Chronology violations and the origin of time

Ettore Minguzzi

Dipartimento di Matematica e Informatica “U. Dini”
Università degli Studi di Firenze
Via S. Marta 3
I-50139 Firenze, Italy

Abstract. We review some results which relate chronology violations to singularities, and explain how the absence of both pathologies implies the existence of a cosmological time. Building on these mathematical ideas we then propose a causality argument in order to solve the homogeneity and entropy problems of cosmology. The solution is based on the replacement of the spacelike Big Bang boundary with a null boundary behind which stays a chronology violating region. This solution requiring a tilting of the light cones near the null boundary is based more on the behavior of the light cones and hence on causality, than on the behavior of the scale factor (expansion). The philosophical connection of this picture with Augustine of Hippo famous discussion on time and creation is commented.

1 Introduction

In Einstein’s general theory of relativity the physical spacetime is mathematically modeled as a connected Hausdorff (paracompact) time-oriented Lorentzian four-dimensional \((C', r \geq 2)\) manifold, and it is usually denoted \((M, g)\) where \(g\) is the Lorentzian metric of signature \((-1, +1, +1, +1)\).

In this work we are going to discuss the role of chronology violations using this model of spacetime. In particular we shall first recall some theorems by Kriele and Hawking on the relation between singularities and chronology violation, and then we shall consider some more recent results by the author, including a model which solves the homogeneity and entropy problems of cosmology through a causality argument. Our treatment will be quite mathematical, and we shall refer the reader to the literature for most proofs.

2 Elements of causality theory

Let us start recalling some basic definitions from causality theory. Let us recall that a tangent vector \(v \in TM\) is called timelike, lightlike (null) or spacelike depending on the values of \(g(v, v)\), respectively negative, null and positive. A curve \(\gamma : I \rightarrow M\) is said to be timelike, lightlike or spacelike depending on the causal nature of its tangent vector, provided it is consistent along the curve.

\(^{a}\)e-mail: ettore.minguzzi@unifi.it
The very definition of time-orientation requires that there is a continuous timelike vector field over $M$. At each point the half of the causal cone which contains the timelike vector of the field is called future, and a causal curve is called future-directed if at every point its tangent vector is contained in the future cone.

Two events $p, q \in M$ are said to be related

- **chronologically**, in formula $p \ll q$, if there is a future-directed timelike curve from $p$ to $q$,
- **causally**, in formula $p \leq q$, if there is a future-directed causal curve from $p$ to $q$ or $p = q$.

We shall also write $p < q$ for $p \leq q$ but $p \neq q$. These relations are respectively denoted

\[
I^+ = \{(p, q) \in M \times M : p \ll q\}, \quad \text{chronology relation},
\]
\[
J^+ = \{(p, q) \in M \times M : p \leq q\}, \quad \text{causal relation}.
\]

We recall that a relation $R \subset M \times M$ is **antisymmetric** if $(x, y) \in R$ and $(y, x) \in R$ implies $x = y$.

A subset $S$ of spacetime is said to be **achronal** if there is no timelike curve which connects a pair of its points. An important theorem establishes that any causal curve which is not an achronal causal curve, and hence a lightlike geodesic, can be deformed to a timelike curve keeping endpoints fixed [1, 2]. As a consequence the set $E^+ = J^+ \setminus I^+$, is made exactly by the pairs which are connected by an achronal lightlike geodesic. This relation is called **horismos** and contrary to $I^+$ and $J^+$ is not transitive. The relation $I^+$ can be shown to be open on $M \times M$, however, $J^+$ need not be closed, contrary to what one could naively expect from the example of Minkowski spacetime. A simple example is provided by Minkowski spacetime minus a point, but more refined examples with a trivial topology can be found. The gravitational plane waves provide such examples [3, 4].

The **chronology violating set** is

\[
vI = \{x : x \ll x\},
\]

and is made by all the events through which there passes a closed timelike curve. The spacetime violates chronology if $vI \neq \emptyset$, or which is the same, if there is a closed timelike curve.

Suppose $vI \neq \emptyset$, then $vI$ can be split into equivalence classes according to Carter’s equivalence relation

\[
x \sim y \Leftrightarrow x \ll y \text{ and } y \ll x.
\]

Two points belong to the same class if there is a closed timelike curve passing through them. The class of $x \in vI$ is denoted $[x]$. Note that $[x] = I^+(x) \cap I^-(x)$, thus $[x]$ is open.

A **time function** is a continuous map $f : M \to \mathbb{R}$ which increases over every causal curve, that is $x < y \Rightarrow f(x) < f(y)$. We shall write $g' > g$ if the light cones of $g$ are contained in the timelike cones of $g'$. Let us observe that $(M, g')$ has more causal curves than $(M, g)$ since a causal curve for $g$ is causal for $g'$. In particular, widening the light cones on spacetime might introduce closed causal curves and hence spoil causality. This observation leads us to the definition of **stably causal** spacetime, which is a spacetime $(M, g)$ for which there exist $g' > g$ such that $(M, g')$ is causal. Equivalently, a spacetime is stably causal if causality is stable according to a topology on Lorentzian metrics generated by the base made of the sets $(g, \overline{g}) := \{g : g < g < \overline{g}\}$ (this is Geroch’s interval topology).

Let us also observe that widening the light cones enlarges the causal relation. Let us introduce Seifert’s relation $J_{S}^+ = \bigcap_{g' > g} J_{g'}^+$. This relation is particularly interesting because it is closed, transitive and contains $J^+$. In fact, under stable causality it is the smallest relation with this property [5]. An important theorem [6–10] establishes that (a) stable causality, (b) the existence of a time function, and (c) the antisymmetry of $J_{S}^+$ (partial order), are equivalent.

A spacetime can satisfy one of several causality conditions. The most important are

- **Chronology**: There are no closed timelike curves (i.e. $I^+$ is antisymmetric),
Figure 1. This figure summarizes the causal ladder (vertical shape) and the transverse ladder (horizontal shape) of causality conditions. The levels of the causal ladder below stable causality protect against causality violation. Those above stable causality protect against influence from infinity.

- **Causality:** There are no closed causal curves (i.e. $J^+$ is antisymmetric),
- **Stable causality:** Causality holds and is stable in a Geroch’s interval topology on Lorentzian metrics (i.e. $J^+_S$ is antisymmetric),
- **Global hyperbolicity:** Causality holds and the causal diamonds $J^+(p) \cap J^-(q)$ are compact.

In fact, there are other interesting causality condition which organize in a hierarchy of properties [5, 11], see Fig. 1. Most researchers are interested in the strongest of these properties, namely global hyperbolicity. One justification could be that under global hyperbolicity many of the partial differential equations which determine the behavior of matter fields admit unique solutions. Another justification is that this is a good working hypothesis which allows analyst to do calculations. However, this strong requirement is not a consequence of Einstein equations which in fact admit exact solutions with causally pathological features (e.g. gravitational plane waves; Gödel and Kerr solutions). As Hawking puts it, the very existence of these solutions might be telling us something physical on the nature of spacetime that we might keep ignoring investigating just the nicer spacetimes. In particular, it seems impossible to provide arguments in favor of nicely behaved spacetimes without studying the bad ones. Here we shall be interested in those spacetimes which are so causally pathological that they do not to enter the displayed ladder of causality conditions. In fact, we shall be mostly interested on those spacetimes which violate chronology.

### 3 The role of lightlike lines in singularity theorems

A curve $\gamma : (a, b) \to M$ on spacetime is said to be *inextendible* if it has no point (endpoint) to which it converges for $t \to a, b$. A *lightlike line* is an inextendible achronal causal curve and hence a lightlike geodesic.
One key strategy to prove a singularity theorem is to look for lightlike lines. If under some assumptions we are able to infer the existence of a lightlike line then we can also infer the incompleteness of spacetime. Let us show more closely this connection discovered by Hawking and Penrose [12].

Let $\gamma: (a, b) \rightarrow M, t \rightarrow \gamma(t)$ be a lightlike geodesic of a congruence (beam). We can consider the behavior of the displacement between two geodesics and infer corresponding evolution equations for the shear $\sigma$ and expansion $\theta$. Let

$$F(t) = \frac{1}{2} R(\gamma', \gamma') + \sigma^2.$$

The Raychaudhuri equation expresses the evolution of the expansion and is usually written in the Riccati form

$$\frac{d\theta}{dt} = -2F - \frac{1}{2} \theta^2.$$

If we define $y$ so that $\theta = \frac{2}{y} \frac{dy}{dt}$, where $y$ represents the (average) size of the beam, the previous equation can be put in the form

$$\ddot{y} = -F(t)y.$$

This is the equation of an elastic spring for which the elastic constant depends on time. The mechanical analogy allows us to grasp very easily that if $F \geq 0$ and $F(t_0) > 0$ for some $t_0$, then there are solution with 2 zeros. Indeed, in the mechanical model, consider the point mass at rest at time $t_0$ at an arbitrary distance from the origin of the spring. Evidently, the spring force will act on the point so as to impart it a velocity towards the origin. The same conclusion is reached looking at this mechanical problem in the backward time direction. The two crosses so obtained of the origin correspond to the two zeros of the found solution. Observe that this zeros exist only if the domain $(a, b)$ of $t$ is sufficiently extended which is assured if $\gamma$ is complete. In our spacetime problem the existence of such solution implies that $\gamma$ admits a pair of conjugate points, namely points where the beam focuses.

Now, it is well known that in a Riemannian manifold an inextendible geodesic minimizes the path length between any of its two points only if it does not admit a pair of conjugate points. Analogously, in a Lorentzian manifold the presence of conjugate points over an inextendible geodesic signals that there is a longer \footnote{Instead of shorter as in the Riemannian case. This is due to the sign in the Lorentzian metric} path between two of its points, namely it is not achronal and hence is not a lightlike line. With these preliminaries we can easily understand the next result.

**Theorem.** [Hawking and Penrose’s (1970) [12]]

If a null geodesically complete spacetime satisfies

(a) $R(n, n) \geq 0$ for every lightlike vector $n$ (null energy condition),

(b) $n^a n^d n^e R_{[bc]} n^e n^f \neq 0$ somewhere over every null geodesics, where $n$ is the tangent to the geodesic (null genericity condition),

then every geodesic has two conjugate points, so it cannot be achronal, and in particular it cannot be a lightlike line.

The connection of this theorem with the argument above is understood noticing that the first condition assures $F \geq 0$, while the second condition assures that $\sigma^2$ becomes positive somewhere, and hence $F$ becomes positive at some $t_0$. These conditions are also physically quite reasonable. Recall that the Einstein equations are

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}.$$
If the energy density for an observer of 4-velocity $u^\mu$ is non-negative then $T_{\mu\nu}u^\mu u^\nu \geq 0$, and since this is a reasonable assumption (in a non-quantum mechanical regime) then this inequality should hold for every timelike vector $u^\mu$ and hence, by continuity, for every lightlike vector $n^\mu$. Using the Einstein equations and $T_{\mu\nu}n^\mu n^\nu \geq 0$ we get the assumption (a) which can therefore be regarded as a consequence of the positivity of the energy density. Actually this assumption can also be weakened to the averaged null convergence condition [13–15] which allows for local violations of the positive energy condition.

In order to better understand the role of the shear, let us consider its evolution equation [2]

$$\frac{d\sigma_{mn}}{dt} = -C_m{}^4 n^4 - \theta \sigma_{mn} - \sigma_{mp}\sigma_{pn} + \delta_{mn}\sigma^2.$$

If the condition “$F(t) > 0$ for some $t$” does not hold then $\sigma_{mn}(t) = 0$ for every $t$ and hence $C_m{}^4 n^4 = 0$ over $\gamma$ (the condition (b) of the above theorem is essentially equivalent to $C_m{}^4 n^4 \neq 0$ somewhere on $\gamma$). That is, if the genericity condition fails, the spacetime is quite special. This is mathematically expressed by the rigidity theorems e.g. [16, Theorem IV.1]. We shall see that this rigidity will play a crucial role in our solution to the entropy problem of cosmology.

### 3.1 Kriele’s theorem

As an application of these ideas let us illustrate Kriele’s theorem on the null geodesic singularity of spacetimes admitting a chronology violating set which has compact closure or has compact boundary.

**Theorem.** [Kriele’s (1989) [17]]

Suppose that $(M, g)$ satisfies the null energy condition and the null genericity condition. If a connected component of the boundary of the chronology violating set $vI$ is compact, then $M$ is null geodesically incomplete.

**Sketch of proof.** Let $b \in vI$, and let us consider a point $r$ on the boundary of the class $[b]$. There is a sequence of closed timelike curves that accumulates on $r$. Each element of the sequence can be considered as an inextendible curve, it suffice to choose the parametrization so as to cover the image of the curve several times.

**Figure 2.** The role of lightlike lines in Kriele’s theorem. A: Through every point of the boundary passes a future or past lightlike ray which stays in the boundary. B: If the boundary is compact then it contains a lightlike line.

By the limit curve theorem [18, 19] there is an inextendible causal curve which passes through $r$ and stays in $[b]$. It is easy to prove that either the half-curve in the future of $r$ is achronal and contained
in $\partial[b]$ or so is the half-curve in its past. Indeed, if not since $I^{+}$ is open there would be two points $p$ and $q$ such that $p \ll r \ll q$ with $p, q \in [b]$, from which we obtain $r \in [b]$ while $r$ belongs to the boundary of this open set, a contradiction (see Fig. 3-A).

We conclude that there is a future (or past) inextendible achronal curve (hence a lightlike geodesic, often called ray) in the boundary $\partial[b]$. Now, if $\partial[b]$ were compact, which is the case if $[b]$ is compact, then the ray accumulates over an inextendible causal curve in $\partial[b]$ (see Fig. 3-B). This causal curve must be achronal, and hence a lightlike line, or one could easily infer that the ray was not achronal, a contradiction. By the null convergence condition and the null genericity condition this curve is incomplete, for if not it could not be a line, which concludes the proof.

Thus creating a region of chronology violation (time machine) implies that either we break the spacetime continuum creating a singularity (e.g. a black hole) or that we cannot control the violating set in a limited spacetime region, that is: there is no way to switch off the time machine.

As an illustration of this theorem, consider the following spacetime example obtained from cut and paste in Minkowski $1+1$ spacetime.

![Identify](Figure 3. Example in $1+1$ Minkowski spacetime after some cut and paste.)

As we can see, the boundary of the shaded chronology violating set is generated by lightlike rays, as the first part of the argument in Kriele’s proof reveals, but there is no lightlike ray there which implies, as we can check, that the boundary of the chronology violating set is not compact.

### 3.2 Hawking’s theorem (1992)

Another result on chronology violation which is worth recalling is a theorem by Hawking which appears in the famous 1992 work [20] where he formulates the chronology protection conjecture.

Let us recall that a partial Cauchy hypersurface $S$ is an acausal edgeless (and hence topologically closed) set. The future Cauchy development $D^{+}(S)$ is the set of all the events $p$ for which any past-inextendible causal curve ending at $p$ intersects $S$. The future Cauchy horizon $H^{+}(S) = \overline{D^{+}(S)} \cap \mathcal{J}^{+}(D^{+}(S))$ represents the boundary of the region whose physical properties are determined from data on $S$. The loss of predictability can have many origins. One of those is the possible formation of closed timelike curves in the future of $S$. If this happens then we expect that this origin of the unpredictable behavior propagates, so to speak, along the generators of the horizon. Hawking make the following crucial observation: if the first closed timelike curves are created by a kind of advanced civilization, then as they were not already there, they must be contained in some compact region $K$ of spacetime where that civilization performed its experiments. If so the generators of $H^{+}(S)$, if followed into the past direction, must enter this compact region where all the unpredictability started from. That is, $H^{+}(S)$ must be compactly generated. However, Hawking was able to prove.
Figure 4. Example in Minkowski spacetime after some cut and paste.

**Theorem. Hawking’s (1992)**

Let $(M, g)$ be a spacetime admitting a non-compact partial Cauchy hypersurface $S$ such that $H^+(S)$ is smooth and compactly generated. Then the null energy condition is violated.

A way out of this result is to assume that $K$ is not compact after all, for instance this could happen if the generators enter a singularity region, that is if the spacetime continuum is broken as the closed timelike curves are created.

Thus, accordingly to Hawking’s theorem, we cannot create a region of chronology violation (even unbounded) without breaking the spacetime (singularity) or violating the classical positive energy condition. Moreover, Hawking claims that, even including quantum mechanics, there is evidence which supports the

**Chronology protection conjecture:** The laws of physics do not allow the appearance of closed timelike curves.

We stress that the conjecture, if true, strongly limits the formation of closed timelike curves, and hence the possibility of building time machines. For, it essentially implies that a technology leading to the creation of time machines would imply either the generation of singularities (under the validity of the strong cosmic censorship one can think of a black hole) or the necessity to use forms of exotic quantum matter, where the ability of this type of source to violate the classical energy inequalities is not yet clear.

In any case, the validity of the chronology protection conjecture does not imply the absence of chronology violating classes which exist since the beginning of the Universe, as it puts limits only on their formation. This fact will play a role in our solution to the entropy and homogeneity problems of cosmology.

**4 A theorem on the origin of time**

In this section I present a theorem, still derived from the study of lightlike lines in spacetimes, which gives new insights on the origin of time and the beginning of the Universe. As mentioned in the previous section Hawking’s theorem leaves open the possibility that there were chronology violating classes since the beginning of the universe. Suppose there is one class $[r]$ which generates the whole Universe, that is, $M = I^+([r])$, then this class is unique. For, suppose $r'$ is such that $M = I^+([r'])$ then
\( r \gg r' \) and \( r' \gg r \), that is \( [r'] = [r] \). I proved [21, 22].

**Theorem** *(Time begins just outside the chronology violating set at the beginning of the universe).*

Let \((M, g)\) be a spacetime which is either chronological or admits a chronology violating class which generates the whole universe, i.e. \( M = I^+(r) \). Assume that on every inextendible geodesic entirely contained in \( M \setminus [r] \), the null energy condition and the null genericity condition hold true. Then either there is a null incomplete geodesic contained in \( M \setminus [r] \), or the spacetime \( M \setminus [r] \) is stably causal and hence admits a time function.

**Sketch of proof.** The proof idea is in the simpler case \([r] = \emptyset\) is as follows: under null completeness there are no lightlike lines. Using a limit curve argument it is then possible to show that the relation \( J^+ \) is transitive and hence coincident with \( J^+_5 \). Furthermore, it is not hard to show that under chronology \( J^+ \) is antisymmetric, thus \((M, g)\) is stably causal and admits a time function.

An inspection of the available singularity theorems shows that none of them compellingly implies the null geodesic past singularity of spacetime. As a consequence, we might well consider the possibility that spacetime is null geodesically complete in some neighborhood of the Big Bang. Philosophically this assumption is quite interesting: it suffices to recall that all matter was in a lightlike state before the Higgs boson underwent the phase transition that made massive all particles. Thus to require null completeness means to ask for an ultimately physically non-singular type of matter.

Then, taking null completeness for granted, the previous theorem suggests a picture for the Universe in which the so called Big Bang is identified with the null hypersurface \( \mathcal{B} := \partial [r] \), namely with the lightlike slice which separates the chronology violating set from the remainder of the Universe which follows causally from it. For simplicity we shall assume that \( \mathcal{B} \) is compact.

![Diagram of a Universe with \( S^1 \) section](image)

**Figure 5.** A Universe with \( S^1 \) section which gives an idea of the cosmological picture presented in this work. The region that admits time functions is causally preceded by the chronology violating set (\( \text{God} \)) as in St. Augustine’s conclusions.
The physical content of this theorem recalls Augustine of Hippo’s conception of time and creation [22, 23] according to which God (\( r \)) created the whole Universe (\( M = I^*([r]) \)) and time (\( t \)) began to exist at the act of creation (\( B \)) and throughout the created world (\( M \setminus (B \cup C) \)), but still outside God which indeed, in the middle age philosophical thought, was usually considered atemporal [22, 24].

### 4.1 Causality solution to the homogeneity and entropy problems of cosmology

This kind of picture allows us to solve almost trivially the homogeneity problem of cosmology. Indeed, if \( \partial [r] \) is compact then it is generated by (almost) closed lightlike geodesic. This fact implies that the events in \( \partial [r] \) share the same chronological future and hence that they influence the same events in the future. One should expected then that at different points of spacetime and looking at different directions, the same temperature in the last scattering hypersurface is measured. Indeed, all the events in the last scattering hypersurface are influenced by the same events at the Big Bang. While homogeneity is immediate one has instead to work out the inhomogeneities, and it seems that they might depend on the model used for evolution.

The entropy problem requires some introduction. This difficulty of standard cosmology arises when considering the huge difference between the entropy of the Universe today with that at the time of the Big Bang. R. Penrose [25] by taking into account also the gravitational entropy, has argued that the Universe at its beginning had probably to be thermalized, to account for the homogeneity problem, but nevertheless it had to be special as the calculation of the entropy shows that it was much smaller than today. In his view the Universe could increase in entropy despite its initial thermalization because in the beginning the gravitational degrees of freedom were almost frozen.

By the way, Penrose reaches this conclusion clarifying a common misconception that attributes the initial small entropy of the Universe to its size. Penrose shows that this position is untenable by considering potentially recontracting universes.

With the aim of justifying the special nature of the beginning of the Universe, Penrose concludes that the gravitational degrees of freedom had to be in a very special state. From his point of view the Universe could increase in entropy despite its initial thermalization because in the beginning the gravitational degrees of freedom were almost frozen.

He also notes that when matter is left to the action of gravity it tends to clump, passing from an homogeneous to an inhomogeneous state. The Weyl tensor increases because of this clumping, and therefore this tensor may quantify in some sense the amount of entropy contained in the gravitational degrees of freedom. Thus Penrose ends suggesting that in the beginning of the Universe the Weyl tensor had to be very small, and possibly zero. This is Penrose’s Weyl tensor hypothesis [25, 26].

In order to accomplish a vanishing Weyl tensor at the boundary Penrose needs to assume that our Big Bang represents, up to oconformal rescaling, the Big Crunch of a previous era, an idea that leads him to formulate a kind of cyclic cosmology.

However, the picture of the beginning of the Universe presented in this work justifies Penrose’s Weyl tensor hypothesis without requiring such identification. Indeed, our idea that the Big Bang has to be the lightlike boundary of a chronology violating region, implies, under smoothness and compactness conditions, that the generators of the Big Bang lightlike slice are achronal. Under our assumption of null completeness we infer that the generators have to be lightlike lines and hence, according to the arguments of previous sections, \( C_{ab4b4} = 0 \) all over \( B \). This equality shows that indeed the Weyl tensor hypothesis holds true, at least for what concerns some degrees of freedom of the Weyl tensor.

Moreover, there is a rigidity result [18] which states that an asymptotically simple vacuum space-times is isometric to Minkowski spacetime in a neighborhood of every achronal lightlike geodesic.
(Galloway’s null splitting theorem [16]). I expect that analogous results should hold for the case considered in this work, that is, I expect the spacetime near the boundary of the chronology violating region to be isometric to some highly symmetric spacetime. This rigidity would clearly fix the Weyl tensor and thus send to zero most of the degrees of freedom contained in it.

References