On the Causal Structure of Spacetime

Course Project for PHY 442: General Relativity

Muhammad Hashir Hassan Khan

2024-10-0111

May 1, 2023

Abstract

A study of the causal structure of spacetime is carried out. The necessary definitions and important properties along with certain illuminating proofs are discussed. Finally, certain causality conditions are imposed and causal ladder of spacetimes is constructed.

Contents

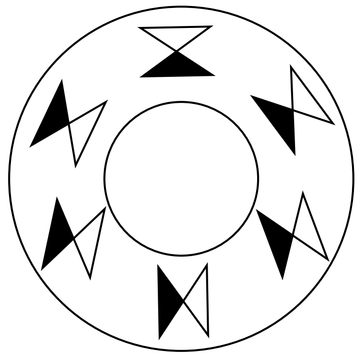

Figure 1: A representation of a Mobius strip as a non-time orientable spacetime.

1 Introduction

Studying the causal structure of spacetime offers a different route to the study of general relativity and Lorentzian geometry as compared to solving Einstein's equations to find results that would hold only for that solution. Instead, causal theory allows one to extract general results valid for not only Einstein's theory of gravity but any theory that involves a Lorentzian manifold. One normally proceeds by demanding some physically reasonable condition to hold and expressing it in precise mathematical terms. Then, one runs the mathematical machinery (primarily that of differential topology) and notes the physical consequences.

This article is divided into two sections. The first section is devoted to building the causal structure by introducing the reader to the various definitions and notations used in the theory. A few results and properties of the sets are proved to give the reader a taste of how one proceeds in such proofs. The second section builds a hierarchy of progressively stronger causal spacetimes based on demanding physically reasonable causality conditions.

2 Building Causal Structure

We begin with the standard manifold structure.

Definition. A Lorentzian manifold M is a real, four-dimensional connected differentiable Hausdorff manifold with a globally defined non-degenerate Lorentzian metric g with signature $(-, +, +, +)$.

The space of all tangent vectors to curves at a point $p \in M$ as T_pM. A vector $v \in T_pM$ is said to be timelike, null or spacelike if $g_{\alpha\beta}v^\alpha v^\beta$ is less than, equal to or greater than 0 respectively. The space of all null vectors at p forms the light cone in T_pM. The interior of the light cone contains timelike vectors.

2.1 Time Orientation

Since a Lorentzian manifold is locally the same as Minkowski space[1], light cones can be interpreted in the same way as in special relativity. In particular, one half of the light cone at a point p can be designated as the "future" and the other half as the "past" of p. The global structure of the manifold, however, may prevent a continuous non-contradictory designation of this sort. For example, consider a representation of the Mobius strip as shown in Figure 1.

Defining the future light cone at a point p and then taking this choice continuously over a closed loop leads us to to designate as future what we had previously labelled the past light cone at p. This is true for every point on the Mobius strip. Therefore, we need to define a notion of time orientation that keeps our physical notions of future and past consistent.

[1]More rigorously, T_pM is isometric to Minkowski space $\mathbb{R}^{3+1}$.

Definition. A Lorentzian manifold M is time-orientable if a consistent continuous choice of a future light cone can be made at every point $p \in M$.

Proposition 2.1.1. *Every Lorentzian manifold has a time-orientable double cover.*

This can be seen by constructing the double cover M' of M. For every point $p \in M$, a point $p' \in M'$ consists of one of the two choices of time orientation at p. We have two points in M' for every point in M, ensuring that M' is a double cover. Moreover, if M is not time-orientable, M' is.[2]

Proposition 2.1.2. *A Lorentzian manifold M is time-orientable if and only if there exists a nowhere-vanishing smooth timelike vector field t^α on M.*

Proof. Clearly, if such a vector field exists, we can arbitrarily label as "future" the direction in which t^α points, establishing a time orientation. Conversely, suppose that a time orientation is given on M. Since M is a metric space, it is paracompact. This implies that we can construct a positive definite Riemannian metric $h_{\alpha\beta}$ on M. At each point $p \in M$, there exists a set of future-directed timelike vectors v^α which have unit norm with respect to $h_{\alpha\beta}$ i.e. $h_{\alpha\beta}v^\alpha v^\beta = 1$. This is a closed and bounded set and hence, compact. The continuous function defined by $g_{\alpha\beta}v^\alpha v^\beta$ thus attains its minimum value over this set. Let t^α be the vector that minimizes the function. This t^α varies smoothly over M, giving us the desired nowhere-vanishing smooth timelike vector field. □

We incorporate this notion of time orientation in our definition of a spacetime.

Definition. A spacetime is a time-orientable Lorentzian manifold.

2.2 Domain of Influence

Once a time orientation has been chosen on M, we can define curves on the manifold which may represent the paths taken by objects through spacetime. A curve is a piecewise differentiable map $\gamma : A \subseteq \mathbb{R} \to M$. We shall take all curves to be suitably parameterized by an affine parameter $\lambda \in \mathbb{R}$. We call a curve *future-directed timelike* (respectively, *causal*) if its tangent vector $\frac{d}{d\lambda}$ at every point is future-directed timelike (respectively, timelike or null). Note that a curve of zero extent is causal since it has a vanishing (and hence, null) tangent vector. Analogous definitions can be written down for past-directed timelike and causal curves.[3] The point $x \in M$ is a future (respectively, past) *endpoint* of a future-directed curve if for all sequences $(a_n) \in A$, $(a_n) \to \sup A$ (respectively, $(a_n) \to \inf A$) implies $\gamma(\sup A) \to x$ (respectively, $\gamma(\inf A) \to x$). A curve with no future (respectively, past) endpoint is called future (respectively, past) *inextendible*.

We now define what are sometimes called the *domains of influence* of an event since they represent the regions that the event can influence or be influenced by.

Definition. The *chronological future* $I^+(p)$ of $p \in M$ is the set of all points $q \in M$ such that there exists a future-directed timelike curve from p to q. In this case, we write $p \ll q$.

Definition. The *causal future* $J^+(p)$ of $p \in M$ is the set of all points $q \in M$ such that there exists a future-directed causal curve from p to q. In this case, we write $p \leq q$.

The chronological and causal futures of a set $S \subseteq M$ are defined as

$$I^+(S) = \bigcup_{p \in S} I^+(p), \qquad J^+(S) = \bigcup_{p \in S} J^+(p).$$

We now attempt to find some properties of these sets. This is done to get us used to the kind of proofs used in causality theory.

[2] A rigorous proof of this can be found in Beem et. al. (1996).

[3] The definitions and results for the "past" cases shall largely be omitted in this discussion and the reader is expected to understand the analogous dual results.

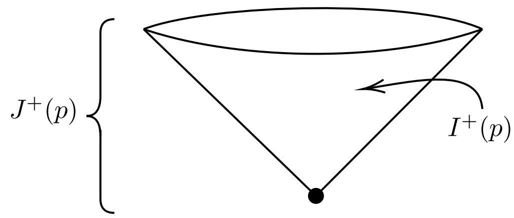

Figure 2: The chronological and causal futures of a point p in Minkowski space.

Proposition 2.2.1. *$I^+(S)$ is open for all $S \subseteq M$.*

Proof. If $q \in I^+(p)$ for some $p \in S$, then there exists a timelike curve from p to q. We can always deform the endpoint of this curve (while letting it remain timelike) to lie in an (sufficiently small) open neighbourhood O of q. Therefore, $O \subseteq I^+(p)$ and since this is true for all $q \in I^+(p)$, we can conclude that $I^+(p)$, being the union of open neighbourhoods, is itself open. Further $I^+(S)$, being the union of the open sets $I^+(p)$, is open as well.[4] $\square$

$J^+(S)$ on the other hand is not necessarily closed for an arbitrary $S \subseteq M$. For example, consider Minkowski space with the point q removed (see figure). All points on the dashed line are in $\overline{J^+(S)}$[5] but none of them are in $J^+(S)$ since no causal curve connects them to p.

Proposition 2.2.2. *$y \in I^+(x)$ if and only if $x \in I^-(y)$ and $y \in J^+(x)$ if and only if $x \in J^-(y)$.*

Proof. Let $y \in I^+(x)$. Then, there exists a future-directed timelike curve γ from x to y. Assume, without loss of generality, that $\gamma : [0,1] \to M$, $\gamma(0) = x$ and $\gamma(1) = y$. Then, define $\gamma' : [0,1] \to M$ as $\gamma'(a) = \gamma(1-a)$. Clearly, $\gamma'(0) = y$ and $\gamma'(1) = x$. The image sets of γ and γ' in M are the same, with the curves running in opposite directions. Therefore, their time orientations are opposite and γ' is a past-directed timelike curve from y to x. Therefore, $x \in I^-(y)$. The other implication can be proven by reversing this argument. The second statement can be proven by replacing timelike with causal everywhere in this proof. $\square$

Proposition 2.2.3. *$I^+(S) = I^+(\overline{S})$.*

Proof. Clearly, $I^+(S) \subseteq I^+(\overline{S})$ since $S \subseteq \overline{S}$. Let $y \in I^+(x)$ and $x \in \overline{S}$. Then, $y \in I^+(\overline{S})$ and $x \in I^-(y)$. Since $I^-(y)$ is open, there exists a point z in the neighbourhood of x such that $z \in I^-(y)$ and $z \in S$. But this implies that $y \in I^+(S)$. Hence, $I^+(\overline{S}) \subseteq I^+(S)$ and therefore, $I^+(S) = I^+(\overline{S})$. $\square$

Proposition 2.2.4. *$I^+(I^+(S)) = I^+(S) \subseteq J^+(S) = J^+(J^+(S))$.*

Proof. If a future-directed timelike curve γ connects x to y and another future-directed timelike curve γ' connects y to z, then the piecewise curve μ defined as γ between x and y and as γ' between y and z is clearly future-directed timelike. In other words, $x \ll y$ and $y \ll z$ implies $x \ll z$. Replacing timelike with causal everywhere leads us to conclude that $x \leq y$ and $y \leq z$ implies $x \leq z$. Let $x \in S$ and $y \in I^+(x)$. Then, $y \in I^+(S)$. Let $z \in I^+(y)$. Then, $z \in I^+(I^+(S))$. But $x \ll z$. Therefore, $z \in I^+(S)$. Hence, $I^+(I^+(S)) \subseteq I^+(S)$. In fact, $I^+(I^+(S)) = I^+(S)$ since $I^+(S) \subseteq I^+(I^+(S))$. The result for $J^+(S)$ can be proven analogously. Let $x \in S$ and $y \in I^+(x)$. Then, $y \in I^+(S)$ and a future-directed timelike curve exists from x to y. But every timelike curve is a causal curve. Therefore, a future-directed timelike curve exists between x and y. Hence, $y \in J^+(x)$, $y \in J^+(S)$ and $I^+(S) \subseteq J^+(S)$. $\square$

We now state a few properties without proof. If $x \ll y$ and $y \leq z$ or $x \leq y$ and $y \ll z$, we have $x \ll z$. Intuitively, this can be thought of as the statement that the interior of the light cone at x contains z, since the

[4] For a more rigorous proof, see Penrose (1972).

[5] The closure of $A \subseteq M$ is denoted by $\overline{A}$ and the boundary of A is denoted by ∂A.

boundary only contains points reached by combinations of null curves. Therefore, if we have a combination of timelike and null curves between two points, we can always find a single timelike curve between the two as well. This can be used to conclude that $\overline{J^+(S)} = \overline{I^+(S)}$, int $J^+(S) \equiv J^+(S) - \partial J^+(S) = I^+(S)$ and $\partial J^+(S) = \partial I^+(S)$. Intuitively, this makes sense since the future light cone at any point $p \in S$ has a boundary that consists of future-directed null curves emanating from p. There will always be timelike cones arbitrarily close to this boundary and therefore, the boundaries of the two sets are the same.

2.3 Future and Past

We begin with a definition of the future and past sets.

Definition. A set $F \subseteq M$ (respectively, $P \subseteq M$ is called a *future set* (respectively, past set) if $F = I^+(S)$ (respectively, $P = I^-(S)$) for some $S \subseteq M$.

By Proposition 2.2.4, $F = I^+(F)$ provides an alternative definition. This also shows that F is an open set. We now investigate some properties of future sets.

Proposition 2.3.1. $\overline{F} = \{p : I^+(p) \subseteq F\}$.

Proof. Let $I^+(p) \subseteq F$. Then, $\overline{I^+(p)} \subseteq \overline{F}$. Since, $p \in \overline{I^+(p)}$, $p \in \overline{F}$. Hence, $\{p : I^+(p) \subseteq F\} \subseteq \overline{F}$. Now, let $p \in \overline{F}$ and $q \in I^+(p)$. Then, $p \in I^-(q)$. Since $I^-(q)$ is open, there are points arbitrarily close to p that lie in $I^-(q)$. In particular, there is such a point $r \in F$. Therefore, $q \in I^+(r) \implies q \in I^+(F) = F$. Hence, $I^+(p) \subseteq F$. Finally, $\overline{F} \subseteq \{p : I^+(p) \subseteq F\}$ and $\overline{F} = \{p : I^+(p) \subseteq F\}$. □

Proposition 2.3.2. $F = I^+(\overline{F})$.

Proof. From Proposition 2.2.3, let $F = S$. Then, $F = I^+(F) = I^+(\overline{F})$. □

Proposition 2.3.3. $\overline{F} = {\sim}I^-({\sim}F)$.[6]

Proof. From Proposition 2.3.1, $\overline{F} = \{p : I^+(p) \subseteq F\}$. Let $p \in \overline{F}$. Then, $p \notin {\sim}F$. Let $q \in {\sim}F$. Then, $q \notin F \implies q \notin I^+(p)$. Therefore, $p \notin I^-(q) \subseteq I^-({\sim}F)$. Hence, $p \in {\sim}I^-({\sim}F)$ and we can conclude that $\overline{F} \subseteq {\sim}I^-({\sim}F)$. Now, let $p \in I^-({\sim}F)$ and $q \notin F$. Then, $q \in I^+(p)$. Therefore, $I^+(p) \nsubseteq F$. Hence, $p \notin \overline{F}$ and ${\sim}I^-({\sim}F) \subseteq \overline{F}$. Combining the two, we get our result. □

Proposition 2.3.4. $\partial F = ({\sim}F) \cap ({\sim}I^-({\sim}F))$.

Proof. $\partial F = \overline{F} - F = \{p : I^+(p) \subseteq F \wedge p \notin F\} = \overline{F} \cap {\sim}F = ({\sim}F) \cap ({\sim}I^-({\sim}F))$. □

Proposition 2.3.5. *If $p \leq q$, then $I^+(q) \subseteq I^+(p)$.*

Proof. Let $r \in I^+(q)$. Then, $q \ll r$. Therefore, $p \ll r$ and $r \in I^+(p)$. Hence, $I^+(q) \subseteq I^+(p)$. □

Proposition 2.3.6. $J^+(S) \subseteq \overline{I^+(S)}$.

Proof. Let $p \in S$. Then, $q \in J^+(S) \implies p \leq q$. Therefore, $I^+(q) \subseteq I^+(p)$ by Proposition 2.3.5. Furthermore, we have $I^+(p) \subseteq I^+(S)$. By Proposition 2.3.1, $I^+(q) \subseteq I^+(S) \implies q \in \overline{I^+(S)}$ since $I^+(S)$ is a future set. Hence, $J^+(S) \subseteq \overline{I^+(S)}$. □

These properties and their proofs were listed to show a glimpse of how being mathematically precise (and specifically, set-theoretic and topological ideas) can lead us to interesting physical properties of spacetime (some of which will be explored in a later section).

[6] ${\sim}A$ denotes the complement of A in M.

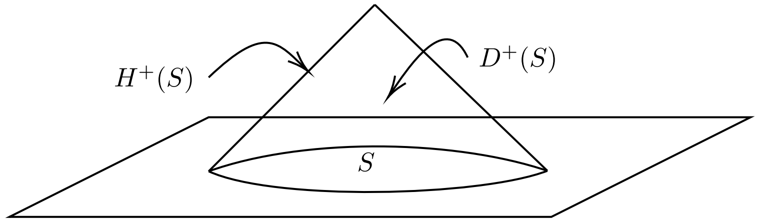

Figure 3: The future domain of dependence and the future Cauchy horizon of an achronal (and spacelike) set S.

2.4 Domain of Dependence

Next, we attempt to look at the regions of spacetime that a set can depend on. For this, we need some preliminary definitions.

Definition. A set S is called *achronal* if no two points in S can be connected by a timelike curve.

Examples of achronal sets include the spacelike surface $t = 0$ and the null line $t = x$, $y = 0$, $z = 0$ in Minkowski space.

Definition. An *achronal boundary* $B \subseteq M$ is boundary of a future set $B = \partial F$.

Proposition 2.4.1. *Every achronal boundary is also an achronal set.*

Proof. The set $I^+(\overline{F}) \cap \partial F$ is the set of points q on the boundary of F for which a future-directed timelike curve exists from $p \in \overline{F}$ to q. This includes any timelike curve among the points on the boundary. By Propositions 2.3.2 and 2.3.4, $I^+(\overline{F}) \cap \partial F = F \cap (\sim F) \cap (\sim I^-(\sim F)) = \emptyset$. Therefore, no timelike curve exists between any two points on ∂F. Hence, it is an achronal set. $\square$

We can now define the domains of dependence.

Definition. The future (respectively, past) *domain of dependence* $D^+(S)$ (respectively, $D^-(S)$) of an achronal set $S \subseteq M$ is the set of all points $p \in M$ such that any past-directed (respectively, future-directed) inextendible timelike curve[7] through p intersects S.

The domain of dependence is then the union of the two aforementioned sets. Therefore, $D(S) = D^+(S) \cup D^-(S)$. There is one more definition we need to state before looking at some properties of $D(S)$.

Definition. The future (respectively, past) *Cauchy horizon* $H^+(S)$ (respectively, $H^-(S)$) of S is the set of all points $p \in D^+(S)$ (respectively, $D^-(S)$) such that no future-directed (respectively, past-directed) timelike curve from p intersects $D^+(S)$ (respectively, $D^-(S)$) at any point other than p. In particular, $H^+(S) = D^+(S) - I^-(D^+(S))$.

The Cauchy horizon $H(S)$ is defined as the union of the two aforementioned sets. Therefore, $H(S) = H^+(S) \cup H^-(S)$. We now investigate some properties of $D^+(S)$.

Proposition 2.4.2. $I^-(p) \cap I^+(S) \subseteq D^+(S)$ *for all* $p \in D^+(S)$.

Proof. Let $q \in I^-(p) \cap I^+(S)$ and let γ be a past inextendible timelike curve from q. Since $q \in I^-(p)$, $p \in I^+(q)$. We can therefore extend γ to p and call this curve γ'. Since $p \in D^+(S)$, γ' intersects S. Since $q \in I^+(S)$ and S is an achronal surface, γ' can only intersect S at one point and this point must lie in the past of q. Thus, γ intersects S and $q \in D^+(S)$. Hence, $I^-(p) \cap I^+(S) \subseteq D^+(S)$. $\square$

Definition. A Cauchy surface Σ is a closed achronal set for which $D(\Sigma) = M$.

[7]We shall follow Penrose and Geroch in our definition, unlike Wald and Hawking and Ellis who use causal curves instead.

Proposition 2.4.3. *Σ is a Cauchy surface if and only if $H(\Sigma)$ is empty.*

Proof. Let Σ be a Cauchy surface. Then, $D(\Sigma) = M$ and hence, $H(\Sigma) = \emptyset$. Now, let $H(\Sigma) = \emptyset$. Then, $H(\Sigma)$ and hence $D(\Sigma)$ are closed. Let $p \in D(\Sigma)$. Then, since $p \notin H^+(\Sigma)$ and $p \notin H^-(\Sigma)$, there are points $q, r \in D(\Sigma)$ where $q \in I^+(p)$ and $r \in I^-(p)$. From Proposition 2.4.2, the open neighbourhood $I^-(q) \cap I^+(r) \subseteq D(\Sigma)$. Since p was arbitrary, $D(\Sigma)$, being a union of open sets, is open. Since $D(\Sigma)$ is both open and closed, $D(\Sigma) = M$. Therefore, Σ is a Cauchy surface. □

The existence of a Cauchy surface Σ allows one to, in principle, calculate the complete future and past evolution of the manifold given initial conditions on Σ. Such is the importance of this condition that many believe that the universe must be globally hyperbolic in light of this definition.

Definition. A spacetime is *globally hyperbolic* if and only if it has a Cauchy surface.

3 Causality Conditions

We now look at some physically reasonable causality conditions and explore the properties of such spacetimes.

3.1 Closed Timelike Curves

Closed timelike curves create big problems for causality. In a spacetime which admits such curves, there would be potential for paradoxical events. In particular, an observer could get on a spaceship and follow a closed timelike curve to stop themselves from getting on the spaceship in the first place! There are other reasons to posit the absence of such curves in physically reasonable spacetimes. One of them is the implications for Maxwell's theory of electromagnetism.

If a person creates electromagnetic waves at point p using some sources, then they have the complete information necessary to compute the evolution of such waves. However, if the waves propagate along a timelike curve, they would influence the sources at p before the wave was created. This would alter the initial conditions for the Maxwell equations and affect the evolution of the waves. However, the current theory of electromagnetism does not place any constraints apart from the Maxwell equations and the initial sources J^α. Thus, the existence of closed timelike curves would necessitate the development of a new theory of electromagnetism to replace the current highly successful one, at least in the region containing the curves.

3.2 The Causal Ladder

To cure such causal problems, we develop a hierarchy of progressively stronger conditions on the causal structure of spacetimes. We begin with the worst-behaved spacetime.

Definition. A *totally vicious* spacetime has a closed timelike curve at every point $p \in M$.

An equivalent condition for total viciousness is:

$$I^+(p) = I^-(p) = M, \quad \forall p \in M$$

Such a spacetime violates causality completely. Surprisingly, Einstein's equations admit such a solution - the Gödel universe with homogenous swirling dust and a negative cosmological constant. Moving towards better behaviour, we reduce the number of closed timelike curves.

Definition. A *non-totally vicious* spacetime does not have a closed timelike curve at some points $p \in M$.

We now completely remove these curves to improve causality.

Definition. A *chronological* spacetime has no closed timelike curves.

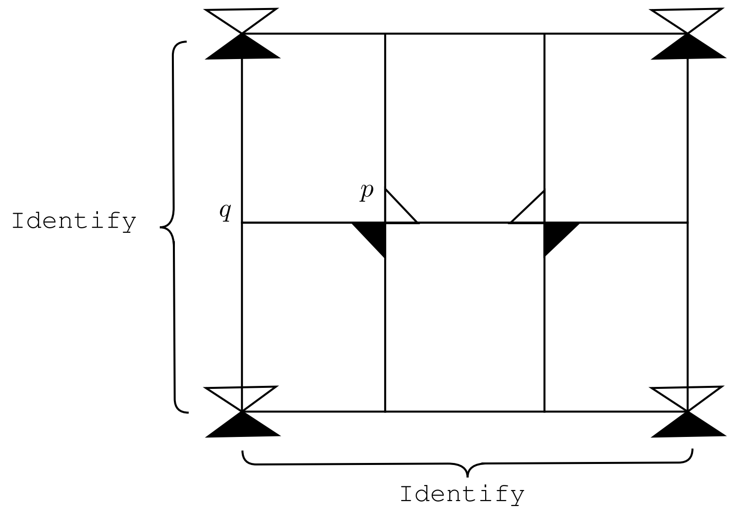

Figure 4: A non-totally vicious, non-chronological spacetime.

An equivalent condition for chronology is:

$$p \ll q \implies p \neq q.$$

Clearly, every chronological spacetime is non-totally vicious, but the converse is not true as can be seen in Figure 4. There are no closed timelike curves at point p, but a closed timelike curve passes through q. Thus, the spacetime is non-totally vicious and non-chronological. Further strengthening causality, we would like to remove all closed null curves as well so that no signal can time-travel.

Definition. A *causal* spacetime has no closed causal curves.

An equivalent condition for causality is:

$$p < q \implies p \neq q.$$

Every causal spacetime is chronological since every timelike curve is causal but the converse is not true. Figure 5 shows a chronological non-causal spacetime. There are no closed timelike curves but the central line is a closed null curve. We would now like to ensure that every point in M has a distinct future and past.

Definition. A *future distinguishing* (respectively, *past distinguishing*) spacetime satisfies the following condition:

$$I^+(p) = I^+(q) \text{ (respectively } I^-(p) = I^-(q)) \implies p = q$$

A distinguishing spacetime is both future distinguishing and past distinguishing. Every distinguishing spacetime is causal. This can be proved using the negation - every non-causal spacetime is non-distinguishing. If p and q lie on a closed causal curve in the non-causal spacetime, then $I^\pm(p) = I^\pm(q)$ and therefore, the spacetime is non-distinguishing. The converse does not hold as Figure 6 shows. Since the point p is removed, there is no closed causal curve in this spacetime although the distinct points q and r have the same future and past and therefore, the spacetime is non-distinguishing. Similarly, one can construct a past distinguishing but non-future distinguishing spacetime with the construction in Figure 7.

The currently mentioned spacetimes do not prevent causal curves that come arbitrarily close to intersecting themselves, thereby creating a closed loop. To remove the possibility of such almost closed causal curves, we define a stricter causality condition.

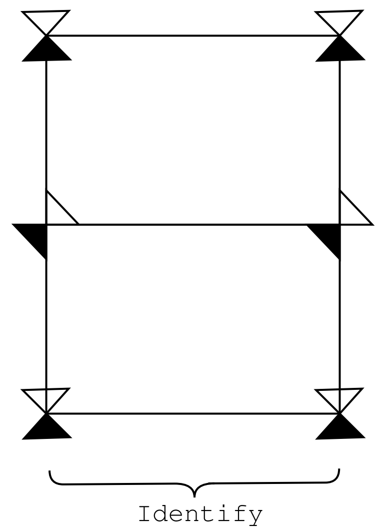

Figure 5: A chronological, non-causal spacetime.

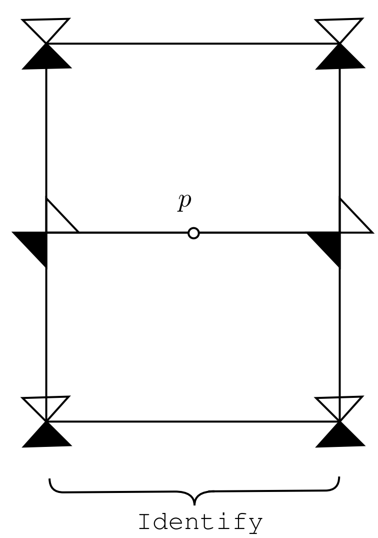

Figure 6: A causal, non-distinguishing spacetime. The point p has been removed from the spacetime.

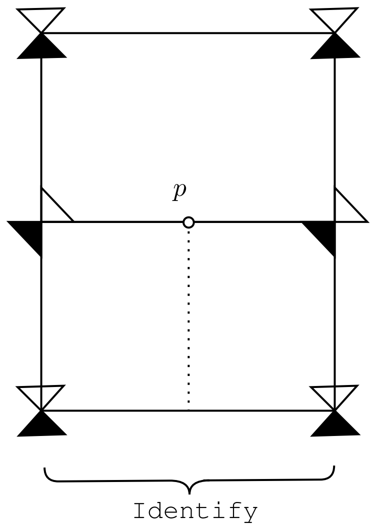

Figure 7: A past-distinguishing non-future distinguishing spacetime. The point p and the dotted line below it have been removed from the manifold.

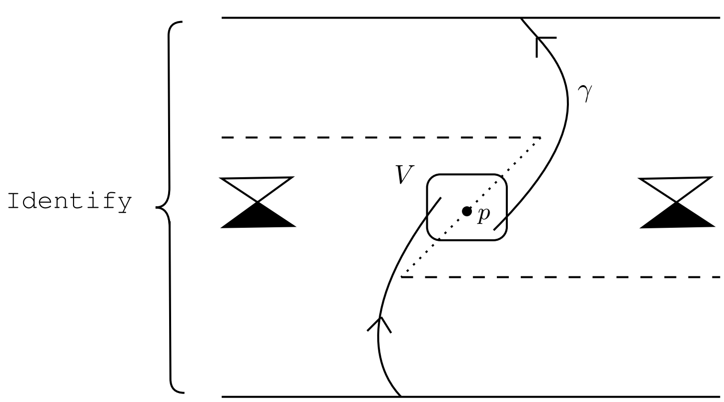

Figure 8: A distinguishing non-strongly causal spacetime. The dashed lines have been removed from the spacetime. The dotted line represents a null line through p.

Definition. A *strongly causal* spacetime has the property that given any neighbourhood U of $p \in M$, there exists a neighbourhood $V \subset U$, $p \in V$ such that any causal curve with endpoints in V lies completely in U.

This captures the required quality that no curve enters the neighbourhood V of p more than once. Thus, it will not come close to violating causality. An example of a distinguishing non-strongly causal spacetime is shown in Figure 8. The removed lines ensure that the spacetime remains distinguishing. The neighbourhood V shown contains the endpoints of the causal curve γ but the entire curve does not lie in an open neighbourhood O of p. Therefore, the spacetime is not strongly causal. An equivalent condition for strong causality is that the Alexandrov topology, constructed using a basis of the open sets $I^+(p) \cap I^-(q)$ for $p, q \in M$, is the same as the manifold topology.[8]

This is still not the strongest causality condition. To develop a stronger condition, we "open up" the light cones slightly at each point $p \in M$. If there are almost closed causal curves anywhere, such a procedure will lead to the existence of closed timelike curves. To implement this "opening up" procedure, we define at every point $p \in M$, a new Lorentzian metric $\tilde{g}_{\alpha\beta} = g_{\alpha\beta} - t_\alpha t_\beta$, where t^α is a timelike vector at p. Consider a vector v^α that was timelike or null according to the original metric. Then, $g_{\alpha\beta} v^\alpha v^\beta \leq 0$. We now find its nature according to the new metric:

$$\begin{aligned}
\tilde{g}_{\alpha\beta} v^\alpha v^\beta &= g_{\alpha\beta} v^\alpha v^\beta - t_\alpha t_\beta v^\alpha v^\beta \\
&\leq 0 - (t_\alpha v^\alpha)^2 \\
&< 0.
\end{aligned}$$

Therefore, v is timelike according to the new metric. We are now in a position to define a stronger causality condition.

Definition. A *stably causal* spacetime has a continuous nowhere-vanishing vector field t^α such that the spacetime equipped with the new metric $\tilde{g}_{\alpha\beta}$ has no closed timelike curves.

Hawking showed that this is equivalent to the existence of a cosmic time function f on M such that $\nabla^\alpha f$ is a past-directed timelike vector field. We show here that the existence of such a function implies stable causality (the converse is also true).

Proof. For every future-directed timelike vector v^α, $g_{\alpha\beta} v^\alpha \nabla^\beta f > 0$. Thus, f strictly increases along every future-directed timelike curve. Therefore, since f can never return to its initial value, there cannot be closed timelike curves in M with the given metric. We now define the new metric as $\tilde{g}_{\alpha\beta} = g_{\alpha\beta} - t_\alpha t_\beta$ where $t^\alpha = \nabla^\alpha f$. The inverse metric is $\tilde{g}^{\alpha\beta} = g^{\alpha\beta} + \frac{t^\alpha t^\beta}{1 - t^\mu t_\mu}$. To confirm,

$$\begin{aligned}
\tilde{g}^{\alpha\beta} \tilde{g}_{\alpha\beta} &= \left(g^{\alpha\beta} + \frac{t^\alpha t^\beta}{1 - t^\mu t_\mu} \right) (g_{\alpha\beta} - t_\alpha t_\beta) \\
&= g^{\alpha\beta} g_{\alpha\beta} - g^{\alpha\beta} t_\alpha t_\beta + \frac{g_{\alpha\beta} t^\alpha t^\beta + t^\alpha t^\beta t_\alpha t_\beta}{1 - t^\mu t_\mu} \\
&= \frac{1 - t^\nu t_\nu - g_{\alpha\beta} t^\alpha t^\beta - g_{\alpha\beta} t^\alpha t^\beta t^\nu t_\nu + g_{\alpha\beta} t^\alpha t^\beta + (t^\nu t_\nu)^2}{1 - t^\mu t_\mu} \\
&= \frac{1 - t^\nu t_\nu - (t^\nu t_\nu)^2 + (t^\nu t_\nu)^2}{1 - t^\mu t_\mu} \\
&= 1.
\end{aligned}$$

[8]For a rigorous proof, see Beem et. al. (1996).

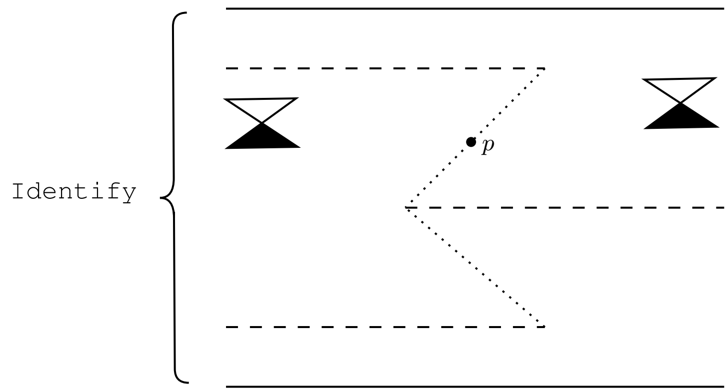

Figure 9: A strongly causal non-stably causal spacetime. The dashed lines have been removed from the spacetime. The dotted lines are null.

Then,

$$\begin{aligned}\tilde{g}^{\alpha\beta}\nabla_\alpha f \nabla_\beta f &= \left(g^{\alpha\beta} + \frac{t^\alpha t^\beta}{1 - t^\mu t_\mu}\right) t_\alpha t_\beta \\ &= t^\alpha t_\alpha + \frac{(t^\alpha t_\alpha)^2}{1 - t^\mu t_\mu} \\ &= \frac{t^\alpha t_\alpha}{1 - t^\mu t_\mu} \\ &< 0\end{aligned}$$

Therefore, $\tilde{g}^{\alpha\beta}\nabla_\beta f$ is a timelike vector under the new metric. As argued before, f is strictly monotone along a directed timelike curve and therefore, no closed timelike curves can exist in this metric. We can conclude that the spacetime is stably causal. □

Every stably causal spacetime is strongly causal but the converse is not true. In particular, the spacetime in Figure 9 is not stably causal but it is strongly causal.

Finally, the strongest causality condition is global hyperbolicity, which we have already stated in Section 2.4.

4 Conclusion

The formalism we have developed to study the causal structure of spacetime and build a causal ladder is extremely powerful. It can be used to study not only the general global properties of the geometry of spacetime but is also useful to probe the exact solutions of the Einstein equations. This study finds applications in the proofs of the singularity theorems and in taking the first steps towards quantum gravity. Must the universe be globally hyperbolic? Are the closed timelike curves in the rotating black holes in the Kerr solution physically acceptable? These questions remain a point of research today.

References

[1] R. Geroch. Domain of Dependence. *J. Math. Phys.*, 11:437–449, 1970.

[2] S. W. Hawking and G. F. R. Ellis. *The large-scale structure of space-time.* Cambridge University Press, 1973.

[3] P. E. E. John K. Beem and K. L. Easley. *Global Lorentzian Geometry.* Marcel Dekker Inc., 1996.

[4] E. Minguzzi and M. Sanchez. The causal hierarchy of spacetimes, 2008.

[5] R. Penrose. *Techniques of Differential Topology in Relativity.* Society for Industrial and Applied Mathematics, 1972.

[6] R. Wald. *General Relativity.* The University of Chicago Press, Ltd., 1984.