

The Class of Causally-Concerned Objects Can Confuse

Brendan Foster

University of Maryland, College Park, MD

bzf@physics.umd.edu

Abstract

I review the classic papers of Hawking, King, and McCarthy, and Malament that demonstrate a close relationship between the topological properties of a spacetime and the causal connectivity of points within it.

1 INTRODUCTION

With the advent of general relativity came the realization that the world might have a non-trivial network of causal connections, or futher, the thought that the world might have *any* causal structure that existed independently of the objects through which the patterns of causation might be observed. A grand prophet of causal structure must be Stephen Hawking, and in 1975, with A. R. King and P. J. McCarthy, he announced ‘A new topology for curved space-time which incorporates the causal, differential, and conformal structures’ [1]. Shortly after, David Malament followed this by announcing, ‘The class of continuous timelike curves determines the topology of space-time’ [2].

Despite the demonstrative title, Malament’s paper, like that of Hawking, King, and McCarthy (HKM), does not forcibly convey the subtleties of its meaning. With this essay, I wish to review these works, with an eye to extruding their content. I make no wild claims; I wish only to convey the original ideas. I begin by discussing two concepts central to the meaning of the results: conformal diffeomorphisms and the meaning of the term ‘causal structure’. From there, I discuss the paper of HKM, followed by that of Malament. Regarding any terminology left undefined in the discussion, I have followed the conventions of HKM and Malament, which derive primarily from conventions in Hawking and Ellis [3].

1.1 Conformal Diffeomorphisms

A spacetime consists of a smooth manifold (with certain other properties, see [3]), and an at least twice-differentiable Lorentzian metric. Now, a diffeomorphism just maps points on one manifold to points on another, while the tangent space at a point (including the metric) is mapped to the tangent space at the image point by a uniquely associated isomorphism (see [3], pp 22-24). Given a diffeomorphism between two spacetime manifolds, one can then meaningfully compare their respective metrics by comparing, in the tangent spaces of one manifold, the metric of that spacetime and the inductively mapped metric of the other.

The common adage, ‘Diffeomorphic spacetimes are physically equivalent,’ implicitly requires that the manifolds be diffeomorphic *and* that the metrics be equivalent via the induced mapping. In parlance, such a diffeomorphism, that ‘preserves the metric’, is an isometry. A homothetic diffeomorphism equates two spacetime manifolds for which the mapped metric of one is equal to the metric of the other up to a globally constant factor. A conformal diffeomorphism equates two spacetimes with metrics equal up to a position dependent factor. Note that a conformal class of spacetimes, all spacetimes equivalent under a conformal diffeomorphism, can include many distinct physical-equivalence classes.

1.2 Causal Structure

By the ‘causal structure’ on a manifold, one often means abstractly a set of binary relations between points, with characteristics that intuitively capture the notion of ‘is lightlike related’ or ‘is timelike related’ or some other suitable

sense (see Kronheimer and Penrose [4] for a precise, technical foray). In this way, one can define the causal structure without the explicit presence of a metric. However, when it comes time to express the *physical* meaning, or even just the geometric meaning in the attempt to associate a metric to the structure, one must conclude as the references cited here do, that ‘p relates to q’ means that p connects to q via a suitable causal curve, as defined by the Lorentzian geometry.

To be precise, HKM and Malament declare a point q to be in the causal future (resp. past) of point p iff there is a smooth, future-(past-)directed, causal curve—that is, a continuously differentiable curve whose tangent vector at every point along the curve does not point out of the future (past) null cone at that point. One can also consider the set of points connected by timelike curves, and in this discussion, unless preceded by ‘causal’, the terms ‘future’ and ‘past’ will designate these sets. There are various other equivalent ways to define these properties (see [5], sec. 2); for instance, any two points connected by a causal curve that is not a null geodesic can also be connected by a timelike curve.

A mapping ‘preserves the causal structure’ iff the image of the future (and past) of a point equals the future (and past) of the image of that point. Sets of the form, $A(p,q) = \{ z : z \text{ is in the future of } p \text{ and the past of } q \}$ constitute a topology on a spacetime manifold (the Alexandrov topology, see [5] pp 33-34), and any one-one, onto causal-structure-preserving mapping is a homeomorphism of this topology. The Alexandrov topology of a spacetime is equivalent to the standard manifold topology iff the spacetime is strongly causal—about every point, there is some open set such that no future- or

past-directed timelike curve leaves the set and then returns, meaning that no path comes arbitrarily close to being a closed timelike curve.

HKM and Malament adapt the notion of ‘causal’ (and ‘timelike’ in the obvious sense) so that one can designate a non-differentiable, but continuous, curve as such: intuitively, at every point on the curve, the immediately preceding points of the curve lie in the causal past of that point, and the immediately succeeding points lie in the causal future. Two points are connected by a non-smooth timelike curve iff they are connected by a smooth timelike curve, so we reach the actual working sense of ‘causal structure’: two points are (time-wise) causally connected iff they are connected by a continuous timelike curve.

With this formulation, we now see that any one-one onto mapping between spacetimes that sends continuous timelike curves to continuous timelike curves, regardless of whether or not smooth curves go to smooth curves, preserves the causal structure (it is probably not obvious that points connected by null geodesics remain so connected under such a mapping, but by the results below, they do). As a result, any such mapping is a homeomorphism of the A-topology; that the converse is false, we shall discuss below.

2 WORKS

2.1 “A New Topology For Curved Spacetime...”

HKM include ‘curved spacetime’ in their title presumably to emphasize that the paper was meant to improve upon results of E. Zeeman [6, 7] involving

flat spacetime. Zeeman—as well as A. Alexandrov [8], although he is not credited by HKM—had shown that the group of one-one, onto mappings of Minkowski space that preserve (or universally reverse) causal relations is just the flat conformal group—the group of homothetic diffeomorphisms of Minkowski space—, consisting of the Lorentz group, translations, and global dilations. Zeeman also established that one could define a non-standard topology on Minkowski space, generalizable to curved space, such that the homothety group is equal to the group of homeomorphisms of this new topology.

HKM construct a ‘new topology’ for which, at least in a strongly causal spacetime, the group of homeomorphisms equals the group of conformal diffeomorphisms of the standard topology. This ‘path’ topology has the defining quality of being the finest such that the induced topology on any timelike path is equal to the topology induced by the standard topology, where a ‘finer’ topology allows as an open set every open set of a ‘coarser’ topology. HKM explicitly define this topology by giving its basis of open sets: for some open convex set U containing point p , take the intersection of all points in U reachable from p by a timelike curve contained entirely in U (the future and past of p in U), and a Euclidean open ball about p contained in U ; then add the point p . The presence of p in a basis set means that that set is not open in the standard topology; however, as shown by HKM, every standard open set is open in the $P(\text{ath})$ -topology, making it strictly finer, hence non-homeomorphic to, the standard topology.

2.2 “...Which Incorporates the Causal, Differential, and Conformal Structures”

As alluded to above, the ultimate result of the paper is the

Theorem HKM: For a strongly causal space-time, the group of homeomorphisms of the path topology is equal to the group of conformal diffeomorphisms of the standard topology.

To demonstrate the equivalence of the topologies, HKM first show that P-homeomorphisms preserve timelike paths, that is, every timelike path is mapped to a timelike path. By the discussion above, the P-homeomorphism is then an A-homeomorphism. Under the assumption that the spacetime is strongly causal, so that the A-topology and the standard topology are equivalent, the P-homeomorphism is a standard homeomorphism.

For the next step, HKM show that in the event of strong causality, P-homeomorphisms preserve null geodesics—essentially because such paths can be expressed in terms of causal relations (p connects to q via a null geodesic iff it connects via a null curve but not a timelike curve), preserved by the mapping. This leads to a proof, originally given ten years earlier by Hawking [9], inspired by a proof of Zeeman [6]:

Theorem H : A homeomorphism of the manifold topology that takes null geodesics to null geodesics is a diffeomorphism¹.

¹Two comments. First, both Zeeman’s and Hawking’s proofs are not valid in two dimensions. Second, in both the original version of this proof and the ‘improvement’ in the HKM paper, the theorem actually appears as a sub-theorem amidst several other statements that explicitly require the assumption of strong causality, and this particular

In both versions of the proof, the respective authors define co-ordinate charts via null geodesics. The proofs demonstrate that a particular smooth, co-ordinate-defining function (such as the parameter of a geodesic, in the earlier case) is mapped to a smooth counterpart so that smooth co-ordinates are mapped to smooth co-ordinates, thus declaring the map a diffeomorphism. Hence, a P-homeomorphism is a standard diffeomorphism; any two spacetimes that possess equivalent P-topologies must be diffeomorphic.

At the next step enters a piece of wisdom (see [3], pg 61), telling that the metric at a point can be inferred, up to a constant, by knowing the separation of vectors in the point's tangent space into spacelike, timelike, and null categories. The tangent null cone can in turn be inferred from the null geodesics on the manifold², so from them the metric can be reconstructed, up to a varying factor. Since, under the assumed conditions, the mapping of interest preserves the cones, it must connect two conformally equivalent metrics. Thus, every P-homeomorphism is a conformal diffeomorphism; two spacetimes with equivalent P-topologies must have conformally equivalent metrics.

clause is proven in this context. However, both versions read abstruse enough that I could not conclusively determine whether or not the assumption is necessary to prove this particular clause. I suspect that it is not; if it is, I sense that it is necessary to ensure that one can define coordinate charts in the way integral to the proofs. I mention this because Malament also makes use of this theorem but without making the assumption. As it seems to me feasible that he is right to do so, I will take his results as correct, but I will make clear where this theorem enters into his arguments.

²...but only in the event that the manifold is geodesically complete?

As a last step, HKM cover a result that holds on a general spacetime. Since a conformal diffeomorphism, g , preserves the standard topology and timelike paths, then the intersection of any standard open set, O , with a timelike path, γ , will again be the intersection of some standard open set with a timelike path. A P-open set, E , is any set such that its intersection with some timelike path, γ , is equal to $O \cap \gamma$, for some standard open O ; thus, $g(E) \cap g(\gamma) = g(O) \cap g(\gamma)$. By the above note, $g(E)$ must be P-open, so a conformal diffeomorphism is a P-homeomorphism; hence, Theorem **HKM**.

2.3 “The Class of Continuous Timelike Curves Determines the Topology of Spacetime”

Malament’s work brings to the foreground a property underlying all of HKM’s results. This is Malament’s

Theorem M1: A one-one, onto mapping between spacetimes that preserves future directed timelike curves is a homeomorphism of the manifold topology.

Lemma and Corollary: Such a mapping also preserves null geodesics, so by **H**, is a conformal diffeomorphism.

Malament works with a notion of continuity based on convergence of sequences: an infinite sequence of points converges to a point p iff every open set containing p contains an infinite number of points of the sequence; a map is continuous at p iff the image of every sequence converging to p converges to the image of p . Malament begins his proof by deriving several properties of the set of points of discontinuity, in the case that this set were

non-empty; in particular, this set would intersect some open set achronally—such that no two points in the intersection were timelike related.

Malament then exploits the lemma on preservation of null geodesics by considering, for some sequence convergent on a ‘bad’ point p , any null geodesic Ω through p and a sequence of null geodesics, one through each point in the sequence, that converges on Ω (‘in the sense that every open set which intersects Ω intersects eventually all [the converging geodesics]’ [2]); the mapping will preserve the null geodesic character of these curves. Malament can then argue that the properties of the ‘bad’ set, the converging geodesics, and the non-converging image sequence imply the existence of distinct causal curves through p contained in the achronal intersection. However, an achronal set can contain at a point at most one null curve and no timelike curves, hence a contradiction, hence the ‘bad’ set is empty.

Since HKM show that in general, a P-homeomorphism preserves timelike paths and a conformal diffeomorphism is a P-homeomorphism, with **M1** we have a generalization of **HKM**:

Theorem HKMM: For a general spacetime, the group of homeomorphisms of the Path topology is equal to the group of timelike path-preserving maps is equal to the group of conformal diffeomorphisms of the standard topology.

2.4 The Causal Structure Alone Does Not Always Determine Such Things

As for the question of when a mapping that preserves the causal structure also preserves continuous timelike paths, Malament proves that if the space-

time is past and future distinguishing (p and q have the same past and future iff $p = q$), then:

Lemma: A one-one, onto mapping, f , between past and future distinguishing spacetimes, such that the image of the future (past) of any point is the future (past) of the image of that point maps continuous timelike curves to continuous timelike curves.

Theorem M2: From **M1**, such a mapping is a homeomorphism of the standard topology.

Corollary: By **H**, such a mapping is a conformal diffeomorphism.

The proof rests on an equivalent characterization of past and future distinguishability—about every point is an open set such that no future or past directed timelike curve *through* p leaves the set but returns to it (thus, strong causality implies pf distinguishability). This allows one to show, essentially, that about every point p on a timelike curve, the images of other arbitrarily close points on the curve must still be arbitrarily close to the image of p : in one of the ‘good’ sets about p , one would have a point that contained in its past p and some of the points on the curve; if the image points were not still close, one would have to have a timelike path connecting q , p , and the curve points, that left the neighborhood of p only to return to pass through q .

Malament illustrates the failure of **M2** to hold in the case of non past and future distinguishing spacetimes with a canonical example (see variations in [5] p 28 and [3] p 193). The 2-d plane is rolled into a cylinder, with the spacelike direction compact; the metric is conformally Minkowskian at $t = -\infty$, tips over at $t = 0$ so that this equator is a closed null geodesic,

then tips back up at $t = \infty$. Every point below the equator has every point above the equator in its future, and vice-versa; the points on the equator contain all points above in their future and all points below in their past. Under any map that leaves the bottom half fixed but rigidly rotates about the cylindrical axis all points on and above the equator, preserves the causal relations while severing every timelike curve passing across the equator. One can delete lines from the cylinder to obtain similarly pathological examples that are past or (but not and) future distinguishing.

This example contrasts with the claim that we should regard the causal structure as most fundamental—the image of a spacetime under such a mapping would fail to be physically equivalent. But since the most physically relevant spacetimes exhibit strong causality, this concern may not be too severe.

References

- [1] S. Hawking, A. King, and P. McCarthy, *J. Math. Phys.* **17** (1976), 174.
- [2] D. Malament, *J. Math. Phys.* **18** (1977) 1399.
- [3] S. Hawking and G. Ellis, *The Large Scale Structure of Space-time* (Cambridge U.P., 1973).
- [4] E. Kronheimer and R. Penrose, *Proc. Camb. Phil. Soc.* **63** (1967) 481.
- [5] R. Penrose, *Techniques of Differential Topology In Relativity* (AMS, 1972).
- [6] E. Zeeman, *J. Math. Phys.* **5** (1964) 490.

- [7] E. Zeeman, *Topology* **6** (1967) 161.
- [8] A. Alexandrov, (1967) I have a copy of the paper, but I don't actually know what journal it came from.
- [9] S. Hawking, *Adam's Prize Essay* (1966) unpublished, p116. Thank you to Dr. Charles Misner for unknowingly providing a copy.