the usual Nambu-Goto action to the WZW action considered for example in [7]. Then the general solution to the classical equations of motion takes the form

$$X(\tau, \sigma) = Y_L(\sigma^+)Y_R(\sigma^-) \quad (43)$$

where X , Y_L , and Y_R all take values in $SL(2, \mathbf{R})$, and for closed strings we demand

$$Y_L(\sigma^+ + \Sigma) = Y_L(\sigma^+)M \quad Y_R(\sigma^- - \Sigma) = M^{-1}Y_R(\sigma^-) \quad (44)$$

for some element $M \in SL(2, \mathbf{R})$. If we stipulate that there is no internal CFT, then the

Virasoro conditions state that the left- and right-moving currents,

$$J_L \equiv (\partial_+ X)X^{-1} \quad J_R \equiv X^{-1}(\partial_- X), \quad (45)$$

are null elements of the $SL(2, \mathbf{R})$ Lie algebra.³ (The $SL(2, \mathbf{R})$ Lie algebra is composed of traceless 2×2 matrices. The null elements are those whose determinant also vanishes.)

The conditions (43)-(45) provide a close analogy to the flat space conditions (1)-(2) and the condition that the tangent vectors to flat space curves $Y_L^\mu(\sigma^+)$ and $Y_R^\mu(\sigma^-)$ are null. So it is not surprising that we can proceed to formulate segmented strings on the $SL(2, \mathbf{R})$ group manifold in fairly close analogy to the flat space case. In particular, the discrete analogs of $Y_L(\sigma^+)$ and $Y_R(\sigma^-)$ are products of forward-directed null elements of $SL(2, \mathbf{R})$, by which we mean elements $g \in SL(2, \mathbf{R})$ with $\text{tr } g = 2$ and $\text{tr } \epsilon g < 0$ where $\epsilon = i\sigma_2$ is the 2×2 anti-symmetric matrix. Thus, forward-directed null elements are a special subset of parabolic elements of $SL(2, \mathbf{R})$. It can be shown that this notion of forward-directed null displacement on $SL(2, \mathbf{R})$ coincides with our earlier one if we write

$$X = \begin{pmatrix} u+x & v+y \\ -v+y & u-x \end{pmatrix} = u\mathbf{1} + vi\sigma_2 + x\sigma_3 + y\sigma_1. \quad (46)$$

Indeed, if g is a forward-directed null element of $SL(2, \mathbf{R})$, then starting at X and moving to gX , or more generally $g^n X$ for any $n > 0$, corresponds to moving along a null geodesic (the same geodesic for any n) in the natural metric on $SL(2, \mathbf{R}) = AdS_3$ inherited from $\mathbf{R}^{2,2}$, and in the positive time direction in the sense of increasing the phase of $u + iv$. The same is true if one starts at X and moves to Xg . Moreover, any forward-directed null displacement can be written *either* as left-multiplication or right-multiplication by a forward-directed null element of $SL(2, \mathbf{R})$. That is, if X_1 is displaced from X_0 on AdS_3 by a forward-directed null displacement, then $X_1 = g_L X_0$ and $X_1 = X_0 g_R$ where g_L and g_R are forward-directed null elements of $SL(2, \mathbf{R})$.

Consider now a serrated slice S . The rules for initial data are the same as we have previously used: $X_{i+1,j} - X_{ij}$ must be forward-directed null whenever ij and $i+1, j$ are in S and likewise all differences $X_{i,j+1} - X_{ij}$ in S must be forward-directed null. To evolve forward in time, in place of (4) or (38), we use

$$X_{11} = g_L X_{00} g_R \quad \text{where} \quad X_{10} = g_L X_{00} \quad \text{and} \quad X_{01} = X_{00} g_R. \quad (47)$$

³Usually there would be a factor of the level k multiplying J_L and J_R , but I suppress it here because it doesn't affect any of the considerations in this section.

Note that g_L and g_R are forward-directed null elements of $SL(2, \mathbf{R})$. Evidently, we can simplify (47) to

$$X_{11} = X_{10} X_{00}^{-1} X_{01}, \quad (48)$$

which is a natural generalization of (4) to $SL(2, \mathbf{R})$. To see that (48) is not the same as (38), it's worth carrying the example (36)-(37) over to AdS_3 . We will quote points on AdS_3 in terms of vectors in $\mathbf{R}^{2,2}$ to facilitate comparison with (38). If

$$X_{00} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad X_{10} = \begin{pmatrix} 1 \\ a \\ -a \\ 0 \end{pmatrix}, \quad \text{and} \quad X_{01} = \begin{pmatrix} 1 \\ b \\ b \\ 0 \end{pmatrix}, \quad (49)$$

with a and b positive, then according to the group theoretic evolution scheme (48) we have

$$X_{11} = \begin{pmatrix} 1 - 2ab \\ a + b \\ -a + b \\ -2ab \end{pmatrix}; \quad (50)$$

whereas according to the minimally coupled evolution scheme (38), choosing $-a$ and $+b$ in (36) we would get the result (37), carried over to AdS_3 by adding a 0 entry as the fourth row of the vector. This is indeed different from (50), and physically the reason is that the WZW model describes a constant field strength $H_3 = dB_2$ which pulls on the string as it moves through AdS_3 .

The evolution rule (48) could be employed with $SL(2, \mathbf{R})$ replaced by any group. If we replace $SL(2, \mathbf{R})$ with a subgroup, for example $SL(2, \mathbf{Z})$, then the causal structure discussed above (46) can be used to determine what null displacements mean within the subgroup. In a general setting it is not so simple to give a natural notion of a causal structure. An obvious class of examples to start with is groups of the form $\mathbf{R}^{d-1,1} \times H$ where H is some general group which we intend as a replacement for transverse spatial directions (compact or otherwise) and we are considering $\mathbf{R}^{d-1,1}$ as a group under addition. The $\mathbf{R}^{d-1,1}$ factor is equipped with a metric which we write as

$$s^2 = -t^2 + \vec{x}^2 \quad (51)$$

Of course, we should be prepared to generalize, for example by replacing $\mathbf{R}^{d-1,1}$ by a module

over a ring and the metric by some appropriate quadratic form. A trivial but valid version of causal structure is to say that N^+ is the direct product of ordinary forward-directed null vectors in $\mathbf{R}^{d-1,1}$ and all of H . An example of this trivial causal structure came up at the end of the discussion of T-duality in section 2.4. More interesting causal structures could be constructed given a spatial metric on H .

3.5 Minkowski space and the Pell equation

It was remarked in section 3.1 that replacing anti-de Sitter space, AdS_D , with a level set of a quadratic form on a more general vector space than $\mathbf{R}^{D-1,2}$ provides an obvious way to generalize the evolution scheme outlined in (38) for classical strings. In order to set up a particular example, let's first review the connection between the Pell equation and Minkowski space. I will do that here in more detail than strictly necessary since the subject is appealing on its own. In two-dimensional Minkowski space, $\mathbf{R}^{1,1}$, the spacetime interval is

$$s^2 = -c^2 t^2 + x^2, \quad (52)$$

where c^2 is a positive constant. Let's require that

$$d \equiv c^2 \quad (53)$$

is a square-free rational number: that is, $d = a/b$, where a and b are relatively prime and neither is divisible by the square of any integer. We disallow $d = 1$. The field extension $\mathbf{Q}[\sqrt{d}]$ comprises all numbers of the form⁴

$$z = x + ct \quad (54)$$

where x and y are rational, and we commit ourselves to always use the positive square root $c = \sqrt{d} > 0$. Evidently, $\mathbf{Q}[\sqrt{d}]$ is a two-dimensional vector space over the rationals, with basis vectors $1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $c = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. To visualize why $\mathbf{R}^{1,1}$ —or, more precisely, $\mathbf{Q}^{1,1}$ —naturally equates to the field extension $\mathbf{Q}[\sqrt{d}]$, think of the map $\begin{pmatrix} t \\ x \end{pmatrix} \rightarrow x + ct$ as taking a point in $\mathbf{Q}^{1,1}$ to an irrational point on the right-moving part of the light-cone. This is a bijection under our assumption that d is a square free rational number, and we can form

⁴Usually one sees field extensions by the square root of an integer, say $\mathbf{Q}[\sqrt{ab}]$ instead of $\mathbf{Q}[\sqrt{a/b}]$. The numbers in these two field extensions are exactly the same since $\sqrt{ab} = b\sqrt{a/b}$; however, the norm defined below is different depending on whether $d = a/b$ or $d = ab$.

another bijection $\begin{pmatrix} t \\ x \end{pmatrix} \rightarrow x - ct$ that we can visualize as projecting onto the left-moving part of the light-cone.⁵

If we write the obvious equality $c(x + ct) = td + cx$ in vector form, it becomes

$$c \begin{pmatrix} x \\ t \end{pmatrix} = \begin{pmatrix} td \\ x \end{pmatrix}. \quad (55)$$

So, when thinking of the action of $\mathbf{Q}[\sqrt{d}]$ on itself by left multiplication, we can represent 1 and c as 2×2 matrices:

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad c = \begin{pmatrix} 0 & d \\ 1 & 0 \end{pmatrix}. \quad (56)$$

Thus $\mathbf{Q}[\sqrt{d}]$ can be represented as all matrices of the form

$$z = x + ct = \begin{pmatrix} x & td \\ t & x \end{pmatrix}. \quad (57)$$

The trace and determinant of the 2×2 matrix are useful constructions:

$$\text{tr}(x + ct) = 2x \quad N(x + ct) = -c^2 t^2 + x^2. \quad (58)$$

The notation $N(x + ct)$ is used because it is the standard mathematical notation for the norm of a field extension.⁶ We may treat the complex numbers in the same way, setting $d = -1$; then one has the usual expression $z = x + y\sqrt{-1}$ and the norm $N(z) = x^2 + y^2$. Pursuing the analogy with complex numbers further, we define

$$\bar{z} = x - ct. \quad (59)$$

Then $N(z) = z\bar{z}$ is equally valid for $d = -1$ or for positive square-free rational numbers d (and, incidentally, for negative square-free rational d).

⁵A related fact is that for square-free $d = c^2$, neither $\mathbf{Q}^{1,1}$ nor $\mathbf{Z}^{1,1}$ has any non-trivial points on the lightcone with respect to the Minkowski metric (52): $N(z) = 0$ implies $z = 0$. By the two-square, three-square, and four-square theorems of Fermat, Legendre, and Lagrange, this situation generalizes, with certain restrictions, to two and three spatial dimensions, but not more. For example, as a consequence of Legendre's theorem, if $d = c^2 = 7$ (or any positive integer congruent to 7 mod 8), then $\mathbf{Z}^{3,1}$ has no points on the lightcone defined by the interval $s^2 = -c^2 t^2 + \bar{x}^2$. Could this observation have anything to do with the observed dimensionality of spacetime, or is it only numerology?

⁶In more general field extensions, $N(\lambda z) = \lambda^n N(z)$ for λ in the base field, where the positive integer n may be different from 2. For example, $n = 4$ for $\mathbf{Q}[\sqrt[4]{2}]$ because the representation matrices one finds are 4×4 .

Rotations in \mathbf{C} (the field of complex numbers) are best represented as $z \rightarrow rz$ where r is on the unit circle: $r\bar{r} = 1$. Generalizing to $\mathbf{Q}[\sqrt{d}]$, we see that maps $z \rightarrow rz$ preserve $N(z)$ precisely when $N(r) = 1$. These norm-preserving linear maps are boosts. Explicitly, if we write $r = p + q\sqrt{d}$, then

$$N(r) = p^2 - dq^2 = 1. \quad (60)$$

The relation (60) is known as Pell's equation. A well-known fact (see for example p. 175 of [8]) is that all solutions of (60) with p and q rational take the form

$$r = \pm \frac{1 + \mu/\sqrt{d}}{1 - \mu/\sqrt{d}} \quad \text{for} \quad \mu \in \mathbf{Q}. \quad (61)$$

We can describe μ as the “half-velocity” of the boost, in the following sense. A boost by velocity v can be described as

$$z \rightarrow \sqrt{\frac{1 + v/c}{1 - v/c}} z \quad \bar{z} \rightarrow \sqrt{\frac{1 - v/c}{1 + v/c}} \bar{z}. \quad (62)$$

(Recall that z and \bar{z} are the light-cone coordinates $x + ct$ and $x - ct$.) If we define μ to be the velocity of a boost which must be iterated twice to obtain (62), then we can rewrite (62) as

$$z \rightarrow \frac{1 + \mu/c}{1 - \mu/c} z \quad \bar{z} \rightarrow \frac{1 - \mu/c}{1 + \mu/c} \bar{z}. \quad (63)$$

Comparing (63) with (61) we see that up to the overall sign on r , μ appearing in (61) indeed has the interpretation of the velocity of a boost which, if iterated twice, leads to the map $z \rightarrow rz$.

The original intent of the Pell equation was to restrict d , p , and q to integers, with d still remaining square-free. Let's make this restriction now. Let's also assume for the moment that p and q are in addition positive. Then there is a fundamental solution $r_1 = p_1 + q_1\sqrt{d}$ of the Pell equation with the property that all other non-trivial solutions (i.e. solutions other than $r = 1$) take the form $r_n = r_1^n$ for $n = 2, 3, 4, \dots$. Furthermore, p_1/q_1 is a convergent of the continued fraction expansion of \sqrt{d} . If we drop the restriction that p and q are positive, then the full set of solutions of $N(r) = 1$ is given by

$$r = \pm r_1^n \quad \text{for} \quad n \in \mathbf{Z}. \quad (64)$$

This is a special case of Dirichlet's unit theorem, whose statement in this case is that the

group of unit elements in $\mathbf{Z}[\sqrt{d}]$ is generated (up to the sign) by a single element. (Generally, a unit of a ring is a member of the ring whose multiplicative inverse is also in the ring.)

Thus $\mathbf{Z}[\sqrt{d}]$ is a lattice version of $\mathbf{R}^{1,1}$ (technically, it is a free \mathbf{Z} -module spanned by two basis vectors) which is preserved by a group of boosts generated by a “fundamental boost” r_1 . If we write a general boost as

$$\begin{pmatrix} x' \\ t' \end{pmatrix} = r \begin{pmatrix} x \\ t \end{pmatrix} \quad \text{where} \quad r = \gamma \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{\gamma v}{c^2} \begin{pmatrix} 0 & c^2 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \gamma = \frac{1}{\sqrt{1 - v^2/c^2}}, \quad (65)$$

then r_1 has the special property that its γ and $\gamma v/c^2$ are integers, and any lattice-preserving boost whose 2×2 matrix form has integer entries can be reached, up to a sign, by applying the boost r_1 or its inverse some finite number of times. Evidently, this situation is more interesting than rotations of $\mathbf{Z}[\sqrt{-1}]$, which are all obtained by multiplying by a power of $\sqrt{-1}$.

3.6 A discrete version of the BTZ black hole

The BTZ black hole [9] is locally AdS_3 , but with identification of points related by a discrete group isomorphic to \mathbf{Z} . Specifically, we start with AdS_3 defined as the locus of points in $\mathbf{R}^{2,2}$ satisfying

$$-c^2(u^2 + v^2) + x^2 + y^2 = -\ell^2. \quad (66)$$

We form

$$z = u + \frac{x}{c} \quad \bar{z} = u - \frac{x}{c} \quad w = v + \frac{y}{c} \quad \bar{w} = v - \frac{y}{c}. \quad (67)$$

It will soon become clear why (67) works better in the current context than the definition (54) which we used in the previous section. Away from extremality we construct the BTZ black hole by identifying

$$z \sim e^{2\pi r_+/\ell} z \quad \bar{z} \sim e^{-2\pi r_+/\ell} \bar{z} \quad w \sim e^{2\pi r_-/\ell} w \quad \bar{w} \sim e^{-2\pi r_-/\ell} \bar{w}, \quad (68)$$

where r_{\pm} are the outer and inner horizon radii, satisfying $0 \leq r_- < r_+$. The construction just outlined is developed more systematically in [10], where the resulting black hole

thermodynamics is also found:

$$\begin{aligned} M &= \frac{c_{\text{Vir}}}{12} \frac{\hbar}{c\ell} \frac{r_+^2 + r_-^2}{\ell^2} & J &= \frac{c_{\text{Vir}}}{12} \hbar \frac{r_+ r_-}{\ell^2} \\ S &= \frac{\pi c_{\text{Vir}}}{3} \frac{r_+}{\ell} & T_+ \pm T_- &= \frac{\hbar c}{\pi \ell^2} r_{\pm} \end{aligned} \quad (69)$$

where

$$c_{\text{Vir}} = \frac{c^3}{\hbar} \frac{3\ell}{2G_3} \quad (70)$$

is the central charge c_{Vir} of the Virasoro algebra of asymptotic symmetry generators.

To pass to a discrete version of BTZ, consider $\mathbf{Z}^{2,2}$ with coordinates (u, v, x, y) equipped with the quadratic form

$$s^2 = -u^2 - v^2 + \frac{1}{c^2}(x^2 + y^2) \quad \text{where} \quad \frac{1}{c^2} = d \quad (71)$$

and d is a square-free integer. Setting $c^2 = 1/d$ contrasts with the discussion of the previous section but is a better choice here, as we will see momentarily. (The interval (71) differs by a factor of c^2 from the left hand side of (66) because I wanted the metric (71) to be expressible wholly in terms of ring operations on \mathbf{Z} .) In terms of the variables z and w defined in (67), the interval (71) can be expressed compactly as $s^2 = -N(z) - N(w)$. Next define

$$X = \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix}, \quad (72)$$

and note that $s^2 = -\det X$. If $\ell = c$, the equation (66) can be rewritten as

$$\det X = N(z) + N(w) = 1. \quad (73)$$

Matrices of the form (72) (with integer u, v, x , and y) satisfying (73) form a group under matrix multiplication. This group can be described as $SL(2, \mathbf{Z})_d$; or it can be thought of as $U(2)_d$ in the sense that its elements satisfy $X^{-1} = X^\dagger$ where \dagger means matrix transposition combined with conjugation, $z \rightarrow \bar{z}$ and $w \rightarrow \bar{w}$. If we allow $d = 1$, then we do not quite recover $SL(2, \mathbf{Z})$; instead, $SL(2, \mathbf{Z})_1$ is the subgroup of $SL(2, \mathbf{Z})$ whose elements $X = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ have $A - D$ and $B - C$ both even. In any case, $SL(2, \mathbf{Z})_d$ is a subgroup of $SL(2, \mathbf{R})$, so it inherits the causal structure introduced above (46).

To connect smoothly with the previous section, one might have expected the definition $z = x + u\sqrt{d}$ in place of $z = u + x\sqrt{d}$. The trouble is, if we set $z = x + u\sqrt{d}$, then the condition

$\det X = 1$ that is preserved under group multiplication defines AdS_3 with signature $+-$, where x and y combine to give the compact timelike direction while u and v are spacelike. Swapping (x, y) for (u, v) , we arrive back at the construction of the previous paragraphs, starting with (66) and (67). In other words, $c^2 = 1/d$ is actually the only choice we can make consistent with the group structure we are interested in.

The identification (68) that defines the BTZ black hole can be rewritten as

$$X \sim \tilde{X} \equiv \begin{pmatrix} r_z z & r_w w \\ -r_w^{-1} \bar{w} & r_z^{-1} \bar{z} \end{pmatrix} \quad (74)$$

where

$$r_z = p_z + q_z \sqrt{d} = r_1^{n_z} \quad \text{and} \quad r_w = p_w + q_w \sqrt{d} = r_1^{n_w} \quad (75)$$

are fixed elements of $\mathbf{Z}[\sqrt{d}]$ with $N(r_z) = N(r_w) = 1$. We also require that the rational part of r_z and r_w is positive. This last requirement is the moral equivalent of stipulating that these boosts should be continuously connected to the identity. It implies that if r_1 is the fundamental solution of $N(r) = 1$ over $\mathbf{Z}[\sqrt{d}]$, then $r_z = r_1^{n_z}$ and $r_w = r_1^{n_w}$ for n_z and n_w both integer. For consistency with the identifications (68) we further require non-negative n_z and n_w . Setting $d = 1$ at this point is a problem, because then no choice of r_z and r_w other than $r_z = r_w = 1$ is allowed. To see this, note that there are elements of $SL(2, \mathbf{Z})_1$ in which any desired one of z, \bar{z}, w , or \bar{w} equals 1. Then for \tilde{X} to be also in $SL(2, \mathbf{Z})_1$, we need r_z, r_z^{-1}, r_w , and r_w^{-1} all to be integers; since we restrict r_z and r_w to be non-negative, they must therefore be 1. We avoid this overly restrictive situation by insisting that $d > 1$ is a square-free integer.

Comparing the identifications (68) and (74), we see that they match provided

$$\frac{r_+}{\ell} = \frac{1}{2\pi} \log r_z \quad \frac{r_-}{\ell} = \frac{1}{2\pi} \log r_w. \quad (76)$$

Plugging into the formula (69) for the entropy and using (75), one finds

$$S = \frac{c_{\text{Vir}}}{6} n_z \log r_1. \quad (77)$$

This formula is striking not only because the permitted values of S are evenly spaced, but because the spacing depends on d in an irregular fashion. For $d = 3$, one finds $r_1 = 2 + \sqrt{3}$, and $\log r_1 \approx 1.32$; but for $d = 61$ (famously), one finds the numerically large result $r_1 = 1766319049 + 226153980\sqrt{61}$, such that $\log r_1 \approx 22.0$. Physically, a large $\log r_1$ means that

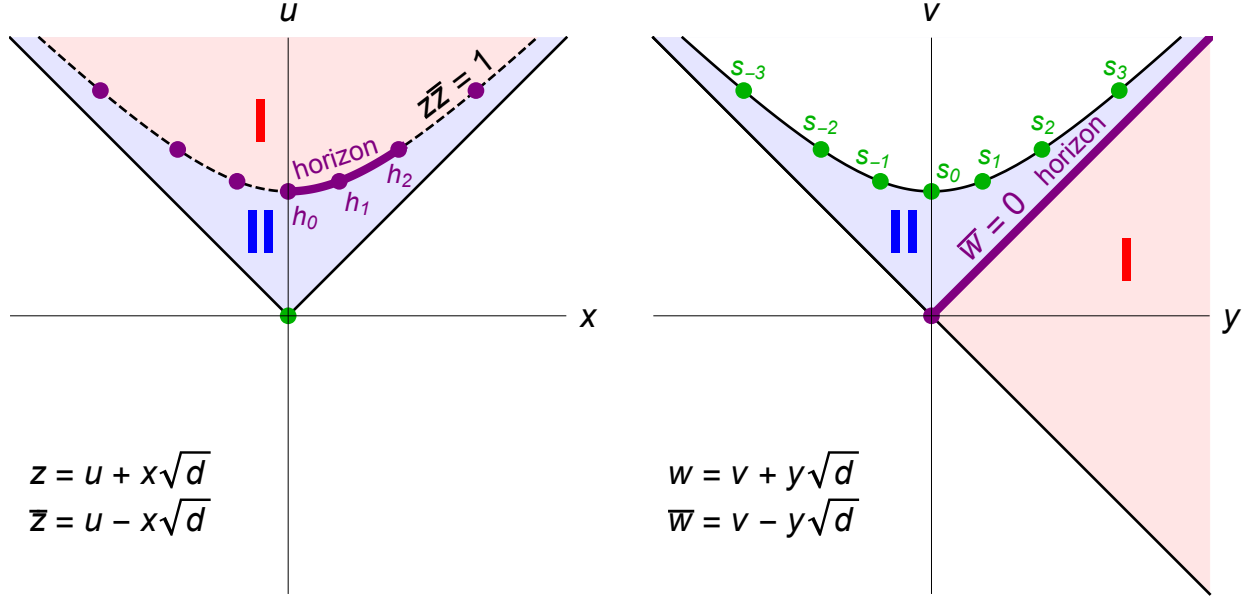


Figure 3: The outer region **I** and inner region **II** of the BTZ black hole, projected onto the z - \bar{z} and w - \bar{w} planes. The points s_n of (79) are shown as green dots, while the points h_n of (81) are shown as dots along the horizon $N(z) \equiv z\bar{z} = 1$. The particular case shown here has an identification with $n_z = 2$ and $n_w = 0$, corresponding to an uncharged black hole. There are two distinct points in $SL(2, \mathbf{Z})_d$ on the horizon, namely h_0 and h_1 , and there are no points in $SL(2, \mathbf{Z})_d$ in the interior of region **II**.

the smallest black hole (above the zero-size black hole associated with the RR vacuum) is very large compared to the AdS_3 radius. Specifically, the horizon area (really a length) scales as $A \propto \ell n_z \log r_1$.

One can also obtain expressions for the dimensionful quantities M , J , and T_{\pm} ; in so doing, it should be noted that the requirement $\ell = c$, which follows from setting $\det X = 1$, is essentially a choice of units. In particular, we are *not* restricting attention to Planck-scale AdS_3 because the Planck scale remains free even after we choose units. The dual statement in field theory is that we have not fixed c_{vir} .

The full geometry of the BTZ construction is somewhat complicated, but the main features that make it a black hole can be understood by focusing on two regions, as illustrated in figure 3:

$$\begin{aligned}
 \text{I (outside):} \quad & N(z) > 1, \quad z + \bar{z} > 0, \quad w > 0, \quad \bar{w} < 0 \\
 \text{II (inside):} \quad & 1 > N(z) > 0, \quad z + \bar{z} > 0, \quad w > 0, \quad \bar{w} > 0.
 \end{aligned} \tag{78}$$

For now, let's regard these regions as defined in the real manifold AdS_3 , with real-valued global coordinates z, \bar{z}, w , and \bar{w} as introduced in (67). The boundary between **I** and **II** is the black hole horizon, in the sense that if $X \in \mathbf{I}$ and $Y \in \mathbf{II}$, then it is possible to have $Y \in N^+(X)$, but one can never have $X \in N^+(Y)$. This is with the notion of causality inherited from $AdS_3 = SL(2, \mathbf{R})$, namely that $N^+(X)$ is all elements of $SL(2, \mathbf{R})$ of the form pX where p is a forward-directed null element of $SL(2, \mathbf{R})_d$. The horizon is a two-dimensional submanifold of $SL(2, \mathbf{R})$ with one direction spacelike and the other null. The identification $X \rightarrow \tilde{X}$ preserves regions **I** and **II** as well as the horizon between them. Region **I** becomes an outside region of the BTZ black hole, while region **II** is an inside region.

Region **II** is empty when we restrict to $SL(2, \mathbf{Z})_d$. The easiest way to see this is that $N(z) = u^2 - dx^2$ is an integer, so it is impossible to satisfy the constraint $1 > N(z) > 0$ of (78). Suppose we expand region **II** to include points with $N(z) = 0$, which immediately implies $N(w) = 1$. There is an infinite sequence of such points, namely

$$s_n = \begin{pmatrix} 0 & r_1^n \\ -r_1^{-n} & 0 \end{pmatrix} \quad \text{for } n \in \mathbf{Z}. \quad (79)$$

The identification map $X \rightarrow \tilde{X}$ as defined in (74) maps each s_n to itself if $n_w = 0$ (the uncharged BTZ black hole), whereas $s_n \rightarrow s_{n+n_w}$ for the charged case. Let's focus on the uncharged case for simplicity. In the geometry of the uncharged BTZ black hole as a real manifold, the locus of fixed points, $z = \bar{z} = 0$ with $N(w) = 1$, is the singularity inside the black hole. Using the explicit coordinate maps of [10], this locus of points corresponds to $r = 0$ in the usual parametrization of the uncharged BTZ black hole as

$$ds^2 = -\frac{r^2 - r_+^2}{\ell^2} c^2 dt^2 + \frac{\ell^2}{r^2 - r_+^2} dr^2 + r^2 d\phi^2. \quad (80)$$

Thus the points s_n are all at $r = 0$. In summary: in the discrete construction, there is nothing inside the uncharged black hole except the singularity, which resolves into an infinite sequence of fixed points s_n of the isometry (74).

A similar discussion can be carried out for the horizon. With the straightforward definition of the horizon between **I** and **II** as the set of points with $N(z) = 1$, $z + \bar{z} > 0$, $\bar{w} = 0$, and $w > 0$, we find immediately that there are no points in $SL(2, \mathbf{Z})_d$ which are on the horizon. That is because $\bar{w} = 0$ implies $w = 0$ in $\mathbf{Z}[\sqrt{d}]$. If we relax the definition slightly

by requiring $w \geq 0$ instead of $w > 0$, then the points in $SL(2, \mathbf{Z})_d$ on the horizon are

$$h_n \equiv \begin{pmatrix} r_1^n & 0 \\ 0 & r_1^{-n} \end{pmatrix} \quad \text{for } n \in \mathbf{Z}. \quad (81)$$

The identification map (74) sends $h_n \rightarrow h_{n+n_z}$, so n_z of these points remain distinct. (This is equally true for the uncharged and charged cases.) Thus we have an appealing picture of a horizon comprising n_z points and carrying an entropy proportional to n_z . It is as if there are

$$W_1 \equiv r_1^{c_{\text{vir}}/6} \quad (82)$$

distinct microstates per horizon point, with the choice of microstate at each point being independent of one another. W_1 as defined in (82) has no reason to be an integer, or even rational; but possibly there is some sense in which it may be used to capture an asymptotic count of the number of states.

The BTZ construction applied to $SL(2, \mathbf{Z})_d$ amounts to a sort of lattice in AdS_3 (namely the points of $SL(2, \mathbf{Z})_d$ itself) which respects the identifications (74) used to construct BTZ black holes of particular masses and angular momentum. But in light of the discussion leading to (48), it should be something more, namely a non-compact discrete set on which strings can propagate consistently, with causal properties inherited from the BTZ black hole over the reals. To see that string propagation on our quotient space is consistent, first note that we can express (74) as

$$X \sim \tilde{X} = g_L X g_R^{-1} \quad \text{where} \quad g_L = \begin{pmatrix} r_L & 0 \\ 0 & r_L^{-1} \end{pmatrix}, \quad g_R = \begin{pmatrix} r_R & 0 \\ 0 & r_R^{-1} \end{pmatrix}. \quad (83)$$

Here $r_L = r_1^{(n_z+n_w)/2}$ and $r_R = r_1^{(-n_z+n_w)/2}$. Plugging into (48), one sees that the relation $X_{11} = X_{10} X_{00}^{-1} X_{01}$ is preserved when passing from X to \tilde{X} . A peculiar point about this argument is that if n_z and n_w have different parity, then r_L and r_R will not be elements of $\mathbf{Z}[\sqrt{d}]$, even though r_z and r_w are. In the current context, this doesn't present a problem since we can pass to a larger ring or field of numbers as an intermediate step in writing the relation (83).

There are two technical points which one must confront in order to frame a fully consistent discussion of segmented strings in the BTZ background:

1. Sometimes, $SL(2, \mathbf{Z})_d$ has no null elements at all, meaning no elements which are null separated from $\mathbf{1}$ in the standard causal structure on $SL(2, \mathbf{R})$. This happens for

$d = 3$, and in fact for any prime d congruent to 3 mod 4. But for many other values of d , for example primes congruent to 1 mod 4, there are null separations in $SL(2, \mathbf{Z})_d$, essentially as a consequence of Fermat’s two-square theorem. In short, we must choose d appropriately to ensure that $SL(2, \mathbf{Z})_d$ has null elements, since our description of segmented strings uses null group elements in an essential way.

2. With an identification of points (74) which are spacelike separated (which is the case in regions **I** and **II**), there isn’t a simple way to distinguish between a translation all the way around the spatial circle and no displacement at all. This is similar to the obstacle discussed in connection with T-duality in section 2.4. As in that case, the most straightforward resolution is to insist on working in the “upstairs” picture, namely the whole of regions **I** and **II** before identification, and restrict attention to configurations which are mapped to themselves by the identification (74). Then one may use the causal structure inherited from $SL(2, \mathbf{R})$, and we have a clear notion on a given serrated slice of whether a string wraps the spatial circle. The evolution law (48) evolves allowed serrated slices to other allowed serrated slices.

Unsurprisingly, the general tendency is for strings to fall into the black hole, i.e. to fail to remain entirely in region **I**. Recall however that region **II** is empty, in the sense that there are no points in $SL(2, \mathbf{Z})_d$ which lie in region **II**, unless we expand it to include the singularity at $N(z) = 0$, comprising for the uncharged case the infinite sequence of points s_n defined in (79). So when a segmented string described in terms of points in $SL(2, \mathbf{Z})_d$ falls into a black hole, the first point that characterizes its presence inside the black hole is already at the singularity. This is even true for uncharged black holes that are large compared to both the Planck scale and the AdS_3 radius. As a result, we avoid the firewall paradox of [11]. The paradox involves entanglement between degrees of freedom inside and outside the black hole, deemed to be necessary to account for Hawking radiation but unacceptable because an outgoing quantum of Hawking radiation cannot simultaneously be entangled with its infalling partner and all previous Hawking radiation. The firewall “resolution” proposed in [11] is to posit the existence of a destructive firewall at the black hole horizon that incinerates any infalling object. In the context of the $SL(2, \mathbf{Z})_d$ version of BTZ black holes, we don’t need a firewall at the horizon because falling into the black hole means hitting the singularity, not just eventually, but at the end of the very first discrete timestep one takes into the black hole.

4 Conclusions

The animating principle of classical segmented strings is to understand the simplest scrap of string which can be propagated forward in time in some specified background. The focus here was on triples of spacetime points (X_{00}, X_{10}, X_{01}) where X_{10} and X_{01} are separated from X_{00} by forward-directed null displacements. Such forward null triples, together with backward null triples (X_{11}, X_{10}, X_{01}) to which they can evolve, are a convenient starting point because we do not need to specify any momentum degrees of freedom, as we would if we considered a segment of string (X_a, X_b) whose endpoints are spacelike separated. Localized momentum is incorporated naturally into the framework of forward and backward null triples. In a forward null triple (X_{00}, X_{10}, X_{01}) , if $X_{10} = X_{01}$, then it means that the scrap of string under consideration is collapsed to a point and moving at the speed of light from X_{00} to X_{10} . A spacetime metric can be used to define null displacements, forward null triples, and backward null triples. However, it is not clear that the reverse is necessarily true. Thus we can perhaps think of forward and backward null triples as capturing coarser information about spacetime.

Joining forward null triples and backward null triples together, we can construct a serrated slice of the worldsheet, and then the local rules for evolving forward null triples into backward null triples becomes a complete (classical) evolution scheme for serrated slices, which accommodates localized momentum (as at least the previous approach of [2] did not). We need not refer explicitly to a spacetime metric to discuss the evolution of segmented strings; instead we only need an appropriate collection of forward-directed null displacements, together with a map from forward null triples to backward null triples. This map can be thought of as a discrete replacement for the second order differential equation for the embedding of the string into spacetime. The elementary analysis of worldsheet fermions in section 2.5 suggests that for the RNS superstring, forward and backward null triples must be augmented by information about the fermions on the legs between sites where bosonic data is located. Possibly further augmentations could be considered that would allow for more general segmented string motions. Another interesting possibility is to pass from real-valued coordinates to coordinates valued in a finite field. Then segmented string evolution can be regarded as a finite-state cellular automaton.

We saw in sections 3.2 and 3.4 that there can be more than one sensible way to propagate forward null triples into backward null triples. In AdS_3 , the most obvious propagation rule, following from the Nambu-Goto action, results in the algebraic evolution law (38). Because this rule involves division, it is difficult to restrict the set of allowed points in the null triples

to lie in a discrete subset of AdS_3 . A different evolution law, (48), involves only group multiplication in $SL(2, \mathbf{R})$, which is the same as AdS_3 (ignoring the issue of global covers). It is therefore easy to restrict (48) to subgroups of $SL(2, \mathbf{R})$. An interesting example is the group $SL(2, \mathbf{Z})_d$, as defined around (72), for square-free integers d . A striking corollary is that we can define a discrete version of BTZ black holes in terms of an identification of elements of $SL(2, \mathbf{Z})_d$ by the action of a map (74) defined in terms of boosts in $\mathbf{R}^{2,2}$. The entropy of the black hole, as computed by embedding it in the real manifold AdS_3 , is then proportional to the number of points of the discrete group on the horizon. However, the constant of proportionality involves the logarithm of the fundamental solution of the Pell equation $p^2 - dq^2 = 1$, and this logarithm varies with d in an irregular fashion. Segmented strings falling into the discrete version of uncharged BTZ black holes make immediate contact with the singularity, obviating any need for a firewall along the lines of [11]. However, I cannot claim to have given a full interpretation of string propagation that includes singular points, like the s_n in the case of the uncharged black hole. Trajectories that continue into regions other than **I** and **II** can eventually explore closed timelike curves, and presumably some appropriate prescription of passing to a covering space will be needed.

If AdS_3 is supported by Neveu-Schwarz three-form flux, then the Nambu-Goto action characterizes D1-brane propagation, whereas a Wess-Zumino-Witten model characterizes fundamental string propagation. Bound states of D1-branes and fundamental strings should lead to a whole family of evolution laws that interpolate between (38) and (48).

It is natural to inquire whether a higher dimensional version of serrated slices can be constructed. For M2-branes, instead of forward null triples, one might try to use forward null quadruples (A, B, C, D) where B , C , and D are all displaced from A in forward null directions: That is, they are all elements of $N^+(A)$. Backward null quadruples can be defined similarly. However, already in flat space one can easily see that the M2-brane worldvolume cannot be filled with unions of the convex hulls of forward and backward null quadruples. In addition one needs tetrahedra formed as the convex hulls of quadruples (A, B, C, D) where C and D are both in $N^+(A) \cap N^+(B)$. I hope to report further on such constructions in the future.

Another natural avenue of thought is to ask how we may quantize segmented strings. It is not natural to impose some notion of infinitesimal time evolution, so a Hamiltonian per se is probably not called for. Instead, perhaps unitary transformations can be defined which advance a serrated slice locally forward. An important first step toward making unitary evolution explicit is to formulate the correct Hilbert space to describe the quantum states of

a segmented string.

Note added

When this paper was complete, we received [12], which includes an equation equivalent to (38) for the AdS_3 case and points out a dual relation to the evolution of normal vectors that formed the basis of [1].

Acknowledgments

I thank P. Landweber, S. Parikh, P. Witaszczyk, and A. Yarom for useful discussions. This work was supported in part by the Department of Energy under Grant No. DE-FG02-91ER40671.

References

- [1] D. Vegh, “The broken string in anti-de Sitter space,” 1508.06637.
- [2] N. Callebaut, S. S. Gubser, A. Samberg, and C. Toldo, “Segmented Strings in AdS_3 ,” 1508.07311.
- [3] D. Vegh, “Colliding waves on a string in AdS_3 ,” 1509.05033.
- [4] X. Artru, “Classical String Phenomenology. 1. How Strings Work,” *Phys. Rept.* **97** (1983) 147.
- [5] J. Polchinski, *String theory. Vol. 1: An introduction to the bosonic string*. Cambridge University Press, 2007.
- [6] A. Ficnar and S. S. Gubser, “Finite momentum at string endpoints,” *Phys.Rev.* **D89** (2014), no. 2 026002, 1306.6648.
- [7] J. M. Maldacena and H. Ooguri, “Strings in $AdS(3)$ and $SL(2,R)$ WZW model 1.: The Spectrum,” *J. Math. Phys.* **42** (2001) 2929–2960, hep-th/0001053.
- [8] S. J. Miller and R. Takloo-Bighash, *An Invitation to Modern Number Theory*. Princeton University Press, Princeton, NJ, 2006.

- [9] M. Banados, C. Teitelboim, and J. Zanelli, “The Black hole in three-dimensional space-time,” *Phys. Rev. Lett.* **69** (1992) 1849–1851, [hep-th/9204099](#).
- [10] M. Banados, M. Henneaux, C. Teitelboim, and J. Zanelli, “Geometry of the (2+1) black hole,” *Phys. Rev.* **D48** (1993) 1506–1525, [gr-qc/9302012](#).
- [11] A. Almheiri, D. Marolf, J. Polchinski, and J. Sully, “Black Holes: Complementarity or Firewalls?,” *JHEP* **02** (2013) 062, [1207.3123](#).
- [12] D. Vegh, “Segmented strings from a different angle,” [1601.07571](#).