Lecture XIV: Global structure, acceleration, and the initial singularity

Christopher M. Hirata
*Caltech M/C 350-17, Pasadena CA 91125, USA**
(Dated: December 5, 2012)

I. OVERVIEW

In this lecture, we will depart from the study of specific spacetimes and ask some more global questions about cosmology in GR. Subject to certain "energy conditions" on the behavior of the matter, we can prove some general theorems about the spacetime that go beyond the highly symmetrical FRW solution.

This lecture will address the most important such issue for cosmology – the existence of an initial singularity. The other classical application of global structure techniques is to black holes (the area theorem, inevitability of forming a singularity).

There is no assigned reading for this lecture. MTW Ch. 34 discusses global structure, but mainly in the context of black holes. A comprehensive but mathematical treatment can be found in Hawking & Ellis, and an abbreviated version in Wald Ch. 9. This treatment provides the mathematical rigor that my "proofs" in class will not (this is a physics class, after all).

These references pre-date the discovery of Λ; fortunately it is easily incorporated into the theory, so I have taken the liberty of doing so.

II. ENERGY CONDITIONS

Most global theorems in GR are based on the assumption of *energy conditions* – inequalities describing the behavior of the stress-energy tensor. Some are exactly valid for all known types of matter, and some may be violated in situations where the cosmological constant is important. The key energy conditions of interest to us are as follows:

- The *strong energy condition* (SEC) – this states that for any observer with a timelike 4-velocity $\boldsymbol{u}$, the observed energy density $\rho = \mathbf{T}(\boldsymbol{u},\boldsymbol{u})$ and the isotropic pressure $p = \frac{1}{3}\sum_{i=1}^3 \mathbf{T}(\boldsymbol{e}_{\hat{i}},\boldsymbol{e}_{\hat{i}}) = \frac{1}{3}(T+\rho)$ satisfy $\rho + 3p \ge 0$.
- The *null energy condition* (NEC) – this states that for any null vector $\boldsymbol{k}$, that $\mathbf{T}(\boldsymbol{k},\boldsymbol{k}) \ge 0$.

For a perfect fluid, the null energy condition is equivalent to $\rho + p \ge 0$.

It is easily seen that "normal" matter, radiation, electromagnetic fields ... satisfy both the strong and null energy conditions. However, a cosmological constant (for $\Lambda > 0$) satisfies the null energy condition but not the strong energy condition.

We note that the strong energy condition can be re-written in the form

$$\mathbf{R}(\boldsymbol{u},\boldsymbol{u}) = \frac{1}{8\pi}\left[\mathbf{T}(\boldsymbol{u},\boldsymbol{u}) - \frac{1}{2}T\mathbf{g}(\boldsymbol{u},\boldsymbol{u})\right] = \frac{1}{8\pi}\left(\rho + \frac{1}{2}T\right) = \frac{1}{8\pi}\left[\rho + \frac{1}{2}(-\rho+3p)\right] = \frac{1}{16\pi}(\rho+3p) \ge 0 \tag{1}$$

for all timelike $\boldsymbol{u}$, and the null energy condition can be written as $\mathbf{R}(\boldsymbol{k},\boldsymbol{k}) \ge 0$ for null $\boldsymbol{k}$. Continuity arguments then tell us that the strong energy condition implies the null energy condition.

Since we appear to have a cosmological constant in the real universe we should consider what happens when there is a Λ but all other matter fields obey the strong energy condition. In this case (which I will denote ΛSEC), we have

$$\mathbf{R}(\boldsymbol{u},\boldsymbol{u}) = \frac{1}{16\pi}(\rho_{\rm tot} + 3p_{\rm tot}) = \frac{1}{16\pi}(\rho_{\rm m} + 3p_{\rm m}) - \Lambda \ge -\Lambda. \tag{2}$$

---

*Electronic address: `chirata@tapir.caltech.edu`

III. CONGRUENCES AND THE RAYCHAUDHURI EQUATION

Let us next consider a set of geodesics that fills the spacetime (this is called a *congruence*). These geodesics are parameterized by their own internal parameters α, β, γ and by an affine parameter λ. We wish to consider the properties of the tangent vector $\boldsymbol{u}$, which we may now regard as a vector field in the spacetime; we assume it is normalized to $\boldsymbol{u} \cdot \boldsymbol{u} = s$.

Two cases will warrant our attention in this class: the case where the congruence is of timelike geodesics, where $s = -1$, and to null geodesics with $s = 0$. We now focus on the case of timelike geodesics ($s = -1$). The same developments can be carried over to the case of null geodesics ($s = 0$) with little difficulty. The main difference is that if $\boldsymbol{u}$ is a timelike vector, then the subspace of $T_{\mathcal{P}}\mathcal{M}$ orthogonal to $\boldsymbol{u}$ is a spacelike vector space. But $\boldsymbol{k}$ is a null vector, then the subspace of $T_{\mathcal{P}}\mathcal{M}$ orthogonal to $\boldsymbol{k}$ is instead spanned by $\boldsymbol{k}$ and two spacelike vectors $\{\boldsymbol{e}_x, \boldsymbol{e}_y\}$. This reduces the 3×3 tensors that we will encounter here to 2×2 tensors. The null geodesic version of the theory plays a critical role in the theory of black holes (e.g. the proof of the horizon area increase theorem). But for cosmology it is the timelike version of the theory that is critical.

We are interested in the evolution of the velocity gradient tensor $H_{\beta\alpha} = u_{\alpha;\beta}$. The geodesic nature of the congruence implies that

$$H_{\beta\alpha} u^\beta = u_{\alpha;\beta} u^\beta = \nabla_{\boldsymbol{u}} u_\alpha = 0. \tag{3}$$

Furthermore the normalization gives

$$H_{\beta\alpha} u^\alpha = u^\alpha u_{\alpha;\beta} = \frac{1}{2} (\boldsymbol{u} \cdot \boldsymbol{u})_{,\beta} = 0. \tag{4}$$

Thus we see that $\boldsymbol{H}$ is orthogonal to $\boldsymbol{u}$ (on both indices – important since $\boldsymbol{H}$ need not be symmetric).

A. Expansion and vorticity

Next define the *expansion* to be the contraction of $\boldsymbol{H}$: $\theta = H^\alpha{}_\alpha$. This is nothing but the trace of the velocity gradient; for timelike geodesics it is simply the generalization of the Newtonian div $\boldsymbol{v}$. We then find

$$\nabla_{\boldsymbol{u}} \theta = u^\alpha H^\beta{}_{\beta;\alpha} = u^\alpha u^\beta{}_{;\beta\alpha} = u^\alpha u^\beta{}_{;\alpha\beta} - u^\alpha R^\beta{}_{\gamma\beta\alpha} u^\gamma = u^\alpha u^\beta{}_{;\alpha\beta} - R_{\gamma\alpha} u^\gamma u^\alpha. \tag{5}$$

The first term simplifies because – using the product rule and the geodesic equation –

$$u^\alpha u^\beta{}_{;\alpha\beta} = (u^\alpha u^\beta{}_{;\alpha})_{;\beta} - u^\alpha{}_{;\beta} u^\beta{}_{;\alpha} = 0 - H^\alpha{}_\beta H^\beta{}_\alpha. \tag{6}$$

Thus we arrive at the result:

$$\nabla_{\boldsymbol{u}} \theta = -H^\alpha{}_\beta H^\beta{}_\alpha - R_{\gamma\alpha} u^\gamma u^\alpha. \tag{7}$$

This is a key result: the expansion changes along a trajectory in accordance with the square of the velocity gradient, and the Ricci tensor component $\mathbf{R}(\boldsymbol{u}, \boldsymbol{u})$.

A second result involves the *vorticity*, defined by $\omega_{\beta\alpha} = H_{[\beta\alpha]}$. We see that

$$\nabla_{\boldsymbol{u}} \omega_{\beta\alpha} = u^\gamma u_{[\beta;\alpha]\gamma} = u^\gamma (u_{[\beta|;\gamma|\alpha]} - R_{[\beta|}{}^\delta{}_{\gamma|\alpha]} u_\delta) = (u^\gamma u_{[\beta|;\gamma})_{;|\alpha]} - u^\gamma{}_{;[\alpha} u_{\beta];\gamma} - u^\gamma u^\delta R_{[\beta|\delta\gamma|\alpha]}. \tag{8}$$

Here the first term is zero due to the geodesic equation, and the last term is zero due to the symmetry of the Riemann tensor. **If** at some particular instant $\boldsymbol{\omega} = 0$ then $H_{\beta\alpha}$ is symmetric and the second term vanishes as well. It follows that if $\boldsymbol{\omega} = 0$ at one point on one of the geodesics, then it remains zero along the entire length of the geodesic. A congruence where this holds for all geodesics in some neighborhood of one of them is said to be *irrotational.* The cases of greatest interest to us will be irrotational flows.

B. Hypersurface-orthogonal congruences

A particular example of an irrotational flow is the set of geodesics projected normal to some spacelike surface Σ (specified by $f = 0$ where f is a smooth scalar field with $\boldsymbol{d}f$ timelike). Then we may write the 4-velocity of the timelike geodesics as a 1-form:

$$\boldsymbol{u} = \alpha\, \boldsymbol{d}f + \boldsymbol{c}, \tag{9}$$

where $\alpha \neq 0$ and $\boldsymbol{c} = 0$ on Σ. If $\boldsymbol{c} = 0$ on Σ (i.e. when $f = 0$) then we may write $\boldsymbol{c} = f\boldsymbol{\beta}$. We then find that the vorticity 2-form is

$$\boldsymbol{\omega} = \boldsymbol{du} = \boldsymbol{d}\alpha \wedge \boldsymbol{d}f + \boldsymbol{d}f \wedge \boldsymbol{\beta} + f\boldsymbol{d\beta}. \tag{10}$$

On Σ, i.e. at $f = 0$, we have $f = 0$ and $\boldsymbol{d}f = \alpha^{-1}\boldsymbol{u}$, so that

$$\boldsymbol{\omega} = \boldsymbol{u} \wedge \left(-\alpha^{-1}\boldsymbol{d}\alpha + \alpha^{-1}\boldsymbol{\beta}\right). \tag{11}$$

But then $\omega_{\beta\alpha} = u_{[\beta}v_{\alpha]}$ for some $\boldsymbol{v}$. Combining this with the orthogonality rule that $\omega_{\beta\alpha}u^\alpha = 0$, we see that $\boldsymbol{\omega} = 0$ (this is most easily shown in a local orthonormal basis where $\boldsymbol{u}$ is the 4-velocity). Thus the congruence of geodesics orthogonal to Σ is indeed irrotational.

Given such a spacelike surface, we may define the *extrinsic curvature* (or *second fundamental form*) as the gradient of the velocity field of the geodesics projected normal to it: $K_{\alpha\beta} \equiv -H_{\beta\alpha} = -u_{\alpha;\beta}$. Since $\mathbf{K}$ is orthogonal to $\boldsymbol{u}$, it is a symmetric rank $\binom{0}{2}$ tensor field on the surface Σ, i.e. at any $\mathcal{P}$ we have $\mathbf{K} \in T^*_{\mathcal{P}}\Sigma \times T^*_{\mathcal{P}}\Sigma$. The *first fundamental form* $\boldsymbol{\gamma}$ is the metric $\gamma_{\alpha\beta}$ induced by the geometry of the manifold $\mathcal{M}$ – that is, $\boldsymbol{\gamma}$ is the dot product $\mathbf{g}$ with its domain restricted to vectors tangent to Σ, i.e. in $T_{\mathcal{P}}\Sigma \subset T_{\mathcal{P}}\mathcal{M}$.

As its name suggests, the extrinsic curvature describes whether observers moving normal to the spatial slice are moving toward each other or away from each other. For the FRW universe, and for the constant-t slices Σ_t, it is easily seen that $K_{\alpha\beta} = -H\gamma_{\alpha\beta}$. But it is also easy to see that a curved 3-surface embedded in flat 4-dimensional spacetime can have $\mathbf{K} \neq 0$.

Some additional intuition can be gained by considering 2-dimensional surfaces embedded in $\mathbb{R}^3$. If we have a surface that is tangent to the xy-plane at the origin, then it is given to quadratic order by:

$$z = \frac{1}{2}c_{xx}x^2 + c_{xy}xy + \frac{1}{2}c_{yy}y^2 + \dots\,. \tag{12}$$

The normal vector $\mathbf{n}$ to such a surface is $(-c_{xx}x - c_{xy}y, -c_{xy}x - c_{yy}y, 1)$ (to linear order) and its gradient (projected onto the xy-plane, i.e. the space tangent to the surface at the origin) is

$$-n_{\alpha;\beta} = \begin{pmatrix} c_{xx} & c_{xy} \\ c_{xy} & c_{yy} \end{pmatrix}. \tag{13}$$

Thus in flat space the extrinsic curvature describes whether the surface "curves up" (positive definite), "curves down" (negative definite), is "saddle-shaped" (one + and one − eigenvalue), or is "cylinder-curved" (one zero eigenvalue). But the construction above generalizes to surfaces Σ embedded in arbitrary manifolds $\mathcal{M}$.

C. The timelike geodesic case

Now that $\boldsymbol{H}$ is a purely spacelike 3×3 tensor on the space orthogonal to $\boldsymbol{u}$, we may decompose it into an expansion, a vorticity, and a shear:

$$H_{\beta\alpha} = \frac{1}{3}\theta h_{\beta\alpha} + \omega_{\beta\alpha} + \sigma_{\beta\alpha}, \tag{14}$$

where $h_{\beta\alpha} \equiv g_{\beta\alpha} + u_\beta u_\alpha$ is the spatial 3-metric orthogonal to $\boldsymbol{u}$, $\boldsymbol{\omega}$ is antisymmetric, and $\boldsymbol{\sigma}$ is symmetric, traceless, and orthogonal to $\boldsymbol{u}$. Squaring this and using (anti)symmetry properties and the tracelessness of $\boldsymbol{\sigma}$, we find that

$$H^\alpha{}_\beta H^\beta{}_\alpha = \frac{1}{3}\theta^2 + \omega_{\alpha\beta}\omega^{\beta\alpha} + \sigma_{\alpha\beta}\sigma^{\alpha\beta}. \tag{15}$$

Returning to Eq. (7), we find that, following along a geodesic,

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \omega_{\alpha\beta}\omega^{\beta\alpha} - \sigma_{\alpha\beta}\sigma^{\alpha\beta} - R_{\gamma\alpha}u^\gamma u^\alpha. \tag{16}$$

This is the *Raychaudhuri equation.*

Here the third term is non-positive since it is the square of a tensor defined on a spatial slice. The second term vanishes if the flow is irrotational. The fourth term is non-positive if the strong energy condition holds. We thus see that

$$\frac{d\theta}{d\tau} \leq -\frac{1}{3}\theta^2 \quad (\text{SEC} + \text{irrotational}). \tag{17}$$

Now $\theta/3$ is the angular-averaged velocity gradient, i.e. the angular-average Hubble rate if an observer following one of our geodesics measures the positions of the neighboring geodesics. Thus Eq. (17) says something about the behavior of the Hubble rate as a function of time – namely that it is non-increasing, and that $d(\theta/3)/d\tau + (\theta/3)^2 \le 0$. For an FRW Universe, since $\theta/3 = \dot{a}/a$, this says that $\ddot{a} \le 0$, i.e. that cosmic expansion is non-accelerating. But Eq. (17) is a generalization of this result to arbitrary geometry. Of course, in the real Universe the presence of the cosmological constant (SEC violation) circumvents this theorem.

If there is a cosmological constant, but other fields obey the SEC, then Eq. (17) should be replaced with

$$\frac{d\theta}{d\tau} \le -\frac{1}{3}\theta^2 + \Lambda \quad (\Lambda\text{SEC} + \text{irrotational}). \tag{18}$$

If $|\theta| \le \sqrt{3\Lambda}$ then Eq. (18) does not say anything very interesting. If θ is outside this range, however – if the local Hubble rate is larger than $\sqrt{\Lambda/3}$ – then there is a profound implication. Let us define the new variable

$$\xi = \sqrt{\frac{3}{4\Lambda}} \ln \frac{\theta + \sqrt{3\Lambda}}{\theta - \sqrt{3\Lambda}}, \tag{19}$$

which is in the range $\infty > \xi > 0$ for $\sqrt{3\Lambda} < \theta < \infty$. Then

$$\frac{d\xi}{d\theta} = \sqrt{\frac{3}{4\Lambda}} \left(\frac{1}{\theta + \sqrt{3\Lambda}} - \frac{1}{\theta - \sqrt{3\Lambda}} \right) = \frac{1}{-\theta^2/3 + \Lambda}, \tag{20}$$

and hence

$$\frac{d\xi}{d\tau} = \frac{d\xi}{d\theta}\frac{d\theta}{d\tau} \ge 1. \tag{21}$$

[Technical note: all of the above considerations apply for $\Lambda = 0$ if we replace $\xi = 3/\theta$, in accordance with l'Hôpital's rule. For $\Lambda < 0$ we would have $\xi = \sqrt{-3/\Lambda}\,\cot^{-1}(\theta/\sqrt{-3\Lambda})$, and there is no restriction on the value of θ for Eq. (21) to be valid and useful.]

Thus if we have an irrotational flow of geodesics, ΛSEC applies, and the local expansion at our position (with $\tau = \tau_0$) is $\theta > \sqrt{3\Lambda}$, then we can find the current value of $\xi < 0$. If we follow the geodesics back to previous epochs, we must reach $\xi = 0$, i.e. $\theta = +\infty$, at some time $\tau_{\rm c}$ in the range

$$\tau_0 - \xi(\tau_0) \le \tau_{\rm c} < \tau_0. \tag{22}$$

What happens at $\tau_{\rm c}$? We haven't said yet, but there are two possibilities. One is an actual physical singularity – a point where the curvature of spacetime diverges, and in the context of classical GR one has no idea how to "integrate back" to before such a singularity. The other is that it is merely an artifact of our suite of geodesics. For example, on the unit sphere the congruence of geodesics formed by the meridians (lines of constant longitude) all converge at the poles, with no ill effect. Their divergence $u_{i;j}$ blows up there (exercise: check the divergence $u^i{}_{;i}$ for this case!) but this is only because the tangent vector $\boldsymbol{u}$ ceases to be a single-valued function of position when the geodesics cross.

The subject of singularity theorems will be to address which of these situations arises.

IV. THE INITIAL SINGULARITY

Let us now consider a spacetime with the following properties:

- There exists a spacelike surface Σ such that all of the inextendible timelike curves intersect Σ. (The idea is that Σ is to fill "all of space.")
- We suppose that the direction of forward time is globally defined and that there are no closed timelike loops (i.e. no time machines!).
- The expansion or negative trace of the extrinsic curvature satisfies $-K \ge \theta_{\rm min}$ everywhere on Σ, where $\theta_{\rm min} > \sqrt{3\Lambda}$ (the last inequality is not necessary if $\Lambda < 0$). This is really a way of saying that the observers on our spacelike slice measure an expanding universe by observing the recession velocities of their neighbors, and that the angle-averaged Hubble rate everywhere exceeds some minimum value. [Recall that $-K$ is the expansion of the geodesics projected normal to Σ.]
- The SEC holds, with the possible exception of a cosmological constant.

Our goal is to show that there is an initial singularity, and hopefully also establish some results about its nature.

A. The maximal geodesic

We first take a set of timelike geodesics, projected normal to Σ and extended back into the past in accordance with the geodesic equation. Without loss of generality we set $\tau = 0$ where they pierce Σ. From the previous discussion, we know that all of these geodesics must encounter either a physical singularity or an orbit-crossing between $\tau = 0$ and $\tau = -\xi_{\min}$, where $\xi_{\min}$ is obtained from $\theta_{\min}$ in accordance with Eq. (19).

How can we tell which possibility occurs? In general we can't. However, it turns out that some of the geodesics projected backward from Σ are special. To see this, consider a point $\mathcal{P}$ in the past of Σ, i.e. such that there is at least one forward-directed timelike trajectory (not necessarily a geodesic) starting from $\mathcal{P}$ and reaching Σ. Then there is a non-empty set $\mathfrak{N}$ of causal (timelike or null) trajectories from $\mathcal{P}$ to Σ. We may define the proper time function $\ell : \mathfrak{N} \to [0, \infty)$. We will next consider for each such $\mathcal{P}$, the curve of maximal proper time.

[Mathematical aside: To be rigorous, we would need to prove that the curve of maximal proper time exists; this is not trivial to prove. The idea is two-fold: one must prove that $\mathfrak{N}$ is compact (time machines would spoil this since one could take the neighborhoods of curves that pass through the time machine $n \in \mathbb{Z}^+$ times), and also that ℓ is upper-semicontinuous, i.e. that $\limsup_{\mathcal{S}\to\mathcal{C}} \ell(\mathcal{S}) = \ell(\mathcal{C})$ for all $\mathcal{C}$. The upper semicontinuity is required since in the vicinity of any timelike curve $\mathcal{C}$, or timelike segment thereof, one can make a neighboring curve that zooms around it at close to the speed of light with a tiny radius. A sequence of such curves may converge to $\mathcal{C}$, but their total proper time remains less than $\ell(\mathcal{C})$ by a finite amount.]

Now this maximal-length curve $\mathcal{C}$ we have obtained from $\mathcal{P}$ to some point $\mathcal{Q} \in \Sigma$ must be a geodesic. Moreover its tangent vector $\boldsymbol{u}$ must be orthogonal to Σ at $\mathcal{Q}$ (otherwise we perturb the endpoint and find a longer curve), so $\mathcal{C}$ is in the aforementioned congruence. If its length L exceeds ξ_0, then the congruence of geodesics projected into the past from Σ must have an orbit-crossing somewhere between $\mathcal{P}$ and $\mathcal{Q}$, say at $\mathcal{A}$.

B. Variations of the proper time

Our next plan of attack is to contradict the result that $\mathcal{C}$ is truly maximal proper time by showing that there is a variation with negative second derivative.

Let us consider now the problem of taking a geodesic $\mathcal{C}$ and finding the length of neighboring curves that start at $\mathcal{P}$ and end on Σ. In particular we consider an extension $\mathcal{E}(t, s)$, $-L < t < 0$, $s \in \mathbb{R}$ of $\mathcal{C}$, such that $\mathcal{E}(t, 0) = \mathcal{C}(t)$, and define the vectors $\boldsymbol{T} = \partial_t \mathcal{E}$ and $\boldsymbol{S} = \partial_s \mathcal{E}$. The trajectory is restricted to $\mathcal{E}(-L, s) = \mathcal{P}$ and $\mathcal{E}(0, s) \in \Sigma$, and with $\boldsymbol{T}$ restricted to be normal to Σ.

The variation of the proper time along the trajectory $\mathcal{E}$ may be found via

$$\frac{d\ell[\mathcal{E}(s)]}{ds} = \int_{-L}^{0} \partial_s \sqrt{-\boldsymbol{T}\cdot\boldsymbol{T}}\, dt = \int_{-L}^{0} \frac{\boldsymbol{T}\cdot(D\boldsymbol{S}/\partial t)}{\sqrt{-\boldsymbol{T}\cdot\boldsymbol{T}}}\, dt, \tag{23}$$

where we recall that $D\boldsymbol{T}/\partial s = \nabla_{\boldsymbol{S}}\boldsymbol{T} = \nabla_{\boldsymbol{T}}\boldsymbol{S} = D\boldsymbol{S}/\partial t$ (by commutation). Integration by parts gives

$$\frac{d\ell[\mathcal{E}(s)]}{ds} = -\int_{-L}^{0} \boldsymbol{S}\cdot\frac{D}{\partial t}\boldsymbol{v}\, dt, \tag{24}$$

where $\boldsymbol{v} = \boldsymbol{T}/\sqrt{-\boldsymbol{T}\cdot\boldsymbol{T}}$ is the unit vector in the direction of $\boldsymbol{T}$. (The surface terms vanish since $\boldsymbol{S} = 0$ at $\mathcal{P}$ and $\boldsymbol{T}\cdot\boldsymbol{S} = 0$ at Σ.)

It turns out that we need one more derivative (we already know the first derivative is zero):

$$\frac{d^2\ell[\mathcal{E}(s)]}{ds^2} = -\int_{-L}^{0} \left[\frac{D\boldsymbol{S}}{\partial s}\cdot\frac{D\boldsymbol{v}}{\partial t} + \boldsymbol{S}\cdot\frac{D}{\partial s}\frac{D\boldsymbol{v}}{\partial t}\right] dt. \tag{25}$$

At $s = 0$ (i.e. expanding around a geodesic $\mathcal{C}$), the first term in brackets vanishes. Moreover, we know that (using that $\boldsymbol{S}$ and $\boldsymbol{T}$ commute):

$$\frac{D}{\partial s}\frac{D\boldsymbol{v}}{\partial t} = \nabla_{\boldsymbol{S}}\nabla_{\boldsymbol{T}}\boldsymbol{v} = \nabla_{\boldsymbol{T}}\nabla_{\boldsymbol{S}}\boldsymbol{v} - \mathbf{R}(\_, \boldsymbol{v}, \boldsymbol{S}, \boldsymbol{T}). \tag{26}$$

Furthermore, we find

$$\nabla_{\boldsymbol{S}}\boldsymbol{v} = \frac{\nabla_{\boldsymbol{S}}\boldsymbol{T}}{\sqrt{-\boldsymbol{T}\cdot\boldsymbol{T}}} + \frac{(\boldsymbol{T}\cdot\nabla_{\boldsymbol{S}}\boldsymbol{T})\boldsymbol{T}}{(-\boldsymbol{T}\cdot\boldsymbol{T})^{3/2}} = \frac{\nabla_{\boldsymbol{T}}\boldsymbol{S}}{\sqrt{-\boldsymbol{T}\cdot\boldsymbol{T}}} + \beta\boldsymbol{T} = \nabla_{\boldsymbol{v}}\boldsymbol{S} + \beta\boldsymbol{T}, \tag{27}$$

where $\beta \in \mathbb{R}$. Now if we apply $\nabla_{\boldsymbol{T}}$ to this, evaluate at $s = 0$ so that $\boldsymbol{T} = \boldsymbol{v} = \boldsymbol{u}$ and $\nabla_{\boldsymbol{u}}\boldsymbol{u} = 0$, and impose the restriction that $\boldsymbol{S} \cdot \boldsymbol{T} = 0$ everywhere (i.e. we consider variations orthogonal to the trajectory):

$$\frac{d^2\ell[\mathcal{E}(s)]}{ds^2} = -\int_{-L}^{0} \boldsymbol{S} \cdot [\nabla_{\boldsymbol{u}}\nabla_{\boldsymbol{u}}\boldsymbol{S} + \mathbf{R}(\_, \boldsymbol{u}, \boldsymbol{S}, \boldsymbol{u})] \, dt. \tag{28}$$

The part in brackets is zero for geodesic deviations $\boldsymbol{S}$, i.e. if the variation of the path under consideration is a perturbation onto a neighboring geodesic.

C. Assembling the pieces

We now return to our maximal geodesic. Suppose that an orbit crossing occurs at $\mathcal{A}$, at time t_1. This means that there is a nonzero first-order perturbation of the trajectory from $\mathcal{Q}$ to Σ with $\boldsymbol{S} = 0$ at $\mathcal{A}$; let us call this geodesic deviation $\boldsymbol{S}^{\#}$ (orthogonality to $\boldsymbol{u}$ is trivially verified). Then let us consider the variation of the path:

$$\boldsymbol{S}(t) = \begin{cases} 0 & t < t_1 - \epsilon \\ \boldsymbol{X} & t_1 \le t \le t_1 + \epsilon \\ \boldsymbol{S}^{\#} & t > t_1 + \epsilon \end{cases}, \tag{29}$$

where $\boldsymbol{X} = 0$ at t_1 and $\boldsymbol{X} = \boldsymbol{S}^{\#}$ at $t_1 + \epsilon$. Now in Eq. (28), the integral has exactly zero contribution from $t < t_1 - \epsilon$ or $t > t_1 + \epsilon$. If we make ϵ small, then we may connect a geodesic from $\mathcal{C}(t_1 - \epsilon)$ to $\mathcal{C}(t_1 + \epsilon)$, and then the only contribution to the integral comes from the δ-function spikes in $\boldsymbol{S}$ at $t_1 \pm \epsilon$:

$$\frac{d^2\ell[\mathcal{E}(s)]}{ds^2} = \boldsymbol{S} \cdot \Delta \frac{D\boldsymbol{S}}{dt}\bigg|_{t_1-\epsilon} + \boldsymbol{S} \cdot \Delta \frac{D\boldsymbol{S}}{dt}\bigg|_{t_1+\epsilon}. \tag{30}$$

The first term here vanishes since $\boldsymbol{S} = 0$, leaving us with the second. But as ϵ becomes small, the neighborhood of $\mathcal{A}$ can be approximated as flat. In a local Lorentz coordinate system then we find

$$\boldsymbol{S}^{\#}(t) = \frac{D\boldsymbol{S}^{\#}}{dt}\bigg|_{t_1} (t - t_1) + ...; \tag{31}$$

straightforward computation then shows that $D\boldsymbol{S}/dt$ and $\boldsymbol{S}$ are spacelike vectors in the same direction, hence with positive dot product. Then the second derivative of the path length is positive so $\mathcal{C}$ could not have been maximal.

The upshot is that given any point $\mathcal{P}$ in the past of Σ, there maximal geodesic from $\mathcal{P}$ to Σ is no longer than $\xi_{\min}$. In fact this must be true of any causal curve from $\mathcal{P}$ to Σ. Therefore **any observer's trajectory, geodesic or not, followed backward in time from Σ, must have originated from some initial singularity within a time $\xi_{\min}$.**