Dirac: Singularities

10/02/19

Day 1: Schwarzschild, time-orientation, causality, geodesic completeness, Singular Spacetime

What is a singularity in physics? The most familiar and simplest example is a stationary point charge in pre-GR E&M, wherein reality is modeled as $\mathbb{R}^3$. The charge generates an electric field $\vec{E}(\vec{x}) = \frac{q}{4\pi\epsilon_0 |\vec{x}-\vec{x}_0|^3}(\vec{x}-\vec{x}_0)$, where $\vec{x}_0 \in \mathbb{R}^3$ is the position of the charge, indicating that test charges placed arbitrarily close to $\vec{x}_0$ feel an unbounded force. This is typically seen as unreasonable, so we say the model breaks down and doesn't apply near $\vec{x}_0$ — $\vec{x}_0$ is called a "singularity" of the model. How does the situation differ in GR?

First, E&M has the benefit of the ambient $\mathbb{R}^3$ structure on which everything is defined. The singularity, $\vec{x}_0$, has an obvious position in this ambient structure. In GR, one faces the challenge that the manifold structure and the physical objects defined on it are intimately intertwined — there is no structure where the physical objects don't make sense and vice-versa. Hence, we cannot simply look for points in the structure at which physical objects aren't defined. To get a sense for this issue and how we might correct our approach, we consider the prototypical GR singularity: the Schwarzschild solution.

Birkhoff's theorem states that every spherically symmetric vacuum solution to Einstein's equation can locally (except at $r=2m$) be given coordinates (t, r, θ, φ) w.r.t. which

$$g = -\left(1-\frac{2m}{r}\right)dt^2 + \left(1-\frac{2m}{r}\right)^{-1}dr^2 + r^2 d\theta^2 + r^2\sin^2(\theta)\,d\varphi^2$$

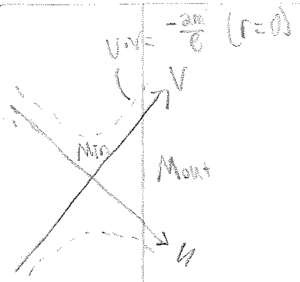
This metric can be given globally equally well to the two spacetimes $M_{out} = \mathbb{R}\times(\mathbb{R}^3 \setminus \overline{B_{2m}(0)})$, $M_{in} = \mathbb{R}\times(B_{2m}(0)\setminus\{0\})$.

Should we call these spacetimes singular, with singularities $\{(t,r,\theta,\varphi) \mid r = 2m, 0\}$?

Defining coordinates $u(r,t) = \sqrt{f(r)}\, e^{-t/4m}$, $v(r,t) = \sqrt{f(r)}\, e^{t/4m}$ $(r > 2m)$, $f(r) = (r-2m)\, e^{r/2m - 1}$, gives

$$g = \frac{8m^2}{r} e^{1-\frac{r}{2m}}(du\otimes dv + dv\otimes du) + r^2 d\theta^2 + r^2\sin^2(\theta)\,d\varphi^2$$

(here $r = r(u,v) = f^{-1}(uv)$). Notice the metric components in these coordinates make sense for all u, v s.t. $u\cdot v > f(0) = -\frac{2m}{e}$, and in



particular none diverge at $U \cdot V = f(2m) = 0$ (the boundary of M_{out} in these coordinates). One can describe M_{in} similarly via $U = -\sqrt{-f(r)}\, e^{-t/4m}$; $V = \sqrt{-f(r)}\, e^{t/4m}$ $(r < 2m)$, giving the same form of the metric. Apparently, then, M_{in} and M_{out} can be embedded in the larger spacetime $M = \{(U, V, \theta, \varphi) \mid U \cdot V > -\frac{2m}{e}\}$, which includes $r(U,V) = f^{-1}(U \cdot V) = 2m$, so there was no singular behavior at $r = 2m$. Our new coordinates still break down at $r = 0$, though. In fact, this could not have been avoided, as $R^{\mu\nu\rho\sigma} R_{\mu\nu\rho\sigma}$ diverges as $r \to 0$, so M can not be extended as a Lorentzian manifold to cover $r = 0$. Is this a problem? How can we tell that $r = 0$ isn't at "infinity"?

Consider the spacetime M with global coordinate chart $V = \{(t, X) \in \mathbb{R}^2 \mid X > 0\}$ wherein the metric is given by

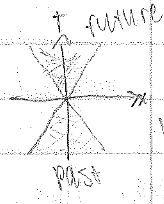
$$g = -dt^2 + \frac{1}{X^2} dX^2$$

the metric apparently breaks down at $X = 0$, which seems to be "finite". However, choosing the new coordinate $y = \ln(X)$ we obtain a new global coordinate chart $\tilde{V} = \mathbb{R}^2 = \{(t, y) \in \mathbb{R}^2\}$ wherein

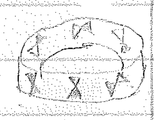
$$dy = \frac{1}{X} dX \implies g = -dt^2 + dy^2$$

That is, M is (isometric to) Minkowski space, and so certainly isn't singular, and $X \to 0$ in V corresponds to $y \to -\infty$ in $\tilde{V}$, i.e. $X = 0$ is "at "infinity"" in the Minkowski space structure. Apparently, coordinate charts aren't sufficient to tell us where "infinity" is.

These examples demonstrate the challenge, but we're getting as many questions as answers. Taking a step back: what should a singularity be? Ultimately, a point or region in which the model breaks down. To know when GR breaks down as a model, we must be precise about what we demand of a spacetime to call it physically reasonable. We've said already that GR requires a Lorentzian 4-manifold (M, g) together with an associated $(0,2)$ stress-energy tensor T for which $G = 8\pi T$. Further, the timelike geodesics of the L-C connection ∇ associated to g then describe the motion of test particles. This is not quite enough.



The tangent space T_pM at each $p \in M$, being a vector space with Lorentzian metric, is isometric to Minkowski space, and hence just like Minkowski space has two connected components of timelike vectors s.t. $V, W \in T_pM$, timelike, are in the same component $\Leftrightarrow \langle V, W \rangle < 0$. In particular, $\langle V, -V \rangle = -\langle V, V \rangle > 0$, so $V, -V$ are always in opposite components. The reduction to special relativity at p leads us to expect one component should represent the future of p and the other the past of p — hence we must require of (M, g) that we may smoothly prescribe the future components of timelike vectors at each $p \in M$ in a consistent manner (smooth in the sense that around each $p \in M$ we may find a smooth timelike vector field in the future component at each point). Such an (M, g) is said to be <u>time-orientable</u>. This is independent of M being orientable as a smooth manifold, which we typically also required (one cannot define the fully antisymmetric Levi-Civita tensor or integration consistently on a non-orientable manifold).


$M = S^1 \times \mathbb{R}$

Given a time-orientation on (M, g), we may speak of smooth timelike curves being future- or past-directed (one or the other is always true), and similarly for smooth causal curves with nonvanishing tangent. Hence each $p \in M$ has a timelike future

$I^+(p) := \{q \in M \mid \exists$ a future-pointing timelike curve from p to $q\}$ and causal future

$J^+(p) := \{q \in M \mid \exists$ " " causal " " $\}$

and similarly for pasts $I^-(p), J^-(p)$. Physically, we expect that any $p \in M$ cannot lie in its own past, or else an observer at p could, in principle, influence their own past. We might go further and require that p cannot be "close" to its own past. There is a hierarchy of these so-called causality conditions that one might impose:

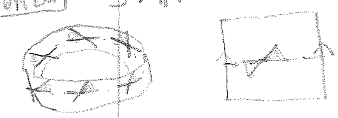
Chronological — M contains no closed timelike curves ($\forall p \; p \notin I^-(p)$)

Causal — M contains no closed causal curves ($\forall p \; p \notin J^-(p)$)

Future/past distinguishing — $I^+(p) = I^+(q) \Rightarrow p = q$

$I^-(p) = I^-(q) \Rightarrow p = q$

Non Ex| $S^1 \times \mathbb{R}$ (cylinder)



(remove a point to make it causal, but not strongly causal).

→ Strongly Causal — $\forall\, p \in M$, $\forall\, U \ni p$ $\exists\, V \subseteq U$, $p \in V$ s.t. no timelike curve γ intersects V more than once (i.e. $\gamma \cap U$ is connected).

Ex| Schwarzschild → Stably Causal — sufficiently small perturbations of g are causal, i.e. $\exists$ a timelike vector field t s.t. the metric $\tilde{g}(X,Y) = g(X,Y) - \langle t, X\rangle \langle t, Y\rangle$ is causal.

Ex| FLRW, Schwarz. → Globally Hyperbolic — $\exists$ a Cauchy hypersurface $S \subseteq M$, i.e. S is a subset met exactly once by every inextendible timelike curve.

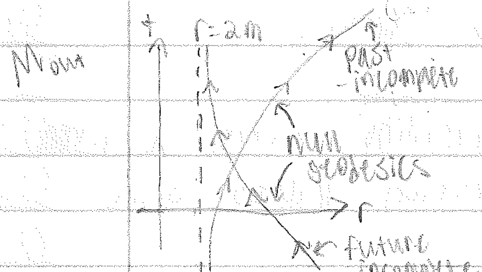
The first two of these are on firm physical footing. Strong causality is rather reasonable and useful in various proofs. Stable causality is expected quantum mechanically — if the metric has quantum fluctuations locally, stable causality ensures it remains causal. Globally hyperbolic is quite strong — it ensures, in fact, that M can be foliated by Cauchy hypersurfaces. In many treatments, strong causality is required for the spacetime to be considered physical.

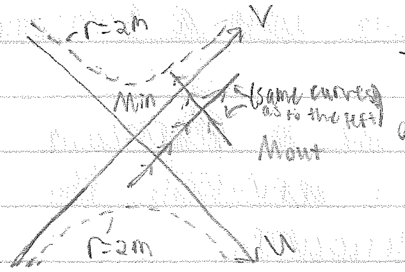
We now address the problem of identifying "infinity". Recall that the geodesic equation $\nabla_{\gamma'(s)} \gamma'(s) = 0$ fixes the parameterization of γ up to a scaling and shift, i.e. $\gamma(s)$ is a geodesic $\Leftrightarrow \gamma(\alpha \cdot s + \beta)$ is, but no other reparameterization is still a geodesic. For timelike geodesics normalized to $\langle \gamma'(s), \gamma'(s)\rangle = -1$, the parameter s is proper time along γ. These observations yield two conclusions:

(1) Traveling along a geodesic for a finite parameter difference is a reasonable notion of moving a finite "distance".

(2) The finiteness of this parameter difference is independent of the choice of geodesic parameterizations.

Hence, a divergence in some physical quantity that obstructs extension of the spacetime should be taken as indicating singular behavior if it occurs in finite parameter time along some geodesic. Notice this implies that such a geodesic cannot be extended past this finite parameter value, i.e. the geodesic is <u>incomplete</u>. What does geodesic incompleteness say about the spacetime? In particular, an incomplete timelike geodesic is a path a <u>non-accelerating</u> observer can take, along which the observer would <u>exit the spacetime</u> in a finite proper time! That is, the spacetime (M, g) cannot model the experience of this observer after the geodesic ends. This is, of course, as strong an indication as one could ask for that (M, g) is breaking down as a physical model. However, geodesic incompleteness alone isn't quite enough to say there is a singularity, as this indicates only a failure of M, not necessarily GR. Consider the Schwarzschild case:





The drawn incomplete null geodesics of M_{out} can be extended past their original maximal parameter ranges in the larger spacetime M. I.e., the incompleteness at $r=2m$ was not a feature of GR, but of the choice of an incomplete model within GR. However, the curves above are still incomplete in M, as they reach $r=0$ in a finite parameter time. Since we've said M cannot be extended to include $r=0$ (since $R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$ diverges), this indicates a singularity of GR as a model for this spacetime. Notice that the incomplete geodesics in M_{out} appear to go to "infinity" in coordinates — this illustrates that coordinate "infinity" is not meaningful! As an aside, the above is precisely the phenomenon that far away stationary observers (those following ∂_t in M_{out}) never see objects falling into a black hole ever actually reach the black hole, even though these objects in fact reach the $r=0$ singularity in finite proper time.

One can take this idea of geodesic incompleteness indicating singular behavior one step further — timelike geodesic incompleteness indicates nonaccelerating observers exiting the spacetime in finite proper time, but it is also problematic if an observer can leave M in finite proper time with bounded acceleration. Physically, this corresponds to an observer with a finite amount of energy resources (e.g. rocket fuel) being able to exit M in finite proper time — again, M cannot model this observer's experience. An example due to Geroch shows this notion, called b-incompleteness, is distinct from geodesic incompleteness.

From all these considerations, we finally arrive at a comprehensive definition of a singular spacetime:

<u>Def</u>

Defⁿ A time-orientable, (+ orientable) strongly causal, connected Lorentzian manifold (M, g) that is maximal with these properties in the sense that it cannot be embedded in a strictly larger such manifold, is <u>singular</u> if it is timelike b-incomplete or b-incomplete with a curvature scalar diverging along a b-incomplete curve.

C^2

Beyond the great difficulty in practically determining whether a given spacetime satisfies this definition, the definition leaves some open questions. In particular,

(1) Given a singular spacetime M, "where" are the singularities? Can we naturally construct a set of points S identified as singularities? Can we put a reasonable topology on $M \cup S$ indicating what points in M are "near" S?

(2) How concerned should we be that the best-fit model from GR to reality is singular? Should we generically expect physically reasonable models to be singular?

(3) How should we interpret a singular model? Is it a problem for GR if the answer to (2) is "yes"?

We will attempt to answer these questions as best as possible in the coming discussions.