Lecture XX: The area theorem

(Dated: November 8, 2019)

I. OVERVIEW

I want to turn now from computations in GR to some global facts about event horizons. You can probably imagine that horizons are messy when black holes form, or when they swallow particles, or when black holes merge. But it turns out there is a very powerful theorem about horizons: their area can only stay the same or increase, it can never decrease. I want to discuss the area theorem here for three reasons:

- It is one of the few practical and general results on black hole evolution.
- It illustrates the usefulness of global techniques for studying spacetimes.
- It relates classical GR to the 2nd law of thermodynamics.

II. THE DESCRIPTION OF HORIZONS WITH NULL GEODESICS

Let's imagine a 4D spacetime, embedded in an asymptotically flat background. We may consider a point $\mathcal{P}$ to be inside the black hole ($\mathcal{P} \in \mathbb{B}$) if there is no causal path from $\mathcal{P}$ to the distant exterior. The boundary of the black hole is the event horizon, $\mathbb{H}$, which is a 3-dimensional surface in the spacetime. We will focus on the smooth part of the horizon first (non-smooth places such as *caustics* where the horizons of two holes merge will be considered later).

A. Tangents and normals to the horizon

Now given any smooth surface, and a point $\mathcal{P}$ on that surface, we note that the surface should be described by a 1-form $\boldsymbol{l} \neq 0$ such that a displacement Δx^μ takes us inside the black hole if $l_\mu \Delta x^\mu < 0$, outside the black hole if $l_\mu \Delta x^\mu > 0$, and tangent to the horizon if $l_\mu \Delta x^\mu = 0$. The 1-form $\boldsymbol{l}$ is thus the normal to the horizon.

We can see that if $\boldsymbol{l}$ must be null. If it were spacelike at $\mathcal{P} \in \mathbb{H}$, then we could go to a local orthonormal coordinate system where $\mathcal{P}$ is at the origin and $\boldsymbol{l}$ is in the 1-direction; then a particle moving at $\dot{x}^1 > 0$ can go from inside the hole to outside, which is a contradiction. If it were timelike, we could set $l_i = 0$; then if $l_0 > 0$ then particles moving forward in time ($\dot{x}^0 > 0$) go from inside to outside (again a contradiction), and if $l_0 < 0$ then all particles go from outside to inside (this is a contradiction since starting from the point $(-\epsilon, 0, 0, 0)$ – nominally outside the hole – all particles must go forward in time and into the hole). So the only possibility is for $\boldsymbol{l}$ to be null. In order for forward-going particles to be able to go from outside the hole to inside, but not the other way around, $\boldsymbol{l}$ must be future-directed, $l^0 > 0$ in a local orthonormal frame for an observer going forward in time.

An interesting feature of this is that $l_\mu l^\mu = 0$, so the null vector $\boldsymbol{l}$ is tangent to $\mathbb{H}$. Really it is both tangent and normal: this is possible only in the Minkowski-signature $(-+++)$ spacetime where a null vector $\boldsymbol{l}$ is orthogonal to itself.

In this way, we can think of the horizon $\mathbb{H}$ as a 3D surface with the null tangent/normal vector $\boldsymbol{l}$. We can take the vector $\boldsymbol{l}$ and draw its *integral curves*, i.e., curves of the form $\mathrm{d}x^\mu/\mathrm{d}\lambda = l^\mu$. The integral curves are 1D curves on the 3D surface $\mathbb{H}$. There are of course many such integral curves; to choose one, I need to give you 2 more numbers (in the stationary black hole case in the usual coordinates, those are θ and ϕ, but in general we can give them a general designation χ and ψ).

The integral curves of the horizon have a special name: they are called *generators* of the horizon. One can think of the generators as the paths of light rays that are just on the boundary of being able to escape the black hole: a light ray just ahead of the generators can escape, one behind is trapped forever.

I have not chosen the normalization of $\boldsymbol{l}$ yet, but of course I can define the integral curves and generators without such a choice. Changing the normalization is merely a re-mapping of λ.

B. Generators are null geodesics

It turns out that the horizon generators are not just null curves, they are null geodesics. To see this, we recall that the horizon – as a smooth surface – can be written as $f = 0$ where f is some scalar function. Then we must have

$l_\mu = \beta \nabla_\mu f$ (the 1-form normal to the surface $f = 0$ is the gradient, with a possible scalar multiplying factor β). Now we may write this as

$$l_\mu = \nabla_\mu(\beta f) - f \nabla_\mu \beta = \nabla_\mu(\beta f) \quad \text{on horizon } (f = 0). \tag{1}$$

So in fact we don't need the separate factor of β, we can just absorb it into the definition of f. This is what I will do in what follows: we have $l_\mu = \nabla_\mu f$ (again, for points on $\mathbb{H}$).

With this said, we find:

$$\frac{\mathrm{D}l_\mu}{\mathrm{d}\lambda} = l^\alpha \nabla_\alpha l_\mu = (\nabla^\alpha f)(\nabla_\alpha \nabla_\mu f) = \frac{1}{2} \nabla_\mu [(\nabla^\alpha f)(\nabla_\alpha f)], \tag{2}$$

where we used the fact that covariant derivatives commute when acting on a scalar. But we know $(\nabla^\alpha f)(\nabla_\alpha f) = l^\alpha l_\alpha = 0$ on $\mathbb{H}$, so Eq. (2) is some scalar times $\nabla_\mu f = l_\mu$. Therefore:

$$\frac{\mathrm{D}l_\mu}{\mathrm{d}\lambda} = q l_\mu \tag{3}$$

for some scalar q. So $\boldsymbol{l}$ is parallel-transported along a generator (aside from possible changes in normalization). By re-definition of λ, we could eliminate q (and thus see that the generators are geodesics), but we won't do that in this lecture as it is more convenient to allow for the fully general parameterization. This way we can use λ as an arbitrary parameter describing when we want to describe the horizon.

C. Caustics

A place where the smooth behavior of the horizon (and hence the single normal vector $\boldsymbol{l}$) breaks down is called a *caustic*. Geometrically, the boundary of a 4D black hole $\mathbb{B}$ is 3D, but it might have sharp "edges" that are 2D, 1D, or even 0D (vertices). They really do exist: for example, in a spherically symmetric collapsing star, the point where the horizon first forms is a caustic! When we go to prove global theorems about the generators of horizons, it seems like the caustics might be a challenge. But as it turns out, generators can only begin at caustics; going forward in time, they cannot hit a caustic.

I won't give a rigorous mathematical proof of this statement, but I will motivate it by going to a local orthonormal frame at an event $\mathcal{P}$ where there is a caustic. You could imagine several ways a caustic might exist at $\mathcal{P}$: $\mathbb{B}$ might be the interior of a ball that pops into existence at $\mathcal{P}$ (i.e., the forward light cone: $\sqrt{x^2 + y^2 + z^2} < t$), or a cylinder ($\sqrt{x^2 + y^2} < t$), or a slab ($|x| < t$), depending on whether the caustic is 0, 1, or 2-dimensional. In each of these cases, new generators are formed; if one starts from a spacetime with no black holes initially, then all horizon generators, if followed back in time, must have come from a caustic. You can analytically continue those generators back before the caustic without any problem; they fly off into space into the distant past.

But clearly these situations do not work in reverse: we can't have the generators come together in the future, have an "edge" to the black hole, and then have them zoom off into space into the distant future. Because we defined $\mathbb{B}$ to be the region where you can't get out going forward in time, there is a fundamental asymmetry in the definition.

III. THE BEHAVIORS OF BUNDLES OF NULL GEODESICS

We now want to turn our attention to the more general behavior of the generators. We will start with the kinematics of the geodesics, and then add some dynamical information. Our strategy will be first to relate the Ricci tensor to the area element of the horizon, and then introduce some inequalities to show that a contracting area element leads to a caustic in the future, and hence to a contradiction.

A. The horizon area and the Ricci tensor

Let us also recall that then the horizon is a 3D surface described by $x^\mu(\lambda, \chi, \psi)$. In what follows, I will choose a special set of coordinates, $x^0 = f$, $x^1 = \lambda$, $x^2 = \chi$, and $x^3 = \psi$. (Recall $f < 0$ inside the black hole, $f = 0$ on the horizon, and $f > 0$ outside the black hole.) Here $l^\alpha = \partial x^\alpha / \partial \lambda \to (0, 1, 0, 0)$: thus $\boldsymbol{l}$ is the $\boldsymbol{e}_\lambda$ coordinate vector. The fact that $\boldsymbol{l}$ is orthogonal to the horizon implies $g_{\lambda\lambda} = g_{\lambda\chi} = g_{\lambda\psi} = 0$. Moreover, the fact that $l_\mu = \nabla_\mu f \to (1, 0, 0, 0)$ implies that $g_{f\lambda} = 1$. By doing coordinate changes of the type

$$\chi \to \chi + f \times F(\lambda, \chi, \psi) \tag{4}$$

and similarly for ψ and λ, we can vary $\boldsymbol{e}_f$ (adding any multiple we want of $\boldsymbol{e}_\chi$, $\boldsymbol{e}_\psi$, and $\boldsymbol{e}_\lambda$) and thus set $g_{ff} = g_{f\chi} = g_{f\psi} = 0$ on $\mathbb{H}$ without relabeling the generators. It follows that – on $\mathbb{H}$ itself (i.e., at $f = 0$) – the metric has the form

$$g_{\mu\nu} \to \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & g_{\chi\chi} & g_{\chi\psi} \\ 0 & 0 & g_{\chi\psi} & g_{\psi\psi} \end{pmatrix} \quad \text{and} \quad g^{\mu\nu} \to \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & g^{\chi\chi} & g^{\chi\psi} \\ 0 & 0 & g^{\chi\psi} & g^{\psi\psi} \end{pmatrix}. \tag{5}$$

In this system, if I take a "snapshot" of the horizon at some given parameter λ (which I may choose arbitrarily), then the physical area element of the horizon is $J\,\mathrm{d}\chi\,\mathrm{d}\psi$, where

$$J = \frac{\mathrm{d}A}{\mathrm{d}\chi\,\mathrm{d}\psi} = \sqrt{g_{\chi\chi}g_{\psi\psi} - (g_{\chi\psi})^2}, \tag{6}$$

i.e., J is the square of the determinant of the 2×2 sub-block of $\mathbf{g}$. One can further show by direct computation that on $\mathbb{H}$ we have:

$$\Gamma^f{}_{\lambda\lambda} = \Gamma^\chi{}_{\lambda\lambda} = \Gamma^\psi{}_{\lambda\lambda} = \Gamma^f{}_{\chi\lambda} = \Gamma^f{}_{\psi\lambda} = 0. \tag{7}$$

It turns out the next useful thing to do is to get the $\lambda\lambda$ component of the Ricci tensor. On $\mathbb{H}$, this is

$$\begin{aligned} R_{\lambda\lambda} &= \Gamma^\alpha{}_{\lambda\lambda,\alpha} - \Gamma^\alpha{}_{\lambda\alpha,\lambda} + \Gamma^\alpha{}_{\delta\alpha}\Gamma^\delta{}_{\lambda\lambda} - \Gamma^\alpha{}_{\delta\lambda}\Gamma^\delta{}_{\lambda\alpha} \\ &= \Gamma^f{}_{\lambda\lambda,f} - \Gamma^f{}_{\lambda f,\lambda} - \Gamma^\chi{}_{\lambda\chi,\lambda} - \Gamma^\psi{}_{\lambda\psi,\lambda} + \Gamma^f{}_{\lambda f}\Gamma^\lambda{}_{\lambda\lambda} + \Gamma^\chi{}_{\lambda\chi}\Gamma^\lambda{}_{\lambda\lambda} + \Gamma^\psi{}_{\lambda\psi}\Gamma^\lambda{}_{\lambda\lambda} \\ &\quad -(\Gamma^f{}_{f\lambda})^2 - (\Gamma^\chi{}_{\chi\lambda})^2 - 2\Gamma^\chi{}_{\psi\lambda}\Gamma^\psi{}_{\chi\lambda} - (\Gamma^\psi{}_{\psi\lambda})^2, \end{aligned} \tag{8}$$

where in the second line we eliminated all the Christoffel symbols that are zero. One can also see by direct computation – and using the fact that at $f = 0$ we have a metric of the form Eq. (5) – that

$$\Gamma^f{}_{\lambda\lambda,f} - \Gamma^f{}_{\lambda f,\lambda} - (\Gamma^f{}_{f\lambda})^2 = -\frac{1}{2}g^{ff}{}_{,f}g_{\lambda\lambda,f} - \frac{1}{4}(g_{\lambda\lambda,f})^2 = \frac{1}{4}(g_{\lambda\lambda,f})^2 = -\Gamma^\lambda{}_{\lambda\lambda}\Gamma^f{}_{\lambda f}, \tag{9}$$

where in the second-to-last step we use the matrix partial derivative formula $g^{\mu\nu}{}_{,\gamma} = -g^{\mu\alpha}g^{\beta\nu}g_{\alpha\beta,\gamma}$ and plug in the metric to get $g^{ff}{}_{,f} = -g_{\lambda\lambda,f}$. The last step then used that $\Gamma^\lambda{}_{\lambda\lambda} = -\frac{1}{2}g_{\lambda\lambda,f}$ and $\Gamma^f{}_{\lambda f} = \frac{1}{2}g_{\lambda\lambda,f}$.

We then see that

$$R_{\lambda\lambda} = -\Gamma^\chi{}_{\lambda\chi,\lambda} - \Gamma^\psi{}_{\lambda\psi,\lambda} + \Gamma^\chi{}_{\lambda\chi}\Gamma^\lambda{}_{\lambda\lambda} + \Gamma^\psi{}_{\lambda\psi}\Gamma^\lambda{}_{\lambda\lambda} - (\Gamma^\chi{}_{\chi\lambda})^2 - 2\Gamma^\chi{}_{\psi\lambda}\Gamma^\psi{}_{\chi\lambda} - (\Gamma^\psi{}_{\psi\lambda})^2. \tag{10}$$

A further simplification arises if we note that

$$\Gamma^\chi{}_{\lambda\chi} + \Gamma^\psi{}_{\lambda\psi} = \sum_{\alpha,\beta\in\chi,\psi} \frac{1}{2}g^{\alpha\beta}g_{\alpha\beta,\lambda} = \frac{1}{2}\partial_\lambda \ln\det\mathbf{g}_{2\times2} = \frac{J_{,\lambda}}{J}. \tag{11}$$

Then

$$R_{\lambda\lambda} = -\left(\frac{J_{,\lambda}}{J}\right)_{,\lambda} + \Gamma^\lambda{}_{\lambda\lambda}\frac{J_{,\lambda}}{J} - (\Gamma^\chi{}_{\chi\lambda})^2 - 2\Gamma^\chi{}_{\psi\lambda}\Gamma^\psi{}_{\chi\lambda} - (\Gamma^\psi{}_{\psi\lambda})^2. \tag{12}$$

B. Some inequalities

As a next step, we note that

$$(\Gamma^\chi{}_{\chi\lambda})^2 + 2\Gamma^\chi{}_{\psi\lambda}\Gamma^\psi{}_{\chi\lambda} + (\Gamma^\psi{}_{\psi\lambda})^2 = \sum_{\alpha\beta\in\{\chi,\psi\}} \Gamma^\alpha{}_{\beta\lambda}\Gamma^\beta{}_{\alpha\lambda} = \frac{1}{4}\sum_{\alpha\beta\gamma\delta\in\{\chi,\psi\}} g^{\alpha\delta}g_{\delta\beta,\lambda}g^{\beta\gamma}g_{\gamma\alpha,\lambda} = \frac{1}{4}\mathrm{Tr}(\mathbf{g}_{2\times2}^{-1}\mathbf{g}_{,\lambda}\mathbf{g}_{2\times2}^{-1}\mathbf{g}_{,\lambda}). \tag{13}$$

Now it turns out that if v_1 and v_2 are the eigenvalues of $\mathbf{g}_{2\times2}^{-1}\mathbf{g}_{,\lambda}$, then the above result is $\frac{1}{4}(v_1^2 + v_2^2)$; but also

$$\frac{J_{,\lambda}}{J} = \frac{1}{2}\mathrm{Tr}(\mathbf{g}_{2\times2}^{-1}\mathbf{g}_{,\lambda}) = \frac{1}{2}(v_1 + v_2), \tag{14}$$

and so it follows that

$$(\Gamma^{\chi}{}_{\chi\lambda})^2 + 2\Gamma^{\chi}{}_{\psi\lambda}\Gamma^{\psi}{}_{\chi\lambda} + (\Gamma^{\psi}{}_{\psi\lambda})^2 = \frac{1}{4}(v_1^2 + v_2^2) = \frac{1}{8}(v_1+v_2)^2 + \frac{1}{8}(v_1-v_2)^2 \geq \frac{1}{8}(v_1+v_2)^2 = \frac{1}{2}\left(\frac{J_{,\lambda}}{J}\right)^2. \quad (15)$$

We therefore have the inequality:

$$R_{\lambda\lambda} \leq -\left(\frac{J_{,\lambda}}{J}\right)_{,\lambda} - \frac{1}{2}\left(\frac{J_{,\lambda}}{J}\right)^2 + \Gamma^{\lambda}{}_{\lambda\lambda}\frac{J_{,\lambda}}{J}. \quad (16)$$

C. The null energy condition

We now come to the question of the left-hand side of Eq. (16): can we say anything in general about $R_{\lambda\lambda}$? It turns out the answer is yes, under fairly general circumstances. We first note that since $g_{\lambda\lambda} = 0$,

$$R_{\lambda\lambda} = G_{\lambda\lambda} = \frac{1}{8\pi}T_{\lambda\lambda} = \frac{1}{8\pi}T_{\mu\nu}l^{\mu}l^{\nu} \quad (17)$$

(this is true with or without a cosmological constant). Now if we go into a local orthonormal frame in the rest frame of the matter ($T_{0i} = 0$) we can rotate $\boldsymbol{l}$ onto the z-axis, and then we have $l^3 = l^0$ and $T_{\mu\nu}l^{\mu}l^{\nu} = \rho + p_{zz}$. We say that a type of matter satisfies the *null energy condition* if its pressure satisfies $p \geq -\rho$ on every axis (or in tensor language: $T_{\mu\nu}l^{\mu}l^{\nu} \geq 0$). Any "normal" type of matter – particles, scalar fields, electromagnetic fields – obeys this rather general condition. If this is true, then the left-hand side of Eq. (16) is non-negative and

$$0 \leq -\left(\frac{J_{,\lambda}}{J}\right)_{,\lambda} - \frac{1}{2}\left(\frac{J_{,\lambda}}{J}\right)^2 + \Gamma^{\lambda}{}_{\lambda\lambda}\frac{J_{,\lambda}}{J}. \quad (18)$$

D. The area theorem

Now let's put this all together. Let's first suppose that we take $q = 0$ (consistent normalization of the horizon generators or $\boldsymbol{\nabla_l l} = 0$); then we would have $\Gamma^{\lambda}{}_{\lambda\lambda} = 0$. The above equation then tells us that

$$0 \leq -\left(\frac{J_{,\lambda}}{J}\right)_{,\lambda} - \frac{1}{2}\left(\frac{J_{,\lambda}}{J}\right)^2 = -\frac{2}{\sqrt{J}}\partial_{\lambda}^2\sqrt{J}, \quad (19)$$

so $\partial_{\lambda}^2\sqrt{J} \leq 0$. This means that if J is initially decreasing ($J_{,\lambda} < 0$), then at finite λ we will have $J \to 0$ and the generators will intersect. This is a contradiction.

The only conclusion, then, is that if the null energy condition is satisfied, J must be non-decreasing, and then the total horizon area (made up of all the little elements) is non-decreasing as well.