May 16, 2009

# An FAQ on Singularity Theorems in Classical General Relativity


Ghanashyam Date[1,*]

[1] *The Institute of Mathematical Sciences*

*CIT Campus, Chennai-600 113, INDIA.*



## Abstract

This is a compilation of some of the 'frequently asked questions' regarding the famous singularity theorems of general relativity. After some introductory remarks, the discussion is organized in a question-answer format. Some references are included at the end for more detailed and precise information.



PACS numbers:



*Electronic address: `shyam@imsc.res.in`

## I. INTRODUCTION

The singularity theorems are invoked in many context and carry an "aura" of mystery around them. These are sharply formulated statements delineating the limitations of classical general relativity. The notion of singularity used is also different from a naive and intuitive extrapolation from the one encountered in the Newtonian context eg: 'singularity as a place where some physical quantity diverges'. The reason for this is that the space-time model is much more sophisticated than the absolute space and time of pre-relativistic era. In particular, there is a non-trivial structure of causality due to a *finite maximum speed* for any propagation of influence. Thus, a singularity is identified with *the existence of features of the general relativistic space-times, which are physically unacceptable.* This is discussed further below.

The appropriate notions of singularities, evolved through various explicit examples which of necessity were highly special, usually with idealized high degree of symmetries. Almost all of the well motivated and physically relevant solutions exhibited a singularity similar in spirit as the Newtonian ones. There are *two* physically relevant contexts viz *an unstoppable gravitational collapse* of localized sources such as massive enough stars and the *Friedmann-Robertson-Walker (FRW) expanding universe* where 'singularities' were encountered. Different sets of people contributed to these in the early stages.

Chandrasekhar discovered that star which has run out of its nuclear fuel and is being supported against a collapse by the quantum mechanical *degeneracy pressure*[1], will fail to be stable if its mass exceeds about $1.4 \times M_{\odot}$ ($M_{\odot} \sim 10^{33}$ grams is the solar mass). This is the famous *Chandrasekhar limit* for the *white dwarfs.* There is a similar upper limit for *neutron stars* which are supported by the degeneracy pressure of the neutrons. This was computed by Oppenheimer and Volkoff. A qualitative reason for the instabilities to set-in for massive bodies is that in general relativity the pressure energy also gravitates and thus encourages

[1] It is a fact of nature that particles called *fermions* (electrons, protons, neutrons, ...) cannot be in the same quantum mechanical state. This is known as the *Pauli principle.* If external conditions try to force the situation (for example by cooling the system and/or applying a squeeze), the resistance offered translates into an effective pressure. This pressure is known as degeneracy pressure. For the white dwarf stars considered by Chandrasekhar, this pressure is provided by the electrons. As the mass of the star increases, the electrons' speeds become comparable to the speed of light which results in decreasing their resistance to the gravitational squeeze and an instability sets in.

collapse. In the Newtonian gravity, only mass gravitates but not energy and stability against collapse is achievable for any mass. These works showed the possibility of an un-stoppable gravitational collapse. Subsequently, Oppenheimer-Snyder solution describing gravitational collapse of a spherical body of uniform density, also exhibited a central region with infinite curvature and density i.e. a 'singularity'.

In the cosmological context, the simplest solution showed a 'beginning of the universe' from an infinitely dense state - the *Big Bang* singularity. These are solutions which are spatially *homogeneous* (same as analyzed from any spatial location) and *isotropic* (same view in every direction)[2]. Raychaudhuri showed that merely including *anisotropies* i.e. allowing the geometry and matter distribution to be different in different directions, does not remove the singularity thereby indicating that the occurrence of singularity may be more general. This was indeed confirmed by more examples. Assuming the existence of a singularity, Belinskii-Khalatnikov-Lifshitz asked whether among the possible singularities are there any which arise without making any particular restrictions on the initial data i.e. are any of the singularities *generic*[3]? They answered in the affirmative. They also formulated a *conjecture* regarding the *nature* of the singularity. However, their analysis did not identify conditions under which singularities are inevitable. This was achieved by the more abstract methods adopted by Penrose-Hawking-Geroch which lead to a set of *singularity theorems*. The equation presented by Raychaudhuri in his analyses, now known as the Raychaudhuri equation, plays a crucial role in the proofs of these theorems. This is discussed further below.

Following is a presentation of some of the main ideas. A few references are included at the end to help get a perspective and some technical level information. The references to the original literature can be found in [1].

[2] There is observational evidence for isotropic distribution of sufficiently distant clusters of galaxies. The homogeneity however has the strongest support from the *cosmological principle* viz "there is nothing special about the location of our galaxy".

[3] This is a technical statement and means that *if* there is some initial data which leads to a singular geometry, *then* almost all initial data in its vicinity will also lead to a singular geometry. It does not imply that *all* initial data lead to a singularity.

1. *What are some explicit examples of singular space-times?*

   (a) The $r = 0$ (a 2-sphere) singularity of the positive mass, Schwarzschild solution. Here the curvature components (and invariants constructed out of these) diverge. By contrast the singularity at $r = 2GM/c^2$ is a coordinate singularity and is absent for negative mass Schwarzschild solution.

   The same feature holds for the Reissner-Nordstrom solution ($m^2 > Q^2$, in the geometrized units: $G = 1 = c$).

   For the Kerr (and the Kerr Newmann solution), the $r = 0$ is a "ring singularity".

   These are the simplest examples of solutions of Einstein equation which represent *localized sources* and have a curvature singularity.

   (b) The FRW solution has a curvature singularity as the scale factor vanishes at a finite synchronous time[4] in the past (implying a finite age for the universe). This is a solution which exhibits the maximal symmetry of *spatial homogeneity and isotropy.*

   There are also spatially homogeneous but anisotropic[5] space-times all of which exhibit curvature singularities. These are classified into *nine Bianchi types.* The simplest of this class of space-times is the *Kasner* (Bianchi-I) solution of vacuum Einstein equation. This has 3 scale factors which behave as $a_i(t) \sim t^{2\alpha_i}$, where the constant exponents satisfy $\sum \alpha_i = 1 = \sum \alpha_i^2$. The volume vanishes as $t \to 0$ such that one scale factor diverges and the other two scale factors vanish. The most complex of this class is the Bianchi-IX type for which an exact solution is not known. It has six metric components and the approach to singularity can be parametrised in terms of 3 scale factors along a a set of 3 directions. The three scale factors behave as in the Kasner solution for some time, then permuting the

---

[4] In general relativity, time is a coordinate and is as arbitrary as the space coordinates. The metric and the coordinates together give the space-time interval by the expression

$$\Delta s^2 = g_{tt}\Delta t^2 + 2\sum_{i=1}^{3} g_{ti}\Delta t \Delta x^i + \sum_{i,j=1}^{3} g_{ij}\Delta x^i \Delta x^j$$

The *synchronous time* refers to a coordinate system such that $g_{ti} = 0$ and $|g_{tt}| = 1$.

[5] An example of a homogeneous but anisotropic situation in the usual space would be (say) a uniform electric field pointing in some direction.

behaviors among themselves and permuting the *directions* as well, *ad infinitum.*

The Belinskii-Khalatnikov-Lifshitz (BKL) conjecture is that *as a space-like singularity is approached, even in a general inhomogeneous model, the inhomogeneous spatial surface can be viewed as an essentially independent collection of approximately homogeneous patches, each of which evolves as a Bianchi-IX model.* The general approach to singularity is thus *chaotic.*

All of these are examples of space-like singularities and as they are approached, curvature components and/or invariants diverge.

There are also solutions with time-like singularities (eg the negative mass Schwarzschild solution) as well as null singularities (eg some of the Weyl class of static, axisymmetric solutions).

2. *What is meant by a 'singularity'?*

It turned out, after a great deal of work that a suitable criterion to characterize a singular space-time is the property of *geodesic incompleteness.* This is *not* the only criterion, but is the most commonly accepted one for which sharp *theorems* are available for their occurrence. The incompleteness property for *causal geodesics* is regarded a physical pathology since these geodesics are world-lines of *freely falling physical observers.* An alternative criterion is stated in terms of the incompleteness of *time-like curves of bounded acceleration* (observers with rockets).

But of course, one could have a mathematically constructed space-time which has other pathologies. For example, it may admit an ill-behaved *causal structure*[6] eg it may admit closed time-like curves (excluded by invoking the *chronology condition*), or such curves are strictly absent but come arbitrarily close to a given point on itself (excluded by invoking *strong causality*), or if the metric is perturbed slightly the strong causality can be violated (excluded by invoking *stable causality*). There are many other notions of bad causality, all such causal pathologies are prevented in a stably causal space-time. An even stronger notion is that of *global hyperbolicity* which incorporates the notion of *determinism* i.e. allows formulation of a Cauchy problem. Globally

[6] The precise definitions can be seen in [3].

hyperbolic space-times are automatically stably causal and thus, free of all causal pathologies.

Such nice space-times would be mathematical fiction if they are not determined by some physical mechanism which in GR means that the metric is a solution of Einstein equation. But these equations should be solved for with matter stress tensor representing *physical matter*. This property of matter is codified in terms of the so-called *energy conditions*. For this purpose, the cosmological constant can be taken as a part of matter stress tensor and the energy conditions can be applied to the new stress tensor.

Thus, finally we can say that a physically acceptable space-time is a globally hyperbolic (or stably causal) space-time which is a solution of Einstein equation with matter satisfying one or more of the energy conditions and in which all causal geodesics are complete. Examples? Minkowski space-time(!), space-times corresponding to stellar bodies (not black-holes) etc.

Most of the hard work went into identifying a precise set of pathologies which could render the general relativistic model of space-time either useless (non-predictive) or irrelevant (physically non-realizable).

3. *What is the connection between the Raychaudhuri equation and singularities?*

   To appreciate this connection, a few terms should be understood at least qualitatively. Recall that a *geodesic* in the usual examples of two dimensional surfaces such as a plane or a sphere are the paths of shortest length i.e. among all curves connecting any two given points, geodesics have the smallest length. When one considers *world-lines* or paths in a space-time, they come in three distinct types: time-like (world-lines of material particles, obey an upper speed limit), light-like (world-lines of light) and space-like (world-lines of as-yet-fictitious particle - 'tachyons' which obey a lower speed limit). The length along a time-like curve is the time elapsed as recorded by clock moving with the material particle (also called *proper time*). A time-like geodesic is a time-like world-line along which the proper time is the largest[7].

[7] This is the basis for the statement that the freely falling twin will be older than the other twin who goes on a rocket tour, when they meet again.

This property of time-like geodesics, of maximizing the proper time between two given events, holds only if the two events are not *conjugate points*. Quite generally, *conjugate points* along a geodesic are the points at which several geodesics meet or focus[8]. The poles on earth are conjugate points along every longitude (which are geodesics), the object and image points in a lens are another example of conjugate points (paths of light are geodesics).

It is important to note that just as the focal point of a convex lens is not a singularity of the ambient space, though it is a "singularity" of the bundle of rays, conjugate points along a time-like geodesic are *not* the locations of space-time singularities.

Now one can ask the questions: do geodesics always have conjugate points, never have conjugate points or have them under some conditions? Interestingly, the answer to these questions depends on the space-time as well as on the properties of *bundles* (technically a congruence) of geodesics. The topological properties of the set of time-like or light-like curves connecting two events, imply that in a *globally hyperbolic* space-time, there always exist a geodesic connecting a pair of points *without* any conjugate points in-between. On the other hand, The Raychaudhuri equation gives precise conditions under which there *will* exist a pair of conjugate points along a time-like geodesic. Now, it is possible to realize the conditions required by the Raychaudhuri equation, in a globally hyperbolic space-time, thus producing a contradiction. The way out is to observe that the geodesic must not be extendible beyond the conjugate points implied by Raychaudhuri equation! The geodesic must be incomplete (not definable for all values of the proper time). The space-time must be singular by the criteria discussed above, under the conditions required by the Raychaudhuri equation[9].

4. *What are singularity theorems?*

Clearly, singularity theorems cannot be implying that every solution of Einstein equation would have some pathology. These theorems isolate *additional properties* of space-times, which if they hold then some physical pathology must also occur. Typically, the physical pathology is taken to be *existence of at least one incomplete causal geodesic.*

[8] This is an intuitive description. The precise technical definition refers to a non-trivial solution of the geodesic deviation equation which vanishes at the two points on the geodesic [3].

[9] I have taken time-like geodesics for simplicity. Analogous statements hold for light-like geodesics [3].

Thus, typically one assumes that (a) space-time is a solution of Einstein equation with matter satisfying one or more energy condition ("gravity is attractive for positive mass-energy"); (b) space-time is free of causal pathologies (globally hyperbolic or stably causal) and then formulates: *if either* space-time also admits *trapped surfaces* (i.e. two dimensional, closed without boundaries, space-like sub-manifold such that both the out-going and in-coming, orthogonally emanating null geodesics have negative *expansion*), *or* there exists a space-like Cauchy surface whose extrinsic curvature for past directed null geodesic congruence is negative and bounded away from zero, *then* there exists at least one causal geodesic which is past or future incomplete.

Different versions of singularity theorems differ in precise articulation of these three features: physically realizable space-time, space-time free of causal pathologies (so that rest of the equations of classical physics remain predictive and meaningful) and a condition characterizing a *sufficiently advanced stage of gravitational collapse* (formation of trapped surface) or *an everywhere expanding universe* (some observed fact of nature). The existence of incomplete causal geodesic(s) is then guaranteed.

The third feature is essential because there certainly exist real space-times which have no pathologies.

5. *What these theorems do NOT imply?*

Firstly, these theorems do *not* imply that every causally acceptable and realizable solution, is necessarily singular - a further property is essential. One could evade the theorems trivially by denying the third set of properties. (Un)Fortunately, there are grounds to believe in the physical existence of black holes which indicate that gravitational collapse *has* proceeded far enough and that universe seems to be expanding everywhere, so the third set of conditions actually hold and hence theorems become non-trivial.

Secondly, the entire formulation is classical and within the framework of Pseudo-Riemannian manifolds. This need not be valid at all scales of observations and this would be one way to evade the theorems. For example, quantum theory (of matter) may render the classical notion of causality inadequate.

Thirdly, Einstein equations are used as linking geometry with gravity. This link is however considered to be relatively weak.

6. *Big bang versus Collapse:*

   It is common to cite the Big bang singularity as an example of a *naked singularity.* It is certainly true that the Big Bang is a singularity (typically the homogeneous, isotropic one) and it is visible to us i.e. we can catch future directed null geodesics from the early universe. Since universe is expanding, it is opposite of a collapse situation. However, the terminology of *naked singularity* is introduced in the context of an unstoppable stage of gravitational collapse *of localized sources* and refers to an end state of collapse *without the formation of an event horizon* (i.e. a black hole). Whether a gravitational collapse will result in a black hole or a naked singularity is addressed by the so-called *cosmic censorship conjecture/hypothesis*, which states that in a gravitational collapse, an event horizon will always form. Again there are variants/refinements of the conjecture and many counter examples have been proposed [5]. Strictly speaking, the two contexts, one of universe and one of collapse are physically very different and it is a little unfair to mix the two and cite the Big Bang as a "naked singularity" [10].

7. *Does time end at a singularity?*

   Physically relevant singularities are those which form, either in the future or in the past, starting from a completely regular physical situation. Thus, one has at the back of the mind a causal and deterministic evolution, in short a globally hyperbolic space-time. In such a context, the singularities will be space-like i.e. can be "approached" in an evolution (by contrast, a time-like singularity will invalidate Cauchy property). For such approachable singularities, it is true that "time will end", since evolution past these is meaningless. The evolution is essentially *incomplete*, in the backward direction for the universe (due to the Big Bang) and in the forward direction "inside" a black hole be it *a few solar masses heavy, massive or super-massive*[11].

[10] If one insists on viewing time reverse of evolution of the universe as a collapse, then the absence of any causal rays reaching us from regions "outside the universe" (and not those coming from the singularity) would correspond to the negation of the censorship hypothesis in the time reverse view!

[11] In general relativity, *Black holes* are regions from which nothing can *ever* escape to very far away regions ("infinity"). Theoretically they can have any mass. In nature they are expected to arise due to collapse of individual massive stars which can grow by a process of accretion. These are expected to have a mass few times the solar mass. Many such black holes could also merge to form *super-massive* black holes (a

8. What is geometry at the Big Bang?

   The Big Bang typically refers to the homogeneous and isotropic (FRW) spatial geometry. There are three such cases: spherical ($K > 0$), flat ($K = 0$) and hyperbolic ($K < 0$). These topological properties are unchanged as the scale factor vanishes at the singularity. So the answer depends which model is supported by observations. For non-FRW models, there is no single "K" since the spatial geometry is not maximally symmetric. The examples of the Bianchi models and the BKL conjecture gives some idea about how the geometry may be viewed close to the singularity.

---


[1] John Earman, The Penrose-Hawking Singularity Theorems: History and Implications, in *The expanding Worlds of General Relativity*, Ed. H Goenner, J Renn, J Ritter, T Sauer (Einstein Studies, Volume 7), pp 235 - 267 (1999).

[2] G L Naber, *Spacetime and Singularities*, Cambridge University Press (1988).

[3] R M Wald, *General Relativity*, Overseas Press India Pvt. Ltd., New Delhi (2006).

[4] S W Hawking and G F R Ellis, *the Large scale structure of space-time*, Cambridge University Press (1999).

[5] P S Joshi, On the genericity of spacetime singularities, *Pramana*, **69**, 119-136, (2007).


few hundred thousand to a billion solar mass). These could also form by direct collapse of massive clouds and are thought to be present at the centres of most galaxies.