The Proof of the Positive Mass Theorem using Minimal Surfaces

Chao Wang

April 4, 2014

1 Basic Setting

The positive mass theorem of general relativity to be addressed here states that the total mass of an isolated gravitational system (viewed from spatial infinity, known as the ADM mass) must be non-negative if its local mass density is non-negative everywhere, and in fact the mass is strictly positive unless the space-time is flat. In this essay we are only concerned with the first part of this theorem. Equivalently, the theorem asserts that the total energy-momentum vector must be non space-like.

Mathematically, the basic setting can be described as follows [1]: Let (M, γ) be a space-time whose local mass density is non-negative everywhere. Suppose M admits an oriented, three-dimensional maximal space-like hypersurface N, with induced Riemannian metric g and second fundamental form h. We assume that there exists a compact subset K of N such that $N \setminus K$ consists of a finite number of ends $N_1, N_2, \ldots, N_r$, with each N_k being diffeomorphic to the exterior of a ball in $\mathbb{R}^3$. We say that the metric g is asymptotically flat if for each end N_k, under the diffeomorphism described above, in the Euclidean coordinates $x = (x^1, x^2, x^3)$ of $\mathbb{R}^3$, g and h have the asymptotic behaviour

$$
\begin{aligned}
g_{ij} &= (1 + \frac{M_k}{2r})^4 \delta_{ij} + h_{ij}, \\
|h_{ij}| &\leq \frac{k_1}{1 + r^2}, \\
|\partial h_{ij}| &\leq \frac{k_2}{1 + r^3}, \\
|\partial\partial h_{ij}| &\leq \frac{k_3}{1 + r^4}, \qquad (1)
\end{aligned}
$$

where k_1, k_2 and k_3 are some positive constants, $r = (\Sigma_{i=1}^3 (x^i)^2)^{\frac{1}{2}}$ and ∂ is the Euclidean gradient. The number M_k is the mass of the end N_k. Let R be the scalar curvature of N, and define

$$\mu = \frac{1}{2}[R - h^{ab}h_{ab} + (h^a_a)^2],$$
$$J^a = \nabla_b(h^{ab} - h^c_c g^{ab}), \tag{2}$$

then the dominant energy condition [2], which requires that the speed of energy flow of matter is always less than the speed of light, says that

$$\mu \geq |J_a J^a|^{\frac{1}{2}}. \tag{3}$$

Here we only consider the special case where $h^a_a = 0$, so that Eqn. 3 implies that the scalar curvature $R \geq 0$. In light of this, the first part of the positive mass theorem states:

Theorem 1.1 *Let g be an asymptotically flat metric on an oriented three-dimensional Riemannian manifold N as described above. If $R \geq 0$ on N, then the mass of each end is non-negative.*

2 Background Knowledge

We refer to the review [3] heavily in this section.

A standard tool in Riemannian geometry which aids greatly in seeing the effect of curvature is geodesics, especially those which minimize arc lengths. If a curve $\gamma : R \to M$ is a geodesic, i.e. the first variation of the arc length functional vanishes at γ, then it minimizes the arc length when the second variation of the arc length functional is non-negative.

Minimal surfaces are the two-dimensional analogue of geodesics, which are the critical points of the surface area functional $A[S]$ (with respect to the metric, as always unless otherwise specified) for surfaces S. By this definition, a two-dimensional surface S embedded in a three-dimensional Riemannian manifold N is minimal if and only if its mean curvature is 0, i.e. $\frac{1}{2}(h_{11} + h_{22}) = 0$, where h_{ij} is the second fundamental form on S induced by the metric of N, expressed in an orthonormal frame field.

The first main ingredient of Schoen and Yau's proof of the positive mass theorem is the second variation inequality. If a minimal surface S actually minimizes the functional $A[S]$ in any one-parameter compactly supported deformation, then it satisfies the second variation inequality:

$$0 \leq \delta^2 A[S] = -\int_S f[\Delta f + f(Ric(\nu) + \|h\|^2)] = \int_S \|\nabla f\|^2 - f^2[Ric(\nu) + \|h\|^2], \tag{4}$$

where f is any C^2 function with compact support on S, $\|h\|^2 = \sum_{i,j=1}^3 h_{ij}^2$, $Ric(\nu)$ is the Ricci curvature of N in the direction ν normal to S. Note that the operators Δ and ∇ are both with respect to the induced metric on S, so that

$$\int_S f\Delta f = -\int_S \|\nabla f\|^2 \tag{5}$$

in the second variation inequality. Using the condition $h_{11}+h_{22}=0$, by standard formulas the second variation inequality can be rewritten as

$$\int_S [R - K + \frac{1}{2}\|h\|^2] f^2 \leq \int_S \|\nabla f\|^2, \tag{6}$$

where R is the scalar curvature of N, and K is the Gaussian curvature of S.

The other main ingredient of the proof is the Gauss-Bonnet theorem, which relates the geometric quantities Gaussian curvature K of a regular region D of an oriented surface S and the geodesic curvature k at the boundary of D in S, to the topological invariant χ of D called the Euler characteristics. For a region D whose boundary is a simple, closed and C^1 curve, the theorem reads [4]

$$\int_D K + \int_{\partial D} k = 2\pi\chi(D). \tag{7}$$

3 Main Idea of the Proof

The main idea of Schoen and Yau's proof [1] goes as follows: Let g be an asymptotically flat metric on N with $R \geq 0$, and let x^1, x^2, x^3 be the asymptotically flat coordinates on N_k, which lie in $\mathbb{R}^3 \setminus B_{\sigma_0}(0)$, where $B_{\sigma_0}(0) = \{|x| < \sigma_0\}$.

Suppose the mass M of some end N_k is negative. First by considering a conformally equivalent metric that is also asymptotically flat, one can assume a stronger condition on the scalar curvature R, namely that $R \geq 0$ everywhere in N and $R > 0$ outside a compact subset of N_k, while the negativity of mass of N_k is still preserved. Then using the negativity of M, a complete area-minimizing surface S (with respect to all compactly supported deformations of the surface) properly embedded in N can be constructed such that $S \cap (N \setminus N_k)$ is compact, and $S \cap N_k$ lies between two parallel Euclidean 2-planes in $\mathbb{R}^3$. Using the second variation inequality Eqn. 6, upon taking the limit for a sequence of bounded subsets of S which form an exhaustion, the conditions on R given in the first step imply that

$$\int_S K > 0. \tag{8}$$

The final step is to use the Gauss-Bonnet theorem to arrive at a contradiction. One can show that there exists an exhaustion D_σ of S $(\sigma > \sigma_0, \sigma \to \infty)$ such that for all sufficiently large σ, D_σ is topologically a disk, and in the x^1, x^2, x^3 coordinates, the projection of ∂D_σ onto the x^1x^2- plane is a circle with radius σ centered at 0. Applying the Gauss-Bonnet theorem to D_σ gives

$$\int_{D_\sigma} K = 2\pi - \int_{\partial D_\sigma} k, \tag{9}$$

where k is the geodesic curvature of ∂D_σ with respect to the inner normal to D_σ in S.

One then shows that there exists a sequence $\sigma_i \to \infty$ such that

$$\lim_{i\to\infty} \int_{\partial D_{\sigma_i}} k \geq 2\pi. \tag{10}$$

The main intuition is that the negativity of the mass not only enables one to construct the complete area-minimizing surface S, but also imposes stringent conditions on S so that very roughly speaking S is "asymptotically larger than" $\mathbb{R}^2$ [5]. Combining Eqn. 10 with Eqn. 9 yields

$$\lim_{i\to\infty} \int_{D_{\sigma_i}} K \leq 0, \tag{11}$$

in contradiction with Eqn. 8. Therefore the minimal surface S cannot exist, and thus the mass of end N_k must be non-negative.

In the next section we fill in the details of this sketch of proof, however, due to the author's limited background, certain important details are skipped, such as the existence proof of the complete area minimizing minimal surface using regularity estimate, and the proof for Eqn. 10.

4 Schoen and Yau's Proof

Once again, let (N, g) be the Riemannian manifold described above, for which the scalar curvature $R \geq 0$. We work with a fixed end N_k, for which we assume the mass M is negative, with the asymptotically flat coordinates $x = (x^1, x^2, x^3)$ as defined above, which lie in $\mathbb{R}^3 \setminus B_{\sigma_0}(0)$. Let $r = |x|$ denote the Euclidean distance from the origin. Note that the asymptotic flatness condition Eqn. 1 implies that the Christoffel symbols Γ^i_{jk} are $O(\frac{1}{r^2})$ and the Riemann curvature tensor is $O(\frac{1}{r^3})$ as $r \to \infty$.

Lemma 4.1 *There exists an asymptotically flat metric $\tilde{g}$ which is conformally equivalent to g, and also has negative mass for the end N_k, such that its scalar curvature $\tilde{R}$ satisfies $\tilde{R} \geq 0$ on N, and $\tilde{R} > 0$ outside a compact subset of N_k.*

Proof Sketch [5]: It is easy to compute and obtain $\Delta \frac{1}{r} < 0$ for sufficiently large r. This allows us to obtain a positive function ϕ which has asymptotic of the form $1 - \frac{M}{4r}$, so that $\Delta\phi < 0$ for sufficiently large r, and by rounding off appropriately we can also obtain $\Delta\phi \leq 0$ on N. Let $\tilde{g} = \phi^4 g$, then the formula for scalar curvature $\tilde{R} = \phi^{-5}(-8\Delta\phi + R\phi)$ gives the desired result on $\tilde{R}$. The rest is easy. ∎

This lemma allows us to assume without loss of generality that $R \geq 0$ on N, and $R > 0$ outside a compact subset of N_k, without compromising either asymptotic flatness or negativity of mass.

For each $\sigma > 2\sigma_0$, let C_σ be the circle of Euclidean radius σ centered at 0 lying in the x^1x^2- plane. Let S_σ be the smooth embedded oriented surface of minimal area among all surfaces with boundary curve C_σ regardless of topological type, which is compact. This known existence result shall be assumed. Due to the fact that N may have multiple ends, we need a lemma which says that S_σ cannot run to infinity in an end other than N_k as $\sigma \to \infty$. To prove this, we need the maximum principle for minimal surfaces: If we translate a convex surface (a surface whose tangent plane always lies on the same side of the surface, such as level sets of a convex function) towards a minimal surface, then the first point of contact must be on the boundary of the minimal surface. More precisely,

Theorem 4.2 *Let E be a convex set bounded by a convex surface H. Suppose an interior point P of a connected minimal surface $S \subset E$ is contained in H. Then all of S is a subset of H.*

Lemma 4.3 *There exists a compact subset $K_0 \subseteq N$ such that for each $\sigma > 2\sigma_0$ we have $S_\sigma \cap (N \setminus N_k) \subseteq K_0$ (K_0 is independent of σ).*

Proof Let $N_{k'}$ be a different end, which is diffeomorphic to $\mathbb{R}^3 \setminus B_{\tau_0}(0)$ with asymptotically flat coordinate system y^1, y^2, y^3 where $B_{\tau_0}(0) = \{y : |y| < \tau_0\}$. We compute the asymptotic behaviour of the covariant Hessian of the function $|y|^2$ using Eqn. 1,

$$D_{ij}|y|^2 = \frac{\partial^2}{\partial y^i \partial y^j}(|y|^2) - (D_{\frac{\partial}{\partial y^i}} \frac{\partial}{\partial y^j})(|y|^2) = 2\delta_{ij} + O(\frac{1}{|y|}), \tag{12}$$

where D_{ij} is the covariant Hessian, D is the Riemannian connection of N. This implies that there exists $\tau_1 > \tau_0$ such that the function $|y|^2$ is convex (i.e., its covariant Hessian is positive definite) in the region $\{y : |y| \geq \tau_1\}$. Suppose when $\sigma \to \infty$, S_σ intersects with $\partial B_{\tau_1}(0)$ first when $\sigma = \sigma_1$ for some $\sigma_1 > 2\sigma_0$. Since $\partial S_\sigma = C_\sigma$ lies in N_k, this gives a contradiction by Theorem 4.2. Therefore we must have $S_\sigma \cap N_{k'} \subseteq B_{\tau_1}(0)$ for all $\sigma > 2\sigma_0$. This is true for any end $N_{k'}$ of N other than N_k, thus the lemma is established. ∎

Next, we show that the height of $S_\sigma \cap N_k$ is in fact bounded in the x^3 direction.

Theorem 4.4 *There exists a number $h > \sigma_0$ so that $S_\sigma \cap N_k \subseteq E_h = \{x \in \mathbb{R}^3 : |x^3| \leq h\}$.*

Proof A direct computation using Eqn. 1 and the asymptotic behaviour of the Christoffel symbols yields

$$D_{ij}x^3 = -\Gamma^3_{ij} = \frac{Mx^j}{r^3}\delta_{i3} + \frac{Mx^i}{r^3}\delta_{j3} - \frac{Mx^3}{r^3}\delta_{ij} + O(\frac{1}{r^3}). \tag{13}$$

Suppose the maximum of x^3 on $S_\sigma \cap N_k$ is H and is achieved at the point $x_0 \in S_\sigma$. If $H \leq \sigma_0$, then the proof is done. Thus we assume $H > \sigma_0$. The tangent space to S_σ at x_0 is spanned by $\frac{\partial}{\partial x^1}(x_0)$, $\frac{\partial}{\partial x^2}(x_0)$. Let ∇ denote the induced Riemannian connection on the submanifold S_σ of N, and let q_{ij} be the induced metric, where $i, j = 1, 2$. Then using the fact that $\nabla_i \frac{\partial}{\partial x^j} = (D_i \frac{\partial}{\partial x^j})^T$, where T denotes orthogonal projection onto the tangent space of S_σ, one can compute

$$\nabla_{ij} x^3 = D_{ij} x^3 - h_{ij} \nu(x^3), \tag{14}$$

where $h_{ij} = \langle D_i \nu, \frac{\partial}{\partial x^j} \rangle (x_0)$ is the second fundamental form, and ν is the unit normal vector field of S_σ. Contracting Eqn. 14 with q_{ij} and using the fact that $q^{ij} h_{ij} = 0$, we obtain

$$q^{ij} \nabla_{ij} x^3 = -\frac{2MH}{r^3} + O(\frac{1}{r^3}). \tag{15}$$

The fact that x^3 achieves a maximum at x_0 implies that $q^{ij} \nabla_{ij} x^3 \leq 0$. Since M is negative, this means that H has to be bounded above by a number h, which is independent of σ. Similarly we can give a lower bound for x^3 on $S_\sigma \cap N_k$. $\blacksquare$

We make no attempt to account for the key step which allows us to extract the desired complete area-minimizing minimal surface S as a stable limit of some sequence S_{σ_i} as $\sigma_i \to \infty$. However, we do note that Lemma 4.3 and Theorem 4.4 are in fact necessary for this important step. Moreover, one key point is that Theorem 4.4 is where we used the negativity of mass of N_k explicitly, so that the construction of the desired minimal surface S depends crucially on this assumption that we eventually would like to disprove.

Since S is non-compact, applying the second variation inequality is not entirely straight-forward. In particular, it is necessary to show that

Lemma 4.5 *$\int_S \|h\|^2 < \infty$, and $\int_S |K| < \infty$, where h is the second fundamental form, $\|h\|^2$ is its norm with respect to the induced metric on S, and K is the Gaussian curvature of S.*

Proof Sketch: For any $\sigma \geq \sigma_0$, let $S_{(\sigma)} = [S \cap [(N \setminus N_k)] \cup [S \cap B_\sigma(0)]$, then $S_{(\sigma)}$ form an exhaustion of S. By using the area minimizing property of S, it is easy to see that there exists a constant C independent of σ, such that $Area(S_{(\sigma)}) \leq C\sigma^2$. Together with Eqn. 1, this results enables one to bound certain integrals on the non-compact surface S, then one can prove the lemma. $\blacksquare$

Note that the second variation inequality Eqn. 4 in fact holds for any Lipschitz function f with compact support in S. Therefore by choosing an appropriately rounded-off Lipshitz function f which takes the value 1 on $S_{(\sigma)}$ and vanishes outside $S_{(\sigma^2)}$ in Eqn. 6, where S is replaced by $S_{(\sigma)}$, and then the desired inequality $\int_S K > 0$ (Eqn. 8) follows upon taking the limit $\sigma \to \infty$.

An interesting and important remark is that by the Cohn-Vossen inequality, $\int_S K \leq 2\pi\chi(S)$, Eqn. 8 implies that S is topologically the plane.

The final step is to arrive at a contradiction by using the Gauss-Bonnet theorem. We wish to find an exhaustion D_σ of S as $\sigma \to \infty$ such that for sufficiently large σ, D_σ is topologically a disk, and the total geodesic curvature on ∂D_σ in S is asymptotically bounded below by that of a Euclidean circle, i.e., there exists a sequence $\sigma_i \to \infty$ so that $\lim_{\sigma_i \to \infty} \int_{\partial D_{\sigma_i}} k \geq 2\pi$ (Eqn. 10, where k is the geodesic curvature on ∂D_σ in S). Then combined with the Gauss-Bonnet theorem Eqn. 9, we arrive at Eqn. 11, which leads to the desired contradiction.

By Lemma 4.5 and the remark above, a result from A. Huber [6] implies that S is conformally equivalent to $\mathbb{C}$, i.e., there exists a conformal diffeomorphism $F : \mathbb{C} \to S$. For any $x \in \mathbb{R}^3$, let $x' = (x^1, x^2, 0)$ and $r' = |x'| = ((x^1)^2 + (x^2)^2)^{\frac{1}{2}}$. Let P_σ be the cylinder $\{x \in \mathbb{R}^3 : r' \leq \sigma\}$. Since S is topologically the plane, it follows that for any $\sigma > \sigma_0$ such that $\partial P_\sigma \cap S$ is transverse, there is a circle in this intersection whose projection onto the x^1x^2- plane is a circle centered at 0 with radius σ. Let D_σ be the connected component of such a circle in $S \cap [(N \setminus N_k) \cup P_\sigma]$. Since S is connected, the D_σ form an exhaustion of S.

Theorem 4.6 *For all σ sufficiently large, D_σ is topologically a disk.*

Proof Consider $F^{-1}(D_\sigma)$, which is a bounded connected region in $\mathbb{C}$. We need to show that $F^{-1}(D_\sigma)$ is simply connected. Suppose not, then there is a bounded domain O contained in $\mathbb{C} \setminus F^{-1}(D_\sigma)$. Since $\partial P_\sigma \cap S$ is transverse, the function $r' - \sigma$ changes sign across each boundary component of $F^{-1}(D_\sigma)$.This implies that $r' = \sigma$ on $\partial F(\mathcal{O}) \subset S$, and $r' > \sigma$ at some points inside $F(\mathcal{O})$. Therefore r' takes a maximum at some point of $F(\mathcal{O})$. It suffices to show that for sufficiently large r' the function $(r')^2$ is subharmonic, which by the maximum principle for subharmonic functions would lead to a contradiction. Using Eqn. 1 and the fact that S is a minimal surface, it can be shown that $\Delta(r')^2 \geq 2 – O(\frac{1}{r})$ as $r \to \infty$, which yields the desired result. ∎

The only remaining task is to show Eqn. 10. In [1] Schoen and Yau gave a 5-page calculation for this, which we do not attempt to repeat here. Very roughly speaking, the key point is that the negativity of mass implies that S is "asymptotically larger than" $\mathbb{R}^2$, so that as $\sigma \to \infty$ the geodesic curvature k on ∂D_σ in S approaches but is bounded below by that of a circle asymptotically. More explicitly, in Schoen and Yau's calculation they used the fact that the complete area-minimizing surface $S \cap N_k$ is bounded in the x^3 coordinate, a fact which as we have seen above indeed depends on the negativity of the mass. Using this important fact they were able to show that in some sense as $\sigma \to \infty$, ∂D_σ in fact approaches a "planar Euclidean circle", more precisely summarized as

Theorem 4.7 *Let ν be the unit normal of S in $\mathbb{R}^3$ relative to the metric g, h be the second fundamental form of S, and let L denote the length of boundary* *curve relative to g. As $\sigma \to \infty$, we have*

$$\begin{aligned}
&\int_{\partial D_\sigma} 1 - \langle \nu, \frac{\partial}{\partial x^3} \rangle \leq O(1), \\
&L(\partial D_\sigma) = O(\sigma) \\
&L(\partial D_\sigma) \geq 2\pi\sigma – O(1), \\
&\int_{\partial D_\sigma} \|h\| \to 0.
\end{aligned} \tag{16}$$

Based on this, one can obtain Eqn. 10. The sketch of proof is now complete.

5 Conclusion

Schoen and Yau's beautiful proof of the positive mass theorem, is an example where one can use the minimal surfaces to study the effect of curvature on a Riemannian manifold. The part of the theorem which states that the mass of each end is strictly positive unless the space-time is Minkowski is not addressed in this essay, nor is the more general case where the space-time M does not necessarily admit a maximal space-like hypersurface, and $h^a_a = 0$ is not assumed. The proofs for these results are much more difficult. Also, there is a generalization of the positive mass theorem, called the Riemannian-Penrose inequality, which gives a higher lower bound for the ADM mass of a space-time in terms of the total area of its black holes. There is clearly so much more to explore on this fascinating topic.

References

[1] R. Schoen and S. T. Yau, "On the Proof of the Positive Mass Conjecture in General Relativity", *Comm. Math. Phys.* **65**, 45-76 (1979).

[2] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[3] J. Kazdan, "Positive Energy in General Relativity", *Séminaire N. Bourbaki* **24**, 315-330 (1981-1982).

[4] M. P. Do Carmo, *Differential Geometry of Curves and Surfaces* (Prentice-Hall, New Jersey, 1976).

[5] T. H. Colding, W. P. Minicozzi II, *A Course in Minimal Surfaces* (American Mathematical Society, Rhode Island, 2011).

[6] A. Huber, "On Subharmonic Functions and Differential Geometry in the large", *Comment. Math. Helv.* **32**, 12-72 (1957).