ADV. THEOR. MATH. PHYS.
Volume 27, Number 4, 1159–1190, 2023

Topology change with Morse functions progress on the Borde–Sorkin conjecture

Leonardo García-Heveling

Topology change is considered to be a necessary feature of quantum gravity by some authors, and impossible by others. One of the main arguments against it is that spacetimes with changing spatial topology have bad causal properties. Borde and Sorkin proposed a way to avoid this dilemma by considering topology changing spacetimes constructed from Morse functions, where the metric is allowed to vanish at isolated points. They conjectured that these Morse spacetimes are causally continuous (hence quite well behaved), as long as the index of the Morse points is different from 1 and $n-1$. In this paper, we prove a special case of this conjecture. We also argue, heuristically, that the original conjecture is actually false, and formulate a refined version of it.

1 Introduction **1160**

2 Proof of Theorem 1 **1164**

3 Towards a full resolution of the Borde–Sorkin Conjecture **1176**

4 Conclusions **1181**

Appendix A Neighborhood spacetimes **1183**

Appendix B Causal continuity **1185**

References **1188**

1. Introduction

In General Relativity, by solving the initial value problem for Einstein's Equations, one finds the time evolution of the spacetime metric. In this picture, the topology of the constant time slices always remains the same. More precisely, the maximal globally hyperbolic developement of some initial data V is, on the level of topology (Geroch [15]) and differentiable structure (Bernal and Sánchez [4, 5]), simply $V \times \mathbb{R}$. The question remains whether this rigid product structure is desirable, or whether we should allow the topology to change over time as well. There are several instances where topology change is desirable. The dynamical creation of a wormhole, for example, is necessarily a topology changing process, as it involves attaching a handle to space. Already in 1957, Wheeler argued that quantum fluctuations of spacetime should modify the topology [31]. Moreover, in certain approaches to Quantum Gravity, instead of considering the deterministic evolution of a spatial slice under the Einstein Equations, the idea is to find the transition probability between two spatial slices V_1, V_2. This is done by computing a path-integral over all cobordisms betweeen V_1 and V_2; that is, manifolds $\mathcal{M}$ with boundary $\partial\mathcal{M} = V_1 \sqcup V_2$. These cobordisms also have to be equipped with a Lorentzian (or, in Euclidean Quantum Gravity, Riemannian) metric, and possibly satisfy some additional conditions. It is then natural to think that the transition probability between V_1 and V_2 might be non-zero also when V_1 and V_2 are not homeomorphic, as long as appropriate cobordisms exist. We refer to [9, 29] for further discussion on the role of topology change within Quantum Gravity.

Which properties should a Lorentzian cobordism satisfy, in order to consider it physically reasonable? In this paper, we will focus on the case of compact cobordisms (i.e. spatially closed universes). Geroch [16] showed that any non-trivial (meaning with $V_1 \neq V_2$) compact Lorentzian cobordism must contain closed timelike curves. Because of this, the only way to have topology change without time travel is by allowing the spacetime metric to degenerate at certain singular points [21, 33]. Notice that the case of non-compact time slices is less restricted, with examples of topology change without closed timelike curves and without singular points obtained by multiple authors (see Sánchez [27] for the most recent ones and for the overview of previous work on p. 16).

One interpretation is to consider the singular points as naked singularities, and not as points in the spacetime manifold. In this paper, however, we do the opposite: we consider the singular points as points in the spacetime, where nothing special happens, except that, in some sense, the topology

change happens there. Our point of view implies that the spacetime metric is not Lorentzian everywhere, but this is not so bad, since the metric is not a physical observable in itself. Indeed, we will show that the causal and length structures can be satisfactorily generalized to include the singular points (some work on the curvature has also been done [23]). Still, allowing degenerate metrics does introduce many new questions and problems (irrespective of our point of view on the singular points). Already in the 1980s, Anderson and DeWitt showed that on their famous "trousers spacetime", quantum fields create infinite bursts of energy in the presence of singular points [1]. This result was later refined and confirmed in Manogue et al. [24] and Buck et al. [7]. The aim of subsequent work was to impose additional conditions that avoid such pathologies.

A concrete and very useful construction of degenerate Lorentzian metrics on cobordisms was given by Yodzis [33, 34] using Morse functions. This idea was further developed by Sorkin [28] and collaborators [6, 10–13, 23], under the name of *Morse geometries*. We continue this approach in the present paper.

The construction of a Morse geometry is as follows. For simplicity, all objects are assumed to be smooth. Let $\mathcal{M}$ be a compact cobordism of dimension n, h a Riemannian metric, $\zeta > 1$ a constant, and f a Morse function. Recall that a smooth function $f\colon \mathcal{M} \to \mathbb{R}$ is called a Morse function if all its critical points (where $df = 0$) are non-degenerate (not to be confused with the (non-)degeneracy of the spacetime metric). This is equivalent to saying that around each critical point, there exist coordinates x^i such that

$$f = \frac{1}{2}\sum_i a_i (x^i)^2 \,, \tag{1}$$

where $a_i \neq 0$ are constants. It follows, in particular, that the critical points are isolated. The *index* of a critical point is defined as the number of negative a_i (see [25] for more details). Louko and Sorkin define the *Morse metric* corresponding to $(\mathcal{M}, h, f, \zeta)$ by

$$g = \|df\|_h^2\, h - \zeta df \otimes df \,, \tag{2}$$

which, in coordinates, gives

$$g_{\mu\nu} := (h^{\alpha\beta}\partial_\alpha f \partial_\beta f) h_{\mu\nu} - \zeta \partial_\mu f \partial_\nu f \,. \tag{3}$$

Let $M = \mathcal{M} \setminus (\partial\mathcal{M} \cup \{p_i\}_i)$, where $\{p_i\}_i$ is the set of critical points of f. By abuse of notation, we call the restriction of g to M also g. Since

df vanishes only at the critical points, g is Lorentzian on M, and the pair (M, g) forms a spacetime in the usual sense. It is clear from (2) that f is a time function on (M, g), when choosing the time orientation to be given by the gradient vector field $\nabla^\alpha f := h^{\alpha\beta}\partial_\beta f$. Following the nomenclature of Borde et al. [6], we call (M, g) a *Morse spacetime* and $(\mathcal{M}, h, f, \zeta)$ a Morse geometry[1]. It is known that for any pair of connected 3-manifolds, there exists a Morse geometry interpolating between them [10, 13].

According to the following two conjectures, the infinite bursts of energy found by Anderson and DeWitt are only present on certain topology-changing spacetimes, but not on others.

Conjecture 1 (Sorkin). *A quantum field propagating on a Morse geometry $(\mathcal{M}, h, f, \zeta)$ has an unphysical singular behaviour if and only if the Morse spacetime (M, g) is causally discontinuous.*

Conjecture 2 (Borde–Sorkin). *The Morse spacetime (M, g) induced by a Morse geometry $(\mathcal{M}, h, f, \zeta)$ is causally continuous if and only if all critical points of f have index different from 1 and $n-1$.*

Recall that causal continuity roughly means that the past and future $I^\pm(p)$ varies continuously with the point p (see Appendix B for details). Causal continuity was introduced in Hawking and Sachs [17] as a minimal requirement for a spacetime to be physically reasonable, for reasons unrelated to quantum theory. Thus Conjecture 2 is also interesting beyond the obvious link to Conjecture 1. Conjecture 1 is mentioned as early as 1990 in Sorkin [28], while the earliest reference for Conjecture 2 is an indirect source (Dowker and Garcia [11] from 1998). Both of them remain open to this day. Conjecture 2 has seen important progress trough the works of Borde, Dowker, Garcia, Sorkin and Surya [6, 12, 13]. In this paper, we contribute to this effort by showing the following special case:

Theorem 1. *Let $(\mathcal{M}, h, f, \zeta)$ be a Morse geometry of dimension n with a single critical point $p_c \in \mathcal{M}$. Suppose that p_c has index $\lambda \neq 0, 1, n-1, n$, and is contained in a coordinate neighborhood where*

$$f = \frac{1}{2}\sum_i a_i (x^i)^2, \quad h = \sum_i (dx^i)^2\,, \tag{4}$$

[1] In [13], the inverted nomenclature is used.

for some real constants $a_i \neq 0$ *satisfying*

$$\frac{1}{\zeta} < \left|\frac{a_i}{a_j}\right| < \zeta \quad \textit{and} \quad \frac{5}{8} \leq \left|\frac{a_i}{a_j}\right| \leq \frac{8}{5} \quad \textit{for all } i\,, j\,. \tag{5}$$

Then the corresponding Morse spacetime (M, g) *is causally continuous.*

In Section 3 (Proposition 1) we will show that one can *always* find coordinates where (4) holds, up to adding a perturbation to h which vanishes at p_c. Moreover, we conjecture that the first part of (5) is sharp, in the sense that its violation leads to causal discontinuity (see Example 2 and Conjecture 3).

Combining Theorem 1 with previous results by other authors (fleshed out below), we can summarize the current progress on Conjecture 2 in the next theorem.

Theorem 2. *Let* $(\mathcal{M}, h, f, \zeta)$ *be a Morse geometry of dimension* $n \geq 2$*, and* (M, g) *the corresponding Morse spacetime. Assume* f *has a single critical point for each critical value.*

(i) If f *has at least one critical point of index* $\lambda = 1, n-1$*, then* (M, g) *is causally discontinuous.*

(ii) If each critical point of f *has index* $\lambda = 0, n$*, or has any index* $\lambda \neq 1, n-1$ *and is contained in a neighborhood as in Theorem 1, then* (M, g) *is causally continuous.*

The case $\lambda = 0, n$ in part (ii) was solved in Borde et al. [6], along with the special case of Theorem 1 corresponding to $|a_i| = 1$ for all i. Part (i) of Theorem 2 was shown in Dowker et al. [13]. Also in [13], it was shown that the case of multiple critical points (as long as there is only one per critical value) reduces to the case of a single critical point: the Morse spacetime is causally continuous if and only if every critical point has a causally continuous neighborhood.

The proof of Theorem 1 is contained in Section 2. In Section 3, we discuss the necessity of our assumptions, and possible generalizations of our proof. Based on this discussion, we propose a modified version of the Borde–Sorkin conjecture in Section 4, where we also give concluding remarks. Appendix A contains results of [6] that we need in our proofs, and Appendix B gives some background on causal continuity.

2. Proof of Theorem 1

Before starting, let us give a brief outline of the proof. Recall from the introduction that the case of $|a_i| = 1$ for all i has already been solved in Borde et al. [6], a result that we build upon. While in the case of $|a_i| = 1$ there are a lot of symmetries, which allow for good coordinate choices (see Appendix A), this is no longer true in the general case. Our strategy is to extend the causal structure from (M, g) to $\mathcal{M}$, in a way that preserves its most important properties: openness of the chronological relation I^+, the push-up principle $J^+(I^+(q)) = I^+(J^+(q)) = I^+(q)$, and the properties of limits of causal curves. Once these properties are proven, causal continuity follows almost immediately, as it would in Minkowski spacetime.

The most difficult to establish, out of the three properties, is the openness of the chronological relation. We do this in Subsection 2.2. The argument is based on reduction to the $|a_i| = 1$ case. Once openness of the chronological relation is established, the rest of the proof can be performed without the need to make any coordinate choices whatsoever, and without further use of the assumptions (4) and (5). This second part of the proof is contained in Subsection 2.3. It requires heavy use of the limit curve theorems of Minguzzi [26].

2.1. Notation and first steps

Throughout this section, we assume that $(\mathcal{M}, h, f, \zeta)$ is a Morse geometry of dimension $n \geq 4$, with a single critical point p_c of index $\lambda \neq 0, 1, n-1, n$ lying in the interior of $\mathcal{M}$. As in the introduction, we write $M := \mathcal{M} \setminus (\partial\mathcal{M} \cup \{p_c\})$, and g denotes the metric (2), which is Lorentzian on M and degenerate-Lorentzian on $\mathcal{M}$. We do not use Einstein's summation convention: all sums are written out, but without making explicit the summation limits. Hence $\sum_i$ means $\sum_{i=1}^n$, and similarly $\max_i$ means $\max_{i=1,\dots,n}$. For convenience, we refer to the hypothesis of Theorem 1 as Condition 1.

Condition 1. *There exists an open set $\mathcal{U} \subseteq \mathcal{M}$ with $p_c \in \mathcal{U}$, an open ball $\mathcal{B} \in \mathbb{R}^n$ around the origin, and a coordinate chart $\phi : \mathcal{U} \to \mathcal{B}$ of $\mathcal{M}$ such that $\phi(p_c) = 0$ and*

$$f \circ \phi^{-1} = \frac{1}{2} \sum_i a_i (x^i)^2, \quad h \circ \phi^{-1} = \sum_i (dx^i)^2 \,, \tag{6}$$

for some real constants $a_i \neq 0$. Moreover, setting

$$\zeta_c := \max_{i,j} \left| \frac{a_i}{a_j} \right| ,$$

we have

$$\zeta_c \leq \frac{8}{5} \quad \textit{and} \quad \zeta_c < \zeta \,. \tag{7}$$

The value of ζ_c does not depend on the choice of coordinates, as long they satisfy (6) (we give a detailed argument for this in Section 3.1). We will usually suppress the coordinate map ϕ from the notation, and whenever we write x^i, it will refer to the coordinates as given by Condition 1. In these coordinates, the metric (2) takes the form

$$g = \sum_{i,j} \left(a_i x^i dx^j\right)^2 - \zeta \left(\sum_k a_k x^k dx^k \right)^2 . \tag{8}$$

An important tool in our proof will be to reduce some computations to the case of *isotropic neighborhoods* as studied in Borde et al. [6] (see also Appendix A). These are metrics where Condition 1 is satisfied, but with the stronger requirement that $|a_i| = 1$ for all i. The following lemma gives us such an isotropic neighborhood metric g_{iso} with lightcones narrower than those of g.

Lemma 1. *Suppose that Condition 1 is satisfied, and consider on $\mathcal{U}$ the linear change of coordinates*

$$x^i \mapsto y^i := \sqrt{|a_i|}\zeta_c^{\frac{1}{4}} x^i \,.$$

Then the tensor given in these new coordinates by

$$g_{\text{iso}} := \sum_{i,j} \left(y^i dy^j\right)^2 - \frac{\zeta}{\zeta_c} \left(\sum_k \operatorname{sign}(a_k) y^k dy^k \right)^2 \tag{9}$$

is a Lorentzian metric on $\mathcal{U} \setminus \{p_c\}$, with lightcones narrower that those of g.

Proof. By (7), $\zeta/\zeta_c > 1$, and hence g_{iso} is a *neighborhood metric* in the sense of Borde et al. (see Appendix A). In particular, g_{iso} is Lorentzian everywhere

except at the origin. In the y-coordinates, the metric (8) takes the form

$$g = \frac{1}{\zeta_c} \sum_{i,j} \left| \frac{a_i}{a_j} \right| \left(y^i dy^j\right)^2 - \frac{\zeta}{\zeta_c} \left(\sum_k \operatorname{sign}(a_k) y^k dy^k \right)^2 .$$

For a vector $V \in T\mathcal{U}$ (with components V^i in the y-coordinates), this means

$$\begin{aligned} g(V,V) &= \frac{1}{\zeta_c} \sum_{i,j} \left| \frac{a_i}{a_j} \right| \left(y^i V^j\right)^2 - \frac{\zeta}{\zeta_c} \left(\sum_k \operatorname{sign}(a_k) y^k V^k \right)^2 \\ &\leq \sum_{i,j} \left(y^i V^j\right)^2 - \frac{\zeta}{\zeta_c} \left(\sum_k \operatorname{sign}(a_k) y^k V^k \right)^2 \\ &= g_{\mathrm{iso}}(V,V) \end{aligned}$$

Hence if $g_{\mathrm{iso}}(V,V) \leq 0$, then also $g(V,V) \leq 0$. In other words, g_{iso} has narrower lightcones than g. □

Another crucial element in the proof will be the extension of the causal relation from M to $\mathcal{M}$. Let $\gamma\colon I \to \mathcal{M}$ be a locally Lipschitz curve. By continuity, $\gamma^{-1}(M)$ is open in $\mathbb{R}$ and hence can be written as a union of intervals $\bigcup_i I_i$. If $\gamma\colon I_i \to M$ is future-directed (f.d.) causal for every i, then we say that $\gamma\colon I \to \mathcal{M}$ is future-directed causal, and analogously for timelike and/or past-directed curves. This gives rise to a notion of futures and pasts $I^{\pm}_{\mathcal{M}}(p), J^{\pm}_{\mathcal{M}}(p)$ in $\mathcal{M}$. Additionally, for $p \neq p_c$, we denote by $I^{\pm}_M(p), J^{\pm}_M(p)$ the usual past and future sets in the spacetime (M,g).

Given a f.d. causal curve $\gamma\colon I \to \mathcal{M}$ as above, since f is a time function on (M,g), $f \circ \gamma$ is strictly increasing on $\gamma^{-1}(M)$. Therefore f is increasing along all of γ and can only cross p_c once, i.e. $\gamma^{-1}(p_c)$ is empty, a point, or a closed connected interval in $\mathbb{R}$. The following lemma tells us that furthermore no causal curve can be imprisoned in a neighborhood of p_c (see [3, pp. 61-62] for the definition of non-imprisonment on non-degenerate spacetimes).

Lemma 2. *Let $\gamma\colon (a,b) \to M$ be a causal curve which is future inextendible in M. Then either $\lim_{s\to b} \gamma(s) = p_c$ or γ runs into $\partial\mathcal{M}$.*

Proof. Because f is a time function on (M,g), (M,g) is strongly causal. Then, by [3, Prop. 3.13], given any compact set $K \subseteq M$, there exists $\delta > 0$ such that $\gamma(s) \notin K$ for all $s \in (b-\delta, b)$ (in other words, γ must leave K and never enter it again). Let $U \subseteq \mathcal{M}$ be any open set (not necessarily connected)

containing p_c and $\partial\mathcal{M}$. Then we can choose $K = \mathcal{M} \setminus U$, and hence there exists $\delta > 0$ such that $\gamma(s) \in U$ for all $s \in (b-\delta, b)$. Since U was arbitrary, we are done. □

2.2. Openness of chronological pasts and futures

In this subsection, we characterize the past $I^-_{\mathcal{M}}(p_c)$ of the critical point p_c. Every statement has a time-reversed analogue for the future $I^+_{\mathcal{M}}(p_c)$ (which we do not write out explicitly). The following condition is very important. It states that if from a point $q \in M$ we can reach p_c via timelike curves, then we can also reach a whole neighborhood of p_c. This is a well-known fact for spacetimes without singular points.

Condition 2 (Openness of $I^+_{\mathcal{M}}$). *For every $q \in I^{\pm}_{\mathcal{M}}(p_c)$ there exists a neighborhood U of p_c such that $U \setminus \{p_c\} \subseteq I^{\mp}_M(q)$.*

An important consequence of Condition 2 is that the chronological relation is not altered by removing p_c.

Lemma 3 ($I^+_M = I^+_{\mathcal{M}} \cap M$). *Suppose Condition 2 is satisfied. Then, for every $p \in M$ it holds that $I^+_M(p) = I^+_{\mathcal{M}}(p) \cap M$.*

Proof. The inclusion $I^+_M(p) \subset I^+_{\mathcal{M}}(p) \cap M$ is trivial. It remains to show th eother direction. Let $q \in I^+_M(p)$, and let $\gamma : [a, b] \to \mathcal{M}$ be a timelike curve from p to q. If γ avoids p_c, there is nothing to prove. Hence suppose that $\gamma(c) = p_c$ for some c. Then $p_c \in I^+_{\mathcal{M}}(p) \cap I^-_{\mathcal{M}}(q)$, so by Condition 2 we can find neighborhoods U, V of p_c such that $U \setminus \{p_c\} \subseteq I^+_M(p)$ and $V \setminus \{p_c\} \subseteq I^-_M(q)$. But then we can find a point $z \in U \cap V \setminus \{p_c\}$, and a timelike curve $\sigma : [a, b] \to M$ from p to q passing though z. □

In Appendix A (Lemma 10), we show that Condition 2 holds for the isotropic metric g_{iso}, which is simpler than g, and has narrower lightcones (Lemma 1). Making use of this fact, we show through the following lemma that Condition 2 also holds for our metric of interest g.

Lemma 4. *Condition 1 implies Condition 2.*

The rest of this subsection is dedicated to proving Lemma 4. We start by discussing coordinate choices. Assume w.l.o.g. that we have ordered our coordinates x^i, where f,h take the form (6), so that $a_i < 0$ for $i = 1, ..., \lambda$ and

$a_j > 0$ for $j = \lambda+1, ..., n$. We then define the following "radial" coordinates

$$r^2 := \frac{1}{2}\sum_{i=1}^{\lambda} -a_i(x^i)^2, \quad \rho^2 := \frac{1}{2}\sum_{j=\lambda+1}^{n} a_j(x^j)^2 .$$

By following the flow of the gradient vector ∇r (by which we mean the gradient taken with respect to h, so that $h(\nabla r, \cdot) = dr(\cdot)$) we get a diffeomorphism from $\mathbb{R}^{n-\lambda} \setminus \{0\}$ to $\mathbb{R} \times S^{\lambda-1}$. This gives us a coordinate system $(r, \theta_1, ..., \theta_{\lambda-1})$ on $\mathbb{R}^\lambda$, where we view $\mathbb{R}^\lambda$ as the subspace spanned by the x^i coordinates with $i = 1, ..., \lambda$. Essentially, all we have done is changing to polar coordinates, but it is important that we have done so in a way that the angular directions are g-orthogonal to the r-direction. We can do the same construction with ρ, obtaining coordinates $(\rho, \phi_1, ..., \phi_{n-\lambda-1})$ on $\mathbb{R}^{n-\lambda}$. Furthermore, we have

$$f = \rho^2 - r^2 , \quad \alpha := (\rho r)^{\frac{1}{p}} , \tag{10}$$

where $p > 0$ is a constant, f is just our Morse function, and α is chosen so that $h(\nabla f, \nabla \alpha) = 0$. Using $(f, \alpha, \theta_1, ..., \theta_\lambda, \phi_1, ..., \phi_{n-\lambda-1})$ as our coordinates, the Euclidean metric h takes the form

$$h = \frac{df^2}{\|\nabla f\|^2} + \frac{d\alpha^2}{\|\nabla \alpha\|^2} + h_\Theta + h_\Phi ,$$

and thus the Morse metric g takes the form

$$g = -(\zeta - 1)df^2 + \frac{\|\nabla f\|^2}{\|\nabla \alpha\|^2} d\alpha^2 + \|\nabla f\|^2 (h_\Theta + h_\Phi) . \tag{11}$$

Here we have used that, by definition, $\|\nabla f\| = \|df\|$. Having chosen our coordinates, we now state a lemma that constitutes the most important step in the proof of Lemma 4.

Lemma 5. *Suppose Condition 1 is satisfied. Let $q \in I^-_{\mathcal{M}}(p_c)$, and let $\gamma\colon [0,1] \to \mathcal{M}$ be any f.d. timelike curve from $\gamma(0) = q$ to $\gamma(1) = p_c$, which we express in coordinates as*

$$\gamma(s) = (f(s) , \alpha(s), \Theta(s), \Phi(s)) . \tag{12}$$

Then, for every $0 < \varepsilon < \alpha(0)$, the curve $\sigma\colon [0, s_\varepsilon] \to \mathcal{M}$ given by

$$\sigma(s) := (f(s) , \alpha(s) - \varepsilon , \Theta(s) , \Phi(s)) \tag{13}$$

is f.d. timelike. Here $s_\varepsilon := \min\{s \in (0,1) \mid \alpha(s) = \varepsilon\}$.

Proof. The statement is trivially true if $\alpha(0) = 0$ (since then there exist no suitable ε), and otherwise s_ε is well-defined (the minimum exists) by continuity of $\alpha(s)$. Moreover, our choice of ϵ and s_ε ensures that $\alpha(s) - \varepsilon \geq 0$ for all $s \in [0, s_\varepsilon]$, so that the curve σ is also well-defined.

Note that shifting α by a constant ε while leaving f, Θ, Φ fixed (as is done in (13)), is equivalent to shifting both ρ^2 and r^2 by a quantity $\epsilon(s)$. We are going to show that $g(\dot\sigma(s), \dot\sigma(s)) \leq g(\dot\gamma(s), \dot\gamma(s)) < 0$. This will be done in multiple steps, corresponding to various terms in (11).

Step 1 (Angular part)**.** Let π_Θ denote the orthogonal projection onto the subspace of the tangent space spanned by the Θ angular directions. We will show that

$$h(\pi_\Theta \dot\sigma, \pi_\Theta \dot\sigma) \leq h(\pi_\Theta \dot\gamma, \pi_\Theta \dot\gamma)\,.$$

An analogous statement holds for π_Φ. We proceed by computing $\pi_\Theta \dot\sigma$. Notice that shifting $r(s)^2$ to $r(s)^2 - \epsilon(s)$ means following the flow $F\colon \mathcal{M} \times \mathbb{R} \to \mathcal{M}$ of the vector field ∇r for a certain time $t(s) > 0$. Then

$$\dot\sigma(s) = DF(\gamma(s), t(s))\dot\gamma(s) + \frac{\partial F}{\partial t}(\gamma(s), t(s))\dot t\,. \tag{14}$$

Similary, shifting ρ^2 means following the flow of $-\nabla\rho$. Notice also that

$$\frac{\partial F}{\partial t}(\gamma(s), t(s)) = \nabla r(F(\gamma(s), t(s)))\,,$$

Hence the second term on the RHS of (14) only adds a contribution to the r component of $\dot\sigma(s)$ (but not to the angular components). We can compute DF by solving the ODE

$$\frac{\partial}{\partial t} DF(x,t) = D(\nabla r)(F(x,t))DF(x,t)$$

with initial condition $DF(x,0) = \mathrm{Id}$. In Cartesian coordinates $D\nabla r$ takes a block diagonal form:

$$D\nabla r_{ij} = \begin{cases} a_i \delta_{ij} & \text{for } i,j = 1,...,\lambda\,, \\ \delta_{ij} & \text{for } i,j = \lambda+1,...,n\,, \\ 0 & \text{otherwise,} \end{cases}$$

hence

$$DF(x,t)_{ij} = \begin{cases} e^{a_i t}\delta_{ij} & \text{for } i,j = 1,...,\lambda\,, \\ \delta_{ij} & \text{for } i,j = \lambda+1,...,n\,, \\ 0 & \text{otherwise.} \end{cases} \tag{15}$$

Moreover, we have that $DF(x,t)\partial_r \propto \partial_r$ because F is the flow of a vector field collinear to ∂r. Therefore

$$\pi_\Theta DF\ V = \pi_\Theta DF \pi_\Theta V \text{ for any } V \in TM\,,$$

and thus we can write

$$\begin{aligned} h(\pi_\Theta \dot\sigma\,, \pi_\Theta \dot\sigma) &= h(\pi_\Theta DF \pi_\Theta \dot\gamma, \pi_\Theta DF \pi_\Theta \dot\gamma) \\ &\le h(DF\pi_\Theta \dot\gamma\,, DF\pi_\Theta\dot\gamma) \\ &\le h(\pi_\Theta\dot\gamma\,, \pi_\Theta\dot\gamma). \end{aligned}$$

Here we have first used that the orthogonal projection π_Θ cannot increase the norm, and then that DF cannot increase the norm either, because it does not increase any of the Cartesian components (15).

Step 2 (α direction)**.** We want to show that $\frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2}$ does not increase when shifting r^2 and ρ^2 by ϵ, so that we do not get a larger contribution in (11). Thus in what follows we view r and ρ as functions of ϵ, in the sense that $r^2 = r_0^2 + \epsilon$ and $\rho^2 = \rho_0^2 + \epsilon$ with respect to some reference values r_0, ρ_0 (but we will omit the subscript 0 from the notation). From this point of view, what we want to show is

$$\frac{\partial}{\partial\epsilon}\Big|_{\epsilon=0} \frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2} \ge 0\,.$$

We begin with some preliminary computations, where $\nu := 2 - \frac{1}{p}$, and all derivatives are evaluated at $\epsilon = 0$.

$$\begin{aligned} \frac{\partial}{\partial\epsilon}\rho^2 &= 1\,, \\ \frac{\partial}{\partial\epsilon}\rho^4 &= 2\rho^2\,, \\ \frac{\partial}{\partial\epsilon}(r\rho)^{2\nu} &= \nu(r\rho)^{2\nu-2}(r^2+\rho^2)\,, \\ \|\nabla f\|^2 &= \|df\|^2 = \|d(r^2)\|^2 + \|d(\rho^2)\|^2\,, \\ \|\nabla\alpha\|^2 &= \|d\alpha\|^2 = \frac{1}{(2p(r\rho)^\nu)^2}\left(\rho^4\|d(r^2)\|^2 + r^4\|d(\rho^2)\|^2\right)\,. \end{aligned}$$

Moreover, we need the following estimates, where $a = 2\min_{i=1,\dots,n} a_i$, $A = 2\max_{i=1,\dots,n} a_i$.

$$a\rho^2 \leq \|d(\rho^2)\|^2 \leq A\rho^2\,,$$
$$a \leq \frac{\partial}{\partial\epsilon}\|d(\rho^2)\|^2 \leq A\,.$$

These are easily proven in Cartesian coordinates.

Applying the chain rule and substituting the previous computations and estimates, we get, after a lengthy but trivial computation, the estimate

$$\frac{\partial}{\partial\epsilon}\frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2} \geq \frac{(\nu a^2 - A^2)(r^6+\rho^6) + \big((3\nu+2)a^2 - 5A^2\big)\,(\rho^4 r^2 + \rho^2 r^4)}{(2p(r\rho)^\nu)^2\|\nabla\alpha\|^4}\,,$$

where the RHS is guaranteed to be positive if

$$\frac{A^2}{a^2} \leq \min\{\nu, \frac{3\nu+2}{5}\}\,.$$

Since $\nu \in (1,2)$ only enters in our choice of coordinates, we can freely choose it. In particular, as long as

$$\frac{A^2}{a^2} \leq \frac{8}{5}\,,$$

we can choose ν close enough to 2 so that $\frac{\partial}{\partial\epsilon}\frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2} \geq 0$.

Step 3. (Final argument)**.** It simply remains to compare $g(\dot\sigma,\dot\sigma)$ to $g(\dot\gamma,\dot\gamma)$, term by term, according to (11). By step 1, we have

$$h_\Theta(\dot\sigma,\dot\sigma) \leq h_\Theta(\dot\gamma,\dot\gamma)\,,$$
$$h_\Phi(\dot\sigma,\dot\sigma) \leq h_\Phi(\dot\gamma,\dot\gamma)\,.$$

Moreover, by computing $\|df\|^2$ in Cartesian coordinates, and using the fact that under the flow of ∇r and $-\nabla\rho$, the Cartesian coordinates are non-increasing (in absolute value), it is easy to check that

$$\|df\|^2(\sigma(s)) \leq \|df\|^2(\gamma(s))\,.$$

By step 2, we have that

$$\frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2}(\sigma(s)) \leq \frac{\|\nabla f\|^2}{\|\nabla\alpha\|^2}(\gamma(s))\,,$$

and since the ∂_α component of $\dot\sigma$ is the same as that of $\dot\gamma$ (because $\alpha_{\gamma(s)}$ and $\alpha_{\sigma(s)}$ only differ by a constant),

$$d\alpha(\dot\sigma) = d\alpha(\dot\gamma)\,.$$

Finally, because $f_{\gamma(s)} = f_{\sigma(s)}$, we have

$$df(\dot\sigma) = df(\dot\gamma)\,.$$

Plugging all of the above into (11), we conclude that

$$g(\dot\sigma,\dot\sigma) \le g(\dot\gamma,\dot\gamma)\,,$$

as desired. □

The following lemma is an easy consequence of Lemma 5, and from it we can derive Lemma 4.

Lemma 6. *Suppose Condition 1 is satisfied, and let $q \in I^-_{\mathcal{M}}(p_c)$. Then there exists a point $\tilde q \in J^+_M(q)$ such that $\alpha_{\tilde q} = 0$ and $f_{\tilde q} < 0$. Equivalently, $\rho_{\tilde q} = 0$ and $r_{\tilde q} > 0$.*

Proof. The equivalence of the two statements follows simply by definition (10). Now for the proof of existence: If $\alpha_q = 0$, choose $\tilde q = q$, and we are done because $q \in I^-_{\mathcal{M}}(p_c)$ implies $f_q < f_{p_c} = 0$. Otherwise, choose a f.d. timelike curve $\gamma\colon [0,1] \to \mathcal{M}$ from $\gamma(0) = q$ to $\gamma(1) = p_c$, and write it in components as in (12). If $\alpha(1/2) = 0$, choose $\tilde q = \gamma(1/2)$, noting that $\gamma(1/2) \in I^-_{\mathcal{M}}(p_c)$ and therefore $f_{\gamma(1/2)} < f_{p_c} = 0$. If $\alpha(1/2) \neq 0$, then since $\gamma(1/2) \in I^+_M(q)$, we can choose $0 < \varepsilon < \alpha(1/2)$ small enough so that $\hat q := (f(1/2), \alpha(1/2) - \varepsilon, \Theta(1/2), \Phi(1/2)) \in I^+_M(q)$. Then, by Lemma 5, there exists a f.d. timelike curve σ from $\hat q$ until some point $\tilde q := \sigma(s_\epsilon)$ such that $\alpha_{\tilde q} = 0$. Moreover, $f_{\tilde q} = f(s_\epsilon) < f(1) = 0$. □

Proof of Lemma 4. Let $q \in I^-_{\mathcal{M}}(p_c)$. Then we can choose $\tilde q \in J^+_M(q)$ as in Lemma 6. We claim that $\tilde q \in I^-_{\mathcal{M}}(p_c)$, not only with respect to our metric g, but even with respect to the metric g_{iso} (see Lemma 1). To prove this claim, simply note that we can reach p_c from $\tilde q$ by following the integral curve of ∇f through $\tilde q$ (which has $\rho = 0$ initially, hence $\rho = 0$ on the whole integral curve, while r must decrease, thus we reach p_c). By Lemma 10, $\tilde q \in I^-_{\mathcal{M}}(p_c, g_{\text{iso}})$ implies that there exists a neighborhood U of p_c such that $U \setminus \{p_c\} \subseteq I^+_M(\tilde q, g_{\text{iso}})$. By Lemma 1, $I^+_M(\tilde q, g_{\text{iso}}) \subseteq I^+_M(\tilde q)$, and since $\tilde q \in J^+_M(q)$, it follows that $U \setminus \{p_c\} \subseteq I^+_M(q)$, as desired. □

2.3. Limit curves, push-up and proof of Theorem 1

Having established the crucial Lemma 4, the rest of the proof of Theorem 1 does not require any computations in coordinates. Yet it follows the same philosophy of showing that the causal relation on $\mathcal{M}$ has some of the same (good) properties that it would have on a non-degenerate spacetime.

The next lemma is a sort of limit curve theorem, but can also be interpreted as telling us that $(\mathcal{M}, h, f, \zeta)$ is causally simple (see [3, p. 65] for causal simplicity of non-degenerate spacetimes).

Lemma 7 ($\overline{I^+_{\mathcal{M}}} \subseteq J^+_{\mathcal{M}}$). *Suppose Condition 2 is satisfied. Let $(p_i)_i$, $(q_i)_i$ be sequences of points in $\mathcal{M}$ such that $q_i \in I^+_{\mathcal{M}}(p_i)$. If $p_i \to p$ and $q_i \to q$, then $q \in J^+_{\mathcal{M}}(p)$.*

Proof. We first show the case $p_i, q_i, p, q \neq p_c$. Then, by Lemma 3, there exists a sequence of f.d. timelike curves $\gamma_i \colon [a_i, b_i] \to M$ such that $\gamma_i(a_i) = p_i$ and $\gamma_i(b_i) = q_i$. The idea is quite simple: we claim that (γ_i), up to a subsequence, converges to a causal curve $\gamma \colon [a, b] \to \mathcal{M}$. We show this by applying the usual limit curve theorem [26, Thm. 3.1] on the spacetime M. It is necessary to distinguish between the case when the limit curve is also in M, and the case when the limit curve crosses over the singular point p_c (then, technically speaking, there are two limit curves in M, which can be joined in $\mathcal{M}$).

- **Case 1:** The sequence γ_i converges uniformly to a causal curve $\gamma \colon [a, b] \to M$, or to a single point. Either way, $q \in J^+_M(p) \subseteq J^+_{\mathcal{M}}(p)$, and we are done.
- **Case 2:** There exist reparametrizations $\gamma_i^p \colon [0, b_i^p) \to M$ of γ_i and a future endless (in M) causal curve $\gamma^p \colon [0, \infty) \to M$ with $\gamma(0) = p$ such that $\gamma_i^p \to \gamma^p$ uniformly on compact subsets. Analogously, there exist reparametrizations $\gamma_i^q \colon (-b_i^q, 0] \to M$ of γ_i and a past endless causal curve $\gamma^q \colon (-\infty, 0] \to M$ with $\gamma^q(0) = y$ such that $\gamma_i^q \to \gamma^q$ uniformly on compact subsets.

In case 2, we claim that $\lim_{t\to\infty} \gamma^p(t) = \lim_{s\to-\infty} \gamma^q(s) = p_c$. This is a direct consequence of Lemma 2 and the fact that f is bounded away from $0, 1$ on γ^p and γ^q, hence γ^p, γ^q cannot run into the boundary $\partial\mathcal{M}$. But if $\lim_{t\to\infty} \gamma^p(t) = \lim_{s\to-\infty} \gamma^q(s) = p_c$, then (after suitable reparametrization) we can extend γ^p, γ^q to p_c and concatenate them, forming a causal curve in $\mathcal{M}$ that joins p with q, as desired.

In case that some of p_i, q_i, p, q equal p_c, we can proceed with an analogous proof, but we have to add a third case, where p_c is an endpoint of the limit curve. $\square$

The next lemma is well-known for non-degenerate spacetimes.

Lemma 8 (Push-up). *Suppose Condition 2 is satisfied. If $q \in J^+_{\mathcal{M}}(p)$, then $I^+_{\mathcal{M}}(q) \subseteq I^+_{\mathcal{M}}(p)$.*

Proof. If $p = q$, the result is trivial, so assume $p \neq q$, and let $\sigma\colon [0,1] \to \mathcal{M}$ be a causal curve from p to q.

Case 1: $q = p_c$**.** Let $\tilde{q} \in I^+_M(p_c)$ (see Figure 1a). Then, by Condition 2, there exists a neighborhood U of p_c such that $U \setminus \{p_c\} \subseteq I^-_M(\tilde{q})$. Because $\sigma(1) = p_c$, there must exist some $0 < s_0 < 1$ such that $\sigma(s_0) \in U \setminus \{p_c\}$. But then $\tilde{q} \in I^+_M(\sigma(s_0))$, and since also $\sigma(s_0) \in J^+_M(p)$, we conclude by the standard push-up lemma in M that $\tilde{q} \in I^+_M(p) \subseteq I^+_{\mathcal{M}}(p)$. Since $\tilde{q}$ was arbitrary, we are finished with this case.

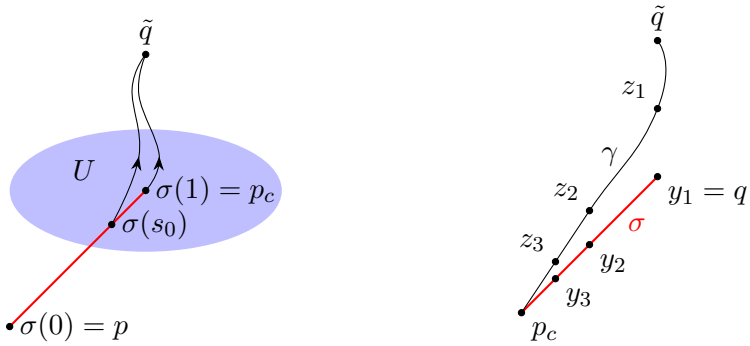
Case 2: $p = p_c$**.** The argument is similar to the one in [2, Prop. 2.1]. Let $\tilde{q} \in I^+_{\mathcal{M}}(q) = I^+_M(q)$ (see Figure 1b). We construct a timelike curve γ from p_c to $\tilde{q}$. Let $y_n := \sigma(1/n)$, and choose a point $z_1 \in I^+_M(y_1) \cap I^-_M(\tilde{q})$. By openness of $I^-_M(z_1)$, and since, by the usual push-up lemma in M, $y_2 \in J^-_M(y_1) \subset I^-_M(z_1)$, we may choose $z_2 \in I^-_M(z_1) \cap I^+_M(y_2) \cap B^h_{1/2}(y_2)$. Here $B^h_{1/2}(y_2)$ denotes the ball of radius $1/2$ around y_2, measured with respect to the Riemannian metric h. Iterating this procedure, we obtain a sequence $(z_l)_l$ such that $z_l \in I^-_M(z_{l-1}) \cap I^+_M(y_l) \cap B^h_{1/l}(y_l)$. Then we construct γ by joining all the timelike segments going from z_l to z_{l+1}. Since, by contruction, $\lim_{l\to\infty} z_l = \lim_{l\to\infty} y_l = p_c$, the timelike curve γ connects p_c and $\tilde{q}$.

Case 3: $p, q \neq p_c$**.** If σ lies entirely in M, the result follows from the standard theory. Therefore, we assume w.l.o.g. that $\sigma(\frac{1}{2}) = p_c$. Then, in particular, $q \in J^+_{\mathcal{M}}(p_c)$, so by case 2, we have that $I^+_{\mathcal{M}}(q) \subseteq I^+_{\mathcal{M}}(p_c)$. But since also $p_c \in J^+_{\mathcal{M}}(p)$, by case 1 it follows that $I^+_{\mathcal{M}}(p_c) \subseteq I^+_{\mathcal{M}}(p)$, and we are done. $\square$

Lemma 9 below, together with Lemma 4, completes the proof of Theorem 1.

Lemma 9. *If Condition 2 is satisfied, then the Morse spacetime (M, g) is causally continuous.*

Proof. By Definition 1 (in Appendix B), (M, g) is causally continuous if it is distinguishing and reflecting. Because f is a time function, (M, g)



(a) Case 1. (b) Case 2.

Figure 1. An illustration of the proof of Lemma 8. The red line represents σ, and the black curves represent future-directed causal curves.

must be distinguishing [6, Sec. 2]. Thus we only need to prove reflectivity. Let $p, q \in M$ be such that $I^-_M(p) \subseteq I^-_M(q)$ (the future case is analogous). We need to prove that $I^+_M(q) \subseteq I^+_M(p)$. By the time-reverse of Lemma 3, $I^-_{\mathcal{M}}(p) \subseteq I^-_{\mathcal{M}}(q)$, and then, since $p \in \overline{I^-_{\mathcal{M}}(p)} \subseteq \overline{I^-_{\mathcal{M}}(q)}$, Lemma 7 tells us that $p \in J^-_{\mathcal{M}}(q)$. But then, by Lemma 8, $I^+_{\mathcal{M}}(q) \subseteq I^+_{\mathcal{M}}(p)$, which again by Lemma 3 implies $I^+_M(q) \subseteq I^+_M(p)$. □

Remark. One may even say that the Morse geometry $(\mathcal{M}, h, f, \zeta)$ is globally hyperbolic. Firstly, it follows from Lemma 2 that f is a Cacuhy time function, in the sense that any causal curve that is inextendible in $\mathcal{M}$, must start at one boundary component and end at the other, crossing each level set exactly once. Secondly, by compactness of $\mathcal{M}$, it is easy to see that the causal diamonds $J^+_{\mathcal{M}}(p) \cap J^-_{\mathcal{M}}(q)$ are compact, for all $p, q \in \mathcal{M}$. However, both arguments are also true when (M, g) is causally discountinuous, such as in the index $1, n-1$ case. The point we would like to make here, is that one should additionally require Condition 2 to hold, and then the causal structure of $\mathcal{M}$ is very well-behaved.

We can turn this remark into a mathematically precise statement by using the language of Lorentzian length spaces, introduced by Kunzinger and Sämann [22] (see also [8], where topology change is discussed in this context). Lorentzian length spaces are topological spaces equipped with a notion of causal order and satisfying a set of axioms which, in particular, imply a version of Condition 2 [22, Lem. 2.12]. A somewhat related point is that (M, g) is semi-globally hyperbolic, meaning that it can be divided into

globally hyperbolic pieces, which in our case are separated by the critical level sets of f. The notion of semi-globally hyperbolic spacetime was introduced by Janssen in [18], with the goal of defining quantum field theories on them (note the connection to Conjecture 1).

3. Towards a full resolution of the Borde–Sorkin Conjecture

Throughout this section, we employ the same notational conventions as in Section 2, except that we allow our Morse functions to have multiple critical points. The current progress on Conjecture 2 is summarized in Theorem 2 in the Introduction. What still remains open is the case when f has critical points of index $\lambda = 2, ..., n-2$, and h is arbitrary. In other words, we do not know what happens if we drop Condition 1.

Notice that Condition 1 is basically telling us two things:

(i) We can find a coordinate neighborhood $\mathcal{U}$ of p_c where both h and f take a specified standard form.

(ii) We have the bounds $\zeta_c < \zeta$ and $\zeta_c \leq 8/5$ (see (7)), which can be interpreted as a bound on how much anisotropy is allowed.

In the first part of this section, we show that the neighborhood $\mathcal{U}$ can always be found, the only difference being that in the general case, we need to add a perturbation to h that vanishes at p_c. In the second part of this section, we give a candidate counterexample to Conjecture 2, which suggests that $\zeta > \zeta_c$ is a necessary condition for causal continuity. We then conclude by proposing a modified version of the conjecture which takes this into account.

3.1. Generalized standard neighborhoods

The statement of the next proposition can be seen as a weaker version of the first part of Condition 1.

Proposition 1. *Let $(\mathcal{M}, h, f, \zeta)$ be a Morse geometry and p_c be a critical point of f. Then there exists an open neighborhood $\mathcal{U} \subseteq \mathcal{M}$ of p_c, an open ball $\mathcal{B} \in \mathbb{R}^n$ around the origin, and a coordinate chart $\phi : \mathcal{U} \to \mathcal{B}$ such that $\phi(p_c) = 0$ and*

$$f \circ \phi^{-1} = \frac{1}{2} \sum_i a_i (x^i)^2, \quad h \circ \phi^{-1} = \sum_i (dx^i)^2 + \sum_{k,l} h^{(1)}_{kl}(x) dx^k dx^l \,,$$

for some real constants $a_i \neq 0$ and some tensor $h^{(1)}$ satisfying $h^{(1)}(p_c) = 0$.

The proof relies on the following two results from the literature.

Simultaneous Diagonalization Theorem [14, Thm. 13.4.3]**.** *Let H, D be two real symmetric $n \times n$ matrices, and let H be positive definite. Then there exists a real non-singular matrix Λ such that both $\Lambda^T H \Lambda$ and $\Lambda^T D \Lambda$ are diagonal.*

Morse Lemma [25, Lem. 2.2]**.** *Let $\mathcal{M}$ be a manifold, $f\colon \mathcal{M} \to \mathbb{R}$ be a Morse function and p_c be a critical point of f. Then there exists an open neighborhood $\mathcal{U} \subseteq \mathcal{M}$ of p_c, an open ball $\mathcal{B} \subseteq \mathbb{R}^n$ around the origin, and a coordinate chart $\phi : \mathcal{U} \to \mathcal{B}$ such that $\phi(p_c) = 0$ and*

$$f \circ \phi^{-1} = \frac{1}{2} \sum_i a_i (x^i)^2 \,,$$

for some real constants $a_i \neq 0$.

Proof of Proposition 1. By the Morse Lemma, we can find coordinates where f already has the desired form. Then we apply a linear change of coordinates in order to simultaneously diagonalize the bilinear forms $h(p_c)$ and $\operatorname{hess} f(p_c)$. Finally, we scale each coordinate, in order to normalize our new basis with respect to $h(p_c)$. □

We use Proposition 1 to formulate a relaxed version of Condition 1. Let $(\mathcal{M}, h, f, \zeta)$ be a Morse geometry, and suppose we have chosen a critical point p_c.

Condition 3. *The constants a_i appearing in Proposition 1 (applied to p_c) satisfy*

$$\zeta > \zeta_c := \max_{i,j} \left| \frac{a_i}{a_j} \right| \,.$$

We do not include the bound $\zeta_c \leq 8/5$, because it will not be relevant in the upcoming examples, and it seems likely to not be a necessary condition for causal continuity. Note also that in this paper, we take the point of view that ζ is specified as part of the Morse geometry. If, instead, we only specify h and f, then we can always choose $\zeta > \zeta_c$ at a given critical point (hence also at any finite number of critical points). In this sense, Condition 3 is not very restrictive. Note, in any case, that ζ_c depends on both h and f in a neighborhood of p_c.

Since the coordinate system that we get from Proposition 1 is not necessarily unique, the question arises whether the truth or falsehood of Condition 3 depends on any coordinate choices. The answer is no. To see this, note that $\max_i a_i$ is the maximum value of the quadratic form $\operatorname{hess} f(p_c)$ applied to the $h(p_c)$-unit ball. Similarly, $\min_i a_i$ is the minimum. These maxima and minima are independent of the basis, so we conclude that the value of ζ_c is the same among all coordinate bases satisfying the properties listed in Proposition 1.

One is left to wonder how much our proofs in Section 2 are affected when adding the perturbation $h^{(1)}$ that appears in Proposition 1. We can say the following:

- If Condition 3 is satisfied, then a generalization of Lemma 1 holds. The only difference is that in (9), we need to replace ζ by $1 < \hat{\zeta} < \zeta$. This makes the lightcones of g_{iso} a bit narrower, compensating for the fact that the perturbation $h^{(1)}$ might have also made the lightcones of g a bit narrower.
- The proofs in Section 2.2 are no longer valid.
- If we can prove that Condition 3 implies Condition 2 (compare with Lemma 4), then causal continuity follows by the same arguments as in Section 2.3.
- On general spacetimes, causal continuity is not stable under perturbations of the metric. This remains true even if we require said perturbations to always widen or narrow the lightcones with respect to the original metric (see Examples 3 and 4 in Appendix B).

3.2. A potential counterexample

The discussion in Section 3.1 leads to a natural question: is Condition 3 necessary in order to have causal continuity? If the answer is yes, it would mean that Conjecture 2 is false in its original form. We believe that this is indeed so. Constructing examples that violate Condition 3 is easy, but showing that they are causally discontinuous is not (and we do not achieve it in this paper).

The following examples are meant to illustrate what happens when Condition 3 is not satisfied. For convenience, we take $\mathcal{M}$ non-compact and without boundary, the idea being that it represents a neighborhood of a critical point in some larger Morse geometry.

Example 1. Let $\mathcal{M}$ be an open ball in $\mathbb{R}^2$, centered around the origin $p_c := 0$, and equipped with coordinates (x, y). Define

$$h := dx^2 + dy^2, \quad f := -\frac{1}{2}\left(x^2 + by^2\right) ,$$

so that

$$g = (x^2 + b^2y^2)(dx^2 + dy^2) - \zeta(xdx + bydy)^2 .$$

The case $b = 1$ is considered in Appendix A: it is a *neighborhood spacetime* in the sense of Borde et al. [6]. We will refer to is as an *isotropic* neighborhood, while in the case of $b \neq 1$, we will talk about *anisotropic* neighborhoods.

Consider a radial line $\gamma(s) = (s, ms)$, for $m \in \mathbb{R}$. Then

$$g\left(\dot\gamma(s), \dot\gamma(s)\right) = \left((1-\zeta)b^2m^4 + (1 + b^2 - 2\zeta b)m^2 + 1 - \zeta\right) s^2 . \tag{16}$$

If $b = 1$, this quantity reduces to

$$g\left(\dot\gamma(s), \dot\gamma(s)\right) = (1-\zeta)\left(m^2 + 1\right)^2 s^2 ,$$

which negative for all m, hence all radial lines are timelike (see Figure 2a). In this case, the past of p_c, which is the whole Morse spacetime $M = \mathcal{M} \setminus \{p_c\}$, can be written as a single TIP, $M = I^-_{\mathcal{M}}(p_c) = I^-_M(\gamma)$, for any future directed timelike curve γ that ends at p_c (see Appendix A).

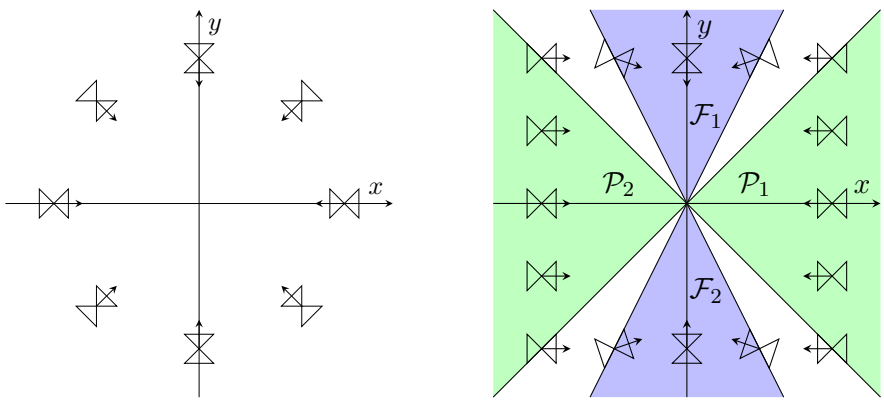
Taking the limit $b \to 0$ in (16), we get an expression that is positive whenever $m^2 > 1 - \zeta$. By continuity, we conclude that for $0 < b < 1$ small enough, $g(\dot\gamma(s), \dot\gamma(s))$ can be negative, zero or positive, depending on m (the dependence, however, is more complicated than in the $b \to 0$ limit). Concretely, for b small (relative to ζ):

- There exist null radial lines, which form the boundaries of the future sets[2] $\mathcal{F}_1, \mathcal{F}_2$ and the pasts sets $\mathcal{P}_1, \mathcal{P}_2$. Intuitively, this happens because the lightcones tilt much faster when moving in the x direction, compared to moving in the y direction (see Figure 2b). Regardless, we still have that $M = I^-_{\mathcal{M}}(p_c)$.
- $I^+_{\mathcal{M}}(q)$ is not open, for any $q \in M$. This is because $p_c \in I^+_{\mathcal{M}}(q)$, but no neighborhood of p_c is entirely contained in $I^+_{\mathcal{M}}(q)$ (compare with Lemma 4).

[2]Here by *future set* we mean a set $\mathcal{F}$ such that $I^+_M(\mathcal{F}) = \mathcal{F}$. Analogously, a *past set* $\mathcal{P}$ satisfies $I^-_M(\mathcal{P}) = \mathcal{P}$.

Note that there also exists an intermediate case, when b has the exact value so that setting (16) equal to zero has degenerate solutions, and then the boundaries of the sets $\mathcal{F}_1, \mathcal{F}_2, \mathcal{P}_1, \mathcal{P}_2$ overlap. The case of $b > 1$ very large can be reduced to the case of $b < 1$ small by rescaling both x and y by a factor of $1/b$ (then also h is rescaled, but this does not affect g).

Since the critical point in this example has index 0, the Morse spacetime (M, g) is causally continuous, no matter how small we choose b (see Theorem 2). In fact, (M, g) is even globally hyperbolic, with f being a Cauchy time function. Note however that in this example, $I^+_{\mathcal{M}}(p_c) = \emptyset$, which simplifies things a lot.



(a) $b = 1$ (b) $b \ll 1$

Figure 2. The causal structure of Example 1.

Building upon the previous example, we propose our candidate counterexample to Conjecture 2.

Example 2. Let $\mathcal{M}$ be an open ball in $\mathbb{R}^4$, centered around the origin $p_c := 0$, and equipped with coordinates (x, y, z, w). Define

$$h := dx^2 + dy^2 + dz^2 + dw^2, \quad f := \frac{1}{2}\left(-x^2 - by^2 + z^2 + w^2\right),$$

and g, as usual, by (3). We claim that reflectivity (see Definition 1) is violated by the pair of points

$$p := (1, 0, 0, 0), \quad q := (m, 0, 1, 0),$$

where $m := \frac{\sqrt{\zeta}-1}{\sqrt{\zeta}-1}$. From the causal analysis of the punctured (x, z)-plane (see Figure 3 and Appendix A) it follows that $I_M^+(q) \subseteq I_M^+(p)$. However, $I_{XZ}^-(q) \not\supseteq I_{XZ}^-(p)$, where the subscript XZ means that we are only considering causal curves in the punctured (x, z)-plane. Nonetheless, it is possible that $I_M^-(q) \supseteq I_M^-(p)$ when also considering causal curves that leave said plane. In particular, if $b = 1$ (or close enough to 1), we see from the analysis in Example 1 (see also Figure A1b) that from p we can reach the negative x-axis via a future-directed timelike curve contained in the (x, y)-plane. Then, from the negative x-axis, we can reach q. If b is too small, however, this construction is no longer possible (see Figure 2b), suggesting that probably $I_M^-(q) \not\supseteq I_M^-(p)$. This is not a bulletproof argument, of course, because we are ignoring all timelike curves that are not contained in any coordinate plane.

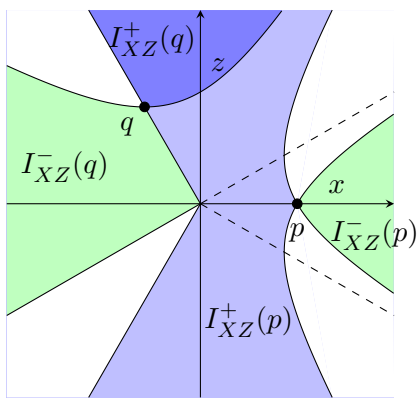


Figure 3. The points p, q of Example 2 and their futures and pasts (restricted to the (x, z)-plane XZ).

4. Conclusions

With the proof of Theorem 1, we have established a new case of the Borde–Sorkin conjecture (Conjecture 2 in the Introduction). In doing so, we have advanced the current progress on the conjecture to that summarized in Theorem 2. Along the way, we have developed a notion of causal structure for Morse geometries that includes the critical points. This supports the view

that degenerate metrics are physically reasonable, and is a first step towards understanding quantum fields on Morse geometries (in view of Conjecture 1).

Let us briefly mention here three recent approaches to quantum field theory that are specially relevant for topology change. The algebraic approach of Janssen [18] has been developed specifically with topology change as one of its applications, but the existence of states in this approach is still an open problem. Another approach is that of Sorkin and Johnston [19, 30]; so far it has been applied to the trousers topology change [7] (confirming the energy divergences there), but, to our knowledge, not to any of the causally continuous Morse geometries (where Conjecture 1 predicts a well-behaved QFT). Lastly, the recent paper of Kontsevich and Segal [20] defines QFTs on the category of cobordisms with certain complex metrics. Real Lorentzian metrics arise as a limit case of these complex metrics, and so do Morse metrics [23, 32]. So far, however, it is only known that a QFT is induced on the limit spacetime when the latter is a non-degenerate and globally hyperbolic [20, Thm. 5.2]. It remains to be seen if this result can be generalized to Morse geometries.

Another important conclusion of the present paper is that we have found a potential counterexample to Conjecture 2 (Example 2 in Section 3.2). Despite being of dimension 4 and having only a critical point of index 2, we believe that our example is causally discontinuous, due to being highly anisotropic (i.e. because it has $\zeta_c > \zeta$). The lack of symmetries and good coordinate choices has prevented us from proving this fully. Regardless, we propose the following refinement of Conjecture 2, which incorporates a bound on the anisotropy (Condition 3).

Conjecture 3. *Let $(\mathcal{M}, h, f, \zeta)$ be a Morse geometry of dimension n, and (M, g) the corresponding Morse spacetime. Assume f has a single critical point for each critical value. Then (M, g) is causally continuous if and only if the following hold:*

(i) None of the critical points has index 1 or $n-1$,

(ii) Condition 3 is satisfied at every critical point of index different from $0, n$.

In order to prove Conjecture 3, two steps remain. One is to show that Condition 3 is really necessary, by showing that Example 2 (where it is violated) is causally discontinuous. The other remaining step is to generalize Theorem 1 by adding a perturbation to h that vanishes at the critical points, and by removing the requirement that $\zeta_c \leq 8/5$. This is nontrivial, because

causal continuity is, in general, not stable under perturbations (see Examples 3 and 4 in Appendix B). Yet the second half of our proof (Section 2.3) is robust under perturbations, and does not require $\zeta_c \leq 8/5$, so it would suffice to prove openness of the chronological relation in $\mathcal{M}$ (Condition 2), and then causal continuity would follow.

Acknowledgements. I would like to thank Elefterios Soultanis and Maximilian Ruep for very interesting discussions, and Annegret Burtscher for comments on the draft. I am also grateful to Bernardo Araneda and Simon Pepin Lehalleur for pointing me to reference [20].

Appendix A. Neighborhood spacetimes

In this appendix, we review the causal structure of isotropic neighborhood spacetimes, as studied in Borde et al. [6]. At the end, we prove Lemma 10, which is new, although it follows quite straightforwardly from the analysis in [6].

Let $\mathcal{M} \subseteq \mathbb{R}^n$ be an open neighborhood of the origin, equipped with a coordinate system $(x^1, ..., x^\lambda, y^1, ..., y^{n-\lambda})$, where $\lambda \neq 0, 1, n-1, n$, and

$$h := \sum_{i=1}^{\lambda} (dx^i)^2 + \sum_{j=1}^{n-\lambda} (dy^j)^2, \quad f := -\frac{1}{2}\sum_{i=1}^{\lambda} (x^i)^2 + \frac{1}{2}\sum_{j=1}^{n-\lambda} (y^j)^2 . \tag{A.1}$$

Note that the origin $p_c = (0, ..., 0)$, is the only critical point of f in this case. We write $M := \mathcal{M} \setminus \{p_c\}$, as usual (here $\mathcal{M}$ has no boundary, but can be thought of as a neighborhood of a critical point in some cobordism). It is convenient to change to polar coordinates (ρ, Θ, r, Φ), where

$$\rho := \sum_{j=1}^{n-\lambda} (y^j)^2, \quad r := \sum_{i=1}^{\lambda} (x^i)^2 ,$$

and where Θ, Φ denote the angular coordinates corresponding to the subspaces $\{r = 0\}$ and $\{\rho = 0\}$ respectively (thus each of Θ, Φ is a collection of angular variables, rather than a single one). The Lorentzian metric (2) is then given by

$$\begin{aligned} g = &\left(r^2 - (\zeta - 1)\rho^2\right) d\rho^2 + \left(\rho^2 - (\zeta - 1)r^2\right) dr^2 + 2\zeta\rho r d\rho dr \\ &+ \left(\rho^2 + r^2\right)\left(\rho^2 d\Theta^2 + r^2 d\Phi^2\right) . \end{aligned} \tag{A.2}$$

Because the coefficients in front of $d\Theta$ and $d\Phi$ are positive, any null geodesic in the $\Theta, \Phi = \text{const.}$ plane (with respect to the restricted metric), must also be a null geodesic in the full spacetime.

Thus we start by commenting on the situation with constant angles. Recall also that in 2 dimensions, any null curve is automatically a null geodesic. The implicit equations for any null geodesic can thus be found from (A.2):

$$\sqrt{\zeta - 1}(x^2 - y^2) = \pm 2xy + c_\pm \,, \tag{A.3}$$

where $c_\pm$ are constants. In particular, the case of $c_\pm = 0$ corresponds to geodesics that bound $I^\pm_{\mathcal{M}}(p_c)$, which are radial lines of a certain slope (depicted as dashed lines in Figure A1a). We can use this information to find the future and past sets of any point (depicted as colored regions in Figure A1a).

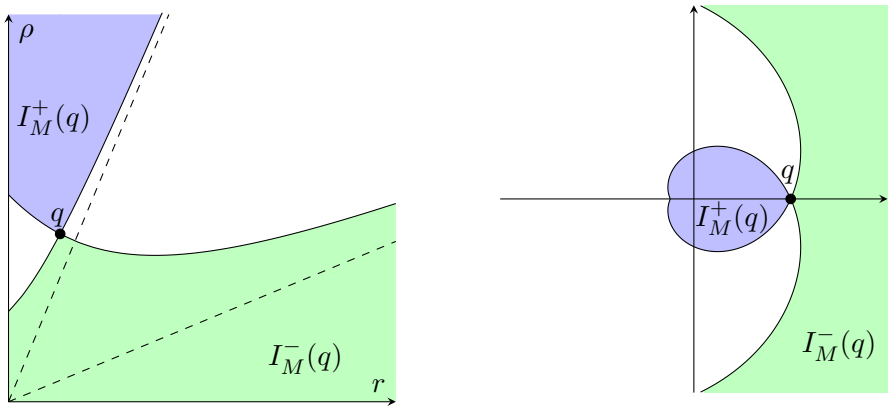
Next we analyse the case when of $\rho = 0$ and $\Theta = \text{const.}$ (the case $r = 0$ and $\Phi = \text{const.}$ is analogous). We restrict to the case where $\Phi = \phi$ is just a single angular variable, hence again reducing the problem to two dimensions. The null geodesics on the $\rho = 0, \Theta = \text{const.}$ plane (with respect to the induced metric) are then given by

$$r(\phi) = r_0 \, \mathrm{e}^{\pm\phi/\sqrt{\zeta-1}} \,.$$

Again, we can use this to find the future and past sets of any point on the plane, with respect to the induced metric (see Figure A1b). In this case, there are no null geodesics going through the origin p_c, and all points with $\rho = 0$ lie in the past of p_c.

Lemma 10. *Let $(\mathcal{M}, h, f, \zeta)$ be an isotropic Morse neighborhood, with h, f given by* (A.1)*. For every $q \in I^\pm_{\mathcal{M}}(p_c)$, there exists a neighborhood U of p_c such that $U \setminus \{p_c\} \subseteq I^\mp_M(q)$.*

Proof. We show the case when $q \in I^-_{\mathcal{M}}(p_c)$. W.l.o.g. we may choose our coordinates such that $q = (\rho_q, 0, r_q, 0)$, where necessarily $\rho_q \neq 0$ (otherwise q cannot be in $I^-_{\mathcal{M}}(p_c)$). We want to find ρ_0, r_0 such that all points $(\rho, \Theta, r, \Phi) \neq p_c$ with $\rho < \rho_0$, $r < r_0$ and Θ, Φ arbitrary, are contained in $I^+_M(q)$. By symmetry, we may choose our coordinates such that at most one of the Θ- and one of the Φ-angles may be different from zero, hence effectively reducing the problem to four dimensions.



(a) (r, ρ)-plane (b) (r, ϕ)-plane

Figure A1. The causal structure of an isotropic neighborhood.

Our argument now resembles the one in the proof of [6, Claim 1]. Let $\ll$ denote the chronological relation in M. If $\rho_q, r_q \neq 0$, then

$$x = (\rho_x, 0, r_x, 0) \ll (\epsilon_1, 0, 0, 0) \ll (\epsilon_1\delta, \theta, 0, 0) \ll (\epsilon_1\delta\epsilon_2, \theta, \epsilon_3, \phi)\,.$$

In every step where we have added an ϵ, we have used our analysis of the causal structure in the case θ, ϕ constant. In the step where we have added δ, it is using our analysis of the $\rho = 0$ and $\theta =$ const. case. In principle, δ depends on θ. We see, however, that the "worst case scenario" (when δ has to be the smallest) is when $\theta = \pi$. Thus we can choose this largest value, so that the procedure works in all cases. Note also that in the last step, since we starting from the origin of the (r, ϕ)-plane, we can choose any value for ϕ that we want. Setting $\rho_0 := \epsilon_1\delta\epsilon_2$ and $r_0 := \epsilon_3$, and again considering the causal structure in the case θ, ϕ constant, we are done. $\square$

Appendix B. Causal continuity

Let (M, g) be a non-degenerate spacetime. We refer to [3, Chap. 3] for the basic concepts and notation of causality theory. The idea is that (M, g) is causally continuous if the set valued functions $q \mapsto I_M^{\pm}(q)$ are continuous. There are various equivalent ways to make this precise [3, pp. 59-71]. In this paper, we use the following definition, which is perhaps the most standard one, even though it does not directly capture the intuition behind the concept.

Definition 1. A spacetime (M, g) is called

(i) *distinguishing* if

$$I_M^-(p) = I_M^-(q) \iff p = q \iff I_M^+(p) = I_M^+(q)$$

for all $p, q \in M$,

(ii) *reflecting* if

$$I_M^-(p) \subseteq I_M^-(q) \iff I_M^+(p) \supseteq I_M^+(q)$$

for all $p, q \in M$,

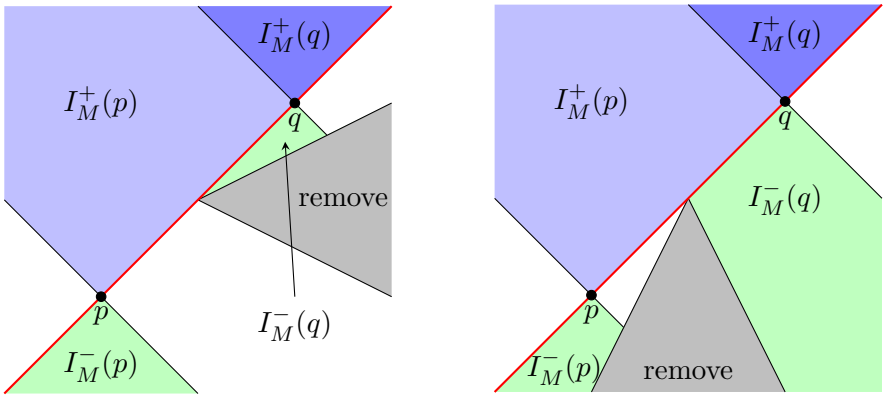
(iii) *causally continuous* if it is distinguishing and reflecting.

The following example shows that causal continuity is not stable under perturbations of the metric g, even if we only allow perturbations that make the lightcones narrower.

Example 3. Let $M := \mathbb{R}^2 \setminus \{(x,t) \mid x \geq 2|t|\}$ and $g_\alpha = -\alpha dt^2 + dx^2$. Then (M, g_α) is causally continuous for $\alpha \geq 2$ and causally discontinuous for $\alpha < 2$. This can be seen in Figure B1a: Reflectivity is violated for pairs of points lying on the diagonal red line, one above and one below the origin (such as the depicted points p, q). The red line has slope $1/\alpha$, hence if $\alpha \geq 2$, half of the red line lies inside the removed wedge, and there is no violation of reflectivity anymore.

The next example shows that causal continuity is not stable under widening of the lightcones, either.

Example 4. Let $M := \mathbb{R}^2 \setminus \{(x,t) \mid t \leq -2|x|\}$ and $g_\alpha = -\alpha dt^2 + dx^2$. Then (M, g_α) is causally continuous for $\alpha \leq \frac{1}{2}$ and causally discontinuous for $\alpha > \frac{1}{2}$. The argument is similar to the one in Example 3 (see Figure B1b).



(a) Example 3 with $\alpha = 1$.

(b) Example 4 with $\alpha = 1$.

Figure B1. A pair of points p, q for which reflectivity is violated.

References

[1] A. Anderson and B. S. DeWitt, *Does the Topology of Space Fluctuate?*, Found. Phys. **16** (1986), 91–105.

[2] L. Andersson, G. J. Galloway, and R. Howard, *The cosmological time function*, Classical Quantum Gravity **15** (1998), 309–322.

[3] J. K. Beem, P. E. Ehrlich, and K. L. Easley, Global Lorentzian Geometry, Monographs and Textbooks in Pure and Applied Mathematics Marcel Dekker, Inc., New York, second edition, 1996.

[4] A. N. Bernal and M. Sánchez, *On smooth Cauchy hypersurfaces and Geroch's splitting theorem*, Comm. Math. Phys. **243** (2003), 461–470.

[5] A. N. Bernal and M. Sánchez, *Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes*, Comm. Math. Phys. **257** (2005), 43–50.

[6] A. Borde, H. F. Dowker, R. S. Garcia, R. D. Sorkin, and S. Surya, *Causal continuity in degenerate spacetimes*, Classical Quantum Gravity **16** (1999), 3457–3481.

[7] M. Buck, F. Dowker, I. Jubb, and R. Sorkin, *The Sorkin–Johnston state in a patch of the trousers spacetime*, Classical Quantum Gravity **34** (2017) 055002.

[8] A. Burtscher and L. García-Heveling, *Time functions on Lorentzian length spaces*, `arXiv:2108.02693`.

[9] F. Dowker, *Topology change in quantum gravity*, Workshop on Conference on the Future of Theoretical Physics and Cosmology in Honor of Steven Hawking's 60th Birthday **6** (2002), 436–452.

[10] F. Dowker and S. Surya, *Topology change and causal continuity*, Phys. Rev. D **58** (1998) 124019.

[11] H. F. Dowker and R. S. Garcia, *A Handlebody calculus for topology change*, Classical Quantum Gravity **15** (1998), 1859–1879.

[12] H. F. Dowker, R. S. Garcia, and S. Surya, *K-causality and degenerate spacetimes*, Classical Quantum Gravity **17** (2000), 4377–4396.

[13] H. F. Dowker, R. S. Garcia, and S. Surya, *Morse index and causal continuity. A criterion for topology change in quantum gravity*, Classical Quantum Gravity **17** (2000), 697–712.

[14] S. R. Garcia and R. A. Horn, A Second Course in Linear Algebra, Cambridge Mathematical Textbooks. Cambridge University Press, 1 edition, 2017.

[15] R. Geroch, *Domain of dependence*, J. Mathematical Phys. **11** (1970), 437–449.

[16] R. P. Geroch, *Topology in general relativity*, J. Math. Phys. **8** (1967), 782–786.

[17] S. W. Hawking and R. K. Sachs, *Causally continuous spacetimes*, Comm. Math. Phys. **35** (1974), 287–296.

[18] D. W. Janssen, *Quantum Fields on Semi-globally Hyperbolic Space-Times*, Commun. Math. Phys. **391** (2022), 669–705.

[19] S. Johnston, *Feynman propagator for a free scalar field on a causal set*, Phys. Rev. Lett. **103** (2009) 180401.

[20] M. Kontsevich and G. Segal, *Wick rotation and the positivity of energy in quantum field theory*, Q. J. Math. **72** (2021), 673–699.

[21] W. Kundt, *Non-existence of trouser-worlds*, Comm. Math. Phys. **4** (1967), 143–144.

[22] M. Kunzinger and C. Sämann, *Lorentzian length spaces*, Ann. Global Anal. Geom. **54** (2018), 399–447.

[23] J. Louko and R. D. Sorkin, *Complex actions in two-dimensional topology change*, Classical Quantum Gravity **14** (1997), 179–204.

[24] C. Manogue, E. Copeland, and T. Dray, *The trousers problem revisited*, Pramana **30** (1988), 279–292.

[25] J. W. Milnor, Morse Theory. Based on lecture notes by M. Spivak and R. Wells, Princeton University Press, 1963.

[26] E. Minguzzi, *Limit curve theorems in Lorentzian geometry*, J. Math. Phys. **49** (2008), 092501–092518.

[27] M. Sánchez, *A class of cosmological models with spatially constant sign-changing curvature*, Port. Math. **80** (2023), 291–313.

[28] R. D. Sorkin, *Consequences of spacetime topology*, Proceedings of the Third Canadian Conference on General Relativity and Relativistic Astrophysics, `https://api.semanticscholar.org/CorpusID:5137407`.

[29] R. D. Sorkin, *Forks in the road, on the way to quantum gravity*, Int. J. Theor. Phys. **36** (1997), 2759–2781.

[30] R. D. Sorkin, *Scalar Field Theory on a Causal Set in Histories Form*, J. Phys. Conf. Ser. **306** (2011) 012017.

[31] J. A. Wheeler, *On the nature of quantum geometrodynamics*, Ann. Phys. **2** (1957), 604–614.

[32] E. Witten, A Note On Complex Spacetime Metrics, Frank Wilczek, (2022) 245–280. `https://www.worldscientific.com/doi/abs/10.1142/9789811251948_0020`.

[33] P. Yodzis, *Lorentz cobordism*, Comm. Math. Phys. **26** (1972), 39–52.

[34] P. Yodzis, *Lorentz cobordism. II*, Gen. Relativity Gravitation **4** (1973), 299–307.

Mathematics Department, Radboud University
Postbus 9010, 6500GL Nijmegen, The Netherlands
Fachbereich Mathematik, Universität Hamburg
Bundesstrasse 55, 20146 Hamburg, Germany
E-mail address: `leonardo.garcia@uni-hamburg.de`