RTG Seminar Lecture Notes: The Positive Mass Theorem of General Relativity

MARCUS A. KHURI

1. Introduction

General Relativity is the study of Lorentz 4-manifolds $(\mathcal{M}^4, g)$ whose metric arises from a solution to the Einstein equations. We will take the signature of the metric to be $(- + ++)$, where the minus sign indicates the time component. Then the tangent space at any point may be divided into three different types of vectors, namely timelike, spacelike, and null vectors whose lengths squared respectively satisfy $g(v, v) < 0$ (> 0) $(= 0)$. By assumption matter particles travel on timelike geodesics, light and radiation travel on null geodesics, and nothing travels faster than the speed of light.

The source for the gravitational field is a symmetric 2-tensor T, referred to as the stress-energy-momentum tensor. It is physically defined in the following way. If $v, w \in T_p\mathcal{M}^4$ are future directed unit timelike vectors, then $T(v, w)$ represents the energy-momentum density in the direction v as measured by an observer moving in the direction w. In order to have a viable physical theory two natural restrictions are usually placed on the stress tensor. The first asserts that

$$T(v, w) \geq 0,$$

and is known as the dominant energy condition. Note that according to the definition this translates into the reasonable statement that all observed energy-momentum densities are nonnegative. The second restriction requires that T be divergence free:

$$\nabla^a T_{ab} = 0.$$

Although this is a strong requirement, it leads to the very natural and desirable property that the matter fields admit local (or approximate) conservation laws. If the spacetime admits killing vector fields the corresponding conservation laws will be global.

With the stress tensor acting as the source term for gravity we may write the Einstein equations as

$$G = 8\pi T,$$

where G is the Einstein tensor. As it was Einstein's basic hypothesis that gravity manifests itself through the curvature of spacetime, the most natural choice for G is the Ricci tensor $Ric(g)$. However the Ricci tensor is not necessarily divergence free, and thus would not be consistent with our desire that the matter fields satisfy the local conservation laws. On the other hand, the twice contracted second Bianchi identity guarantees that

$$Ric(g) - \frac{1}{2}R(g)g$$

is a tracefree version of the Ricci tensor, where $R(g)$ denotes the scalar curvature. The Einstein tensor is then chosen to be this tracefree Ricci tensor.

There are two important explicit solutions to the vacuum Einstein equations which will have special significance for this talk. Here the vacuum equations refer to the case in which there are no matter fields, that is $T = 0$. Note that by taking a trace the vacuum equations may simply be written as

$$Ric(g) = 0.$$

The first solution is Minkowski space

$$\mathbb{M}^4 = (\mathbb{R}^4, -dt^2 + dx^2 + dy^2 + dz^2).$$

This is the spacetime of Special Relativity and corresponds to an empty universe with no matter fields and no gravity. The second solution is the exterior Schwarzchild spacetime

$$\mathbb{SC}^4 = (\mathbb{R} \times (\mathbb{R}^3 - B_{2m}(0)), -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2 + r^2 d\Omega^2),$$

where $d\Omega^2$ represents the round metric on the 2-sphere and $m \geq 0$ is a parameter. This represents the region exterior to a static black hole of mass m in a vacuum. One motivation for referring to m as the mass of the black hole comes from our intuition associated with Newton's theory of gravity. More precisely, if a test particle is released in this spacetime, it will accelerate towards the origin with respect to the $t = 0$ slice at a rate proportional to $\frac{m}{r^2}$. We also note that the apparent singularity at the black hole horizon $r = 2m$, is actually only a coordinate singularity, whereas the singularity at $r = 0$ is a true spacetime singularity.

The Minkowski and Schwarzchild solutions will play the role of the extremal spacetimes for the Positive Mass Theorem and Penrose Inequality respectively. They are also the only spherically symmetric vacuum solutions (Birkhoff's Theorem [16]). Moreover, Bunting and Masood-ul-Alam [5] have shown that they are also the only complete, asymptotically flat, vacuum solutions with no boundary and black hole boundary respectively. The notion of asymptotic flatness refers to the situation in which every spacelike hypersurface (timelike normal) has certain fall-off conditions for its metric and second fundamental form at spatial infinity. To make this precise, we will say that a spacelike hypersurface (M^3, h, k) (with one end) is asymptotically flat, if outside a compact set M^3 is diffeomorphic to $\mathbb{R}^3 - B_1(0)$, and in the asymptotic coordinate system provided by the diffeomorphism the following fall-off conditions hold:

$$\partial^\alpha(h_{ij} - \delta_{ij}) = O(|x|^{-1-|\alpha|}), \quad |\alpha| \leq 2, \qquad \partial^\beta k_{ij} = O(|x|^{-2-|\beta|}), \quad |\beta| \leq 1.$$

Here h is the induced Riemannian metric on M^3 and k is its second fundamental form. In general asymptotically flat manifolds model isolated gravitating systems, by which we mean that there is a finite amount of matter and gravity being modeled with everything else assumed to be infinitely far away.

It turns out that for many problems in General Relativity, including the Positive Mass Theorem and Penrose Inequality, knowledge of the entire spacetime is not necessary, rather attention may be focused solely on a spacelike slice. That is we will be interested in Cauchy data (M^3, h, k), consisting of an asymptotically flat manifold M^3 with Riemannian metric h, and a symmetric 2-tensor k. The terminology Cauchy (or initial) data is used for this triple, since if appropriately chosen the metric h and 2-tensor k may be interpreted as the initial position and initial velocity for the solution along the initial hypersurface M^3, in analogy with the standard Cauchy problem for hyperbolic equations. Note that the requirement that h be Riemannian guarantees that the initial surface M^3 is noncharacteristic (or spacelike), and that from the PDE perspective the assumption of asymptotic flatness can be interpreted as imposing a boundary condition at spatial infinity.

The data h and k cannot be chosen arbitrarily if the corresponding triple is to be a valid initial data set for the Einstein equations, that is if it is to be isometrically embedded into a spacetime. Rather these data must satisfy certain constraints. These are analogous to the constraints that appear, in the perhaps more familiar situation, when solving Maxwell's equations. Recall that the electric and magnetic fields are encoded in an antisymmetric 2-tensor

$$F = \frac{1}{2}\sum_{a,b=0}^{3} F_{ab} dx^a \wedge dx^b, \qquad F_{ab} = -F_{ba},$$

where the electric field is given by $E_i = F_{0i}$ and the magnetic field is given by $(*B)_{ij} = F_{ij}$, $1 \leq i, j \leq 3$, or equivalently

$$F = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix}.$$

The Maxwell equations in Minkowski space may then be written

$$dF = 0 \Leftrightarrow F_{ab,c} + F_{bc,a} + F_{ca,b} = 0,$$

$$(\text{div} F)_b = \nabla^a F_{ab} = -J_b, \qquad J = \begin{pmatrix} \rho \\ \mathbf{j}_1 \\ \mathbf{j}_2 \\ \mathbf{j}_3 \end{pmatrix},$$

or in standard vector notation (with $t = x^0$)

$$\text{div}B = 0, \quad \partial_t B + \nabla \times E = 0, \quad \text{div}E = \rho, \quad \nabla \times B - \partial_t E = \mathbf{j},$$

where the 4-current J is composed of the charge density ρ and the current density $\mathbf{j}$. If a solution exists, then on the $t = 0$ slice

$$\text{div}E = \sum_{i=0}^{3} F_{0a,a} = J_0 = \rho, \quad (\text{div}B) dx^1 \wedge dx^2 \wedge dx^3 = d(F|_{t=0}) = (dF)|_{t=0} = 0.$$

These are the constraint equations. Note that they consist of two equations for six unknowns, and so are underdetermined. By solving the constraints and then choosing a gauge for the vector potential, one may obtain a solution for the Maxwell system.

In the setting of General Relativity, the constraints are obtained from the Gauss and Codazzi equations. If n denotes the future directed unit normal to the spacelike hypersurface M^3, then $\mu := T(n,n)$ denotes the energy density and $J := T(n,\cdot)$ denotes the momentum density, and the Gauss-Codazzi equations respectively yield:

$$2\mu = R_h + (\text{Tr}_h k)^2 - \parallel k \parallel_h^2, \quad J_i = \nabla^j (k_{ij} - (\text{Tr}_h k) h_{ij}),$$

where R_h is the scalar curvature of the induced metric h. These constraints are a necessary and sufficient condition for the initial data (M^3, h, k) to arise from a spacelike slice of a spacetime. Furthermore, we point out that the dominant energy condition implies that $\mu \geq \parallel J \parallel_h$. In the case of maximal slicing ($\text{Tr}_h k = 0$) this is equivalent to the condition of nonnegative scalar curvature.

2. Lagrangian Formulation of General Relativity

The appropriate notions of total energy-momentum for an asymptotically flat spacetime shall be derived from the ADM Hamiltonian formulation of General Relativity. However we must first recall the Lagrangian formulation, based on the Einstein-Hilbert action or total scalar curvature functional

$$S_{\text{GR}}(g) = \int_{\mathcal{M}^4} R_g d\omega_g.$$

By varying the metric and ignoring boundary terms, a calculation shows that the vacuum Einstein equations arise as the Euler-Lagrange equations for this functional

$$\delta S_{\text{GR}} := \frac{d}{d\lambda} S_{\text{GR}}(g + \lambda \delta g)|_{\lambda=0} = \int_{\mathcal{M}^4} G_{ab} \delta g^{ab} d\omega_g.$$

The full action is obtained by appending the Lagrangians for all matter fields onto the scalar curvature. For example consider the Lagrangian of a scalar field ϕ of mass m

$$\mathcal{L}_{\text{KG}} = (|\nabla_g \phi|^2 + m^2 \phi^2) \sqrt{|g|},$$

which gives rise to the Klein-Gordon equation, and the Lagrangian for the electromagnetic field

$$\mathcal{L}_{\mathrm{EM}} = \| F \|_g^2 \sqrt{|g|},$$

which gives rise to the homogeneous Maxwell equations by varying the vector potential. Here $|g|$ denotes the absolute value of $\det g_{ab}$. The stress-enery tensor for these two fields may be obtained by varying the metric in their corresponding actions

$$(T_{\mathrm{KG}})_{ab} = \phi_{,a}\phi_{,b} - \frac{1}{2}(|\nabla_g \phi|^2 + m^2\phi^2)g_{ab}, \qquad (T_{\mathrm{EM}})_{ab} = F_a^c F_{cb} - \frac{1}{4} \| F \|_g^2 \, g_{ab}.$$

Observe that if $n = \partial_t$ is the future directed unit normal to the $t = 0$ slice in Minkowski space, then the corresponding energy densities are given by

$$\mu_{\mathrm{KG}} = T_{\mathrm{KG}}(n,n) = \frac{1}{2}(\phi_t^2 + |\overrightarrow{\nabla}\phi|^2 + m^2\phi^2), \qquad \mu_{\mathrm{EM}} = T_{\mathrm{EM}}(n,n) = \frac{1}{2}(|E|^2 + |B|^2),$$

which respectively correspond to the well-known notions of energy density for solutions of the wave equation and for the electromagnetic field. Moreover, the full action for the coupled Einstein-Klein-Gordon-Maxwell system is given by

$$S = S_{\mathrm{GR}} + \alpha_{\mathrm{KG}} S_{\mathrm{KG}} + \alpha_{\mathrm{EM}} S_{\mathrm{EM}},$$

for some constants α_{KG}, α_{EM}.

We also point out that the divergence free conservation property of the stress-energy tensor for matter fields follows directly from the diffeomorphism invariance of the action. To see this let $f_\lambda : \mathcal{M}^4 \to \mathcal{M}^4$ be a one-parameter family of diffeomorphisms, then

$$S_{\mathrm{M}}[g,\psi] = S_{\mathrm{M}}[f_\lambda^* g, f_\lambda^* \psi]$$

where S_{M} is the action for a matter field ψ. It follows that

$$0 = \frac{d}{d\lambda} S_{\mathrm{M}} = \int_{\mathcal{M}^4} \frac{\delta \mathcal{L}_{\mathrm{M}}}{\delta g^{ab}} \delta g^{ab} + \int_{\mathcal{M}^4} \frac{\delta \mathcal{L}_{\mathrm{M}}}{\delta \psi} \delta\psi,$$

where our notation includes the volume form in the Lagrangian density $\mathcal{L}_M$. If ψ satisfies the field equations then the second integral on the right-hand side vanishes. Furthermore if $X = \frac{d}{d\lambda} f_\lambda$ is the vector field associated with the flow $\lambda \mapsto f_\lambda$, then $\delta g^{ab} = L_X g^{ab} = \nabla^{(a} X^{b)}$ and $\delta\psi = L_X \psi$ (here L_X denotes Lie differentiation) so that

$$0 = \int_{\mathcal{M}^4} T_{ab} \nabla^{(a} X^{b)} d\omega_g = -2 \int_{\mathcal{M}^4} \nabla^a T_{ab} X^b d\omega_g,$$

assuming that X has compact support. Thus since X is otherwise arbitrary we find that T is conserved.

3. Hamiltonian Formulation of General Relativity

Given a Lagrangian formulation of a field theory, there exists a general prescription for obtaining a Hamiltonian formulation based on the Legendre transformation of classical mechanics. The first step in this process is to divide the spacetime manifold into space and time. We assume that the spacetime metric is smooth. A function $t \in C^\infty(\mathcal{M}^4)$ is referred to as a "time function" if its gradient is everywhere timelike. We may choose local coordinates (t, x) adapted to the level sets Σ_t of t, such that $x = (x^1, x^2, x^3)$ are coordinates for each Σ_t. In terms of these coordinates the normal vector field takes the lapse-shift form

$$n = N^{-1}(\partial_t - X^i \partial_{x^i}),$$

where N is the lapse function and $X = X^i \partial_{x^i}$ is the shift vector. Equivalently the flow of the time evolution vector field

$$\partial_t = Nn + X,$$

when timelike, may be interpreted as a congruence of spacetime observers. The spacetime metric may then be expressed as

$$g = -N^2 dt^2 + h_{ij}(dx^i + X^i dt)(dx^j + X^j dt),$$

where $h_{ab} = g_{ab} + n_a n_b$ is the induced metric on Σ_t, so that $\sqrt{|g|} = N\sqrt{|h|}$.

It seems appropriate to choose the configuration space to consist of the triples $q = (h, N, X)$. However as we will see this configuration space is too large. In order to determine the generalized momentum

$$\pi = \frac{\partial \mathcal{L}_{\mathrm{GR}}}{\partial \dot{q}},$$

where $\dot{q} = \frac{dq}{dt}$, the Lagrangian must be written in terms of the configuration space variables. Observe that

$$\mathcal{L}_{\mathrm{GR}} := R_g\sqrt{|g|} = 2(G_{ab}n^a n^b - R_{ab}n^a n^b)N\sqrt{|h|},$$

and by the Gauss equation

$$G_{ab}n^a n^b = \frac{1}{2}(R_h - \parallel k \parallel_h^2 + (\mathrm{Tr}_h k)^2).$$

A similar calculation yields

$$R_{ab}n^a n^b = (\mathrm{Tr}_h k)^2 - \parallel k \parallel_h^2 - \nabla_a(n^a \nabla_c n^c) + \nabla_c(n^a \nabla_a n^a),$$

as well as

$$k_{ab} = \frac{1}{2}N^{-1}(\dot{h}_{ab} - D_a X_b - D_b X_a),$$

where D is covariant differentiation with respect to h. It follows that

$$\mathcal{L}_{\mathrm{GR}} = (R_h + (\mathrm{Tr}_h k)^2 - \parallel k \parallel_h^2) N\sqrt{|h|}$$

if boundary terms are ignored, and the momentum conjugate to h is given by

$$\pi_h^{ab} := \frac{\partial \mathcal{L}_{\mathrm{GR}}}{\partial \dot{h}_{ab}} = (k^{ab} - (\mathrm{Tr}_h k) h^{ab})\sqrt{|h|}.$$

The momentum conjugate to N and X vanishes. Therefore we cannot solve for $\dot{N}$ or $\dot{X}$ in terms of q and π. This suggests that N and X should not be viewed as dynamical variables, so we redefine the configuration space to consist solely of h. Then the Hamiltonian density is given by

$$\begin{aligned}
\mathcal{H}_{\mathrm{GR}} &= \pi^{ab}\dot{h}_{ab} - \mathcal{L}_{\mathrm{GR}} \\
&= -|h|^{1/2} N R_h + N|h|^{-1/2}(\parallel \pi \parallel_h^2 - \frac{1}{2}(\mathrm{Tr}_h \pi)^2) + 2\pi^{ab} D_a X_b \\
&= -|h|^{1/2}\{N(R_h + \frac{1}{2}|h|^{-1}(\mathrm{Tr}_h \pi)^2 - |h|^{-1} \parallel \pi \parallel_h^2) + 2X_b D_a(|h|^{-1/2}\pi^{ab})\},
\end{aligned}$$

if boundary terms are ignored. As N and X are not part of the newly defined phase space, their respective coefficients in the Hamiltonian density should vanish. For this reason we append the equations

$$R_h + \frac{1}{2}|h|^{-1}(\mathrm{Tr}_h \pi)^2 - |h|^{-1} \parallel \pi \parallel_h^2 = 0, \qquad D_a(|h|^{-1/2}\pi^{ab}) = 0,$$

which are easily seen to be the constraint equations for the vacuum Einstein equations, to Hamilton's equations

$$\dot{h} = \frac{\delta H_{\mathrm{GR}}}{\delta \pi}, \qquad \dot{\pi} = -\frac{\delta H_{\mathrm{GR}}}{\delta h},$$

where the full Hamiltonian is given by

$$H_{\mathrm{GR}} = \int_{\Sigma_t} \mathcal{H}_{\mathrm{GR}}.$$

These four equations yield a constrained Hamiltonian formulation and are equivalent to the vacuum Einstein equations.

The above construction has ignored boundary terms which would otherwise contribute to the full Hamiltonian. We assume that the spacetime foliation Σ_t consists of asymptotically flat hypersurfaces so that the following limits exist,

$$E = \frac{1}{16\pi} \lim_{r\to\infty} \int_{S_r} \sum_{i,j=1}^{3} (h_{ij,i} - h_{ii,j})\nu^j d\sigma, \tag{3.1}$$

$$P_i = \frac{1}{8\pi} \lim_{r\to\infty} \int_{S_r} \sum_{j=1}^{3} (k_{ij} - (\mathrm{Tr}_h k) h_{ij}) \nu^j \, d\sigma, \tag{3.2}$$

where S_r denotes a coordinate sphere (in the given asymptotic coordinate system) of radius r with unit outer normal ν. If the phase function and shift vector are chosen so that the following limits exist

$$\overline{N} = \lim_{r\to\infty} N|_{S_r}, \qquad \overline{X}_i = \lim_{r\to\infty} X_i|_{S_r},$$

then inclusion of all boundary terms in the Hamiltonian produces

$$H_{\mathrm{GR}} = \alpha(E\overline{N} - P^i \overline{X}_i) \tag{3.3}$$

for some positive constant α. Here we have used the fact that the constraint equations persist under the Hamiltonian flow, eliminating the interior terms of the full Hamiltonian.

The 4-vector (E, P^i) may be interpreted as the 4-momentum of the spacelike hypersurface Σ_t, while (3.3) may be interpreted as the energy of Σ_t as measured by an observer at spatial infinity moving with velocity $(\overline{N}, \overline{X}_i)$. The Lorentz length of the 4-momentum

$$M = \sqrt{E^2 - |P|^2},$$

is a conserved quantity which is invariant under the choice of foliation. It is the analogue of the rest mass for a test particle, and is referred to as the total mass of the spacetime. The quantities E, P, and M are often referred to as the ADM energy, linear momentum, and mass, as they arise from the Hamiltonian formulation first propsed by Arnowitt, Deser, and Misner in 1962 [1].

4. The Positive Mass Theorem

The fact that expressions (3.1) and (3.2) arise from a Hamiltonian is significant evidence supporting the claim that these are the correct notions of total energy-momentum in General Relativity. However this assertion can only be established by showing that under appropriate hypotheses the 4-momentum (E, P^i) is timelike, that is $E \geq |P|$, with $M = 0$ if and only if the spacetime is trivial. This is the content of the positive mass theorem, which will be stated in more detail below.

In order to gain better intuition we consider a simple class of examples. Let $(\mathbb{R}^3, u^4(x)\delta_{ij})$ be a time symmetric ($k = 0$) slice of some asymptotically flat spacetime, where $u \in C^\infty(\mathbb{R}^3)$ is positive and δ denotes the Euclidean metric. A well-known formula for the scalar curvature yields

$$R = -8u^{-5}\Delta u.$$

Assuming for simplicity that $R \equiv 0$ outside a compact set, we may expand the conformal factor in spherical harmonics to obtain

$$u(x) = a + \frac{b}{|x|} + O\left(\frac{1}{|x|^2}\right) \quad \text{as} \quad |x| \to \infty.$$

It is readily seen that in this case $E = 2ab$. Moreover since the dominant energy condition ensures that $R \geq 0$, we may apply the maximum principle to find that $\min_{x \in \mathbb{R}^3} u(x) = a$. Therefore $b \geq 0$ which implies that $E \geq 0$. We remark that a nice lemma of Schoen and Yau [14] shows that an asymptotically flat metric of nonnegative scalar curvature can be perturbed to obtain the asymptotic expansion above, all while preserving nonnegativity of the scalar curvature and changing the energy by an arbitrarily small amount. Thus we may view the energy as the $\frac{1}{r}$ rate at which the metric becomes flat. Note that the Minkowski and Schwarzchild solutions fall into this class of examples, and that their respective masses are $M = 0$, $M = m$ where m is the mass of the black hole.

We point out that the ADM energy is more subtle than it may appear. For instance, there exist nonconformally flat examples of time symmetric initial data $(M^3, h, 0)$ with zero scalar curvature and strictly positive energy. Note that a simple conformally flat example is the $t = 0$ slice of the Schwarzchild spacetime. Since the scalar curvature represents the energy density of matter fields for time symmetric initial data, the question arises as to where this extra energy comes from. It is interpreted as arising from the gravitational field itself. Unfortunately this phenomenon is not well understood, as there is no natural notion of energy density for the gravitational field. This also shows why the naive definition

$$\text{Energy} = \int_{M^3} \mu d\omega_h$$

of total energy is inadequate. More precisely, this expression only includes contributions from the matter fields, but neglects the contribution from the gravitational field.

In the late 1970s and early 1980s Schoen and Yau ([12], [13], [14]) established the result which placed the ADM notions of energy-momentum on a firm footing. Shortly thereafter Witten [17] offered a different proof of the same result. This is the

Positive Mass Theorem. *Let (M^3, h, k) be a complete, asymptotically flat initial data set for the Einstein equations, satisfying the dominant energy condition $\mu \geq \| J \|_h$. Then $E \geq |P|$ and $E = 0$ if and only if (M^3, h, k) arises from a spacelike slice of Minkowski space.*

5. Proof of the Positive Mass Theorem

The first proof given by Schoen and Yau is nonintuitive from a physical perspective, but is highly motivated by geometric considerations. Recall that a geodesic

$\gamma : \mathbb{R} \to M^3$ is a critical point of the arclength functional, that is $\delta L(\gamma) = 0$. If γ actually minimizes arclength, then the second variation yields the inequality $\delta^2 L(\gamma) \geq 0$. From this inequality one is led to the Jacobi equation, which has been a productive tool for examining the effects of curvature on geodesics. The Schoen/Yau proof follows a similar philosophy, with geodesics replaced by minimal surfaces. A minimal surface $S : \mathbb{R}^2 \to M^3$ is a critical point of the area functional, that is $\delta A(S) = 0$, which by a well-known formula shows that the surface S is minimal if and only if its mean curvature vanishes. If S minimizes area (such minimal surfaces are referred to as stable), then upon taking normal variations with speed $\psi \in C_c^\infty(S)$ the second variation formula yields

$$0 \leq \delta^2 A(S) = \int_S |\nabla_S \psi|^2 - (Ric(n,n) + \parallel k \parallel^2)\psi^2, \tag{5.1}$$

where k is the second fundamental form of S and n is a unit normal vector field. It is this inequality which reveals the primary consequences of the dominant energy condition.

For the time being we restrict ourselves to the time symmetric case, so that the Positive Mass Theorem reads: if (M^3, h) is an asymptotically flat Riemannian 3-manifold with nonnegative scalar curvature, then $E \geq 0$ and $E = 0$ if and only if (M^3, h) is isometric to Euclidean space $(\mathbb{R}^3, \delta)$. Under the given hypotheses, Schoen and Yau show that if $E < 0$ then a properly embedded stable minimal hypersurface S exists. By the Gauss equation

$$Ric(n,n) + \parallel k \parallel^2 = \frac{1}{2}(R_{M^3} - R_S + \parallel k \parallel^2),$$

so that according to the second variation inequality we have

$$\int_S \frac{1}{2}(R_{M^3} - R_S + \parallel k \parallel^2)\psi^2 \leq \int_S |\nabla_S \psi|^2.$$

As $R_{M^3} \geq 0$, a small perturbation argument shows that we can assume $R_{M^3} > 0$ some where, then

$$\int_S K_S > 0$$

by choosing $\psi \equiv 1$ (this of course requires an approximating sequence and some finer analysis, since (5.1) only holds for ψ with compact support) and recalling that the Gauss curvature is $K_S = \frac{1}{2} R_S$. In order to obtain a contradiction, let $B_r \subset S$ be the interior region of a large coordinate circle ∂B_r of radius r, and apply the Gauss-Bonnet Theorem to conclude that

$$0 < \lim_{r \to \infty} \int_{B_r} K_S = \lim_{r \to \infty} \left(2\pi\chi(B_r) - \int_{\partial B_r} \kappa \right) \leq 2\pi - 2\pi = 0,$$

where κ is the geodesic curvature. Note that we have used the fact that $\chi(B_r) \leq 1$ for open surfaces, and the fact that as intuition suggests the total geodesic curvature converges to 2π in analogy with circles in $\mathbb{R}^2$. This contradiction shows that $E \geq 0$. We point out that this proof can be combined with an inductive argument to yield the desired result for dimensions $n \leq 7$. However for higher dimensions the existence of a regular properly embedded stable minimal hypersurface is not guaranteed.

To show the rigidity of Euclidean space with respect to energy, assume that $R_{M^3} \geq 0$ and $E = 0$. If $R_{M^3} > 0$ some where, then a standard argument yields a conformal deformation to an asymptotically flat scalar flat metric $u^4 h$, with strictly negative energy. However this cannot happen since we must have $E \geq 0$. If $Ric(h) \neq 0$ some where, then define a 1-parameter family of asymptotically flat metrics by $h_\lambda = h + \lambda Ric(h)$. Analysis of the first variation yields $R_{h_\lambda} \geq 0$ and $E_\lambda < 0$ for sufficiently small $|\lambda|$. Thus we get a contradiction again, which implies that $Ric(h) \equiv 0$. However in three dimensions the Ricci tensor vanishes if and only if the Riemann tensor vanishes. It follows that $(M^3, h) \cong (\mathbb{R}^3, \delta)$.

Having established the Riemannian (time symmetric) version of the Positive Mass Theorem, it remains to show the full spacetime version ($k \neq 0$). This situation, in which the time symmetric version of a result has been proven first, often occurs in General Relativity. It also happens to be the case that often the full spacetime version can simply be reduced to the time symmetric version by some deformation of the initial data (M^3, h, k). In the particular instance of the Positive Mass Theorem, this deformation occurs in the form of a graph $t = f(x)$ lying inside the product manifold $(\mathbb{R} \times M^3, dt^2 + h)$. The idea here is to choose this graph in such a way that the induced metric has scalar curvature which is "as positive as possible". Note that the primary difficulty in extending the arguments of the time symmetric proof to the spacetime setting, is the lack of nonnegative scalar curvature. It turns out that an appropriate choice for the graph requires the function f to satisfy the following quasilinear elliptic equation

$$H(f) = \mathrm{Tr}_{\overline{h}} k,$$

where $H(f)$ is the mean curvature which in coordinates takes the form

$$\overline{h}^{ij} \frac{D_{ij} f}{\sqrt{1 + |\nabla_h f|^2}} = \overline{h}^{ij} k_{ij},$$

with D representing the connection on (M^3, h). This equation was first posed by the physicist P. Jang [7]. An adequate discussion of the technical existence proof would be much too long to discuss here, so we refer the reader to [13].

In contrast to the Schoen/Yau proof of the Positive Mass Theorem, the Witten proof is well motivated from the physical perspective. In most field theories the energy is naturally expressed as the sum of squares, and therefore the corresponding positivity and rigidity statements are trivial. In the case of General Relativity such an expression is by no means obvious. However Witten did succeed in finding the correct expression. The flavor of his proof is similar to most "vanishing theorems" in geometry. More precisely, let (M, h) be a Riemannian manifold possibly with boundary, and let L be a differential operator of the form

$$L\psi = \nabla^*\nabla\psi + Q\psi$$

where ψ is a section of some bundle with covariant derivative ∇, ∇^* is its formal adjoint, and Q is a self-adjoint endomorphism of the bundle. If $L\psi = 0$ then taking the inner product with ψ and integrating by parts produces

$$\int_M (|\nabla\psi|^2 + (\psi, Q\psi)) = \int_{\partial M} (\text{something}). \tag{5.2}$$

In practice the operator L is a natural operator, such as the Hodge Laplacian on differential forms, or the square of the Dirac operator on spinors, and Q is expressed in terms of the curvature of (M, h). Equation (5.2) is often called a Weitzenböck formula. If one assumes that $Q > 0$ and M is compact without boundary, then (5.2) implies that $\psi = 0$, that is, $\ker L = 0$. If M is not compact or has a boundary, then one must impose either growth conditions or boundary conditions on ψ to control the boundary integral. For the proof of the Positive Mass Theorem, the operator L is $\mathcal{D}^2$, where $\mathcal{D}$ is the Dirac operator on spinors, and the endomorphism Q is positive because of the dominant energy condition. Moreover, the boundary integral is exactly the expressions (3.1) and (3.2) for the difference $E - |P|$. Thus, the identity (5.2) is the desired expression of the mass as a sum of squares. This proof requires the existence of harmonic spinors with the appropriate fall-off behavior at spatial infinity. It can be shown [10] that if (M, h) is asymptotically flat, then given any constant spinor field ψ_0 there is a unique spinor field ψ such that

$$\mathcal{D}\psi = 0, \quad \text{and} \quad \psi = \psi_0 + o(|x|^{\varepsilon - 1}).$$

Note that the arguments of this proof continue to hold in all dimensions as long as (M, h) admits a spin structure.

It should be pointed out that a new proof of the Riemannian Positive Mass Theorem has recently been put forth by Huisken and Ilmanen [6], and is based on the techniques of inverse mean curvature flow. In fact they prove a much stronger and refined version known as the Penrose Inequality, which is suited to initial data which contain black holes. For a beautiful exposition of this work, including the alternative conformal flow approach [2] of Bray (which is based on the Positive Mass Theorem), we refer the reader to [3].

Lastly we mention that although the Positive Mass Theorem is primarily motivated by its physical origins, it has had several important applications to pure geometry. These include finding obstructions to positive scalar curvature [15], as well as aiding in the existence and compactness results for the Yamabe problem [9], [11]. Moreover the refined version (Penrose Inequality) has been used to calculate the

Yamabe invariants of compact 3-manifolds [4], and thus aids with the classification of such objects.

References

[1] R. Arnowitt, S. Deser, C. Misner, *The dynamics of General Relativity*, Gravitation (L. Witten, ed.), Wiley, NY, 1962, 227-265.

[2] H. Bray, *Proof of the Riemannian Penrose inequality using the positive mass theorem*, J. Differential Geom., **59** (2001), 177-267.

[3] H. Bray, *Black holes, geometric flows, and the Penrsoe inequality in General Relativity*, Notices of the AMS, **49** (2002), 1372-1381.

[4] H. Bray, A. Neves, *Classification of prime 3-manifolds with σ-invariant greater than* $\mathbb{RP}^3$, Ann. Math., **159** (2004), 407-424.

[5] G. Bunting, A. Masood-ul-Alam, *Nonexistence of multiple black holes in asymptotically Euclidean static vacuum space-time*, Gen. Rel. and Grav., **19** (1987), 147-154.

[6] G. Huisken, T. Ilmanen, *The inverse mean curvature flow and the Riemannian Penrose inequality*, J. Differential Geom., **59** (2001), 353-437.

[7] P. Jang, *On the positivity of energy in General Relaitivity*, J. Math. Phys., **19** (1978), 1152-1155.

[8] J. Kazdan, *Positive energy in General Relativity*, Séminaire Bourbaki, **593** (1981/82), 315-330.

[9] M. Khuri, F. Marques, R. Schoen, *A compactness theorem for the Yamabe problem*, preprint (2007).

[10] T. Parker, C. Taubes, *On Witten's proof of the positive energy theorem*, Comm. Math. Phys., **84** (1982), 223-238.

[11] R. Schoen, *Conformal deformation of a Riemannian metric to constant scalar curvature,* J. Differential Geom., **20** (1984), 479-495.

[12] R. Schoen, S.-T. Yau, *On the proof of the positive mass conjecture in General Relativity*, Comm. Math. Phys., **65** (1979), 45-76.

[13] R. Schoen, S.-T. Yau, *Proof of the positive mass theorem II*, Comm. Math. Phys., **79** (1981), 231-260.

[14] R. Schoen, S.-T. Yau, *The energy and the linear momentum of space-times in General Relativity*, Comm. Math. Phys., **79** (1981), 47-51.

[15] R. Schoen, S.-T. Yau, *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Ann. Math., **110** (1979), 127-142.

[16] R. Wald, *General Relativity*, University of Chicago Press, Chicago, 1984.

[17] E. Witten, *A simple proof of the positive energy theorem*, Comm. Math. Phys., **80** (1981), 381-402.