Abstract:
A brief overview of the foundational ideas of General Relativity and then a more in depth overview on the basic idea of a singularity theorem. A basic background and necessary needed ideas for the proof of Hawking's singularity theorem is given, including the derivation of the Raychunduri equation for time like geodesics, and then the proof is done in full. The outline of Penrose's singularity theorem is given, though in less depth, with the null Raychunduri equation described. The consequences of these theorems are touched on. The level of mathematical rigor is moderate, with some ideas being described physically. A prerequisite basic idea of tensors along with basic concepts in mathematical analysis, such as that of a neighborhood, is needed. An idea of special relativity is also assumed.

THE SINGULARITY THEOREMS OF HAWKING (AND PENROSE)

The singularity theorems are some of the most important results in general relativity. There are many versions for different types of curves, but they all share the same structure. A energy condition is imposed on the matter content in the universe and an assumption is made on the global structure of spacetime, which is the mathematical model used to represent the universe. These assumptions lead to a contradiction about certain aspects of the nature of certain curves on this spacetime called geodesics if we assume that there is no singularity. So the singularity theorems can be understood if each of these aspects are fully understood. The result of the singularity theorems show that singularities are not just a mathematical coincidence, but must arise naturally in our universe due to the nature of the universe. And since Einstein's equations, which are the equations of motion in general relativity, break down at these singularities, this shows there are parts of our universe that GR cannot fully describe. So GR needs to be modified or added to in some way to account for these singularities.

So, the starting assumption is that spacetime can be modeled by a topological manifold, which is defined as a topological space, on which every region is locally homeomorphic to Euclidean space. This means a bijective function exists between every region of the manifold and Euclidean space. This means that every point on the entire manifold can be described by a corresponding point in Euclidean space. So a chart can literally be placed over each region of the manifold, meaning each point on the manifold can be described by a coordinate. Now, we want the ability to do calculus on the manifold, so we require that the manifold be a differentiable manifold, meaning the transition functions, the functions between different charts that have a part that point to the same region on the manifold, are smooth (infinitely differentiable). So any smooth function defined on the manifold can be safely taken the derivative off (since then all corresponding points in Euclidean space smoothly transition to each other in regions of overlap on the manifold). So, since every point is similar to Euclidean space, this manifold can be thought of as having a tangent vector space at each point. Then, there also exists a dual vector space, along with a countably infinite number of tensor product spaces. Now, since these spaces exist at each point in spacetime, we can actually talk about them as vector fields, dual vector fields, and tensor product fields. As we have a differentiable manifold, we use partial derivatives ∂_u =

$\frac{\partial}{\partial x^u}$ as the basis for each vector space and the differential elements dx^u as the basis for the dual space. Then, $<dx^u, \partial_v> = \delta^u_v$ (defined by how it acts on a test function f), where the brackets just denote a linear function, either dx^u or ∂_v . Since we are on a four dimensional manifold, there are four such linearly independent vectors and dual vectors that span each vector and dual space. Now a rank (0, 2) tensor $g_{uv}\, dx^u \otimes dx^v$ is chosen as the inner product in a given vector space. Since a vector space exists at each point, $g_{uv} = g_{uv}(x)$, where $x = x(x^0,x^1,x^2,x^3)$ are the coordinates of the point, is actually a tensor field that exists over all the spacetime. So this inner product can give a sense of distance on the space-time, and so g_{uv} is called the metric tensor. The actual value of this metric, which depends on the nature of our spacetime and the nature of the matter that is present in the spacetime, is actually what is solved for in Einstein's equations, the equations that postulate how spacetime itself reacts (meaning essentially curves) due to the matter present. The equation is $G_{uv} = \frac{(8\pi G)}{c^4} T_{uv}$, where G_{uv} is the Einstein tensor, which describes how spacetime is curved, and T_{uv} is the energy momentum tensor that describes how matter is distributed on the spacetime. In fact, the Einstein tensor can be written as $G_{uv} = R_{uv} - \frac{1}{2} R\, g_{uv} + \lambda\, g_{uv}$ where R_{uv} is the Ricci tensor, R is the Ricci scalar, and λ is the cosmological constant. Actually the Ricci tensor and scalar come from a more general tensor, the Riemann curvature tensor $R^p{}_{uvq}$, a rank (1,3) tensor that describes essentially the curvature of the spacetime at every point. The Riemann tensor contracted $R^q{}_{uvq} = R_{uv}$ gives the Ricci tensor which is then just a lesser measure of curvature used along curves, and the Ricci tensor contracted with the metric tensor gives the Ricci scalar $R = g^{uv} R_{uv}$, which is then just a real number measure of curvature at each point in spacetime. The cosmological constant λ is just added in to account for dark matter and is taken to be 0 on scales up to that of a galaxy.

Now come the posulates of general relativity. Einstein established the Einstein Equivalence Principle, which states that all reference frames are equivalent for observers in free fall. This includes accelerating reference frames where the acceleration is caused by gravity (so here free fall means no external forces other than gravity are acting on the observer). This means for each observer in free fall, the laws of physics have the same form meaning the speed of light is the same and two observers will see each other accelerating away at the same rate, will always have the same relative velocity. So the spacetime locally is Minkowski, the spacetime of special relativity, for each observer. The metric of Minkowski space, the Minkowski metric, is $g_{00} = -1$, $g_{11} = g_{22} = g_{33} = 1$, $g_{uv} = 0$ for $u \neq v$. This metric has the possibility of being negative, so spacetime will actually be a pseudo-Riemannian manifold. Another postulate is that objects in free fall will move along geodesics, paths on the manifold that can be parameterized such that the tangent vector field along the curve does not change on the manifold (this is equivalent to saying objects with no forces on them move at constant velocity on a straight line classically).

Now, a problem arises in that we cannot compare vectors at different points on the manifold as it is because they belong to different tangent spaces. So, we define the Levi-Cevita connection, Γ^m_{np}, which gives a notion of how to compare vectors in different vector spaces by allowing for a vector in one space to be parallel transported to a different space. It is in fact used to define the covariant derivative, which gives a measure of how a vector changes on the manifold. The covariant derivative for a vector x = x^u existing at a point in spacetime in the direction of another vector v = v^α at the same point is defined as $\nabla_v x = v^\alpha \partial_\alpha (x^u) + v^\alpha \Gamma^u_{B\alpha} x^B$, where the first term is just the ordinary directional derivative while the second term is the change in the vector due to moving it, meaning parallel transporting it, in the direction of v. Now, the connection on the spacetime is not necessarily unique. However, the one we do have, the Levi-Cevita connection is unique because it comes from the fact we assert that the connection be a metric connection, meaning it should be defined by the metric. We also assert that spacetime is torsion free, meaning the torsion tensor, defined as $T^q{}_{uv}$, a rank (1,2) tensor which acts on a dual vector field v_a and two vector fields x^u, y^v, as <v,T(x,y)> is 0. T is the torsion map defined as T(x,y) = $\nabla_x y$ - $\nabla_y x$ - [x,y], where [x,y] is the Lie Bracket of the two vector fields, which just means the commutator of the two vector fields. $T^q{}_{uv}$ measures essentially how much the spacetime twists out of being a flat space. Torsion being 0 ends up meaning that the metric is symmetric in the lower indices: $\Gamma^m_{np} = \Gamma^m_{pn}$. So the actual definition is given by stating the metric tensor should not change under parallel transport meaning $\nabla_v g_{uv}$ = 0. This leads to, along with the given symmetric property of the connection, Γ^m_{np}= $\frac{1}{2} g^{um}(\partial_n (g_{pu}) + \partial_p (g_{un}) - \partial_u (g_{np}))$. So if the metric gives kind of a zeroth notion of curvature, the connection gives a first notion of curvature. They do not describe the actual curvature of the spacetime however, because they depend on the metric itself or the first derivative of the metric, and the metric depends on the chosen coordinate system. So essentially they represent a curvature at the point or the curvature because of a change in some direction from the point. However, it turns out the Riemann tensor is actually defined in terms of the metric, and gives the notion of curvature based on going in an infinitesimal 4 dimensional “circle”, which gives a true notion of the curvature of the spacetime. Now, consider an one dimensional curve on the spacetime, parameterized by a variable t, a(x(t)). This curve has a vector field on it made up of all the tangent vectors, $a'(t) = \frac{d}{dt} a(x(t))$, at each point on the curve. Now, say a’’(t) = 0, meaning that the tangent vector field does not change along the curve. This gives the notion of a straight line on the manifold. Now, this condition means $\nabla_{a'} a'$ = 0, meaning a’ does not change in its own direction. A curve that has this property for all t is called a geodesic. The condition written more explicitly leads to the geodesic equation: $\frac{d^2 a^u}{dt^2} + \Gamma^u_{np} \frac{da^n}{dt} \frac{da^p}{dt} = 0$. (note the connection is a function of t also since it’s values on the curve are being used). Solving for a(t) with a given position and velocity at t = 0 gives the geodesic a(t).

So now we can look at the different types of geodesics that lead to different singularity theorems. There are actually three types of geodesics: spacelike, timelike, and null geodesics. By the Einstein Equivalence principle, since spacetime is flat for any freely falling observer locally, one can always write the metric, by some change of coordinates, as the Minkowski metric. Then, a differential length of a given geodesic at any given point on it can be written as
$ds = \sqrt{(dx)^2 + (dy)^2 + (dz)^2 - (cdt)^2} = \sqrt{g_{uv}\, dx^u\, dx^v}$. Now say a regular object with some rest mass is free falling. It will be falling in a path traced by a geodesic and, in the proper time reference frame, there will be no change in spatial coordinates, so $ds^2 = -c^2\, d\tau^2$ will be negative. So the convention

$ds = \sqrt{-g_{uv}\, dx^u\, dx^v}$ is used to allow for the square root to be taken. Such a path, taken by any physically moving object, is called a timelike geodesic. Now, say the particle has no mass, meaning it is a light particle. Then, $r =$ ct, meaning the path of the photon is on the light cone, so ds = 0. In this case, the geodesic path is defined alternatively and such a geodesic is called a null geodesic. Also, we need the concept of a congruence of curves, specifically geodesics. Consider a solution to the geodesic equation, a(t), where the initial velocity is given. Then, the position can be varied to basically any point on the manifold, so essentially the curve can be transported around the spacetime and the velocity vectors of each of these curves at each point of the curve defines a vector field on the spacetime. The set of all these curves is defined as a congruence, and since these are geodesics, this is called a geodesic congruence of curves.

Now, we discuss the needed assumptions for the proof of any singularity theorem. First, one needs to put an energy condition on the matter content in the universe. This basically gets rid of any possible unphysical solutions that the Einstein equations can give us and so applies for most observable matter in the universe. These have to do with essentially putting conditions on allowed values of the energy momentum tensor on curves, or more accurately, the eigenvalues and eigenvectors of this tensor. Specifically, for a null vector field x^a and the set of curves defined by this vector field, one can assert the null energy condition $T_{uv}\, x^u\, x^v \geq 0$, essentially saying the "contracted" trace of the energy momentum tensor with respect to the velocity vector of the geodesic is positive definite. This ensures that observers on the geodesics do not observe a negative matter content (so this would not apply for dark matter). For a timelike vector field, one can assert the strong energy condition $\left(T_{uv} - \frac{1}{2}\, T\, g_{uv}\right) x^u\, x^v \geq 0$ where the metric term is added to ensure that what the observer truly sees is that the contracted energy momentum tensor is positive definite (the effects of the non-zero metric are accounted for). We will do fully this case for timelike geodesics with the strong energy condition.

Secondly, we discuss the conditions put on the global nature of the spacetime. We start off by asserting that the spacetime is causal, meaning that no causal curve, meaning timelike or null curve, goes in a closed loop. This just asserts that all events that occur on the spacetime follow causality. A stronger condition is strongly causal which states that every open neighborhood at any given point has a casual curve with both endpoints in that neighborhood, then the curve lies entirely in that neighborhood. This leads to asserting that the manifold is globally hyperbolic, meaning that it is strongly causal and the intersection of the future light cone with the past light cone at any point in the space is compact. This essentially ensures that one cannot find a way to cover up the intersection between the future and past such that it takes an infinite cover, meaning loosely there are no singularities in the space. It turns out that such a manifold has Cauchy surfaces, which are subspaces of the manifold such that each inextendible time like curve intersects the Cauchy surface at exactly one point. A time like curve $(a, b) \longleftrightarrow M$ is inextendible if there is no such extension such $[a,b] \longleftrightarrow M$, meaning the curve must end right before those points. So a Cauchy surface essentially describes a subsection of space at a given point in time. Finally, we need the idea of an mean, extrinsic curvature H, defined on a Cauchy surface. This essen-

tially gives an idea of how the Cauchy surface is curved in a direction of time, hence the notion of "extrinsic".
There are also some more needed concepts. Consider a spacetime with a singularity at a point. As there are geodesics from any point in any direction, there will be geodesics that will will end at that point from all directions. These geodesics started at some other point, so they have finite length. If such a singularity did not exist, then these geodesics will just keep going and they will have no reason to stop. This notion is called geodesic completeness meaning any geodesic, a(t), can be defined for any t. So the spacetime can be said to have a singularity if it has the property of geodesic incompleteness, meaning there is some geodesic a(t) that cannot be extended past a finite interval of t. So in looking for singularities, we will instead look for the existence of geodesics that have this property.
One must also understand the concept of conjugate points. On a geodesic a : [p,q] $\longleftrightarrow$ M , a point a[c] is a conjugate point relative to a[p] if there is a non-zero Jacobi field along a that vanishes at p and c. A Jacobi field is just a vector field defined on a geodesic that is just the tangent vector field of a congruence of curves that connect each point that geodesic to corresponding points on nearby geodesics that have the same parameter value(all the geodesics can share the same parameter). The rate of change of these vectors give an idea of how geodesics are accelerating away from each other, so it gives a notion of geodesic deviation. So, a conjugate point just means that the vector, which is part of the Jacobi vector field, at that point is 0 so the geodesic next to this one is intersecting this one there, roughly. So the stationary(meaning maximal for timelike geodesics) distance between the two points is not uniquely determined by one geodesic.

Now, there are are 3 mini theorems that are used in the proof for this singularity theorem called the Hawking singularity theorem (Landsman).
(1) If a manifold is globally hyperbolic, than for any point on the spacetime x and for any point on the positive light cone at that point y, from the set of all possible curves between x and y, there is a smooth time-like geodesic of maximal length between those two points.
Since global hyperbolicity essentially means that there are no singularities in time and timelike curves do not really extend out at all(past the smallest neighborhood that contains their end points- strongly casual), then considering the smallest possible neighborhood that contains x and y, one can think of a maximal length geodesic out of all the geodesics that connect x and y.
(2) Between any two points x and y on the spacetime, there is a time-like geodesic B that locally maximizes the length of all curves from x to y iff there is no conjugate point on B between x and y.
$L = \int ds = \int \sqrt{-g_{uv}\, dx^u\, dx^v}$ is the length of a curve between points x and y if integrated from x to y. A geodesic always tries to minimize $g_{uv}\, dx^u\, dx^v$, the spacetime distance between two infinitesimal points. But for time like paths, $g_{uv}\, dx^u\, dx^v = - c\tau$ (at a given point) is negative, so τ must be maximized for the spacetime distance to be minimized on a geodesic. But the length of the geodesic has a negative sign in order for the square root to make sense. So L is maximized when τ is maximized. Then, time like geodesics are the curves of maximum length between two points. But if there is a conjugate point on this geodesic, then the uniqueness of the geodesic as the maximizing distance between 2 points is

taken away, as essentially, there are other geodesics that are nearby enough that another path can be taken which may be bigger than this one.
(3) Consider a timelike congruence of curves with velocity x^a in free fall. These congruence of curves, by virtue of all their tangent vectors, can be treated as a fluid flow. The change in the volume of these curves is defined by $\theta = \nabla_a x^a$. The shear of these curves is given by $\sigma_{uv} = \nabla_{(u,v)} x - \frac{1}{3}\theta h_{uv} - x'_{(a} x_{b)}$. The vorticity of these curves is given by $w_{uv} = \nabla_{[u,v]} x - x'_{[a} x_{b]}$.
Then, we can say $\nabla_v x_u = \sigma_{uv} + w_{uv} + \frac{1}{3}\theta h_{uv} + (x')_u x_b$
Now, from the properties of the Riemann tensor, $\nabla_w \nabla_v x^u - \nabla_v \nabla_w x^u = R^u{}_{zvw} x^z$. Contraction of the u and v indices and then multiplying by x^w, we get $x^w \nabla_w \nabla_u x^u - x^w \nabla_u \nabla_w x^u = R^u{}_{zuw} x^z x^w \longleftrightarrow$
$x^w \nabla_w \Theta - x^w \nabla_u \nabla_w x^u = R_{zw} x^z x^w$
Since $x^w \nabla_u \nabla_w x^u = \nabla_u (x^w \nabla_w x^u) - \nabla_u x^w \nabla_w x^u = \nabla_u (\dot{x}^u) - \nabla^u x^v \nabla_v x_u$.
Now, it can be noted that since $\sigma_{uv} \sigma^{uv} = 2\sigma^2$ and same for w, $\nabla^u x^v \nabla_v x_u = 2\left(\sigma^2 - w^2\right) + \frac{1}{3}\Theta^2$
So, overall, $x^w \nabla_w \Theta - \nabla_u (\dot{x}^u) + 2\left(\sigma^2 - w^2\right) + \frac{1}{3}\Theta^2 = R_{zw} x^z x^w$
Now, for a perfect fluid, the stress energy tensor is $T_{uv} = (\rho + p) x_u x_v - p g_{uv}$, so
$R_{uv} x^u x^v = -4\pi(\rho + 3p)$
Then, denoting $\dot{\Theta} = x^w \nabla_w \Theta$, $\dot{\Theta} = \nabla_u (\dot{x}^u) - 2\left(\sigma^2 - w^2\right) - \frac{1}{3}\Theta^2 - 4\pi(\rho + 3p)$
which is the Raychaudhuri equation (Dadhich).
Now assuming no acceleration, no rotation, the strong energy condition (ρ+3p > 0, where ρ+3p is interpreted as the active gravitational density, meaning the regular density plus the pressure), $\dot{\Theta} \leq -\frac{1}{3}\Theta^2 \longleftrightarrow (\Theta(\tau))^{-1} \geq (\Theta(\tau_0))^{-1} + \frac{1}{3}\tau$
So Θ can approach infinity if $0 \geq (\Theta(\tau_0))^{-1} + \frac{1}{3}\tau \longleftrightarrow \tau \leq \frac{3}{|\Theta(\tau_0)|}$
So it is possible for the divergence of a congruence of geodesics to go to infinity in a finite amount of time given the strong energy condition if the geodesic can be extended up till $\tau_s = \frac{3}{|\Theta(\tau_0)|}$.
Now, consider that $\theta(\tau_0) \leq 0$. A negative divergence implies that some of the velocity vectors are pointing inward towards a point. This implies that, at the point τ_0 along a given geodesic γ where this is the case, there is another velocity vector of another nearby geodesic that is pointing towards it. This implies the Jacobi field along γ is 0 at τ_0. Then, since the same applies at the point of infinite divergence, the Jacobi field is also 0 there. Now, since $(\Theta(\tau))^{-1} \geq -(\Theta(\tau_0))^{-1} + \frac{1}{3}\tau$ in this case and $\tau \leq \frac{3}{|\Theta(\tau_0)|}$, $(\Theta(\tau))^{-1}$ must be bigger than some negative number, so it must go to infinity. And since $\theta(\tau)$ can be positive for other values of τ and so the Jacobi field is non-zero, τ_0 and τ_s are conjugate points. So we can conclude the following theorem: Say a geodesic β is part of a congruence of time like geodesics that have 0 vorticity. If $R_{uv} u^u v^v \geq 0$ and $\theta(\tau_0)$ is negative somewhere on β, then β has a later congruence point relative to τ_0, τ_s, provided that the geodesic can be extended between those 2 points.

Now, we are ready to state the singularity theorem for timelike geodesics. Say our spacetime M is globally hyperbolic with Cauchy surface S. Let's assume the strong energy condition, which states that for every timelike vector field $x^u(x)$, the momentum energy tensor has the following property: $\left(T_{mn} - \frac{1}{2}\,T g_{mn}\right) x^m x^n \geq 0$, holds on all time like geodesics on the manifold . Then, let's assume H of S is positive in the past direction, so S is kind of bent towards the future. If those conditions hold, then all geodesics going through S have finite length. This means the manifold M has a timelike singularity. This can be proved by contradiction.

Proof:

Say there is a geodesic from a point x on S to a point y on M not on S that is much further away. Then, there is one β of maximum length too between those points (1). Then, β does not have conjugate points (2). But β does have conjugate points if it can be extended far enough (3). This is a contradiction so such a geodesic cannot exist.

Since this is true for any Cauchy surface S, no matter what the given moment of time is for S, then one can consider each Cauchy surface that exists in the past timelike direction as far back as possible. Then, the whole universe, for which the path of any object is timelike, can be traced back to kind of a hole in the spacetime where all the timelike vectors come from. This is the Big Bang, the timelike singularity from which our universe came.

The null case is very similar. All the same assumptions are used, such as the existence of Cauchy surfaces and so the global hyperbolicity. Only main difference is that the Raychaudhuri equation is written in its null form along null geodesics (Walters): $\dot{\bar{\theta}} = -2\left(\bar{\sigma}^2 - \bar{w}^2\right) - \frac{1}{2}\,\bar{\theta}^2 - T_{uv}\,x^u\,x^v$. Since this is light, there is no acceleration term and $\frac{1}{2}$ comes from the fact that each quantity is taken along the geodesics in the perpendicular (as defined by the metric) direction of travel as indicated by the - sign (so 2 dimensional geodesic deviation instead of 3). This is because light does not diverge in the direction it is going in as it is essentially always going at the same speed. So, along with the assumptions of 0 vorticity and the null energy condition, $\dot{\bar{\theta}} \leq -\frac{1}{2}\,\bar{\theta}^2$, which leads to $(\theta(\tau))^{-1} \geq (\theta(\tau_0))^{-1} + \frac{1}{2}\,\tau$. Then, θ can approach infinity if $\tau \leq \frac{2}{|\theta(\tau_0)|}$. So, similarly with the global assumptions, one can conclude that all null geodesics that radiate from any Cauchy surface must have finite length, so one can conclude null geodesic incompleteness. This is called Penrose's singularity theorem. And since this is true for any Cauchy surface, so any moment frozen in time, then all light must come from or go towards what is called a trapped null surface, a closed surface on which all light rays are converging. An example of this is the apparent horizon which surrounds a black hole. So this essentially shows that spacelike singularities, such as black holes, are inevitable in the universe.

In conclusion, we have shown that spacetime, with some reasonable physical assumptions, must necessarily have singularities. They are not removable(like that of the Schwarzschild radius for a Schwarzschild black hole which depends on the coordinate system chosen), but necessarily exist because of the physical conditions of the universe. At these points, general relativity breaks down, meaning the Einstein equations result in a metric that blows up to infinity and a point where the curvature becomes infinite. So, either GR is incomplete or it needs to be integrated with another theory (like Quantum Field Theory) in order to describe those regions.

Citations:

Dadhich, Naresh. “Derivation of the Raychaudhuri Equation.” ArXiv.org, 28 Nov. 2005. <arxiv.org/abs/gr-qc/0511123v2. >

Landsman, Klaas. “The Foundations of General Relativity.”24 May 2018. PDF File. <https://www.math.ru.nl/~landsman/FGRfinal.pdf>

Walters. “The Schwarzschild Solution.” PDF File. <https://www.math.ru.nl/~walters/Schwarzschild.pdf>