1

OF/DE

1

PM-1 3 1/2 x 4" PHOTOGRAPHIC MICROCOPY TARGET
NBS 1010a ANSI/ISO #2 EQUIVALENT

PRECISIONSM RESOLUTION TARGETS

**VARIATIONAL METHODS IN GENERAL RELATIVITY
AND THE
POSITIVE ENERGY THEOREMS**

by
JEFF McPHAIL, B.Sc.

A thesis submitted to
the Faculty of Graduate Studies and Research
in partial fulfillment of
the requirements for the degree of

Master of Science

Department of Mathematics and Statistics
Carleton University
Ottawa, Ontario

August 19

© Copyright

1996, Jeff McPhail

National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395, Wellington Street
Ottawa (Ontario)
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file - Votre référence

Our file - Notre référence

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-612-19544-9

Name JEFF McPHAIL

Dissertation Abstracts International is arranged by broad, general subject categories. Please select the one subject which most nearly describes the content of your dissertation. Enter the corresponding four-digit code in the spaces provided.

MATHEMATICS

SUBJECT TERM

0405

U-M-I

SUBJECT CODE

Subject Categories

THE HUMANITIES AND SOCIAL SCIENCES

COMMUNICATIONS AND THE ARTS

Architecture 0729
Art History 0377
Cinema 0900
Dance 0378
Fine Arts 0357
Information Science 0723
Journalism 0391
Library Science 0399
Mass Communications 0708
Music 0413
Speech Communication 0459
Theater 0465

EDUCATION

General 0515
Administration 0514
Adult and Continuing 0516
Agricultural 0517
Art 0273
Bilingual and Multicultural 0282
Business 0488
Community College 0275
Curriculum and Instruction 0727
Early Childhood 0518
Elementary 0524
Finance 0277
Guidance and Counseling 0519
Health 0480
Higher 0745
History of 0520
Home Economics 0278
Industrial 0521
Language and Literature 0279
Mathematics 0280
Music 0522
Philosophy of 0998
Physical 0523

Psychology 0525
Reading 0535
Religious 0527
Sciences 0714
Secondary 0533
Social Sciences 0534
Sociology of 0340
Special 0529
Teacher Training 0530
Technology 0710
Tests and Measurements 0288
Vocational 0747

LANGUAGE, LITERATURE AND LINGUISTICS

Language
 General 0679
 Ancient 0289
 Linguistics 0290
 Modern 0291
Literature
 General 0401
 Classical 0294
 Comparative 0295
 Medieval 0297
 Modern 0298
 African 0316
 American 0591
 Asian 0305
 Canadian (English) 0352
 Canadian (French) 0355
 English 0593
 Germanic 0311
 Latin American 0312
 Middle Eastern 0315
 Romance 0313
 Slavic and East European 0314

PHILOSOPHY, RELIGION AND THEOLOGY

Philosophy 0422
Religion
 General 0318
 Biblical Studies 0321
 Clergy 0319
 History of 0320
 Philosophy of 0322
Theology 0469

SOCIAL SCIENCES

American Studies 0323
Anthropology
 Archaeology 0324
 Cultural 0326
 Physical 0327
Business Administration
 General 0310
 Accounting 0272
 Banking 0770
 Management 0454
 Marketing 0338
Canadian Studies 0385
Economics
 General 0501
 Agricultural 0503
 Commerce Business 0505
 Finance 0508
 History 0509
 Labor 0510
 Theory 0511
Folklore 0358
Geography 0366
Gerontology 0351
History
 General 0578

Ancient 0579
Medieval 0581
Modern 0582
Black 0328
African 0331
Asia, Australia and Oceania 0332
Canadian 0334
European 0335
Latin American 0336
Middle Eastern 0337
United States 0585
History of Science 0587
Law 0398
Political Science
 General 0611
 International Law and Relations 0616
 Public Administration 0617
Recreation 0810
Social Work 0452
Sociology
 General 0626
 Criminology and Penology 0627
 Demography 0938
 Ethnic and Racial Studies 0631
 Individual and Family Studies 0628
 Industrial and Labor Relations 0629
 Public and Social Welfare 0630
 Social Structure and Development 0700
 Theory and Methods 0344
Transportation 0709
Urban and Regional Planning 0999
Women's Studies 0452

THE SCIENCES AND ENGINEERING

BIOLOGICAL SCIENCES

Agriculture
 General 0473
 Agronomy 0285
 Animal Culture and Nutrition 0475
 Animal Pathology 0476
 Food Science and Technology 0359
 Forestry and Wildlife 0478
 Plant Culture 0479
 Plant Pathology 0480
 Plant Physiology 0817
 Range Management 0777
 Wood Technology 0746
Biology
 General 0306
 Anatomy 0287
 Botany 0308
 Botany 0309
 Cell 0379
 Ecology 0329
 Entomology 0353
 Genetics 0369
 Immunology 0793
 Microbiology 0410
 Molecular 0307
 Neuroscience 0317
 Oceanography 0416
 Physiology 0433
 Radiation 0821
 Veterinary Science 0778
Zoology 0472
Biophysics
 General 0786
 Medical 0760

Geodesy 0370
Geology 0372
Geophysics 0373
Hydrology 0388
Mineralogy 0411
Paleobotany 0345
Paleocology 0426
Paleontology 0418
Paleozoology 0985
Palmatology 0427
Physical Geography 0368
Physical Oceanography 0415

HEALTH AND ENVIRONMENTAL SCIENCES

Environmental Sciences 0768
Health Sciences
 General 0566
 Audiology 0300
 Chemotherapy 0992
 Dentistry 0567
 Education 0350
 Hospital Management 0789
 Human Development 0738
 Immunology 0982
 Medicine and Surgery 0564
 Mental Health 0347
 Nursing 0569
 Nutrition 0570
 Obstetrics and Gynecology 0380
 Occupational Health and Safety 0354
 Ophthalmology 0581
 Pathology 0571
 Pharmacology 0419
 Pharmacy 0572
 Physical Therapy 0382
 Public Health 0573
 Radiology 0574
 Recreation 0575

Speech Pathology 0460
Toxicology 0383
Home Economics 0386

PHYSICAL SCIENCES

Pure Sciences
 Chemistry
 General 0485
 Agricultural 0749
 Analytical 0486
 Biochemistry 0487
 Inorganic 0488
 Nuclear 0738
 Organic 0490
 Pharmaceutical 0491
 Physical 0494
 Polymer 0495
 Radiation 0754
 Mathematics 0405
 Physics
 General 0605
 Acoustics 0986
 Astronomy and Astrophysics 0606
 Atmospheric Science 0608
 Atomic 0748
 Electronics and Electricity 0607
 Elementary Particles and High Energy 0798
 Fluid and Plasma 0759
 Molecular 0609
 Nuclear 0610
 Optics 0752
 Radiation 0756
 Solid State 0611
 Statistics 0463
Applied Sciences
 Applied Mechanics 0346
 Computer Science 0984

Engineering
 General 0537
 Aerospace 0538
 Agricultural 0539
 Automotive 0540
 Biomedical 0541
 Chemical 0542
 Civil 0543
 Electronics and Electrical 0544
 Heat and Thermodynamics 0348
 Hydraulic 0545
 Industrial 0546
 Marine 0547
 Materials Science 0794
 Mechanical 0548
 Metallurgy 0743
 Mining 0551
 Nuclear 0552
 Packaging 0549
 Petroleum 0765
 Physics
 Sanitary and Municipal 0554
 System Science 0790
 Geotechnology 0428
 Operations Research 0796
 Plastics Technology 0793
 Textile Technology 0994

PSYCHOLOGY

General 0621
Behavioral 0384
Clinical 0622
Developmental 0623
Experimental 0624
Industrial 0625
Personality 0626
Physiological 0989
Psychobiology 0347
Psychometrics 0637
Social 0451

EARTH SCIENCES

Biogeochemistry 0425
Geochemistry 0996

The undersigned recommend to
the Faculty of Graduate Studies and Research
acceptance of the thesis

**VARIATIONAL METHODS IN GENERAL RELATIVITY
AND THE
POSITIVE ENERGY THEOREMS**

submitted by

Jeff McPhail, B.Sc.

in partial fulfillment of the requirements for
the degree of Master of Science

Chairman

Department of Mathematics and Statistics

Thesis Supervisor

CARLETON UNIVERSITY

Date September, 1996

Abstract

Variational methods may be applied to general relativity to obtain Einstein's equations. This is done through the Lagrangian formulation of the geometrodynamical action. Spacetime may be "sliced" into "space" and "time" in such a way that the information about "time" contained in the action may be characterized by the extrinsic curvature of the spacelike slice. This allows for a Hamiltonian formulation of the action which leads to a well defined initial value problem for Einstein's equations and to a meaningful definition of energy momentum of asymptotically Minkowskian spacetimes. It had been conjectured that this energy, the so called ADM mass (energy), is non-negative for spacetimes possessing certain physically reasonable properties and is zero only for Minkowski spacetime. Several special cases were proven, the most notable being Choquet Bruhat and Marsden's proof of the local version. Schoen and Yau first proved the global case assuming the existence of maximal slices and shortly thereafter proved the most general case through a perturbation method. In this thesis much of the theory leading to the definition of energy is developed and the above proofs are outlined with an emphasis on the first proof of Schoen and Yau.

Contents

0	Introduction	1
1	Preliminaries and Conventions	8
1.1	Tensors and Differential Geometry	8
1.2	Calculus of Variations	12
1.3	Energy Conditions	13
1.4	Causal Structure of a Spacetime	14
2	Variational Methods In Relativity	16
2.1	Variational Methods in Particle Dynamics	16
2.2	The Geometrodynamical Action in General Relativity	20
2.3	The Hamiltonian Formulation of Geometrodynamics	25
3	The Energy of a Spacetime	39
3.1	The Hamiltonian Definition	39
3.2	The Stress-Energy Definition	42
3.2.1	Backgrounds that are Minkowskian at Infinity	46
4	The Positivity of Mass	50
4.1	The Local Version of the Positive Mass Conjecture	51

1.1.1	Function Spaces	51
1.1.2	Asymptotically Minkowskian Manifolds	52
1.1.3	The Proof of the Local Version of the Positive mass Conjecture	53
1.2	The Global Version of Positive Mass Conjecture	55
1.2.1	The Existence of an Area Minimizing Surface	57
1.2.2	The Existence of a Class of Metrics Conformal $\mathbf{g}$	64
1.2.3	The Contradictory Estimates for Gaussian Curvature of S	62
1.2.4	The Only Spacetime with Zero Mass is Minkowski Spacetime	65
1.2.5	The Full Conjecture	66
5	Discussion and Further Developments	68

Chapter 0

Introduction

Since the beginning of the general theory of relativity, variational methods have played a key role in its development. Probably the most celebrated consequences of the theory were the singularity theorems of Hawking and Penrose which predict the “big bang” and the existence of black holes (see [19]). It was variational techniques applied to certain families of geodesics that were paramount to proving these theorems.

However, it was a different and much more general variational theory that played the most fundamental role in the development of general relativity (GR). Here it is a family of spacetimes themselves that are the objects subjected to variation. The contributions of this type of variational theory to GR are outlined in a flow chart in Figure 0.

The first significant contribution was the formalism of *geometrodynamics* (GMD). This first involved the Lagrangian derivation of Einstein’s equations from the *principle of least action*. The next step was to develop a *slicing* of spacetime, into “space” and “time”, necessary for the Hamiltonian treatment of GR.

Figure 0 Flowchart characterizing the involvement of variational methods in general relativity

The second major contribution of variational methods in GR was that this Hamiltonian formulation provided the framework for a well defined IVP for GR. This allows one to solve for the metric of a 3-dimensional slice of spacetime, given certain initial data that can in principle be measured, and integrate along “time” to solve for the metric of spacetime itself. This is equivalent to solving Einstein’s field equations.

Finally, this Hamiltonian theory provided a well defined notion of energy of a spacetime associated with the gravitational field. This lead to one of the major questions in GR in the last half of this century: “Is this energy always positive?”. This was very significant to the question of validity of GR as a true theory of nature.

In all physical theories it is desirable to define a notion of energy for the associated field that governs the system. It proved to be a particularly difficult problem to define the energy of the gravitational field in GR. This is because there is no well defined definition of local energy-momentum density associated with the gravitational field. This is due to the fact that the field variables for gravity are the metric components themselves, which determine the spacetime in which matter resides. Thus there is no “background” space in which to “test” the field. Thus the acceleration of a particle due to the gravitational field can only be measured by observing the deviation of its trajectory from that of another nearby particle subjected to the same field.

However, a global notion of energy is possible when the spacetime admits a time-like Killing vector field. In this case we can define a covariant energy-momentum current density along this vector field with respect to some background spacetime, and hence can define an energy-momentum 4-vector associated with the spacetime as a flux integral at spacelike infinity in much the same way as was done for gravitational waves. This may only be given for spacetimes that asymptotically approach the background spacetime at spatial infinity. Of course the most natural background

to choose is Minkowski spacetime. Thus only spacetimes that are asymptotically Minkowskian (in a sense to be defined later) will qualify for this definition of energy.

In the 1950's Arnowitt, Deser, and Misner [3] (ADM) introduced a Hamiltonian formulation of general relativity. They found that for asymptotically Minkowskian spacetimes the associated Hamiltonian, whose non zero part is a surface integral at spacelike infinity, gives a non-trivial definition of energy-momentum. Also, as mentioned above, one may define the energy-momentum of a spacetime as a flux integral of a certain energy-momentum current density. This density is defined with respect to the linearization of the stress-energy tensor around the metric of some background space. This background space for asymptotically Minkowskian spacetimes is of course the Minkowski spacetime itself. This definition depends on the existence of timelike Killing vector fields, whereas the Hamiltonian definition does not.

It turns out that these two definitions of energy-momentum coincide and the magnitude of this quantity is $16\pi m$ where m is the Schwarzschild mass. Because asymptotically Minkowskian spacetimes represent isolated gravitating systems, this energy-momentum definition is very promising as a viable physical quantity. Another property of the energy of a physical system is that the total energy should be non-negative.

This is because physical systems have a tendency to decay to lower energy states, so a "ground state" solution of the equations governing them is necessary for stability. The existence of a negative energy solution precludes the existence of a ground state. This is because the magnitude of the energy, of a negative greatest lower bound solution, may be made arbitrarily large through rescaling.

Thus the ground state should be the zero-energy solution, and all other solutions should have positive energy. Since we have every reason to believe that the universe is stable we should expect that any energy that we define for a spacetime must be

positive. This is known as the *positive mass conjecture* of GR. If this conjecture were not true then it would be reasonable to become suspicious of GR as a realistic theory of gravity.

The positive mass conjecture was first suggested by ADM (see [3] and the references therein) in the late 1950's. Although several special cases were proven along the way, it was not until some twenty years later that a proof was available for the general case.

In 1959 Brill [7] proved the result for time-symmetric ($\mathbf{K} \equiv 0$) and axially-symmetric spacetimes, and Araki [2] for a certain class of time-symmetric spacetimes constructed by conformal methods. In 1968 Brill and Deser [8] established the result for up to second-order perturbations from Minkowski spacetime. The problem was solved for the local case by Choquet-Bruhat and Marsden [11] in 1976 who proved that there exists a neighbourhood of Minkowski space (with respect to a certain class of weighted Sobolev topologies) for which the conjecture holds. Although this proof relied on the existence of maximal slices ($\text{tr}(\mathbf{K}) = 0$) they were able to prove the existence for spacetimes in such a neighbourhood as mentioned above. After this, the problem was divided into several special cases, one of which assumes the existence of a slicing of spacetime with Cauchy surface, Σ , diffeomorphic to $\mathbb{R}^3$. The proof relies on the fact that metrics, with non-negative scalar curvature in $\mathbb{R}^3$ and which are Euclidean outside a compact set, are flat.

In 1979 Schoen and Yau [31] came up with the first general (global) proof, obtaining the result for spacetimes admitting maximal slices. This proof is largely geometric and calls upon a version of the Gauss-Bonnet theorem. Also they appeal to geometric measure theory to show the existence of area-minimizing surfaces and the regularity thereof, where the competing class of "surfaces" includes those with possible singularities. (The standard reference for Geometric Measure Theory is Federer [14] but there is a much more readable text by Morgan [24]). The next year

York [37] showed that the conjecture had to be generalized somewhat, for physical reasons, to cover a larger class of spacetimes. Schoen and Yau [29] responded in 1981 and later that year proved the full conjecture relaxing the existence of maximal slices to slices with $\text{tr}(\mathbf{K}) = O(r^{-1})$ and $|K_{ij}| = O(r^{-2})$ which conforms to York's definition of physically reasonable. These later results were basically a perturbation argument on their previous paper.

Their work was truly a remarkable achievement. The arguments were mainly geometric and variational in nature, but also relied on results from topology, analysis, and geometric measure theory. That same year Witter [36] came up with a totally independent proof using supergravity, but Schoen and Yau are still credited with solving the global positive mass conjecture in its most general form. This was one of the major milestones for the relativists.

In Chapter 1 we recall some topics from differential geometry, relativity, and variational calculus. It is assumed that the reader is familiar with most of the material so only brief descriptions, sufficient to specify notation, are given.

In Chapter 2 we discuss the variational techniques used in general relativity for obtaining a Hamiltonian formulation. These include: the Lagrangian derivation of Einstein's equations, the splitting of spacetime into "space" and "time" in order to have a proper Hamiltonian set up, and the geometric interpretation of "time" flow that provides a well defined initial value problem (IVP) with the resulting Hamilton's equations. Although these concepts apply more directly to the Cauchy problem — the problem of formulating a well defined and practically feasible IVP for Einstein's equations — they are instrumental in obtaining a natural definition of the energy of a spacetime, to which the positive mass conjecture pertains.

In Chapter 3 we derive and reconcile two definitions of energy for asymptotically Minkowskian spacetimes. The first definition is due to the usual Hamiltonian treatment of field theories. The second definition assigns an energy-momentum 4-vector to certain spacetimes derived from a flux integral. This flux integral is of an energy-momentum flux (current) density defined from a stress-energy tensor linearized around Minkowski spacetime.

In Chapter 4 we state and outline the proofs of the two versions of the positive mass conjecture. In Section 1 we define what is meant for a spacetime to be asymptotically Minkowskian. The next section outlines ideas used in the proof of Choquet-Bruhat and Marsden for the local version of the conjecture. Section 2 gives a fairly detailed outline of Schoen and Yau's 1979 proof of the global version discussing in detail some of the main ideas. Only brief outlines of their two extensions are included since these are both perturbation arguments whose details are more technical than insightful.

Chapter 1

Preliminaries and Conventions

It is assumed that the reader is familiar with general relativity (GR) and differential geometry. There are three standard texts in GR; *Gravitation* by Misner, Thorne and Wheeler [22], *General Relativity* by Wald [35] and *The Large Scale Structure of Spacetime* by Hawking and Ellis [19]. These are listed in the order of increasing sophistication and decreasing readability. The physicist will prefer Misner *et al.*, the mathematicians will prefer Hawking and Ellis, and both will like Wald.

The reader unfamiliar with differential geometry should find a sufficient coverage of the subject in Misner *et al.* to get by. Otherwise there are virtually countless texts on the subject. A knowledge of the calculus of variations is helpful but not essential. A good reference for this is given at the end of Section 1.4. In the following sections we very briefly recall some important definitions and properties and specify notation.

1.1 Tensors and Differential Geometry

Tensors

A tensor, $\mathbf{T}$, will be in bold and its components given by $T^{a_1 \dots a_k}_{b_1 \dots b_l}$ where the upper indices are contravariant and lower indices are covariant. Usually a tensor will be referred to as T^{ab} or T_{ab} etc. instead of $\mathbf{T}$ so that its type will be understood.

The dual of a vector $\mathbf{v}$ will be denoted by $\tilde{\mathbf{v}}$. A 3-vector will be denoted by $\tilde{\mathbf{v}}$. The usual Einstein summation convention is used where summation is assumed on repeated indices, one of which is upper and the other is lower. Greek indices run from 0 to 3, where 0 is the index with respect to the timelike basis vector and Latin indices ($\geq i$) run from 1 to 3. When the dimension is arbitrary we will use Latin indices from a to h.

A metric ${}^{(4)}g$ refers to a pseudo-Riemannian metric on a 4-manifold. This is an indefinite, nondegenerate bilinear form with signature $\{-, +, +, +\}$. A metric ${}^{(3)}g$ refers to a Riemannian metric on a 3-manifold with signature $\{+, +, +\}$. Both of these may be referred to as simply g when it is clear from the context. For instance when the components of ${}^{(4)}g$ are written out they will have Greek indices so there will be no confusion. However, ${}^{(4)}g_{ij} = {}^{(3)}g_{ij}$ but ${}^{(4)}g^{ij} \neq {}^{(3)}g^{ij}$ as we will see, thus in certain instances there will be a need to distinguish. The determinants of ${}^{(4)}g$ and ${}^{(3)}g$ will be denoted by ${}^{(4)}g$ and ${}^{(3)}g$, respectively, or simply g if the context is clear.

The trace of a 2-tensor T^i_i will be denoted by T (not in bold) or, if there may be some confusion with another symbol, by $\text{tr}(\mathbf{T})$. For instance, the trace of $\mathbf{K}^2 \equiv K^{ij}K_{ij}$ must be denoted by $\text{tr}(\mathbf{K}^2)$ because $K^2 \equiv \text{tr}(\mathbf{K})^2$.

Manifolds

A manifold in this thesis will refer to a smooth semi-Riemannian manifold when referring to a 4-manifold and to a smooth Riemannian manifold for 2- or 3-manifolds. They will have coordinates defined from the usual atlas definition of a smooth manifold and will have a coordinate basis at each point denoted by $\{\mathbf{e}_0, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ or $\{\mathbf{e}_1, \mathbf{e}_2\}$ and dual basis $\{\omega^0, \omega^1, \omega^2, \omega^3\}$ etc. for 4-, 3-, and 2-manifolds, respectively, with $\mathbf{e}_0$ timelike and $\mathbf{e}_i$ spacelike, $i = 1, 2, 3$. Unless otherwise stated a 4-manifold will be denoted by M , a 3-manifold by Σ , and a

2-manifold (surface) by S (with the possibilities of indices).

A connection on M or S will be denoted by ∇ (as no confusion will occur) and on Σ by D . The covariant derivative of a tensor $T^{\alpha\beta}$ on M with respect to e_γ will be

$$\nabla_\gamma T^\alpha_\beta \equiv T^\alpha_{\beta;\gamma} = T^\alpha_{\beta,\gamma} + {}^{(4)}\Gamma^\alpha_{\beta\gamma} T^\alpha_\sigma - {}^{(4)}\Gamma^\sigma_{\alpha\gamma} T^\alpha_\beta \quad (1.1.1)$$

and

$$D_k T^i_j \equiv T^i_{j|k} = T^i_{j,k} + {}^{(3)}\Gamma^i_{kj} T^k_l - {}^{(3)}\Gamma^l_{ij} T^i_k \quad (1.1.2)$$

on Σ , where “,” denotes partial differentiation, “;” and “|” denote covariant differentiation, and ${}^{(4)}\Gamma$ and ${}^{(3)}\Gamma$ (or simply Γ) are the connection coefficients. The Hessian operator is defined as $\nabla_{\alpha\beta} \equiv \nabla_\alpha \nabla_\beta$ and $D_{ij} \equiv D_i D_j$, on M and Σ respectively. The exterior derivative will be denoted by d . The Laplacian is denoted by Δ .

The Riemann curvature tensor is the 4-tensor

$$\begin{aligned} R(u, v, w, \tilde{s}) &= \langle R(u, v)w, \tilde{s} \rangle \\ &= R^a_{bcd} w^b u^c v^a s_d \end{aligned}$$

where $R(u, v)$ is the operator

$$R(u, v)w = \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u, v]} w.$$

The vanishing of the last term follows from the fact that we only consider torsion free coordinates.

The Ricci tensor is the contraction of the Riemann tensor on its first and third indices and has components $R_{ab} \equiv R^c_{acb}$. The scalar curvature, R , is the contraction of the Ricci tensor, i.e. $R \equiv R^a_a$. An equation that we will need is the classical formula for $\tilde{R}(\tilde{g})$, where $\tilde{g} = \phi^4 g$ is a metric conformal to g , which is $\tilde{R} = \phi^{-5}[-8\Delta\phi + R\phi]$.

The extrinsic curvature tensor K on a manifold, M , is the negative of the projection of the differential of the unit normal vector field on M . Thus its components are

$$K_{ij} = -n_{i;j}. \quad (1.1.3)$$

From this quantity we get the *mean curvature* $K_m = \frac{1}{n} \text{tr}(K)$ (where $n = \dim M$) and the *Gaussian curvature* $\kappa = \det(K)$. (Actually this is only called the Gaussian curvature for $n = 2$, and for a general n , it is called the Gauss-Kronecker curvature). The global version of the Gauss-Bonnet theorem states that $\int_S \kappa dS = 2\pi \chi(S)$, where $\chi(S)$ is the Euler characteristic which is defined as *Faces - Edges + Vertices* in any triangulation of S .

The Lie derivative, $\mathcal{L}_v$, of a tensor T_{ab} , with respect to the vector v is defined by

$$\mathcal{L}_v T_{ab} = \lim_{t \rightarrow 0} \frac{\phi_{-t}^* T_{ab} - T_{ab}}{t} = \lim_{t \rightarrow 0} \frac{T_{ab} - \phi_t^* T_{ab}}{t} \quad (1.1.4)$$

where $\{\phi_t : M \rightarrow M\}$ is a one-parameter family of diffeomorphisms along the flow of v . That is, ϕ_t maps a point, $p \in M$, to the point $\phi_t(p) \in M$ which is a parameter distance equal to t away from p along the integral curve generated by v at p . $\phi_t^* T_{ab}$ denotes the pull-back of T_{ab} which is defined as follows. For a function f the pull back of f is

$$\phi^* f|_{\phi(p)} = f \circ \phi|_p$$

and the push-forward of a vector v is

$$\phi_* v(f)|_{\phi(p)} = v(\phi^* f)|_p, \quad \forall f \text{ on } M.$$

The pull-back of a covariant tensor T_{ab} is then defined by

$$\phi^* T(u, v)|_{\phi(p)} = T(\phi_* u, \phi_* v)|_p.$$

The Lie derivative of a vector u is just the Lie bracket,

$$\mathcal{L}_v u = [v, u],$$

and the Lie derivative of a scalar function is just

$$\mathcal{L}_v f = v(f) = f_{;i} v^i.$$

The components of the Lie derivative for a 2-tensor are

$$\mathcal{L}_{\mathbf{v}} T_{ab} = T_{ab;c} v^c + T_{cb} v^c{}_{;a} + T_{ac} v^c{}_{;b}.$$

A Killing vector $\mathbf{v}$ is one for which $\mathcal{L}_{\mathbf{v}} g_{ab} = 0$. Thus it satisfies Killing's equation,

$$g_{ab;c} v^c + g_{cb} v^c{}_{;a} + g_{ac} v^c{}_{;b} = v_{b;a} + v_{a;b} = 0.$$

This means that the metric is conserved along the flow of a Killing vector field and that a Killing vector field generates a one-parameter family of isometries. This is why the existence of timelike Killing vector fields is so important — it allows for “time” translation symmetry.

1.2 Calculus of Variations

Let $\mathcal{F}$ be a function (vector) space and $J : \mathcal{F} \rightarrow \mathbb{R}$ be a functional mapping into the reals. We define the *Gâteaux derivative* of $J[\mathbf{u}]$ with respect to $\mathbf{v} \in \mathcal{F}$ (or in the direction of $\mathbf{v}$) by

$$\begin{aligned} \delta J[\mathbf{u}]\mathbf{v} &\equiv \lim_{t \rightarrow 0} \frac{J[\mathbf{u} + t\mathbf{v}] - J[\mathbf{u}]}{t} \\ &\equiv \left. \frac{d}{dt} J[\mathbf{u} + t\mathbf{v}] \right|_{t=0}. \end{aligned} \quad (1.2.5)$$

When $\mathcal{F}$ is a finite dimensional tensor space, $\delta J[\mathbf{u}]\mathbf{v}$ corresponds to the covariant derivative $\nabla_{\mathbf{v}} \mathbf{u}$, sometimes denoted $\delta_{\mathbf{v}} \mathbf{u}$. In this case, when $\mathbf{v}$ is not specified, “ δ ” acts as a total derivative in the specified independent variables, i.e.,

$$\delta J[\mathbf{u}, \mathbf{v}, \mathbf{w}] \equiv \frac{\delta J}{\delta \mathbf{u}} \delta \mathbf{u} + \frac{\delta J}{\delta \mathbf{v}} \delta \mathbf{v} + \frac{\delta J}{\delta \mathbf{w}} \delta \mathbf{w}, \quad (1.2.6)$$

where if $J[\mathbf{u}, \mathbf{v}, \mathbf{w}] = \int_M L(\mathbf{u}, \mathbf{v}, \mathbf{w})$,

$$\frac{\delta J}{\delta \mathbf{u}} \equiv \int_M \frac{\partial L}{\partial \mathbf{u}}.$$

We call δJ the *variation* of J and (1.2.6) the variation of J with respect to $\mathbf{u}$, $\mathbf{v}$, and $\mathbf{w}$ etc.

The *second variation* of a functional J is $\delta^2 J[\mathbf{u}](\mathbf{v}, \mathbf{w}) \equiv \delta(\delta J[\mathbf{u}]\mathbf{v})\mathbf{w}$. (Some authors use the notation $J'[\mathbf{u}]\mathbf{v}$ and $J''[\mathbf{u}](\mathbf{v}, \mathbf{w})$ for the first and second variations respectively.)

One can form a calculus from the Gateaux derivative which is called the *calculus of variations*. This calculus has a critical point analysis directly analogous to that of the usual multi variable calculus.

For a functional, J , to be stationary, (i.e., a critical point) we require that $\delta J = 0$ and for $\delta^2 J > 0$ or $\delta^2 J < 0$ also J will be a minimum or a maximum, respectively. Of course J will only be extremal at such a point with respect to the "basis" that the total variation (derivative) was taken on.

Although the mechanics of the variational procedures and definitions are straight forward and seem to follow directly as a generalization of multi-variable calculus, the analysis required to show these results is far from trivial. For a thorough treatment of the foundations of the calculus of variations see [17].

1.3 Energy Conditions

The stress energy tensor, T^{ab} , is a symmetric 2-tensor which characterizes the matter energy source distribution with the following properties:

$$T^{00} = \text{density of mass energy,}$$

$$T^{0i} = i\text{th component of momentum density,}$$

$$T^{i0} = i\text{th component of energy flux, and}$$

$$T^{ij} = j\text{th component of flux of } i\text{-momentum}$$

$$= \text{flux of } i\text{-momentum across surface of } x^j = \text{constant.}$$

For $\mathbf{v}$, a 4-velocity vector, $J_\alpha \equiv T_{\alpha\beta}v^\beta$ (sometimes denoted by p_α), is the mass-energy current density along $\mathbf{v}$ and, $\mu \equiv T_{\alpha\beta}v^\alpha v^\beta$, is the mass-energy density with respect

to $\mathbf{v}$. The *weak energy condition* is the requirement that $\mu \geq 0$ for all timelike $\mathbf{v}$. The *strong energy condition* is $\mu \geq -1/2T$ for all unit timelike $\mathbf{v}$. The *dominant energy condition* is $\mu \geq |J^\alpha J_\alpha|^{1/2}$ for all unit timelike $\mathbf{v}$.

The conservation of energy-momentum (density) of a field is a consequence of the fact that its associated stress-energy tensor satisfies $\nabla \cdot \mathbf{T} = T^{ab}{}_{;b} = 0$. The conservation of energy along the flow of some vector field $\mathbf{v}$ would require that $\nabla \cdot \mathbf{J} = J^a{}_{;a} = (T^{ab}v_b)_{;a} = 0$. But

$$(T^{ab}v_b)_{;a} = \cancel{v^{ab}{}_{;a}v_b} + T^{ab}v_{b;a} = T^{ab}v_{b;a}. \quad (1.3.7)$$

Since T^{ab} is symmetric, the contraction of T^{ab} with an anti-symmetric tensor will be identically zero. Thus only for $v_{(a;b)} = \frac{1}{2}(v_{a;b} + v_{b;a}) = 0$ will we have energy conservation along its flow. But this is the requirement that $\mathbf{v}$ be a Killing vector. Therefore before we attempt to define energy as a flux integral we must be sure of the existence of Killing vector fields.

1.4 Causal Structure of a Spacetime

A curve, $\alpha(s)$, is said to be a future (past) directed causal curve if at each s_0 , $\alpha'(s_0)$ (the tangent vector to α at s_0) is either future (past) directed and timelike or null. The future (past) domain of dependence of a set $S \subseteq M$ is

$$D^{+(-)}(S) \equiv \{p \in M \mid \text{all past (future) inextendable causal curves through } p \text{ intersect } S\}$$

and the *domain of dependency* of S is

$$D(S) \equiv D^+(S) \cup D^-(S).$$

An *achronal* set, S , is one for which there does not exist a future directed timelike curve with two elements of S as its endpoints. A *Cauchy surface* is an achronal set $\Sigma \subseteq M$ such that $D(\Sigma) = M$. A spacetime M is said to be *globally hyperbolic* if it

contains a Cauchy surface. (A Cauchy surface is sometimes called a spacelike slice of M).

A spacetime is said to be *stably causal* if its metric g has a neighbourhood V such that there are no closed timelike curves in any metric belonging to V . For such a spacetime it can be shown that there exists a “global time function”, t , which is such that its gradient is everywhere timelike. Thus all level surfaces of a global time function are Cauchy. All the spacetimes that we consider in this thesis will be stably causal and globally hyperbolic.

Chapter 2

Variational Methods In Relativity

The variational techniques that are necessary for the formalization of a field theory have multi-dimensional spaces as the family of competing objects. The dynamical variables, called *field variables*, are tensors that describe a physical interaction instead of a generalized coordinate. For general relativity the objects to be varied are spacetimes and the field variables are the metric components of the spacetime.

Before we discuss this field variational procedure it will be helpful to recall the basic variational techniques from particle dynamics. This was the main motivation for the development of variational calculus and is very instructive because of its intuitive nature.

2.1 Variational Methods in Particle Dynamics

A typical problem in particle dynamics is to find the trajectory of a particle between two points that lie in some force field. One could solve the problem by force considerations but often it is more economical to look at the energy of the system. This involves minimizing an integral of a function, $L(\mathbf{x}, \dot{\mathbf{x}}, t)$ (which is related to the energy), between the endpoints:

$$I = S[\mathbf{x}, \dot{\mathbf{x}}] = \int_{(t_0, \mathbf{x}_0)}^{(t_1, \mathbf{x}_1)} L(\mathbf{x}, \dot{\mathbf{x}}, t) dt . \quad (2.1.1)$$

In order for this integral to be minimal it must be stationary with respect to $\mathbf{x}$ and $\dot{\mathbf{x}}$ so that the variation,

$$\delta S = \frac{\delta S}{\delta \mathbf{x}} \delta \mathbf{x} + \frac{\delta S}{\delta \dot{\mathbf{x}}} \delta \dot{\mathbf{x}}, \quad (2.1.2)$$

is zero. Computing this variation gives rise to the familiar Euler-Lagrange equations:

$$\begin{aligned} \delta S &= \int_{(t_0, x_0)}^{(t_1, x_1)} \left(\frac{\partial L}{\partial \mathbf{x}} \delta \mathbf{x} + \frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right) dt \\ &= \cancel{\frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \bigg|_{(t_0, x_0)}^{(t_1, x_1)}} + \int_{(t_0, x_0)}^{(t_1, x_1)} \left[\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) \right] \delta \mathbf{x} dt \\ &\Rightarrow \frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) = 0. \quad \text{(Provided } \frac{\partial L}{\partial \dot{\mathbf{x}}} \text{ is differentiable).} \end{aligned} \quad (2.1.3)$$

If there is some (generalized) force acting on the system external to the fields incorporated into L , these equations will take the form,

$$\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) = \mathbf{Q}(\mathbf{x}, \dot{\mathbf{x}}, \dots, t), \quad (2.1.4)$$

where $\mathbf{Q}$ is the force. This force may also have a Lagrangian associated with it which may be added to the original field Lagrangian. The variation of this quantity will turn out to be $\mathbf{Q}$. (Later we will see that this is how the other matter fields are incorporated into Einstein's field equations).

If we allow for a variation of the second endpoint the variation of I must also be taken with respect to t ;

$$\begin{aligned} I &= S[\mathbf{x}, \dot{\mathbf{x}}, t] = I = \int_{(t_0, x_0)}^{(t, x)} L(\mathbf{x}, \dot{\mathbf{x}}, s) ds, \\ \delta S &= \frac{\delta S}{\delta t} \delta t + \frac{\delta S}{\delta \mathbf{x}} \delta \mathbf{x} + \frac{\delta S}{\delta \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \\ &= L \delta t + \int_{(t_0, x_0)}^{(t, x + \Delta x)} \left(\frac{\partial L}{\partial \mathbf{x}} \delta \mathbf{x} + \frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \right) ds \\ &= L \delta t + \cancel{\frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} \bigg|_{(x_0, t_0)}^{(x + \Delta x, t)}} + \int_{(t_0, x_0)}^{(t, x + \Delta x)} \left[\frac{\partial L}{\partial \mathbf{x}} - \frac{d}{ds} \left(\frac{\partial L}{\partial \dot{\mathbf{x}}} \right) \right] \delta \mathbf{x} ds \end{aligned}$$

$$\begin{aligned}
&= L\delta t + \frac{\partial L}{\partial \dot{\mathbf{x}}} \Delta \mathbf{x} \\
&= L\delta t + \frac{\partial L}{\partial \dot{\mathbf{x}}} (\delta \mathbf{x} - \dot{\mathbf{x}} \delta t) \\
&= \frac{\partial L}{\partial \dot{\mathbf{x}}} \delta \mathbf{x} - \left(\dot{\mathbf{x}} \frac{\partial L}{\partial \dot{\mathbf{x}}} - L \right) \delta t \\
&= \mathbf{p} \delta \mathbf{x} - H \delta t.
\end{aligned} \tag{2.1.5}$$

This is the well known dispersion relation where $\mathbf{p}$ is the momentum of the particle and H is the Hamiltonian, which is equal to the total energy of the particle see Figure (2.1.1). Now we can express L as

$$L = \dot{\mathbf{x}} \frac{\partial L}{\partial \dot{\mathbf{x}}} - H = \dot{\mathbf{x}} \mathbf{p} - H, \tag{2.1.6}$$

to get

$$I = S[\dot{\mathbf{x}}, \mathbf{p}, t] = \int_{t_0}^t L ds = \int_{t_0}^t (\dot{\mathbf{x}} \mathbf{p} - H) ds \tag{2.1.7}$$

and

$$\delta S = \frac{\delta S}{\delta \dot{\mathbf{x}}} \delta \dot{\mathbf{x}} + \frac{\delta S}{\delta \mathbf{p}} \delta \mathbf{p} = \int_{t_0}^t \left[\left(\mathbf{p} - \frac{\partial H}{\partial \dot{\mathbf{x}}} \right) \delta \dot{\mathbf{x}} + \left(\dot{\mathbf{x}} - \frac{\partial H}{\partial \mathbf{p}} \right) \delta \mathbf{p} \right] ds = 0. \tag{2.1.8}$$

Upon integration by parts we obtain

$$I = \int (-\dot{\mathbf{x}} \mathbf{p} - H) dt \quad \text{and} \quad \delta I = \int_{t_0}^t \left[\left(-\dot{\mathbf{p}} - \frac{\partial H}{\partial \dot{\mathbf{x}}} \right) \delta \mathbf{x} + \left(-\dot{\mathbf{x}} - \frac{\partial H}{\partial \mathbf{p}} \right) \delta \mathbf{p} \right] dt = 0 \tag{2.1.9}$$

as the boundary conditions are such that $\delta \mathbf{x} = 0$ at $s = t_0$ and $s = t$. This shows that

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}} \quad \text{and} \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}}. \tag{2.1.10}$$

Lagrangian Configuration

Hamiltonian Configuration

Figure 2.1.1: These graphs represent the trajectories of a particle subjected to some field. In the Lagrangian configuration the endpoints are fixed, in the Hamiltonian configuration the second endpoint is allowed to vary by an amount $\Delta x = \delta x - \dot{x} \delta t$.

These first-order equations known as Hamilton's equations are equivalent to the second-order Euler-Lagrange equations. The order of these differential equations is reduced at the cost of doubling the number of equations. Notice that we now have a well defined initial value problem (IVP). If we specify initial data, $(\mathbf{x}, \mathbf{p}, t)$, we may integrate equations (2.1.10) for all t to obtain $(\mathbf{x}, \mathbf{p})$ for all t . The set of all $(\mathbf{x}, \mathbf{p})$ satisfying (2.1.10) is called the *phase space*. These ideas can be generalized beyond the notion of particle dynamics to any physical system. Given some manifold with a quantity (tensor) $q_{\alpha_1 \dots \alpha_k}$ defined on it and a function $L(q_{\alpha_1 \dots \alpha_k}, \dot{q}_{\alpha_1 \dots \alpha_k}, t)$ whose integral over the manifold must be extremized, we will obtain upon variation second order differential equations or twice as many first-order equations.

Also there will be a dispersion relation of the same form as above with a "conjugate momentum" $p^{\alpha_1 \dots \alpha_k}$ and a Hamiltonian H that will represent (in a sense) the energy of the system. This procedure of minimizing the Lagrangian function L , of the dynamic variables of a system, $q_{\alpha_1 \dots \alpha_k}$, to obtain equations that govern this system is called the *principle of least action*. This principle permeates all of physics and is considered a fundamental notion. It was the driving force behind quantum field theory and sets the stage for the Cauchy problem of general relativity and the definition of mass-energy of a spacetime. This chapter will develop in detail the Hamiltonian formulation of general relativity which allows a natural definition of the energy-momentum of spacetime as an energy-momentum flux integral that is directly related to the conjugate momentum of the system and the Hamiltonian.

2.2 The Geometrodynamical Action in General Relativity

The problem of geometrodynamics (GMD) is to obtain the equations that relate the geometry of spacetime with the gravitational field (and more generally with the other

fields as well). All of the information about spacetime is contained in the metric, since it is a (pseudo-) Riemannian manifold, so these dynamical equations should relate the metric to quantities that describe the fields. In keeping with this notion the action principle should have the metric coefficients as its dynamic variables. Once the spacetime metric is determined, then all of the relativistic dynamics may be calculated.

The end result of a Lagrangian formulation of this action principle is of course to obtain Einstein's equations. Later we will see that the Hamiltonian formulation provides a more practical framework for the determination of the geometry of spacetime — the Cauchy problem.

The geometrodynamical action is given by $I = I_G + I_F$ where

$$I_G = \int_M \mathcal{L}_G d^4x = \int_M L_G \sqrt{-g} d^4x = \frac{1}{16\pi} \int_M R \sqrt{-g} d^4x \quad (2.2.11)$$

and $I_F = \int_M \mathcal{L}_F d^4x = \int_M L_F \sqrt{-g} d^4x$,

where R is the scalar curvature (contraction of the Ricci tensor $R_{\alpha\beta}$) and g is the determinant of the metric. There are several reasons why we have chosen $\frac{1}{16\pi}R$ as the Lagrangian (the constant $1/16\pi$ is included to maintain consistency of physical units).

First of all, we want to obtain a relationship between the geometry of spacetime and the mass distribution (which is the gravitational source) so the quantity to be extremized should be a geometrical invariant. This geometrical quantity must be a function of the metric coefficients and their derivatives since the field equations, which will govern the dynamics of general relativity, are completely determined by the geometry of the spacetime manifold which is in turn determined by the metric. This quantity must also contain enough information to describe the geometry, otherwise the subsequent field equations will be incomplete. The Riemann curvature tensor, $R^\alpha{}_{\beta\gamma\delta}$, completely determines the geometry of a manifold, and, of its 14 independent components, R is the only one that is linear in the second derivatives

of g and contains no higher derivatives. Allowing for nonlinearity or higher derivatives of g would jeopardize the Newtonian limit of GR. Also there is no geometrical invariant that can be made out of g and its first derivatives alone.

Finally one would expect that the curvature of spacetime would be minimized (extremized) for a given mass distribution in the spirit of least action (why should spacetime curve more than it has to?). So it seems that R is uniquely qualified (modulo a scaling factor) to be the geometrical Lagrangian of GR.

The other term L_F is the contribution to the Lagrangian of the fields present in the spacetime other than gravity (i.e. the mass-energy source distribution) such as electromagnetism, etc.

Hilbert was the first to set up such a variational derivation for determining the field equations of GR. He varied only the $g_{\alpha\beta}$'s so his action integral was

$$S[g_{\alpha\beta}] = I = \int_M L_G \sqrt{-g} d^4x + \int_M L_F \sqrt{-g} d^4x . \quad (2.2.12)$$

Later Palitini (see [22], Chp. 21) also included the connection coefficients as dynamic variables. There are 40 independent $\Gamma_{\beta\gamma}^\alpha$'s and when they are included in the variation greatly reduce the computation. We will vary the Palitini action assuming that the $\Gamma_{\beta\gamma}^\alpha$'s are free. We will find that, just as in the Hamiltonian formulation of particle dynamics where the momentum p was varied as a free variable and later found to coincide with the physical momentum, the $\Gamma_{\beta\gamma}^\alpha$'s turn out to be

$$\Gamma_{\beta\gamma}^\alpha = \frac{1}{2} g^{\alpha\sigma} [g_{\sigma\gamma,\beta} + g_{\beta\sigma,\gamma} - g_{\beta\gamma,\sigma}] ,$$

which is the precise formula for the connection coefficients of g .

The Palitini action is

$$\begin{aligned} I &= S[g_{\alpha\beta}, \Gamma_{\beta\gamma}^\alpha] = \frac{1}{16\pi} \int_M R \sqrt{-g} d^4x + \int_M L_F \sqrt{-g} d^4x , \\ \delta S &= \frac{1}{16\pi} \int_M [\delta R \sqrt{-g} + R \delta \sqrt{-g}] d^4x \\ &\quad + \int_M [\delta L_F \sqrt{-g} + L_F \delta \sqrt{-g}] d^4x . \end{aligned} \quad (2.2.13)$$

The variation of g is,

$$\delta g = \frac{\partial g}{\partial g_{\alpha\beta}} \delta g_{\alpha\beta} = \frac{\partial(g_{\alpha\beta} C^{\alpha\beta})}{\partial g_{\alpha\beta}} \delta g_{\alpha\beta} = g^{\alpha\beta} \delta g_{\alpha\beta}, \quad (2.2.14)$$

where $C^{\alpha\beta}$ is the cofactor matrix of g . Thus

$$\delta \sqrt{-g} = \frac{-1/2}{\sqrt{-g}} \delta g = -\frac{1}{2} \sqrt{-g} g^{\alpha\beta} \delta g_{\alpha\beta} = -\frac{1}{2} \sqrt{-g} g_{\alpha\beta} \delta g^{\alpha\beta}. \quad (2.2.15)$$

The Ricci tensor is

$$R_{\alpha\beta} = \Gamma_{\alpha\beta,\gamma}^{\gamma} - \Gamma_{\alpha\gamma,\beta}^{\gamma} - \Gamma_{\alpha\beta}^{\sigma} \Gamma_{\gamma\delta}^{\gamma} - \Gamma_{\alpha\gamma}^{\sigma} \Gamma_{\beta\sigma}^{\gamma} \quad (2.2.16)$$

but we may work in a coordinate system where the Γ 's vanish and $R_{\alpha\beta}$ becomes

$R_{\alpha\beta} = \Gamma_{\alpha\beta,\gamma}^{\gamma} - \Gamma_{\alpha\gamma,\beta}^{\gamma}$. Now we compute its variation;

$$\begin{aligned} \delta R_{\alpha\beta} &= \delta \Gamma_{\alpha\beta,\gamma}^{\gamma} - \delta \Gamma_{\alpha\gamma,\beta}^{\gamma} = (\delta \Gamma_{\alpha\beta}^{\gamma})_{;\gamma} - (\delta \Gamma_{\alpha\gamma}^{\gamma})_{;\beta} \\ &= (\delta \Gamma_{\alpha\beta}^{\gamma})_{;\gamma} - (\delta \Gamma_{\alpha\gamma}^{\gamma})_{;\beta}. \end{aligned} \quad (2.2.17)$$

The last equality holds because the non-tensorial part of the Γ 's cancel out in the variation so $\delta \Gamma_{\alpha\beta}^{\gamma}$ is a tensor. Finally

$$\begin{aligned} \sqrt{-g} g^{\alpha\beta} \delta R_{\alpha\beta} &= \sqrt{-g} \left[g^{\alpha\beta} (\delta \Gamma_{\alpha\beta}^{\gamma})_{;\gamma} - g^{\alpha\beta} (\delta \Gamma_{\alpha\gamma}^{\gamma})_{;\beta} \right] \\ &= \sqrt{-g} \left[g^{\alpha\beta} (\delta \Gamma_{\alpha\beta}^{\gamma})_{;\gamma} - g^{\alpha\gamma} (\delta \Gamma_{\alpha\mu}^{\mu})_{;\gamma} \right] \\ &= \sqrt{-g} \nabla_{\gamma} \left[g^{\alpha\beta} (\delta \Gamma_{\alpha\beta}^{\gamma})_{;\gamma} - g^{\alpha\gamma} (\delta \Gamma_{\alpha\mu}^{\mu})_{;\gamma} \right] \\ &\equiv \sqrt{-g} \nabla_{\gamma} V^{\gamma}, \end{aligned} \quad (2.2.18)$$

where V^{γ} is the term in brackets. The variation of the action is then

$$\begin{aligned} \delta I &= \frac{1}{16\pi} \delta \int_M R \sqrt{-g} d^4x + \delta \int_M L_F \sqrt{-g} d^4x \\ &= \frac{1}{16\pi} \delta \int_M R_{\alpha\beta} g^{\alpha\beta} \sqrt{-g} d^4x + \int_M \delta(L_F \sqrt{-g}) d^4x \\ &= \frac{1}{16\pi} \int_M [R_{\alpha\beta} \delta g^{\alpha\beta} + g^{\alpha\beta} \delta R_{\alpha\beta}] \sqrt{-g} d^4x + R \delta(\sqrt{-g}) d^4x \end{aligned}$$

$$\begin{aligned}
 & + \int_M \delta L_F \sqrt{-g} + L_F \delta(\sqrt{-g}) d^4x \\
 & = \frac{1}{16\pi} \int_M \sqrt{-g} \left[R_{\alpha\beta} \delta g^{\alpha\beta} + V_{;\gamma}^\gamma - \frac{1}{2} g_{\alpha\beta} R \delta g^{\alpha\beta} + 16\pi \delta L_F - 8\pi g_{\alpha\beta} \delta g^{\alpha\beta} L_F \right] d^4x \\
 & = \frac{1}{16\pi} \int_M \sqrt{-g} \left[\left(R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R \right) + 8\pi \left(2 \frac{\delta L_F}{\delta g^{\alpha\beta}} - L_F g_{\alpha\beta} \right) + V_{;\gamma}^\gamma \right] \delta g^{\alpha\beta} d^4x \\
 & = 0.
 \end{aligned} \tag{2.2.19}$$

For this to be zero we require

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R + 8\pi \left(2 \frac{\delta L_F}{\delta g^{\alpha\beta}} - L_F g_{\alpha\beta} \right) = 0 \quad \text{and} \quad \int_M V_{;\gamma}^\gamma \sqrt{-g} d^4x = 0. \tag{2.2.20}$$

The second equality can be shown to be equivalent to

$$\Gamma_{\alpha\beta}^\gamma = \frac{1}{2} g^{\gamma\mu} [g_{\alpha\mu,\beta} + g_{\mu\beta,\alpha} - g_{\alpha\beta,\mu}]. \tag{2.2.21}$$

But this is a requirement of the connection coefficient so this term must vanish requiring that the first equation holds. The first equation is usually written

$$G_{\alpha\beta} \equiv R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = 8\pi \left(L_F g_{\alpha\beta} - 2 \frac{\delta L_F}{\delta g^{\alpha\beta}} \right) \equiv 8\pi T_{\alpha\beta} \tag{2.2.22}$$

which are the well known Einstein's equations.

These are Einstein's equations in their standard form which are a system of 10 independent second-order differential equations. There are a few known exact solutions to these equations [3] but none of them agree with the known universe. This is because these equations are only solvable when one appeals to a high degree of symmetry, which clearly is not the case for our universe. The equations in this form do not give us an IVP so there is no way to determine the spacetime that we live in without imposing constraints. In the Hamiltonian picture of geometrodynamics we can, given certain initial data (called Cauchy data) on an initial spacelike slice of M , integrate Hamilton's equations forward or backward in "time" to obtain (M, g) (at least in principle).

2.3 The Hamiltonian Formulation of Geometrodynamics

To obtain a Hamiltonian formulation we require:

- (1) An initial data set, defined on an "initial space",
- (2) A well defined "evolution" of this data and space along some flow that may be characterized by a single parameter that is usually denoted by t and referred to as "time",
- (3) A "time derivative" that characterizes the rate of change of the data along this flow.

In this situation it is natural to have all of "space", $\Sigma_{t=t_0}$ (our 3-dimensional spatial universe), "at one time" t_0 , to be the initial space. By "at one time" we mean the value of a *global time function*, t . This is a function on M , such that each of its level surfaces, Σ_t ; is closed, has a spacelike tangent bundle, and contains no two points that are causally related.

Each of the quantities on Σ_t are 3-dimensional, and in particular the $^{(3)}g_{ij}$'s will be the dynamical field variables. Thus we will have to express the Lagrangian and hence R , in terms of 3-dimensional quantities.

Both the dynamic variables and their conjugate momenta depend on the parameter t . In addition to this the action integral must be an integral "along t " (in this generalized version it will be along the flow generated by a vector field which is parametrized by t instead of just along a curve parametrized by t) starting at a fixed point in the dynamic variable space to some other point that is allowed to vary.

Once the conjugate momenta and Hamiltonian are determined, the generalized form of Hamilton's equations can be found through variation and the generalized dispersion relation. These will be 20 independent first-order equations involving

the dynamic variables, their conjugate momenta, and the Hamiltonian, which are analogous to those in (2.1.10).

Thus, knowing these quantities at the "initial" (i.e. $t = t_0$) fixed point we can integrate forward or backward in time to obtain their values for all t .

In geometrodynamics the dynamic variables are the spacelike components of the metric, g_{ij} , on some spacelike slice, Σ_{t_0} (which is the initial fixed "point"), of our yet-to-be-determined spacetime. The slicing of spacetime must be such that each Σ_t is a Cauchy surface as outlined in Section 1.4. This requirement is necessary to guarantee the existence of a smooth timelike vector field on M normal to Σ_t and a global time function t . This time function will parametrize the flow of the timelike vector field leaving in its path a one-parameter family of Cauchy surfaces Σ_t . These surfaces may be defined *via* a one-parameter family of diffeomorphisms, $\phi_t : \Sigma_0 \rightarrow M$, from some initial Cauchy surface Σ_0 to the spacetime M .

Let $\mathbf{n}$ be a unit timelike vector field on M such that its restriction to Σ_t for any t is a normal vector field on Σ_t . The "time" lapse, $N = N(t, x^1, x^2, x^3)$, is the parameter distance along the flow of $\mathbf{n}$ between neighbouring slices Σ_t and Σ_{t+dt} . For a given point $p \in \Sigma_t$, $\mathbf{n}|_p$ in general will not pierce Σ_{t+dt} at the point, p_2 , corresponding to the trajectory of p at parameter value $t + dt$. This is because there will also be a shift in the spacelike coordinates by $N^i = N^i(t, x^1, x^2, x^3)$ due to the curvature of the slices, which in general, will change as Σ_t evolves into Σ_{t+dt} . Furthermore, when calculating the proper distance from p to an arbitrary point p_3 on Σ_{t+dt} , a correction term $N^i dt$ must be added to dx^i to compensate for this shifting.

The evolution of Σ_t is defined by the fact that each Σ_t is a level surface of the global time function t . Thus the gradient of t , dt , is everywhere orthogonal to Σ_t . This means that the trajectory of each point in Σ_t will be an integral curve of the vector field dual to dt . Denote this vector by, $\mathbf{t}$.

(b) The Evolution of Σ along t

Figure 2.3.1: This shows two cauchy surfaces (with two dimensions suppressed) that are, (a) an infinitesimal infinitesimal parameter distance, dt , apart and (b) a parameter distance, $t - t_0$ apart. The vector $\tilde{t}$ generates the evolution of the family of Cauchy surfaces Σ_t . It is the dual to the gradient of the global time function, t , and has components (N, N^i) in the coordinates adapted to the evolution structure. Also (a) shows the slicing of spacetime and (b) represents the evolution of Σ_{t_0} into Σ_t . This is analogous to the particle mechanics construction in Figure (2.1.1). The slicing is required because in this case the "endpoints" have three degrees of freedom instead of none and the trajectory space between "endpoints" is a 4-dimensional flow instead of a space-curve.

By definition, $\mathbf{n}$ is orthogonal to the three spacelike basis vectors so $\{\mathbf{n}, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ is another basis for M . In this new basis $\mathbf{t}$ has components $\mathbf{t} = (N, N^1, N^2, N^3)^T$ since $\mathbf{t} = N\mathbf{n} + \vec{N}$ as one can see from Figure (2.3.1(a)).

More formally this is just a change of coordinates from $\{x^0, x^1, x^2, x^3\}$ to $\{t, \hat{x}^1, \hat{x}^2, \hat{x}^3\} = \{t, x^1, x^2, x^3\}$. From this we obtain

$$\begin{aligned} d\mathbf{x}^0 &= \frac{\partial x^0}{\partial t} d\mathbf{t} + \frac{\partial x^0}{\partial x^i} d\mathbf{x}^i = \frac{\partial x^0}{\partial t} d\mathbf{t} \quad \text{and} \\ d\mathbf{x}^i &= \frac{\partial x^i}{\partial t} d\mathbf{t} + \frac{\partial x^i}{\partial x^j} d\mathbf{x}^j = \frac{\partial x^i}{\partial t} d\mathbf{t} + \delta^i_j d\mathbf{x}^j = \frac{\partial x^i}{\partial t} d\mathbf{t} + d\mathbf{x}^i. \end{aligned} \quad (2.3.23)$$

As Σ_t evolves with t , the tangent to the flow will be $\mathbf{t} = (N, N^i)$ with respect to the new coordinates. From the previous definition of lapse and shift functions, we have

$$N = \frac{\partial x^0}{\partial t} \quad \text{and} \quad N^i = \frac{\partial x^i}{\partial t}.$$

As the names suggest, N is the “lapse” — the timelike change with respect to t and $\vec{N} = (N^1, N^2, N^3)^T$ is the “shift” — the spacelike change with respect to t of a given point in spacetime.

The separation between p and p_1 is

$$\begin{aligned} ds^2 &= g_{\alpha\beta} d\mathbf{x}^\alpha d\mathbf{x}^\beta = \tilde{g}_{nn} d\mathbf{t} d\mathbf{t} + \tilde{g}_{ij} (d\mathbf{x}^i + N^i d\mathbf{t}) (d\mathbf{x}^j + N^j d\mathbf{t}) \\ &= -N^2 d\mathbf{t}^2 + \tilde{g}_{ij} N^i d\mathbf{t} d\mathbf{x}^j + \tilde{g}_{ij} N^j d\mathbf{t} d\mathbf{x}^i + \tilde{g}_{ij} d\mathbf{x}^i d\mathbf{x}^j \end{aligned} \quad (2.3.24)$$

where $\tilde{g}_{\alpha\beta}$ are the components of $\mathbf{g}$ with respect to the new basis. In terms of the original metric, the new metric components are such that

$$\tilde{g}_{ij} = g_{ij}, \quad N_j = \tilde{g}_{ij} N^i = g_{j0} \quad \text{and}$$

$$\tilde{g}_{ij} N^i N^j - N^2 = N^i N_i - N^2 = \|\vec{N}\|^2 - N^2 = g_{00}.$$

The inverse elements of $\tilde{g}$ are $\tilde{g}^{ij} = g^{ij} + N^i N^j / N^2$; also $\tilde{g}^{00} = -1/N^2$ and $\tilde{g}^{0i} = N^i / N^2$. From now on we will denote $\tilde{g}_{ij}$ as g_{ij} and $g_{\alpha\beta}$ as $^{(4)}g_{\alpha\beta}$ when there might be

confusion. In this configuration, $\tilde{\mathbf{n}} = (-N, 0, 0, 0)$ and $\mathbf{n} = (1/N, -N^i/N)^T$. The new basis and dual basis are

$$\begin{aligned} \mathbf{e}_n &= \mathbf{n} & \mathbf{w}^n &= N d\mathbf{t} \\ \mathbf{e}_i &= \partial_i & \mathbf{w}^i &= d\mathbf{x}^i + N^i d\mathbf{t} . \end{aligned} \quad (2.3.25)$$

These are called ADM coordinates (due to Arnowitt, Deser, and Misner) [5, 6; and the references therein] which provide a sort of splitting of “space” and “time”. By this we mean that at any given parameter, t , of the flow of $\mathbf{t}$, Σ_t is all of “space” at “time”, t . The GMD action may now be expressed as

$$\begin{aligned} S[(^{(4)}g^{ij})] &= \frac{1}{16\pi} \int_{\Sigma_{t_0}}^{\Sigma_t} R \sqrt{-^{(4)}g} d^4x = \frac{1}{16\pi} \int_{\Sigma_{t_0}}^{\Sigma_t} R \sqrt{g} (N dt) d^3x \\ &= \frac{1}{16\pi} \int_{t_0}^t \left[\int_{\Sigma_t} R \sqrt{g} d^3x \right] N dt \end{aligned} \quad (2.3.26)$$

which will support a Hamiltonian formulation. The next step is to express R without explicit dependence on t , allowing us to have an initial data set without having to specify the “time” of the initial slice. This is done *via* the extrinsic curvature of the spacelike slices.

Recall that the extrinsic curvature tensor is just the familiar Weingarten map, which is now referred to as the second fundamental form. Its components on Σ_t are the spacelike components of the negative of the differential of $\mathbf{n}$.

$$\mathbf{K} \equiv P(-d\mathbf{n}) = -n_{ij} d\mathbf{x}^i \otimes d\mathbf{x}^j , \quad (2.3.27)$$

where P is the projection operator from M to Σ_t . Note that this is the covariant derivative of $\mathbf{n}$ with respect to the connection on M not Σ_t . Thus $\mathbf{K}$ not only depends on Σ_t , but also the spacetime that it is embedded in.

All the information about “time” which is realized by M ’s constituents as the evolution of Σ_t along the flow of $\mathbf{n}$, is therefore contained in $\mathbf{K}$. In order to obtain a splitting of “space” and “time” necessary for a Hamiltonian formulation we will

express $\mathbf{R}$ in terms of ${}^{(3)}\mathbf{R}$ (the Riemann tensor for Σ_t) and $\mathbf{K}$. First of all note that $\mathbf{K}$ is symmetric,

$$\begin{aligned}
 K_{ij} &= g_{ik} K^k_j = (\mathbf{e}_i \cdot \mathbf{e}_k) K^k_j = \mathbf{e}_i \cdot (K^k_j \mathbf{e}_k) \\
 &= \mathbf{e}_i \cdot (-n^k_j \mathbf{e}_k) \\
 &= -\mathbf{e}_i \cdot \nabla_j \mathbf{n} \\
 &= \mathbf{n} \cdot \nabla_j \mathbf{e}_i \\
 &= \mathbf{n} \cdot \nabla_i \mathbf{e}_j = \dots = K_{ji}
 \end{aligned} \tag{2.3.28}$$

and we can express the covariant derivative of the spacelike basis vectors as

$$\nabla_i \mathbf{e}_j = \Gamma^\alpha_{ij} \mathbf{e}_\alpha = \Gamma^0_{ij} \mathbf{e}_0 + \Gamma^k_{ij} \mathbf{e}_k = K_{ij} \frac{\mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} + \Gamma^k_{ij} \mathbf{e}_k = -K_{ij} \mathbf{n} + \Gamma^k_{ij} \mathbf{e}_k, \tag{2.3.29}$$

since $K_{ij} = \mathbf{n} \cdot \nabla_i \mathbf{e}_j = \mathbf{n} \cdot \Gamma^\alpha_{ij} \mathbf{e}_\alpha = \mathbf{n} \cdot \Gamma^0_{ij} \mathbf{e}_0$. Thus we can calculate the spatial components of Riemann as follows

$$\begin{aligned}
 \nabla_j \nabla_k \mathbf{e}_i &= \nabla_j (-K_{ik} \mathbf{n} + \Gamma^l_{ik} \mathbf{e}_l) \\
 &= \nabla_j (-K_{ik}) \mathbf{n} + K_{ik} \nabla_j \mathbf{n} + \nabla_j (\Gamma^l_{ik} \mathbf{e}_l) \\
 &= -K_{ik,j} \mathbf{n} + K_{ik} (-K^m_j \mathbf{e}_m) + \Gamma^l_{ik,j} \mathbf{e}_l + \Gamma^l_{ik} \Gamma^m_{jl} \mathbf{e}_m \\
 &\quad + [-K_{mj} \mathbf{n} + \Gamma^l_{mj} \mathbf{e}_l] \Gamma^m_{ik},
 \end{aligned} \tag{2.3.30}$$

$$\nabla_j \nabla_k \mathbf{e}_j - \nabla_k \nabla_j \mathbf{e}_i = [K_{ij|k} - K_{ik|j}] \mathbf{n} + [K_{ik} K^m_j - K_{ij} K^m_k + {}^{(3)}R^m_{ijk}] \mathbf{e}_m.$$

The Riemann tensor may be expressed in terms of $\mathbf{K}$ and ${}^{(3)}R^m_{ijk}$ the Riemann tensor on Σ_t :

$$R^m_{ijk} = K_{ik} K^m_j - K_{ij} K^m_k + {}^{(3)}R^m_{ijk}. \tag{2.3.31}$$

We have already constructed a 3-metric, ${}^{(3)}\mathbf{g}$, from the spatial components of ${}^{(4)}\mathbf{g}$ but it is useful to define another tensor, $\mathbf{h}$, on M that coincides with ${}^{(3)}\mathbf{g}$ on Σ_t ,

$$h_{\alpha\beta} \equiv g_{\alpha\beta} + n_\alpha n_\beta, \text{ so that } h_{ij} = g_{ij} = {}^{(3)}g_{ij}.$$

Keep in mind that the indices of $h_{\alpha\beta}$ are raised by $^{(4)}g$ not $^{(3)}g$ but $h^{ij} = g^{ij} + n^i n^j = g^{ij} + \frac{N^i N^j}{N^2} = ^{(3)}g^{ij}$. Now consider

$$\begin{aligned}
 R_{\alpha\beta\gamma\delta} h^{\alpha\gamma} h^{\beta\delta} &= R_{\alpha\beta\gamma\delta} (g^{\alpha\gamma} + n^\alpha n^\gamma) (g^{\beta\delta} + n^\beta n^\delta) \\
 &= R_{\alpha\beta\gamma\delta} g^{\alpha\gamma} g^{\beta\delta} + R_{\alpha\beta\gamma\delta} (g^{\alpha\gamma} n^\beta n^\delta + g^{\beta\delta} n^\alpha n^\gamma) \\
 &\quad + R_{\alpha\beta\gamma\delta} n^\alpha n^\beta n^\gamma n^\delta \\
 &= R + R_{\beta\delta} n^\beta n^\delta + R_{\alpha\gamma} n^\alpha n^\gamma + \cancel{R^{0000} n_0 n_0 n_0 n_0} \\
 &= R + 2R_{\beta\delta} n^\beta n^\delta.
 \end{aligned} \tag{2.3.32}$$

Also we can express R^m_{ijk} as

$$h_\alpha{}^m h_i{}^\beta h_j{}^\gamma h_k{}^\delta R^\alpha{}_{\beta\gamma\delta} = h^{\alpha m} h_i{}^\beta h_j{}^\gamma h_k{}^\delta R_{\alpha\beta\gamma\delta}$$

which gives

$$\begin{aligned}
 R^j{}_i &= h^{\alpha j} h^{i\beta} h_j{}^\gamma h_i{}^\delta R_{\alpha\beta\gamma\delta} \\
 &= h^{\alpha\gamma} h^{\beta\delta} R_{\alpha\beta\gamma\delta}.
 \end{aligned} \tag{2.3.33}$$

So equation (2.3.31) gives us

$$\begin{aligned}
 K^i{}_i K^j{}_j - K^i{}_j K^j{}_i + ^{(3)}R &= R + 2R_{\alpha\beta} n^\alpha n^\beta, \\
 R &= ^{(3)}R + K^i{}_i K^j{}_j - K^i{}_j K^j{}_i - 2R_{\alpha\beta} n^\alpha n^\beta, \quad \text{or} \\
 R &= ^{(3)}R + K^2 - \text{tr}(K^2) - 2R_{\alpha\beta} n^\alpha n^\beta.
 \end{aligned} \tag{2.3.34}$$

Finally we calculate the last term:

$$\begin{aligned}
 R_{\alpha\beta} n^\alpha n^\beta &= R^\gamma{}_{\alpha\gamma\beta} n^\alpha n^\beta \\
 &= n^\alpha (\nabla_\beta \nabla_\alpha - \nabla_\alpha \nabla_\beta) n^\beta \\
 &= n^\alpha (\nabla_\beta (\nabla_\alpha n^\beta) - \nabla_\alpha (\nabla_\beta n^\beta))
 \end{aligned}$$

$$\begin{aligned}
&= \nabla_\beta(n^\alpha \nabla_\alpha n^\beta) - (\nabla_\beta n^\alpha)(\nabla_\alpha n^\beta) + (\nabla_\alpha n^\alpha)(\nabla_\beta n^\beta) - \nabla_\alpha(n^\alpha \nabla_\beta n^\beta) \\
&= K^\alpha_\alpha K^\beta_\beta - K^\alpha_\beta K^\beta_\alpha + \nabla_\beta(n^\alpha \nabla_\alpha n^\beta) - \nabla_\alpha(n^\alpha \nabla_\beta n^\beta) \\
&= K^2 - \text{tr}(\mathbf{K}^2) + \nabla_\beta(n^\alpha \nabla_\alpha n^\beta) - \nabla_\alpha(n^\alpha \nabla_\beta n^\beta)
\end{aligned} \tag{2.3.35}$$

which provides our reformulation of the Lagrangian density¹

$$\mathcal{L}_G = \frac{1}{16\pi} R \sqrt{-g} = \frac{1}{16\pi} R N \sqrt{{}^{(3)}g} = \frac{1}{16\pi} [{}^{(3)}R + \text{tr}(\mathbf{K}^2) - K^2] N \sqrt{{}^{(3)}g}. \tag{2.3.36}$$

The Hamiltonian can now be obtained from the generalized dispersion relation

$$\delta S = \pi^{ij} \delta g_{ij} - \mathcal{H}_G \delta t \quad \text{where} \quad \mathcal{H}_G = \pi^{ij} \dot{g}_{ij} - \mathcal{L}_G,$$

$$I = S[g_{ij}, \pi^{ij}] = \int_M \mathcal{L}_G d^4x = \int_{t_0}^t \left[\int_{\Sigma_t} [\pi^{ij} \delta g_{ij} - \mathcal{H}_G \delta t] d^3x \right] dt, \tag{2.3.37}$$

$$\pi^{ij} \equiv \frac{\delta S}{\delta g_{ij}} = \frac{\partial \mathcal{L}_G}{\partial \dot{g}_{ij}}. \tag{2.3.38}$$

We still have not defined $\dot{g}_{ij}$ explicitly other than that it is the “time” derivative of g_{ij} in the sense that it is the rate of change of g_{ij} along the flow of $\mathbf{t}$. The natural derivative to describe this is the Lie derivative which is the rate of change of a tensor field along some flow. Since $\mathbf{t} = N\mathbf{n} + \vec{N}$ we may write $\mathbf{n} = \frac{1}{N}(\mathbf{t} - \vec{N})$. The Lie derivative of $g_{\alpha\beta}$ with respect to $\mathbf{n}$ is

$$\mathcal{L}_{\mathbf{n}} g_{\alpha\beta} = \frac{g_{\alpha\beta;\gamma} n^\gamma + g_{\gamma\beta} n^\gamma{}_{;\alpha} + g_{\alpha\gamma} n^\gamma{}_{;\beta}}$$

¹The second last line in equation (2.3.35) follows from the fact that $\mathbf{n}$ is a unit vector and therefore its derivative must be orthogonal to itself; i.e. $\nabla \mathbf{n} \cdot \mathbf{n} = 0$.

Also the last two terms have been dropped because they are divergences. This is the usual practice in Lagrangian mechanics where the boundary conditions are such that the variation of quantities with respect to the dynamic variables vanish on the boundary of the manifold in question. Because divergences become surface integrals through the use of Stokes theorem they have no contribution. It is also a common practice in Hamiltonian mechanics, but here one must be careful because the divergences may not have the necessary boundary conditions to vanish. As we shall see later, these divergences must not be excluded in certain cases otherwise we do not have a well defined Hamiltonian system (remember that the “endspace” is not fixed in the Hamiltonian picture). However for now we shall discard them.

$$= g_{\gamma\beta} n^\gamma_{;\alpha} + g_{\alpha\gamma} n^\gamma_{;\beta} \quad (2.3.39)$$

$$= n_{\beta;\alpha} + n_{\alpha;\beta}.$$

Looking at just the spacelike components of g one has

$$\mathcal{L}_{\mathbf{n}} g_{ij} = n_{i;j} + n_{j;i} = -K_{ij} - K_{ji} = -2K_{ij}. \quad (2.3.40)$$

The extrinsic curvature of a hypersurface is just the $-1/2$ rate of change of its metric along the flow of its normal vector field. This may be re-written as

$$\begin{aligned} K_{ij} &= -\frac{1}{2} \mathcal{L}_{\mathbf{n}} g_{ij} = -\frac{1}{2} \left[g_{\alpha j} n^\alpha_{;i} + g_{i\alpha} n^\alpha_{;j} \right] \\ &= -\frac{1}{2N} \left[g_{i\alpha} \left[(N n^\alpha)_{;j} + \cancel{\not{g}^\alpha(N)_{;j}} \right] + g_{\alpha j} \left[(N n^\alpha)_{;i} + \cancel{\not{g}^\alpha(N)_{;i}} \right] \right] \\ &= -\frac{1}{2N} [\mathcal{L}_{(N\mathbf{n})} g_{ij}] \\ &= -\frac{1}{2N} [\mathcal{L}_{(\mathbf{t}-N)} g_{ij}] \\ &= -\frac{1}{2N} [\mathcal{L}_{\mathbf{t}} g_{ij} - \mathcal{L}_N g_{ij}] \\ &= -\frac{1}{2N} [\mathcal{L}_{\mathbf{t}} g_{ij} - \cancel{\not{g}_{ij;k} N^k} - g_{ik} N^k_{;j} - g_{kj} N^k_{;i}] \\ &= \frac{1}{2N} [N_{i|j} + N_{j|i} - \mathcal{L}_{\mathbf{t}} g_{ij}] \\ &= \frac{1}{2N} [N_{i|j} + N_{j|i} - \dot{g}_{ij}] \end{aligned} \quad (2.3.41)$$

$$\text{or } \dot{g}_{ij} = -2NK_{ij} + N_{i|j} + N_{j|i} = 2(N_{(i|j)} - NK_{ij}).$$

The momentum conjugate to g_{ij} is

$$\pi^{ij} \equiv \frac{\delta S}{\delta g_{ij}} = \frac{\partial \mathcal{L}_G}{\partial \dot{g}_{ij}} = \frac{N}{16\pi} \sqrt{-g} \left[\frac{\partial \cancel{\not{g}^{ij}}}{\partial \dot{g}_{ij}} + \frac{\partial(\text{tr}(\mathbf{K}^2))}{\partial \dot{g}_{ij}} - \frac{\partial K^2}{\partial \dot{g}_{ij}} \right]. \quad (2.3.42)$$

From above,

$$\frac{\partial K_{ij}}{\partial \dot{g}_{ij}} = -\frac{1}{2N},$$

which will allow us to compute the other terms:

$$\begin{aligned}
 \frac{\partial K}{\partial \dot{g}_{ij}} &= \frac{\partial(g^{ij} K_{ij})}{\partial \dot{g}_{ij}} = -\frac{1}{2N} g^{ij} \Rightarrow \frac{\partial K^2}{\partial \dot{g}_{ij}} = -\frac{1}{N} g^{ij} , \\
 \frac{\partial(\text{tr}(\mathbf{K}^2))}{\partial \dot{g}_{ij}} &= \frac{\partial(K^{ij} K_{ij})}{\partial \dot{g}_{ij}} \quad (2.3.43) \\
 &= \frac{\partial K^{ij}}{\partial \dot{g}_{ij}} K_{ij} - \frac{1}{2N} K^{ij} \\
 &= \frac{\partial(g^{ik} g^{jl} K_{kl})}{\partial \dot{g}_{ij}} K_{ij} - \frac{1}{2N} K^{ij} \\
 &= g^{ik} g^{jl} \left(-\frac{1}{2N} \delta^k_i \delta^l_j \right) K_{ij} - \frac{1}{2N} K^{ij} \\
 &= g^{ij} g^{jj} \left(-\frac{1}{2N} \right) K_{ij} - \frac{1}{2N} K^{ij} \\
 &= -\frac{1}{2N} (K^{ij} + K^{ij}) \\
 &= -\frac{1}{N} K^{ij} .
 \end{aligned}$$

Therefore the conjugate momentum is

$$\pi^{ij} = \frac{1}{16\pi} \sqrt{g} [K^{ij} - g^{ij} K] . \quad (2.3.44)$$

The dispersion relation for GMD is

$$\begin{aligned}
 \delta S &= \pi^{ij} \delta g_{ij} - \mathcal{H}_G \delta t \\
 &= \pi^{ij} \dot{g}_{ij} \delta t - \mathcal{H}_G \delta t \\
 &= (\pi^{ij} \dot{g}_{ij} - \mathcal{H}_G) \delta t \quad (2.3.45) \\
 &= \frac{\delta S}{\delta t} \delta t \\
 &= \mathcal{L}_G \delta t .
 \end{aligned}$$

The Hamiltonian density in terms of the Lagrangian density is then

$$\mathcal{H}_G = \pi^{ij} \dot{g}_{ij} - \mathcal{L}_G$$

$$\begin{aligned}
&= \pi^{ij} [2N K_{ij} + N_{j|i} + N_{i|j}] - \sqrt{g} N [^{(3)}R + K_{ij} K^{ij} - K^2] \\
&= \pi^{ij} \left[2N g^{-1/2} \left(\pi_{ij} - \frac{1}{2} g_{ij} \pi \right) + N_{j|i} + N_{i|j} \right] \\
&\quad - \sqrt{g} N \left[^{(3)}R + g^{-1} \left(\pi_{ij} - \frac{1}{2} g_{ij} \pi \right) \left(\pi^{ij} - \frac{1}{2} g^{ij} \pi \right) - g^{-1/4} \pi^2 \right] \\
&= 2N g^{1/2} \left(\pi^{ij} \pi_{ij} - \frac{1}{2} \pi^2 \right) + 2\pi^{ij} N_{(i|j)} - g^{1/2} N ^{(3)}R \\
&\quad - g^{-1/2} N \left(\pi^{ij} \pi_{ij} - \pi^2 + \frac{3}{4} \pi^2 \right) + N g^{-1/2} \pi^2 / 4 \\
&= -g^{1/2} N ^{(3)}R - N g^{-1/2} \left(\pi^{ij} \pi_{ij} - \frac{1}{2} \pi^2 \right) + 2\pi^{ij} N_{i|j} \\
&= g^{-1/2} \left[N \left(^{(3)}R - g^{-1} (\pi^{ij} \pi_{ij} - \frac{1}{2} \pi^2) \right) + 2 \left[(g^{-1/2} \pi^{ij} N_i)_{|j} - N_i (g^{-1/2} \pi^{ij})_{|j} \right] \right] \\
&= g^{1/2} \left[N \left(^{(3)}R - g^{-1} (\pi^{ij} \pi_{ij} - \frac{1}{2} \pi^2) \right) - 2N_i (g^{-1/2} \pi^{ij})_{|j} \right]
\end{aligned} \tag{2.3.46}$$

Because N and N^i are just quantities that determine how spacetime is sliced and may be adjusted to suit the problem at hand, the value of the Hamiltonian $H_G = \int_{\Sigma_t} \mathcal{H}_G d^3x$ should be independent of them. Therefore the variation of H_G with respect to N and N_i should vanish requiring that

$$^{(3)}R - g^{-1} (\pi^{ij} \pi_{ij} - \frac{1}{2} \pi^2) = 0 \tag{2.3.47}$$

and

$$D_j (g^{-1/2} \pi^{ij}) = (g^{-1/2} \pi^{ij})_{|j} = 0. \tag{2.3.48}$$

So it appears that the GMD system is a constrained Hamiltonian system with Lagrange multipliers N and N^i and Hamilton's equations

$$\dot{g}_{ij} = \frac{\delta H_G}{\delta \pi^{ij}} = 2g^{-1/2} N \left(\pi_{ij} - \frac{1}{2} g_{ij} \pi \right) + 2\pi^{ij} N_{(i|j)} \tag{2.3.49}$$

and

$$\dot{\pi}^{ij} = -\frac{\delta H_G}{\delta g_{ij}} = -N g^{-1/2} (^{(3)}R^{ij} - \frac{1}{2} g^{ij} ^{(3)}R) + \frac{1}{2} N g^{-1/2} g^{ij} (\pi^{kl} \pi_{kl} - \frac{1}{2} \pi^2)$$

$$\begin{aligned}
& -2Ng^{-1/2}(\pi^{ik}\pi_k^j - \frac{1}{2}\pi^{ij}\pi) + g^{1/2}(N^{[ij} - g^{ij}N^{lk}) \\
& + g^{1/2}(g^{-1/2}N^k\pi^{ij})_{|j} - 2\pi.
\end{aligned} \tag{2.3.50}$$

These (20 independent) equations may be shown to be equivalent to Einstein's equations (2.2.22) and will lead to a well defined initial value problem. Solving these equations for a given initial data set, called Cauchy data $(\Sigma, g_{ij}, K^{\nu})$, subject to constraints (2.3.47) and (2.3.48), is known as the Cauchy problem of geometrodynamics.² This may be formalized by the following theorem that we quote from Wald [35; Thm 10.22].

Theorem 2.3.1 *Let Σ be a three-dimensional C^∞ manifold, let g_{ij} be a smooth Riemannian metric on Σ , and let K_{ij} be a smooth symmetric tensor field on Σ . Suppose $^{(3)}g_{ij}$ and K_{ij} satisfy the constraint equations (2.3.47), (2.3.48). Then there exists a unique C^∞ spacetime $(M, g_{\alpha\beta})$ called the maximal Cauchy development of $(\Sigma, ^{(3)}g_{ij}, K_{ij})$, satisfying the following four properties:*

- (i) $(M, g_{\alpha\beta})$ is a solution of Einstein's equations (2.2.22);
- (ii) $(M, g_{\alpha\beta})$ is globally hyperbolic with Cauchy surface Σ ;
- (iii) The induced metric and extrinsic curvature are $^{(3)}g_{ij}$ and K_{ij} , respectively;
- (iv) Every other spacetime satisfying (i)–(iii) can be mapped isometrically into a subset of $(M, g_{\alpha\beta})$.

Furthermore, $(M, g_{\alpha\beta})$ satisfies the desired domain of dependency property in the following sense. Suppose $(\Sigma, ^{(3)}g_{ij}, K_{ij})$ and $(\Sigma', ^{(3)}g'_{ij}, K'_{ij})$ are initial data sets with maximal developments $(M, g_{\alpha\beta})$ and $(M', g'_{\alpha\beta})$, respectively. Suppose there is a diffeomorphism between $S \subset \Sigma$ and $S' \subset \Sigma'$ which carries $(^{(3)}g_{ij}, K_{ij})$ on S into

²Here we have only considered $\mathcal{L}_G$ instead of doing the calculation with the full Lagrangian, $\mathcal{L} = \mathcal{L}_G + \mathcal{L}_F$. For this case the calculation follows directly so long as the sources, $T^{\mu\nu}$, are integrable, but is more cumbersome so we omit it.

$(^{(3)}g'_{ij}, K'_{ij})$ on S' . Then $D(S)$ in the spacetime $(M, g_{\alpha\beta})$ is isometric to $D(S')$ in the spacetime $(M', g'_{\alpha\beta})$. Finally, the solution $g_{\alpha\beta}$ on M depends continuously on the initial data $(^{(3)}g_{ij}, K_{ij})$ on Σ .

When one actually goes about solving the system for a given initial data set, the calculation is facilitated by a judicious choice of lapse and shift functions since they may be arbitrarily chosen. It is conceivable, then, that we will be able to obtain an initial data set of the universe through measurement, to a sufficient approximation to solve the Cauchy problem numerically. This is one of the major topics in the discipline of numerical relativity.

A field theory that is described by a constrained Hamiltonian system such as this has more degrees of freedom than are required dynamically. That is, there is an additional “gauge” freedom in the phase space (g_{ij}, π^{ij}) that will allow a whole class of solutions to (2.3.49) and (2.3.50) that give an identical physical configuration. A familiar example of this is in electromagnetic field theory where the solution to the field equations is the potential $\vec{A}$. The physical “fields” $\vec{E}$ and $\vec{B}$ only depend on $d\vec{A}$ so for $\vec{A}' = \vec{A} + d\xi$, for some function ξ , $d\vec{A}' = d\vec{A}$. Thus either of the two potentials are equally well qualified to describe the system.

A solution of the field equations is actually a whole equivalence class of solutions and may be represented by any of its members. This is called a *gauge field theory* where a given specification of solution is called a *gauge*.

A common gauge to work in for electromagnetism is the Lorentz gauge where $\nabla \cdot \vec{A} = 0$. Also this gauge invariance must imply that $\int \vec{\pi} \cdot [\delta\vec{A} - \nabla(\delta\xi)] = \int \vec{\pi} \cdot \delta\vec{A}$ which in turn implies $\nabla \cdot \vec{\pi} = 0$.

Similarly in GMD field theory there is an equivalence class of solutions with the relation $g'_{ij} \sim g_{ij}$ if there exists a diffeomorphism φ of Σ_t such that $g'_{ij} = \varphi^* g_{ij}$. This phase space of equivalence classes in general relativity is referred to as *superspace*. The relation that must be satisfied in this case is $\int \pi^{ij}(\delta g_{ij} - v_{(i;j)}) = \int \pi^{ij} \delta g_{ij}$ which

means that $D_i(g^{-1/2}\pi^{ij}) = 0$ which is constraint (2.3.48). Unlike the electromagnetic case we still have one constraint left which may be viewed as the gauge freedom in choosing a slicing of spacetime.

It is interesting to note that in an unconstrained theory one may specify a gauge to obtain a constraint. This is in effect choosing a parametrization. The resulting constraint turns out to be linear in the momentum conjugate to the parameter characterizing the gauge. This allows one to solve the constraint for the momentum and substitute it back into the Lagrangian in effect deparametrizing the system.

In the GMD case, constraint (2.3.47) is quadratic in the momentum so this is not possible. Thus as yet no one has been able to remove this constraint from the theory using this deparametrization procedure. This is one of the major problems in formulating a quantum theory of gravity.

Another important consequence of the Hamiltonian formulation of relativity is that it leads to a meaningful (global) definition of the energy of a certain class of spacetimes. In the next chapter we discuss this and show that it is equivalent to another definition of energy obtained from a stress energy tensor linearized around Minkowski spacetime.

Chapter 3

The Energy of a Spacetime

3.1 The Hamiltonian Definition

As in all Hamiltonian formulations, the energy of a spacetime should be defined as the Hamiltonian for the GMD action. The numerical value of this Hamiltonian is defined by

$$H_G = \int_{\Sigma} N\mathcal{H} + N^i \mathcal{H}_i d^3x \quad (3.1.1)$$

where

$$\mathcal{H} \equiv g^{1/2} {}^{(3)}R - g^{-1/2} (\pi^{ij} \pi_{ij} - 1/2 \pi^2) = 0$$

and

$$\mathcal{H}^i \equiv -2g^{1/2} \pi^{ij}{}_{;j} = 0^i$$

are identically zero. Thus it appears that in this definition the energy of any source-free spacetime is zero.

However, recall that throughout the derivation of the above formula all divergences that appeared were discarded because it was assumed that variations on the

¹When sources are present ($\mathcal{L}_F \neq 0$) then it can be shown that $\mathcal{H} = \mu$ and $\mathcal{H}_i = J_i$. The weak energy condition $\mu \geq 0$ with $\mu = 0$ for Minkowski spacetime, is an assumption in the positive mass conjecture. Thus $\int_{\Sigma} \mu d^3x \geq 0$ and therefore its inclusion makes no difference in proving positivity of energy - the Hamiltonian with an extra term to be added later. So we only need consider the case where $\mathcal{H} = 0$. (The $N^i \mathcal{H}_i$ term has no contributions because of the assumed asymptotic behavior of N^i .)

boundary vanish. (See (2.3.35))

In the case of a closed spacetime there is no spatial boundary; thus such considerations are unnecessary and the total energy is zero. On the other hand, for asymptotically flat spacetimes this is far from clear. In fact for such spacetimes the absence of all of these boundary terms will produce a Hamiltonian formulation that is not well defined. The phase space (g_{ij}, π^{ij}) of the above Hamiltonian will not contain the extremal trajectories for the system. It can be shown that the variation of H_G is (without dropping any terms),

$$\begin{aligned} \delta H_G = & \int_{\Sigma} \frac{\delta H_G}{\delta g_{ij}} \delta g_{ij} + \frac{\delta H_G}{\delta \pi^{ij}} \delta \pi^{ij} d^3x \\ & - \int_{\partial\Sigma} \frac{1}{2} g^{1/2} (g^{ii} g^{jl} + g^{il} g^{jk} - 2g^{ij} g^{kl}) (N \delta g_{ij|k} - N_{,k} \delta g_{ij}) d^2 S_l \quad (3.1.2) \\ & - \int_{\partial\Sigma} \left[2N_k \delta \pi^{kl} + (2N^k \pi^{jl} - N^l \pi^{jk}) \delta g_{jk} \right] d^2 S_l, \end{aligned}$$

where $dS_l = 1/2 g^{1/2} \varepsilon_{lmn} dx^m \wedge dx^n$. For Hamilton's equations to be well defined, the variation of the Hamiltonian must be of the form

$$\delta H = \int A^{ij} \delta g_{ij} + B_{ij} \delta \pi^{ij} d^3x \quad (3.1.3)$$

where, in this case,

$$A^{ij} = \frac{\delta H_G}{\delta g_{ij}} \text{ and } B_{ij} = \frac{\delta H_G}{\delta \pi^{ij}}.$$

Thus if H_G is the true Hamiltonian then the two surface integrals must vanish. However the boundary conditions at infinity for a general (physically) meaningful asymptotically flat spacetime will give a non-zero result for the first surface integral. It can be shown that the surviving term of this surface integral at infinity² is

$$- \int_{\partial\Sigma} \frac{1}{2} g^{1/2} (g^{ik} g^{jl} + g^{il} g^{jk} - 2g^{ij} g^{kl}) N \delta g_{ij|k} d^2 S_l = -\delta \int_{\partial\Sigma} (g_{ik,l} - g_{il,k}) d^2 S_k. \quad (3.1.4)$$

²By surface integral at infinity we mean an integral taken over a 2-sphere as the radius goes to spacelike infinity.

Now define the new Hamiltonian as

$$H = H_G + E \equiv H_G + \int_{\Sigma} (g_{ik,i} - g_{ii,k}) d^2 S_k \quad (3.1.5)$$

which will give us the variation,

$$\delta H = \int_{\Sigma} \left(\frac{\delta H_G}{\delta g_{ij}} \delta g_{ij} + \frac{\delta H_G}{\delta \pi^{ij}} \delta \pi^{ij} \right) d^3 x, \quad (3.1.6)$$

that we require. E is known as the ADM energy.

One may also obtain the linear and angular momentum vectors from the action by considering spatial translation invariance and rotation invariance of the action, respectively. These quantities are

$$P^k = -2 \int_{\partial \Sigma} \pi^{ik} d^2 S_i \quad (3.1.7)$$

and

$$L_i = 2 \int_{\partial \Sigma} \varepsilon_{ijk} \pi^{lj} x^k d^2 S_l. \quad (3.1.8)$$

It is interesting to note that when Arnowitt, Deser and Misner first derived this quantity [3] they did assume (incorrectly) that all of the boundary terms vanished. They fixed coordinates in the transverse traceless gauge and worked with a reduced Hamiltonian.

Recall from the last chapter that one has extra gauge freedom (in addition to the true dynamical degrees of freedom) in the GMD action so one can restrict the coordinate system accordingly as one wishes. After doing this, they found that if they waited until after fixing coordinates to discard the divergences, the whole Hamiltonian would have to be discarded. They tried to resolve this by assuming that the asymptotic boundary conditions were insufficient to make δE vanish when in the canonical coordinates of the transverse traceless gauge, which was false. For a more thorough account of this and the arguments for modifying the Hamiltonian system, see [28].

³Here the summation convention is extended to include *all* repeated indices regardless of their type.

3.2 The Stress-Energy Definition

The class of spacetimes that have a non-trivial definition of energy are the asymptotically flat spacetimes which represent isolated gravitating systems. We can also define another energy for these spacetimes with respect to a background spacetime. The natural spacetime to use as the background is of course Minkowski spacetime. This procedure involves defining an energy-momentum 4-vector for the whole spacetime via the linearization of the stress-energy tensor around the metric of the spacetime in question.

Let (M, g) and $(M, \hat{g})$ be two spacetimes. Also assume that $(M, \hat{g})$ satisfies the empty-space Einstein's equations

$$G(\hat{g}) = 0 \quad \text{or} \quad \hat{G}^{\alpha\beta} \equiv \hat{R}^{\alpha\beta} - \frac{1}{2}\hat{g}^{\alpha\beta}\hat{R} = 0. \quad (3.2.9)$$

The stress-energy $\hat{g}$ -tensor of (M, g) is defined as

$$t(g, \hat{g}) \equiv G'(\hat{g}) \cdot (g - \hat{g}) - G(g) \quad (3.2.10)$$

where

$$G'(\hat{g}) \equiv \left. \frac{dG}{dg} \right|_{\hat{g}}$$

and (M, g) is a solution of Einstein's equations with sources, i.e., $G(g) = T$ and $G^{\alpha\beta} \equiv R^{\alpha\beta} - \frac{1}{2}g^{\alpha\beta}R = T^{\alpha\beta}$ where T is the stress-energy tensor of (M, g) . So we can describe t as the difference between T and the linearization of T around $\hat{g}$.

We may rewrite this as

$$\begin{aligned} t + T &= G'(\hat{g})(g - \hat{g}) \\ &= \left(\lim_{g \rightarrow \hat{g}} \frac{G(g) - G(\hat{g})}{g - \hat{g}} \right) (g - \hat{g}). \end{aligned} \quad (3.2.11)$$

Since we have $\nabla \cdot G(g) \equiv 0$ and $G(\hat{g}) = 0$, then $\hat{\nabla} \cdot (t + T) = 0$ or $\hat{\nabla}_\alpha (t^{\alpha\beta} + T^{\alpha\beta}) = 0$.

We call this tensor, $\mathcal{T} \equiv t + T$, the stress-energy $\hat{g}$ -tensor of (M, g) . From above it has the conservation property $\hat{\nabla}_\alpha \mathcal{T}^{\alpha\beta} = 0$ and $\mathcal{T}^{\alpha\beta}$ is symmetric so it has all

the properties of a stress-energy tensor. Given a vector field $\mathbf{v}$ we can define the energy-momentum flux density of $\mathcal{T}^{\alpha\beta}$ along its flow (current density):

$$p_\alpha = \mathcal{T}_{\alpha\beta} v^\beta. \quad (3.2.12)$$

This will have the conservation law $\widehat{\nabla}_\alpha p^\alpha = 0$ of an energy-momentum 4-vector provided $\mathbf{v}$ is a $\widehat{\mathbf{g}}$ -Killing vector ($\widehat{\nabla}_\alpha v_\beta + \widehat{\nabla}_\beta v_\alpha = 0$).

The following integral must then vanish

$$\int_M \widehat{\nabla}_\alpha p^\alpha dM = \int_{\partial M} p^\alpha n_\alpha d(\partial M) = 0. \quad (3.2.13)$$

Now consider a subset D of M where $M = \Sigma \times \mathbb{R}$ and $D = \Sigma \times (0, t)$. D is a “world tube” of the whole (spatial) universe. The above integral becomes

$$\int_{\partial D} p^\alpha w_\alpha d(\partial D) = 0 \quad (3.2.14)$$

where $\mathbf{w}$ is the unit normal to ∂D . We can break the integral up into two integrals at the “ends” of the tube and one on the (hyper) surface of the tube. The outward normal on the “top” end ($\Sigma_t \equiv \Sigma \times \{t\}$) will be $\mathbf{n}$, on the “bottom” end ($\Sigma_0 \equiv \Sigma \times \{0\}$) will be $-\mathbf{n}$, and along the (hyper) surface ($\partial\Sigma \times (0, t)$) will be $\vec{\nu}$ (see Figure 3.2.1). Thus

$$\begin{aligned} \int_{\partial D} p^\alpha w_\alpha d(\partial D) &= \int_{\Sigma_t} p^\alpha n_\alpha d\Sigma_t + \int_{\Sigma_0} -p^\alpha n_\alpha d\Sigma_0 \\ &\quad + \int_0^t \int_{\partial\Sigma_t} p^i \nu_i d(\partial\Sigma_t) (N dt) \\ &= \int_{\Sigma_t} N p^0 d\Sigma_t - \int_{\Sigma_0} N p^0 d\Sigma_0 + \int_0^t \int_{\partial\Sigma_t} N p^i \nu_i d(\partial\Sigma_t) dt \\ &= 0. \end{aligned} \quad (3.2.15)$$

Now we differentiate both sides with respect to t to get,

$$\frac{d}{dt} \left(\int_{\Sigma_t} N p^0 d\Sigma_t \right) = \int_{\partial\Sigma_t} N p^i \nu_i d(\partial\Sigma_t), \quad (3.2.16)$$

Figure 3.2.1: The boundary of the world-tube, ∂D , consists of three sections; Σ_0 with outward normal $-n$, Σ_t with outward normal n , and $\partial\Sigma \times (0, t)$ with outward normal $\vec{v}$. The net energy-momentum flux passing through ∂D will be zero. We can extend this construction to all of M by letting $(0, t) \sim \mathbb{R}$.

which from Gauss's law (that equates the time derivative of energy to the flux of momentum passing through a surface) we conclude that

$$E = \int_{\Sigma_t} N p^0 d\Sigma_t = \int_{\Sigma_t} N p^\alpha n_\alpha d\Sigma_t . \quad (3.2.17)$$

Thus a general expression for the $\hat{g}$ -energy relative to $\mathbf{v}$ is

$$\begin{aligned} \mathbf{P} &= \int_{\Sigma} p^\alpha n_\alpha d\Sigma \quad \text{or} \\ \mathbf{P} &= \int_{\Sigma} \star \mathbf{p} \quad \text{where} \quad \star \mathbf{p} = \frac{1}{3!} \varepsilon_{\alpha\beta\gamma\delta} p^\alpha \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma \wedge \mathbf{d}x^\delta \end{aligned} \quad (3.2.18)$$

is the dual to $\mathbf{p}$. Recall that $\nabla_\alpha p^\alpha = 0$ which implies that $\star \mathbf{p}$ is an exact differential form, i.e.,

$$\begin{aligned} d(\star \mathbf{p}) &= \frac{1}{3!} \varepsilon_{\alpha\beta\gamma\delta} p^\alpha{}_{;\sigma} \mathbf{d}x^\sigma \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma \wedge \mathbf{d}x^\delta \\ &= \frac{1}{3!} \varepsilon_{\alpha\beta\gamma\delta} \cancel{p^\alpha{}_{;\alpha}} \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \wedge \mathbf{d}x^\gamma \wedge \mathbf{d}x^\delta \\ &= 0 . \end{aligned} \quad (3.2.19)$$

This means that locally there exists a two-form $\star \mathbf{q}$ such that $\star \mathbf{p} = d(\star \mathbf{q})$. If we can show that such a $\star \mathbf{q}$ exists globally then

$$\begin{aligned} \mathbf{P} &= \int_{\Sigma} \star \mathbf{p} = \int_{\partial \Sigma} \star \mathbf{q} \\ &= \int_{\partial \Sigma} \frac{1}{4} \varepsilon_{\alpha\beta\gamma\delta} q^{\gamma\delta} \mathbf{d}x^\alpha \wedge \mathbf{d}x^\beta \\ &= \int_{\partial \Sigma} q^{0i} dS_i \end{aligned} \quad (3.2.20)$$

where $dS_i = 1/2 \hat{g}^{1/2} \varepsilon_{ijk} \mathbf{d}x^j \wedge \mathbf{d}x^k$. This last equality holds because in each Σ_t , $x^0 = \text{constant}$.

It can be shown [10] that such a $\star \mathbf{q}$ exists and

$$q^{\alpha\beta} = (\widehat{\nabla}_\gamma K^{\alpha\beta\delta\gamma}) v_\delta - K^{\alpha\gamma\delta\beta} \widehat{\nabla}_\gamma v_\delta , \quad (3.2.21)$$

where

$$\begin{aligned} K^{\alpha\beta\gamma\delta} &= \frac{1}{2}(\hat{g}^{\alpha\delta}H^{\beta\gamma} + \hat{g}^{\beta\delta}H^{\alpha\gamma} - \hat{g}^{\alpha\gamma}H^{\beta\delta} - \hat{g}^{\beta\gamma}H^{\alpha\delta}), \\ H^{\alpha\beta} &= h^{\alpha\beta} - \frac{1}{2}\hat{g}^{\alpha\beta}\text{tr}(\mathbf{h}), \text{ and} \\ h^{\alpha\beta} &= g^{\alpha\beta} - \hat{g}^{\alpha\beta}. \end{aligned} \quad (3.2.22)$$

3.2.1 Backgrounds that are Minkowskian at Infinity

A Minkowski manifold is a semi-Riemannian 4-manifold with metric

$$\eta_{\alpha\beta} = \text{diag}\{-1, 1, 1, 1\}.$$

This is the manifold that represents the flat spacetime of special relativity. This will be the background space with which the energy of certain other non-flat spacetimes will be defined.

Previously we discussed the fact that there is no notion of energy density associated with the gravitational field so there can therefore be no notion of local energy. Thus we would expect that for a definition of total energy, one would have to appeal to global considerations.

Also any definition of total energy in closed spacetimes (in the sense that there exists a slicing of spacetime where $M = \Sigma \times \mathbb{R}$ and Σ is a Cauchy surface that is compact — see Wald [35]) becomes trivial as its numerical value will be zero due to vanishing boundary terms of the Hamiltonian. In the case of asymptotically flat spacetimes, however, the boundary conditions allow the numerical value of the surface integral in the Hamiltonian to be expressed as a limit with respect to space-like infinity. This quantity characterizes the deviation of the spacetime in question from Minkowski spacetime providing a definition of energy with respect to the flat background.

Let $C \subset M$ be a compact set where $M = \mathbb{R}^3 \times \mathbb{R}$. If ${}^{(4)}g$, the metric on M , approaches η on $(\mathbb{R}^3 \setminus C)$, then outside of C , M will have the character of flat space.

In general $M = \Sigma \times \mathbb{R}$ will not be diffeomorphic to $\mathbb{R}^4$, so we make a more formal definition of a spacetime that is "Minkowskian at infinity".

Definition 3.2.1 Let $(M, \eta) = (\Sigma \times \mathbb{R}, \hat{\eta})$ with Σ a Cauchy surface. If $\exists$ compact $C \subset M$ such that $\Sigma \setminus C$ is the disjoint union of a finite number of open sets $\{U_\alpha\}_{\alpha=1}^n$ and $\exists$ a diffeomorphism $\varphi_\alpha : U_\alpha \rightarrow \mathbb{R}^3$ such that $\varphi_\alpha(U_\alpha) = \mathbb{R}^3 \setminus B_\alpha$ where B_α is a closed ball of $\mathbb{R}^3$, then $(M, \hat{\eta})$ is said to be **Minkowskian at infinity** if in each $U_\alpha \times \mathbb{R}$, $\hat{\eta} = \eta$ in the φ_α coordinates. The U_α 's are called the "ends" of Σ .

The motivation for this is that we want to be able to appeal to a manifold structure for the asymptotic part of M . For an isolated system, by definition, we can enclose all the sources inside a compact set. As we get further outside this set, the space begins to resemble flat space.

We cannot assume that a single coordinate system exists to describe the exterior region, however it will have a manifold structure. Thus the ends, $(U_\alpha, \varphi_\alpha)$, just correspond to the components of $\Sigma \setminus C$ since in general $\Sigma \setminus C$ will not be connected.

Definition 3.2.2 A spacetime $(M, {}^{(4)}g)$ is **asymptotically Minkowskian** if M admits a metric $\hat{\eta}$ such that $(M, \hat{\eta})$ is Minkowskian at infinity and ${}^{(4)}g - \hat{\eta}$ approaches zero at infinity.

To obtain the ADM energy-momentum we define the energy-momentum around $\hat{\eta}$ which is Minkowskian at infinity. The $\hat{\eta}$ -energy-momentum, for some $\hat{\eta}$ -Killing vector $\boldsymbol{v}$ is

$$\boldsymbol{P}(\boldsymbol{v}) = \sum_{A=1}^n \int_{\Omega_A} q^{0i} dS_i \quad (3.2.23)$$

which is a sum of the energies in each end U_A . The integrals are over Ω_A which is the limit as the radius tends to infinity of the two-spheres in U_A . The vector $\boldsymbol{v}$ is defined on M but will have different coordinates in each $U_A \times \mathbb{R}$, say v_A^i which will be constant for each A (i.e., Killing vector in flat space). We may now write the

$\hat{\eta}$ -energy-momentum as

$$\begin{aligned} \mathbf{P}(\mathbf{r}) &= \sum_{A=1}^n P_A^{\alpha} (r_A)_{,\alpha} \\ &= \sum_{A=1}^n (r_A)_{,\alpha} \int_{\Omega_A} (-\eta^{\alpha k} H^{\alpha 0}_{,k} + \eta^{\alpha k} H^{\alpha 0}_{,k} \\ &\quad + \eta^{\alpha 0} H^{\alpha k}_{,k} - \eta^{\alpha 0} H^{\alpha k}_{,k}) dS_k \end{aligned} \quad (3.2.24)$$

which after some calculation will show [10]

$$P_A^0 \equiv \int_{\Omega_A} H^{lk}_{,k} dS_k = - \int_{\Omega_A} (g_{lk,l} - g_{lk,l}) dS_k \quad (3.2.25)$$

and

$$\begin{aligned} P_A^l &= \int_{\Omega_A} (-H^{l0}_{,k} + \delta^l_k H^{l0}_{,k} - H^{lk}_{,0}) dS_k \\ &= -2 \int_{\Omega_A} (\delta^{lj} K_{,j} - K^{lj}) dS_l \\ &= -2 \int_{\Omega_A} \pi^{lj} dS_l. \end{aligned} \quad (3.2.26)$$

It is interesting to note that the reason we could not define the energy as a flux integral was the lack of a "background" space, and that in the Hamiltonian formulation a non-trivial notion of energy only arises for those spacetimes that we can represent as asymptotic to a background. When we define a new stress energy tensor which is the difference between the usual stress energy tensor and the linearization of this tensor around the flat metric, we are effectively providing a local energy density. This should not be surprising then that it was only these spacetimes that had a non-trivial definition of energy and that it was equal to the flux integral definition of energy with respect to this new stress energy tensor coincides with the Hamiltonian definition.

These asymptotically flat spacetimes will correspond to isolated gravitating systems which will behave like Schwarzschild spacetimes. The Schwarzschild metric (in

isotropic coordinates) defined by

$$g_{ij} = \left(1 + \frac{m}{2r}\right) \delta_{ij} \text{ , } g_{0i} = g_{i0} = 0, \text{ and } g_{00} = \left(\frac{1 - m/2r}{1 + m/2r}\right) \text{ for } r > R > 2m$$

has mass $m = \frac{1}{16\pi} P^0$ and $P^k = 0$. Thus the ADM mass for this case is

$$m = \frac{1}{16\pi} \|\mathbf{P}\| = \frac{1}{16\pi} |P^\alpha P_\alpha|^{1/2} \text{ .}$$

Chapter 4

The Positivity of Mass

In the previous chapter we defined the mass-energy of spacetimes that asymptotically approach Minkowski space. Such spacetimes correspond to isolated gravitating systems whose metric becomes Minkowskian at spacelike infinity. Their mass-energy, which will from now on be referred to simply as mass, corresponds to the Schwarzschild mass, m (modulo the 16π scaling factor that maintains consistency of units). It is an important property that the energy of physical systems be non-negative. This allows for the existence of a stable (absolute) minimum energy solution. In this case the $m = 0$ (Minkowski spacetime) would then be the minimum energy solution if the mass must be non-negative. The positive mass conjecture (PMC) asserts that given certain physically reasonable assumptions, the ADM mass-energy is positive except for Minkowski spacetime.

First we consider the local case where the class of spacetimes in the conjecture is restricted to a weighted Sobolev neighbourhood of Minkowski space. We outline the main ideas of the proof of Choquet-Bruhat and Marsden. This approach was first suggested by Brill and Deser [8].

Although these proofs contain many sophisticated mathematical concepts, the main idea is nothing more than the second derivative test for functionals.

Next we take a somewhat closer look at the results from Schoen and Yau's proof of the global version for spacetimes admitting maximal slices [31]. Then we briefly

sketch the extensions of this proof that lead to the full conjecture.

4.1 The Local Version of the Positive Mass Conjecture

The first proof of the positive mass conjecture for general spacetimes was provided by Choquet-Bruhat and Marsden [11]. They showed the conjecture holds for spacetimes in a neighbourhood of Minkowski space in a functional sense to be defined. More specifically they show that there exists only one critical point for $m(g)$ which is zero for flat space and that the second variation of m is positive.

Most of the arguments used by Choquet-Bruhat and Marsden are functional analytic in nature. They restrict the initial data to certain classes of function spaces, that allow them to appeal to existing theorems for the appropriate estimates. Because of this, we just briefly describe the setting for their local theorem and state the main results.

Before we outline the arguments of Choquet-Bruhat and Marsden we have to define what it means for a spacetime to be asymptotically Minkowskian.

4.1.1 Function Spaces

We will require two different types of function spaces. Both of these are tensor spaces for which there is appropriate fall off at spacelike infinity for the tensor and its derivatives. This is needed to characterize a notion of flatness for a definition of an asymptotically Minkowskian spacetime. Let C^s_δ consist of tensor fields, $\mathbf{T}$, on Σ which are C^s and satisfy

$$\|\mathbf{T}\|_{C^s_\delta} \equiv \sum_{k=0}^s \sup_{x \in \Sigma} \sigma(x)^{\delta+k} |\partial^k \mathbf{T}| < \infty. \quad (4.1.1)$$

Let H_δ^s consist of tensor fields, $\mathbf{T}$, on Σ which are C^∞ and satisfy

$$\|\mathbf{T}\|_{H_\delta^s} \equiv \left(\int_\Sigma \sum_{k=0}^s \sigma^{2(s+k)} |\partial^k \mathbf{T}|^2 \right)^{1/2} < \infty. \quad (4.1.2)$$

The function $\sigma(x) \equiv \sqrt{1+r^2}$ (where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$) characterizes the distance to a given point $x \in \Sigma$ outside some compact set from some fixed point inside. Recall that the space we are measuring these tensor fields on is essentially $\mathbb{R}^3$ minus a compact set. Thus σ is the natural weight function to use. The “1” under the square root sign incorporates the fact that we start measuring r outside this compact set. The fact that H_δ^s is a Hilbert space (of the Sobolev type) is exploited by Choquet-Bruhat and Marsden to define a neighbourhood of Minkowski spacetime for which the positive mass conjecture holds.

4.1.2 Asymptotically Minkowskian Manifolds

There are two definitions of ${}^{(4)}\mathbf{g} - \hat{\mathbf{e}}$ “approaching infinity”, one classical and the other restricts $\mathbf{g}$ and $\mathbf{K}$ to certain function spaces. An initial data set $(\Sigma, \mathbf{g}, \mathbf{K})$ is said to be *asymptotically Euclidean* in the classical sense if;

$$\mathbf{g} - \hat{\mathbf{e}} = O(r^{-1}), \quad \partial^k \mathbf{g} = O(r^{-(1+k)})$$

$$\mathbf{K} = O(r^{-2}), \quad \text{and} \quad \partial^k \mathbf{K} = O(r^{-(2+k)})$$

and in the functional sense if,

$$\mathbf{g} - \hat{\mathbf{e}} \in H_\delta^s \quad \text{and} \quad \mathbf{K} \in H_{\delta+1}^{s-1}$$

for some δ such that $-3/2 < \delta < -1/2$ and $s \geq 2$. In the latter case one says that $(\Sigma, \mathbf{g}, \mathbf{K})$ is H_δ^s -asymptotically Euclidean. In the first case we say that $(\Sigma, \mathbf{g}, \mathbf{K})$ is *asymptotically flat* and that $(M, {}^{(4)}\mathbf{g})$ is asymptotically Minkowskian in the classical sense.

We need a slightly more precise definition of asymptotically Minkowskian in the functional sense. This is because the arguments leading to the local version of the

positive mass conjecture require estimates that depend upon theorems on specific function spaces.

Let I be a closed interval of $\mathbb{R}$ and let ${}^{(4)}\mathbf{g}$ be defined on $\Sigma \times I$. Denote by $C^k(I, H_\delta^s)$ the space of tensor fields $\mathbf{T}_t$ on Σ depending on $t \in I$ for which $\mathbf{T}_t \in H_\delta^s$ and the mapping $t \mapsto \mathbf{T}_t$ is C^k . Now we define another space E_δ^s by

$$E_\delta^s \equiv \bigcap_{0 \leq k \leq s} C^k(I, H_{\delta+k}^{s-k})$$

with norm

$$\|\mathbf{T}\|_{E_\delta^s} = \sup_{t \in I} \sum_{k=0}^s \left\| \frac{\partial^k \mathbf{T}}{\partial t^k} \right\|_{H_\delta^k},$$

where the norm $\|\cdot\|_{H_\delta^k}$ is as defined in Section 4.1.1.

Now let $\hat{\mathbf{e}} = \delta_{ij} \mathbf{d}x^i \otimes \mathbf{d}x^j$ be the Euclidean metric in the φ_a coordinates on $\Sigma \times I$. We say that $(\Sigma \times I, {}^{(4)}\mathbf{g})$ is E_δ^s -asymptotically Minkowskian if

$$\mathbf{g} - \hat{\mathbf{e}} \in E_\delta^s, \quad N - 1 \in E_\delta^s, \quad \text{and} \quad \vec{N} \in E_\delta^s$$

where N is the lapse function and $\vec{N}$ is the shift vector defined in Chapter 2. By requiring the lapse and shift functions to belong to these spaces we are restricting the gauge so that only spacetimes with behaviour sufficiently close to Minkowski spacetime remain. Also ${}^{(4)}\mathbf{g}$ is said to be H_δ^s -asymptotically Minkowskian on Σ_t if

$$\frac{\partial^k (\mathbf{g} - \hat{\mathbf{e}})}{\partial t^k} \in H_\delta^s, \quad \frac{\partial^k (N - 1)}{\partial t^k} \in H_\delta^s, \quad \text{and} \quad \frac{\partial^k \vec{N}}{\partial t^k} \in H_\delta^s \quad \text{at time } t.$$

4.1.3 The Proof of the Local Version of the Positive mass Conjecture

The result of Choquet-Bruhat and Marsden relies on the existence of maximal slices of spacetime. A maximal slice is a Cauchy surface Σ such that its volume is a maximum. (Maximum because Σ is spacelike, it would be a minimum if M was a proper Riemannian manifold.) This can be shown to be equivalent to the requirement that $\text{tr}(\mathbf{K}) = 0$.

They use a result of O'Murchadha and York [10] that shows that, for any asymptotically Minkowskian spacetime, $(M, {}^{(4)}g)$, that satisfies the weak energy condition and admits a maximal slice, there exists an empty, $(\bar{R}(\bar{g}) \equiv 0)$, time-symmetric ($K \equiv 0$) spacetime, $(M, {}^{(4)}\bar{g})$, of equal or lesser mass. Thus they need only show that for all time-symmetric spacetimes (in a neighbourhood of Minkowski spacetime), the mass is positive.

First Choquet-Bruhat and Marsden proved the following theorem:

Theorem 4.1.1 *Let $(M, {}^{(4)}\tilde{g})$ be a space-time satisfying the strong energy condition, asymptotically Minkowskian in the sense that $M = \Sigma \times \mathbb{R}$, $(\Sigma, \hat{e})$ is Euclidean at infinity, and ${}^{(4)}\tilde{g} - \hat{\eta} \in E_k^2$ for some δ such that $-3/2 < \delta < -1/2$, which admits a maximal slice Σ . Then there exists a number $c > 0$ such that each neighbouring spacetime ${}^{(4)}g$, such that $\|{}^{(4)}\tilde{g} - {}^{(4)}g\|_{E_k^2} < c$, admits a maximal slice.*

A corollary of this theorem is that there exists a neighbourhood of Minkowski spacetime such that each of its elements admits a maximal slice.

Now all that is left to show is that the only solution that allows for the first variation of mass to vanish is the Minkowski solution and that the second variation of mass is positive when the variations are restricted to time-symmetric spacetimes.

The mass functional in this setting is,

$$\begin{aligned} \bar{m}(g, K) &\equiv \frac{H}{16\pi} = \frac{1}{16\pi} \int_{S_t} (g_{ij,i} - g_{ji,i}) dS_t + \frac{1}{16\pi} \int_{\Sigma} N \mathcal{H} + N_i \mathcal{H}^i d\Sigma \\ &= \frac{1}{16\pi} \left[\int_{\Sigma} (g_{ij,i} - g_{ji,i}) d\Sigma(\hat{e}) + \int_{\Sigma} R + K^{ij} K_{ij} + K^2 d\Sigma(g) \right], \end{aligned}$$

since $N = 1$, $\tilde{N}$, and $g - \hat{e}$ belong to H_{-1}^2 . Also, since we are only assuming the strong energy condition, $\mathcal{H} \neq 0$ in general. For the time-symmetric case we get

$$\bar{m}(g, 0) = \frac{1}{16\pi} \left[\int_{\Sigma} (g_{ij,i} - g_{ji,i}) d\Sigma(\hat{e}) + \int_{\Sigma} R d\Sigma(g) \right]$$

for the mass. The variation of this with respect to $h \in H_{-1}^2$ is

$$\delta \bar{m}(g, 0) h = \frac{1}{16\pi} \int_{\Sigma} R_{ij} h^{ij} d\Sigma(g).$$

For this to vanish for all such $\mathbf{h}$, R_{ij} must be zero.

The second variation may be reduced to

$$\delta^2 \bar{m}(\mathbf{g}, 0)(\mathbf{h}, \mathbf{h}) = \frac{1}{16\pi} \int_{\Sigma} \frac{1}{2} h^i{}_{|k} h_{i,j}{}^{|k} d\Sigma(\mathbf{g})$$

when we reduce to the transverse-traceless gauge ($\text{tr}(\mathbf{h}) = 0$ and $D \cdot \mathbf{h} = 0$) and assume $R_{ij} = 0$. This may be shown to be a positive quadratic form on $\mathbf{h}$. Hence $\bar{m}(\mathbf{g}, 0)$ has only one extremum which is a minimum and occurs when $R_{ij} = 0$. For a 3-manifold, $R_{ij} = 0$ implies flatness.

The above arguments show that there exists an E_δ^2 -neighbourhood ($-3/2 < \delta < -1/2$) of Minkowski spacetime such that all the spacetimes in this neighbourhood have positive mass except for Minkowski spacetime, which has zero mass.

Choquet-Bruhat and Marsden suggested possible approaches for solving the global version of the PMC with ideas of a similar flavour, but it was Schoen and Yau who solved the problem using geometric methods.

4.2 The Global Version of Positive Mass Conjecture

In their first version of the theorem Schoen and Yau restrict themselves to spacetimes $(M, \mathbf{g})$ where in U_α ,

$${}^{(3)}g_{ij} = \left(1 + \frac{m_\alpha}{2r}\right)^4 \delta_{ij} + h_{ij}, \text{ with } h_{ij} = O(r^{-2}), \text{ and} \quad (4.2.3)$$

$(M, \mathbf{g})$ admits a maximal slice and satisfies the weak energy condition. The metric $\mathbf{g}$ is said to be asymptotically flat if each boundary component of Σ has positive mean curvature (with respect to the outward normal), and on each U_A there exists a coordinate system where $\mathbf{g}$ is as above with

$$|h_{ij}| \leq \frac{k_1}{1+r^2}, |\partial h_{ij}| \leq \frac{k_2}{1+r^3}, |\partial^2 h_{ij}| \leq \frac{k_3}{1+r^3},$$

$$K_{ij} = O(r^{-2}) \text{ and } \partial K_{ij} = O(r^{-3}).$$

Schoen and Yau's configuration of the positive mass conjecture is:

Theorem 4.2.1 *Let (Σ, g) be an oriented 3-manifold having an asymptotically Euclidean metric so that in each end U_α the metric is*

$$g_{ij} = \left(1 + \frac{m_\alpha}{2r}\right)^4 \delta_{ij} + h_{ij}.$$

- (1) *If $\mathbf{h}$ is as above (or equivalently $\mathbf{h} \in C^2_2$) then each m_α is non-negative.*
- (2) *If in addition $\mathbf{h} \in C^5$, $\sigma^5(|\partial^3 \mathbf{h}| + |\partial^4 \mathbf{h}| + |\partial^5 \mathbf{h}|) < \infty$, and $\exists \alpha$ such that $m_\alpha < 0$, then (Σ, g) is isometric to $(\mathbb{R}^3, C)$.*

To prove this they assume that for some α , say $\alpha = A$, in U_A the mass, m_A , is negative. Let U_A be diffeomorphic to $\mathbb{R} \setminus B_{\sigma_0}(0)$ via φ_A in the coordinates $\{x^1, x^2, x^3\}$, where $B_\sigma(0) = \{x : |x| < \sigma\}$. There are three basic steps to arrive at a contradiction.

First of all one must show the existence of a complete area-minimizing surface S embedded in Σ in such a way that $S \cap \{\Sigma \setminus U_A\}$ is compact and $S \cap U_A$ is contained in a region between two parallel 2-planes of $\mathbb{R}^3$. This surface is the limit of surfaces S_σ , with least area of the oriented surfaces with boundary $C_\sigma \equiv \{\text{the circle of radius } \sigma \text{ and center } O \text{ in the } x^3 = 0 \text{ plane}\}$. Schoen and Yau show that $\exists$ a compact set $K_A \subseteq \Sigma$ such that

$$S_\sigma \cap (\Sigma \setminus U_A) \subseteq K_A \quad \forall \sigma > 2\sigma_0, \quad (4.2.4)$$

which means that S_σ cannot run to infinity in any end other than U_A .

The second step is to show that, for $\mathbf{g}$ asymptotically flat on Σ such that $R \geq 0$ and $m_A < 0$, there exists an asymptotically flat metric $\tilde{\mathbf{g}} = \phi^4 \mathbf{g}$ conformally equivalent to $\mathbf{g}$ with $\tilde{R} \geq 0$ on Σ and $\tilde{R} > 0$ outside a compact subset of U_A having a mass $\tilde{m}_A < 0$ for U_A .

The last step involves showing that first,

$$\int_S \kappa_s ds > 0 \quad \text{and then} \quad \int_S \kappa_s ds \leq 0, \quad (4.2.5)$$

(where κ_g is the Gaussian curvature of S), precluding the existence of S .

The proof of the second inequality is a consequence of the Gauss-Bonnet theorem which relates topological and geometric properties of 2-surfaces.

The non-existence of S contradicts the fact that $m_A < 0$ proving that $m \geq 0$. Finally they show that $m = 0$ implies that g is the flat metric.

4.2.1 The Existence of an Area Minimizing Surface

The authors appeal to some important results from geometric measure theory for the existence of the area minimizing surfaces S_σ . They make use of the generalization of surfaces to rectifiable currents to allow for the possibility of singularities.

A rectifiable set is basically a generalized manifold that allows for the mappings in charts to be Lipschitz instead of smooth. Also the Hausdorff measure [14], $\mathcal{H}^m$, is used on the sets instead of that of Lebesgue, to extend the class of possible candidates from the m -Borel sets to all m -dimensional subsets of $\mathbb{R}^n$. A $(\mathcal{H}^m, m)$ -rectifiable set, S , is such that $\mathcal{H}^m(S) < \infty$ and there exists a countable Lipschitz chart to S from bounded subsets of $\mathbb{R}^m$. A current is defined as the dual to a smooth differential. In particular a rectifiable current is a linear functional of a smooth differential form on a rectifiable oriented set of finite area, integer multiplicity, and compact support;

$$S(\varphi) = \int_S \varphi . \quad (4.2.6)$$

Considering sets as functionals in this way induces a topology, dual to the topology of differential forms, that provides useful properties of these sets, not the least of which being a "boundary" operator, ∂ , from m -dimensional rectifiable currents to $(m - 1)$ -dimensional currents such that

$$(\partial S)(\varphi) = S(d\varphi) . \quad (4.2.7)$$

With this we can obtain a generalized Stokes theorem [14], with a normal vector field defined $\mathcal{H}^m$ -a.e. on S .

Section 5.1.6 of [14] shows the existence of an area minimizing current, S_σ , with boundary C_σ . Such currents turn out to be smooth manifolds for the case of $m \leq 6$ [24], so there will be no singularities in the interior of $S_\sigma \cap \Sigma$.

Theorem 4.2.2 *There exists a complete area minimizing (with respect to g) surface S properly embedded in Σ so that $S \cap (\Sigma \setminus U_A)$ is compact, and $S \cap U_A$ lies between two parallel Euclidean planes (in the x^3 coordinates).*

Sketch of Proof.

S will be constructed out of a family of surfaces $\{S_\sigma\}$ each area minimizing among oriented surfaces with boundary $C_\sigma = \{\text{circle of center } O \text{ in the } x^3 = 0 \text{ plane of } \mathbb{R}^3, \text{ and radius } \sigma \text{ large enough that } C_\sigma \subset U_A\}$. The existence of such surfaces was just shown above. From this family we construct the sequence $\{S_{\sigma_i}\}$ such that $S_{\sigma_i} \rightarrow S$ as $\sigma_i \rightarrow \infty$.

Now we show that there exists a compact set $K_A \subseteq \Sigma$ such that

$$S_\sigma \cap (\Sigma \setminus U_A) \subseteq K_A \quad \forall \sigma > 2\sigma_0. \quad (4.2.8)$$

This means that S_σ cannot run to infinity in any other end than U_A . Let $\{y^i\}$ be a coordinate system on some other end U_B that is diffeomorphic to $\mathbb{R}^3 \setminus B_{\tau_0}(O)$ where $B_\rho(O) = \{y : |y| < \rho\}$. The Hessian of $|y|^2$, $D_{ij}|y|^2$, under the asymptotic boundary conditions, can be shown to be $D_{ij}|y|^2 = 2\delta_{ij} + O(1/|y|)$ for large $|y|$. Thus $|y|^2$ is convex for sufficiently large τ such that $|y| \geq \tau > \tau_0$. Because $C_\sigma \subset U_A$ the maximum principle implies that $\sup_{x \in (S_\sigma \cap U_B)} \{|y|\} \in C_\sigma$ and hence $S_\sigma \cap U_B \subset B_\tau(O)$.

Now we let $E_h = \{x \in \mathbb{R}^3 : |x^3| \leq h\}$ and show that there exists a number $h > \sigma_0$ so that

$$S_\sigma \cap U_A \subseteq E_h \quad \forall \sigma > 2\sigma_0. \quad (4.2.9)$$

To accomplish this, the maximum principle is used on x^3 restricted to $S_\sigma \cap U_A$. Let h_{max} be the maximum of x^3 (the third coordinate) on $S_\sigma \cap U_A$. If $h_{max} < \sigma_0$ we

have the result. Assume $h_{max} > \sigma_0$ and that it occurs at $x_0 \in S_\sigma$. Let $\{v_1, v_2\}$ be a basis for the tangent space to S_σ at x_0 , then

$$\nabla_{ij}x^3 = D_{ij}x^3 + h_{ij}\nu(x^3), \quad (4.2.10)$$

where $\vec{\nu}$ is the normal vector field to S_σ and ∇ is the associated connection for the metric $\mathbf{g}$ on S_σ induced from $\mathbf{g}$. One can show that

$$D_{ij}x^3 = -\Gamma_{ij}^3 = \frac{m_A x^j}{r^3} \delta_{i3} + \frac{m_A x^i}{r^3} \delta_{j3} - \frac{m_A x^3}{r^3} \delta_{ij} + O(1/r^3) \quad (4.2.11)$$

and since S_σ is minimal, $q^{ij}h_{ij}\nu(x^3) = 0$. Then

$$q^{ij}\nabla_{ij}x^3 = \frac{-2m_A h_{max}}{r^3} + O(1/r^3) > 0 \quad (4.2.12)$$

at x_0 for h_{max} sufficiently large when $m_A < 0$. This is contrary to there being a maximum at x_0 . Thus $h_{max} \leq \sigma_0$ and we can form a similar argument to obtain a lower bound for x^3 on $S_\sigma \cap U_A$ showing that $S_\sigma \cap U_A$ lies between the planes $x^3 = h_{max}$ and $x^3 = h_{min}$.

Now define the set,

$$W_\rho \equiv (\Sigma \setminus U_A) \cup \{x : |x| \geq \sigma_0, (x^1)^2 + (x^2)^2 \leq \rho^2\}, \quad (4.2.13)$$

for $\rho > 2\sigma_0$. Equations (4.2.8) and (4.2.9) imply that,

$$S_\sigma \cap W_\rho \subseteq (K_A \cup E_h) \cap W_\rho, \quad (4.2.14)$$

for any $\sigma > \rho$.

Schoen and Yau then appeal to a result from Federer [14] that shows that the intersection of S_σ with a geodesic ball is the graph of a C^3 function over the tangent plane to S_σ . Furthermore this function has a regularity estimate that bounds the derivatives of this function up to order three. But first one must show that there are no singularities in the interior of $S_\sigma \cap \Sigma$.

To show that S_σ is regular (more precisely that the quotient $S_\sigma/\mathbb{Z}_2$ is a regular surface), Schoen and Yau prove that $S_\sigma \cap \partial\Sigma$ is empty. Let B be a component of $\partial\Sigma$. Define

$$\mathcal{O}_t \equiv \{x \in \Sigma : \text{dist}(x, B) < t\}$$

and

$$B_t \equiv \{x \in \Sigma : \text{dist}(x, B) = t\} = \partial\mathcal{O}_t.$$

The B_t 's are smooth surfaces diffeomorphic to $B = B_0$. Also $\mathcal{O}_t$ is smoothly retractable to $B = B_0$, for all $0 < t < \varepsilon$.

Now let $\vec{\nu}$ be the outward unit normal vector field to B . This field may be extended to $\mathcal{O}_t$ by parallel translation along geodesics normal to B . Thus this is also a normal vector field to B_t for $0 < t < \varepsilon$.

The positive mean curvature of Σ implies that $D_i \nu^i > 0$ on B , and thus by continuity there exists a $t \in (0, \varepsilon)$ such that $D_i \nu^i > 0$ on $\mathcal{O}_t$. There exists an open set $V_\sigma \subset \Sigma$ such that,

$$S_\sigma = \partial V_\sigma,$$

where " ∂ " is the boundary of geometric measure theory. We may now apply the generalized divergence theorem when $\mathcal{O}_t \cap V_\sigma$ is non empty,

$$\begin{aligned} 0 &< \int_{\mathcal{O}_t \cap V_\sigma} D_i \nu^i d\mathcal{L}^3 \\ &= \int_{\partial(\mathcal{O}_t \cap V_\sigma)} \nu^i n_i d\mathcal{H}^2 \\ &= \int_{(\partial\mathcal{O}_t \cap V_\sigma) \cup (\mathcal{O}_t \cap \partial V_\sigma)} \nu^i n_i d\mathcal{H}^2 \\ &= \int_{\mathcal{O}_t \cap S_\sigma} \langle \vec{\nu}, \boldsymbol{\eta} \rangle d\mathcal{H}^2 + \int_{B_t \cap V_\sigma} \nu^i \nu_i d\mathcal{H}^2 \\ &= \int_{\mathcal{O}_t \cap S_\sigma} \langle \vec{\nu}, \boldsymbol{\eta} \rangle d\mathcal{H}^2 - \int_{B_t \cap V_\sigma} d\mathcal{H}^2, \end{aligned} \tag{4.2.15}$$

where $\mathcal{H}^2$ is the Hausdorff 2-dimensional measure and $\mathcal{L}^3$ is the Lebesgue 3-dimensional measure. Also $\vec{n}$ is the outward normal to $\mathcal{O}_t \cap V_\sigma$, and $\vec{\eta}$ is the outward normal to V_σ on S_σ . The negative in the last line comes from the fact that even though $\vec{\nu}$ is the outward normal with respect to Σ on B , it is inward with respect to $\mathcal{O}_t$ on B_t . Because both $\vec{\nu}$ and $\vec{\eta}$ are spacelike unit vectors, $0 \leq \langle \vec{\nu}, \vec{\eta} \rangle \leq 1$. This implies that $\text{Area}(V_\sigma \cap B_t) < \text{Area}(S_\sigma \cap \mathcal{O}_t)$. This is a contradiction to the fact that S_σ is area minimizing. Thus $\mathcal{O}_t \cap V_\sigma = \emptyset$ and hence $S_\sigma \cap \partial\Sigma = \emptyset$.

Thus with the regularity estimate above and (4.2.14), there exists a sequence $\sigma^{(r)} \rightarrow \infty$, such that $S_{\sigma^{(r)}} \cap W_\rho$ converges in the C^2 topology. We may also take a sequence $\rho_k \rightarrow \infty$ such that $S_{\sigma_k^{(r)}} \cap W_{\rho_k}$ converges for each k , as long as $\rho > 2\sigma_0$. Now we extract a diagonal sequence which is such that $S_{\sigma_k^{(r)}} \rightarrow S$ uniformly in the C^2 norm on compact sets of Σ . Equations (4.2.8) and (4.2.9) give us that $S \cap (\Sigma \setminus U_A)$ is compact and that $S \cap \Sigma_A \subseteq E_h$, respectively.

4.2.2 The Existence of a Class of Metrics Conformal g

The existence of the appropriate class of metrics necessary to prove the conjecture is summarized in the following theorem.

Theorem 4.2.3 *If g is an asymptotically flat metric on Σ with $R \geq 0$ and $m_A < 0$ then there exists another metric $\tilde{g} = \phi^4 g$ which is conformally equivalent to g with $\tilde{R} \geq 0$ on Σ and $\tilde{R} > 0$ outside a compact subset of U_A with mass $\tilde{m}_A < 0$ for U_A .*

Proof:

The g -Laplacian of $1/r$ on U_A is

$$\begin{aligned} \Delta(1/r) &= \sum_{i=1}^3 \frac{\partial}{\partial x^i} \left[\left(1 + \frac{m_A}{2r} \right)^2 \frac{\partial}{\partial x^i} \left(\frac{1}{r} \right) \right] + O(1/r^5) \\ &= \left(1 + \frac{m_A}{2r} \right) \frac{m_A}{r^4} + O(1/r^5) \\ &= \frac{m_A}{r^4} + O(1/r^5) \end{aligned} \tag{4.2.16}$$

so that for $\sigma > \sigma_0$, $\Delta(1/r) < 0$ for $r \geq \sigma$.

Now we construct the function ϕ such that $\Delta\phi < 0$. Let

$$\phi(x) = \begin{cases} 1 + 3t_0/2 & \text{on } \Sigma \setminus U_A \\ 1 + \zeta(-m_A/4r) & \text{on } U_A \end{cases} \quad (4.2.17)$$

where $r = |x|$, $\zeta(t) = \begin{cases} t, & t \leq t_0 \\ 3t_0/2, & t \geq 2t_0 \end{cases}$ is C^∞ with $\zeta'(t) \geq 0$ and $\zeta''(t) < 0$ on $(0, \infty)$, and $t_0 = -\frac{m_A}{8\sigma_0}$. Thus from the estimate for $\Delta(1/r)$ we have $\Delta\phi < 0$ on Σ and $\Delta\phi < 0$ for $r > 2\sigma$. The conformal metric $\hat{g} = \phi^4 g$ is asymptotically flat because on U_A

$$\begin{aligned} \hat{g}_{ij} &= \left(1 - \frac{m_A}{4r}\right)^4 \left(1 + \frac{m_A}{2r}\right)^4 \delta_{ij} + O(1/r^2) \\ &= \left(1 + \frac{m_A}{4r}\right)^4 \delta_{ij} + O(1/r^2) \end{aligned} \quad (4.2.18)$$

so $\tilde{m}_A = m_A/2$ and is a constant multiple of g_i on any end other than U_1 . The scalar curvature for $\hat{g}$ is

$$\tilde{R} = \phi^{-5}[-8\Delta\phi + R\phi], \quad (4.2.19)$$

where $\phi > 0$ and $\Delta\phi \leq 0$, so that $\tilde{R} \geq 0$. When $r > 2\sigma$, $\Delta\phi < 0$, so $\tilde{R} > 0$.

4.2.3 The Contradictory Estimates for Gaussian Curvature of S

Finally, this last step brings us to a contradiction. Define a collection of sets

$$S_{(\sigma)} = (S \cap (\Sigma \setminus U_A)) \cup (S \cap B_\sigma(O)) \quad (4.2.20)$$

for $\sigma \geq \sigma_0$. These surfaces exhaust S , and are such that $\text{Area}(S_{(\sigma)}) \leq C_1 \sigma^2$. To show this Schoen and Yau use the fact that $S \cap \partial B_\sigma(O)$ will be the union of C^2 Jordan curves¹ that bound $S_{(\sigma)}$. They appeal to the minimal properties of $S_{(\sigma)}$ to obtain the estimate.

¹This is true for transverse intersections only, but happens a.e.

Now we look at the second variation inequality for the area of S .

$$\begin{aligned}\delta^2 \text{Area}(S)(f, f) &= - \int_S f(\Delta_S f + f \text{Ric}(\boldsymbol{\nu}) + f \|(^{(2)}\mathbf{K}\|^2) dS \\ &= \int_S (\|\nabla f\|^2 - f^2(\text{Ric}(\boldsymbol{\nu}) + \|(^{(2)}\mathbf{K}\|^2)) dS\end{aligned}\quad (4.2.21)$$

$$\geq 0 \quad (4.2.22)$$

for any $f \in C^2$ with compact support, where $\boldsymbol{\nu}$ is the unit normal vector to S , $\text{Ric}(\boldsymbol{\nu}) = R^j \nu_i \nu_j$, $^{(2)}\mathbf{K}$ is the extrinsic curvature of S , and $\|\cdot\|$ is the norm of $\mathbf{q}$ on S . Thus

$$\int_S f^2(\text{Ric}(\boldsymbol{\nu}) + \|(^{(2)}\mathbf{K}\|^2) dS \leq \int_S \|\nabla f\|^2 dS \quad (4.2.23)$$

and when $\text{tr}(^{(2)}\mathbf{K}) = 0$ (S is a minimal surface) we have

$$\text{Ric}(\boldsymbol{\nu}) - R = -1/2 \|(^{(2)}\mathbf{K}\|^2 - \kappa_S$$

and

$$\int_S (R - \kappa_S + 1/2 \|(^{(2)}\mathbf{K}\|^2) f^2 \leq \int_S \|\nabla f\|^2, \quad (4.2.24)$$

where we have now omitted the volume element dS as it is understood.

For $\sigma > \sigma_0$, let

$$\varphi = \begin{cases} 1 & \text{on } S_{(\sigma)} \\ \frac{\log(\sigma^2/r)}{\log \sigma} & \text{on } S_{(\sigma^2)} \setminus S_{(\sigma)} \\ 0 & \text{outside } S_{(\sigma^2)} \end{cases} \quad (4.2.25)$$

and let g be a Lipschitz function on S with $|g| \leq 1$ and $g = 1$ outside a compact subset of S .

We may re-write (4.2.24) as

$$\begin{aligned}\int_S (R - \kappa_S + 1/2 \|(^{(2)}\mathbf{K}\|^2) \varphi^2 g^2 &\leq 2 \int_S g^2 \|\nabla \varphi\|^2 + 2 \int_S \varphi^2 \|\nabla g\|^2 \\ &\leq \frac{2}{(\log \sigma)^2} \int_{S_{(\sigma^2)} \setminus S_{(\sigma)}} \frac{\|\nabla r\|^2}{r^2} + 2 \int_S \varphi^2 \|\nabla g\|^2.\end{aligned}\quad (4.2.26)$$

Since there exists a constant C_2 such that $\|\nabla r\|^2 \leq C_2$ we get

$$\int_{S_{(\sigma)}} \|(^{(2)}\mathbf{K}\|^2 g^2 \leq 2C_1 C_2 (\log \sigma)^{-1} + 2 \int_S \|\nabla g\|^2 + \int_S |\text{Ric}(\boldsymbol{\nu})| g^2 \quad (4.2.27)$$

and letting $\sigma \rightarrow \infty$ we see that

$$\int_S \|(2)K\|^2 g^2 \leq 2 \int_S \|\nabla g\|^2 + \int_S |\text{Ric}(\nu)|g^2. \quad (4.2.28)$$

From (4.2.3) we have that $\text{Ric}(\nu) = O(1/r^3)$ and see that $\int_S \|(2)K\|^2 < \infty$ which can be used to show that $\int_S |\kappa| < \infty$.

Now let $f = \varphi$ in (4.2.21) and let $\sigma \rightarrow \infty$ as before to obtain

$$\int_S (R - \kappa_S + \frac{1}{2} \|(2)K\|^2) \leq 0. \quad (4.2.29)$$

But $R \geq 0$ and $R > 0$ outside a compact subset S , thus

$$\int_S \kappa_S > 0. \quad (4.2.30)$$

The last step of the proof uses a version of the Gauss-Bonnet theorem to show

$$\int_S \kappa_S \leq 0. \quad (4.2.31)$$

Schoen and Yau give two proofs of this, both of which rely on certain versions of Gauss-Bonnet theorem. We present a sketch of one of these.

Let D_σ denote a disk of radius σ in $\mathbb{C}$ and let C_σ be a circle of radius σ . Because $\int_S |\kappa_S| < \infty$ and the Cohn-Vossen inequality;

$$\int_S |\kappa_S| \leq 2\pi \chi(S), \quad (4.2.32)$$

where $\chi(S)$ is the Euler characteristic of S , we have [15] that S is conformally equivalent to the complex plane. This means that there exists a diffeomorphism $F : \mathbb{C} \rightarrow S$. We may now apply the simply connected version of a theorem by Huber and Finn [15] related to the Gauss-Bonnet theorem that states

$$\int_S K_S = 2\pi - \lim_{t \rightarrow \infty} \frac{L_t^2}{2A_t}, \quad (4.2.33)$$

where $L_t = \text{length}(F(C_t))$ and $A_t = \text{Area}(F(D_t))$. Thus we need to show that

$$\lim_{t \rightarrow \infty} \frac{L_t^2}{4\pi A_t} \geq 1. \quad (4.2.34)$$

For this Schoen and Yau construct a sequence of surfaces, $\tilde{\Omega}_i$, of least Euclidean area, $\tilde{A}_i \equiv \tilde{A}(\tilde{\Omega}_i)$ and use the well-known inequality,

$$\tilde{A}(\tilde{\Omega}_i) \leq \frac{\tilde{L}_i^2}{4\pi}, \quad (4.2.35)$$

to show that

$$A_i \leq A(\tilde{\Omega}_i) \leq (1 + o(1))\tilde{A}_i \leq (1 + o(1))\frac{\tilde{L}_i^2}{4\pi} \leq (1 + o(1))\frac{L_i^2}{4\pi} \quad (4.2.36)$$

for i sufficiently large, proving the result.

4.2.4 The Only Spacetime with Zero Mass is Minkowski Spacetime

To show that if $m = 0$ the spacetime is Minkowskian, Schoen and Yau prove the following theorem;

Theorem 4.2.4 *Let Σ be an oriented 3-manifold with asymptotically flat metric $\mathbf{g}$. Suppose for some end U_A that*

$$|\partial^3 K_{ij}| + |\partial^4 K_{ij}| + |\partial^5 K_{ij}| \leq \frac{C}{1 + r^5}$$

for some constant $C > 0$, and $m_A = 0$. If $R \geq 0$ on Σ then $\mathbf{g}$ is flat and Σ is isometric to $(\mathbf{R}^3, \mathbf{e})$.

This involves a series of lemmas providing estimates for the integral of R on Σ and the existence of a function φ such that $\tilde{\mathbf{g}} = \varphi^4 \mathbf{g}$ is an asymptotically flat metric conformal to $\mathbf{g}$ with total mass

$$\tilde{m} = -\frac{1}{32\pi} \int_{\Sigma} R\varphi d\Sigma. \quad (4.2.37)$$

A corollary that follows from this states that, when $m = 0$, $R \geq 0$ and $R \not\equiv 0$, there exists a metric conformally equivalent to $\mathbf{g}$ with negative mass in contradiction to

theorem (4.1.1). Thus $R \equiv 0$ on Σ . Schoen and Yau assume that $\text{Ric}(\mathbf{g}) \neq 0$ and then construct a one-parameter family of metrics $\mathbf{g}_t$ such that $R(\mathbf{g}_t) = 0$ and

$$\mathbf{g}_t = \varphi_t^4(\mathbf{g} + t \text{Ric}(\mathbf{g}))$$

with mass

$$m(t) = -\frac{1}{32\pi} \int_{\Sigma} R_t \varphi_t \sqrt{g_t} d^3x.$$

They show the differentiability of $m(t)$ and that

$$m'(0) = \frac{1}{32\pi} \int_{\Sigma} |\text{Ric}(\mathbf{g})|^2 \sqrt{g} d^3x > 0,$$

when $\text{Ric}(\mathbf{g}) \neq 0$. This means that there would exist a $t < 0$ such that $m(t) < 0$, a contradiction. So $\text{Ric}(\mathbf{g}) \equiv 0$ and Σ is flat whenever $m = 0$.

4.2.5 The Full Conjecture

A paper by York [13] the year following Schoen and Yau's proof pointed out that the assumptions in the conjecture were slightly too restrictive to include all physically reasonable possibilities. They relaxed the asymptotic behaviour by a factor of r in each of the quantities in Schoen and Yau's definition of asymptotic flatness except for K . Also the statement of the conjecture has been somewhat generalized as well. The metrics are now assumed to be of the form

$$g_{ij} = \delta_{ij} + h_{ij}$$

and the mass for each end is defined to be

$$m_A = \frac{1}{16\pi} \int_{U_A} (g_{ij,j} - g_{j,i,j}) dS_i$$

which is the ADM mass (in the end U_A).

The theorem now states [29] that the energy-momentum 4-vector $\mathbf{P}$ (for which $P^0 = m$) is future directed and timelike, unless M is Minkowski spacetime. These

vectors are invariant under Lorentz transformations so it is then enough to show that $m_A > 0$ for the non-Minkowskian case. This is done by proving the existence of a metric $\tilde{g}_{ij} = \varphi^4 \delta_{ij}$ with conformal factor

$$\varphi^4 = \left(1 + \frac{m}{2r}\right) + O(r^{-2}), \quad |\partial\varphi| = O(r^{-2}), \quad |\partial^2\varphi| = O(r^{-3}),$$

and such that $\tilde{R} = 0$ and $\tilde{m}_A \leq m_A + \varepsilon$ on each U_A . Thus the previous theorem may be applied.

A much more significant generalization is the removal of the requirement that $\text{tr}(\mathbf{K}) = 0$. This was a crucial step in obtaining a full conjecture as there is no physical reason for the existence of maximal slices in a spacetime. Schoen and Yau also handled this case [30]. Their proof is too technical and labour intensive to sketch here but basically involves proving the existence of a slice $\tilde{\Sigma}_t$ of $M = \Sigma \times \mathbf{R}$ with $t = f(x)$ satisfying [27]

$$\left(K^{ij} - \frac{D^i D^j f}{1 + |Df|^2}\right) \left(g^{ij} - \frac{D_i f D_j f}{1 + |Df|^2}\right) = 0,$$

and such that $m = \tilde{m}$. Such a slice may be conformally deformed to another with zero scalar curvature and smaller or equal mass (which is covered under their previous result).

It turns out that the existence of such slices is related to the existence of apparent horizons. Schoen and Yau show the existence of $\tilde{\Sigma}$ which is asymptotic to zero (scalar curvature) at infinity outside a finite family of apparent horizons. This means that the class of spacetimes satisfying the conjecture include those with singularities hidden inside these apparent horizons, which include black holes.

Later that year Witten [36] came up with a very elegant and relatively simple proof based on supersymmetry and the spinor formulation of the GR.

Chapter 5

Discussion and Further Developments

Variational methods are becoming more and more prevalent in many areas of physics and engineering. In general relativity they are an especially indispensable tool. It is doubtful that many of the results of GR would be attainable without this theory. For instance, the basic idea in proving the singularity theorems by showing that reasonable spacetimes must be geodesically incomplete, relies on the existence of a maximal geodesic crossing a certain Cauchy surface orthogonally. But nowhere is the need for variational calculus more evident than in the formulation of the Cauchy problem in general relativity, and the definition of gravitational energy of a spacetime that follows.

It is a most important property that this energy be positive. Allowing for negative energy would cause the universe to be unstable. This is because of the tendency of all physical systems to decay to lower energy states. Thus for a system to be stable there must exist a “ground state”. One can always scale the metric by a positive constant, $g \rightarrow \Omega^2 g$, which affects the energy such that, $E \rightarrow \Omega E$, without changing the physical configuration of the system. If negative energy states exist, then there is no lower bound as we can always scale the metric to a lower energy state. Thus for stability we need the ground state to be of zero energy. The existence of a ground state in the energy spectrum will be important in the quantization of gravity since this is a necessity of any quantum theory.

The ADM energy was defined as an integral over a surface at spacelike infinity. It is also possible to define another energy where the integral is taken at null infinity. This is called the Bondi mass (see [35]). Here the mass functional is taken over the limit of 2 spheres at null infinity. Thus this mass does not have time-translation symmetry as the ADM mass does. It turns out that for a given time-translation the Bondi energy associated with the cross-section at the “later” time will be less than that of the “earlier” cross-section. This “energy loss” is the radiation of gravitational energy. Because this gravitational radiation extracts positive energy out of a gravitating system, we would expect that the Bondi mass must be positive. Otherwise a gravitating system could radiate away an infinite amount of energy, which is clearly an undesirable property.

A year after their final paper for the ADM proof, Schoen and Yau [32] extended their previous results to obtain the positivity of the Bondi mass as well. One can also define ADM and Bondi energies of asymptotically anti-de Sitter spacetimes. The positivity of these expressions was established by Abbot and Deser [1]. The final paper of Schoen and Yau [30] in proving the PMC, included the possibility of singularities if hidden behind apparent horizons and thus, includes black holes. The following year Gibbons, Hawking, and Horowitz showed that the mass (both Bondi and ADM) of a black hole is greater than the modulus of its electromagnetic charge [18].

Although extensions of Schoen and Yau’s ideas are sufficient to prove the various versions of the PMC, the supersymmetric results due to Witten [36] are used more widely. Also, it is worth noting that Nester [25] came up with another covariant expression whose integral at spacelike infinity also gives the ADM energy. The positivity of the divergence of this expression over a spacelike slice is easy to show, which allows for a very simple proof of the PMC.

Over the last fifteen years there have been two main focal points in the contin-

uation of the study of the positive energy theorems. One is the investigation into finding the appropriate conditions in the theorem that do not exclude any physical possibilities, the other is the investigation of the consequences of the PMC in its various forms.

The first condition of the theorem that was addressed was the asymptotic boundary conditions. This problem was solved in the mid 1980's. For a discussion of the weakest possible boundary conditions for which the PMC holds see [6] and [13].

More recently, (1995) Shiromizu and Sugai [33] investigated the possibility of relaxing the local energy conditions. They note that if we take into consideration certain quantum effects, the weak energy condition is no longer satisfied. Thus for the PMC to include spacetimes with the possibility of the quantum effects, the weak energy condition must be relaxed. They showed that one can relax the weak energy condition for static asymptotically flat spacetimes. They also suggest, but do not prove, that this condition may also be relaxed to include spacetimes that admit black holes that emit Hawking radiation, general non-static asymptotically flat spacetimes, and spacetimes with the existence of conjugate points of null geodesics which is required for the singularity theorems. The reason for their optimism is that the local (weak) energy condition can be shown to be equivalent to a global condition on the boundary of Σ .

Probably the most significant usage of the PMC is its role in determination of the validity of various extensions of GR. In the formulation of quantum and unified theories of gravity the two major conditions that must be satisfied are, the positivity of energies at spacelike and null infinity (extensions of the ADM and Bondi energy theorems respectively), and the agreement with the dynamics of GR in the classical regime. Corresponding versions of the PMC have been a useful criterion in considering these generalized theories.

Both the ADM and Bondi definitions of mass may be extended to the Kaluza

Klein generalization of GR. In this theory the spacetime $P = M \times A$ is of dimension $4 + n$. With the addition of each dimension one may introduce a new field (such as electromagnetism etc.) into the Ricci tensor by requiring some conformal symmetry. The motivation for this is to obtain a unified theory for gravity and the other fields of nature.

In this theory there are actually extra spacelike dimensions that we cannot detect (at least at lower energies). Thus this extra space may be considered to be "internal" to each point, $p \in M$, in that it can be represented as a fibre at p . That is, A is a fibre bundle on M . With this in mind it is assumed that the 4-dimensional field equations hold and the dynamics of gravity are the same as for classical GR.

The major question in this theory is whether or not gravity is still strictly an attractive interaction. The answer to this question is yes if the PMC' holds in P . This was indeed established by Moreschi and Sparling [23].

Another family of unified field theories which can be shown to satisfy the PMC are certain classes of scalar-tensor theories that incorporate other fields by coupling scalar fields to gravity [5].

There has been some criticism of the restricted nonlinear theories of gravity (i.e. nonlinear metric). These theories have been shown to be equivalent, through the use of the Legendre transform, to GR with additional fields. The problem with this formulation is that it introduces a new scalar dynamical variable which must be dealt with. Although such a system agrees with Einstein gravity (GR) mathematically, this scalar field may be incorrectly coupled to gravity which would cause instability and the possibility of a self-exciting system. Also there is freedom in how one may couple this field with gravity which gives rise to a whole class of allowable theories which, although mathematically consistent, may not have the same physical characteristics.

The PMC' resolves this controversy. In fact it is instrumental in deciding whether

or not a generalized gravity (unified) theory corresponds to Einstein gravity physically. There are three basic types of these theories: dimensionally reduced multi-dimensional Einstein theories, the restricted nonlinear theories, and the scalar tensor field theories with non-minimal coupling to gravity. In each of these theories there is a unique conformal mapping from the system onto a minimally coupled scalar-Einstein gravity theory. These are Legendre transformations which may be interpreted physically as a transition to a system with definite energy. Conformal but non-Legendre transformations will give an indefinite energy momentum tensor associated with the scalar field. Thus the PMC restricts the class of allowable theories to a unique physical theory inherited from its parent theory through a Legendre transformation.

For general nonlinear theories the energy of the scalar field is indefinite for such transformations and it turns out that these theories are mathematically incompatible with GR as well. For a detailed discussion of this see [34].

As mentioned above a requirement of all quantum theories is the existence of a lower bound in the energy spectrum, which allows for a ground state of the system. Thus any theory of gravitation which leads to a quantization of gravity must satisfy a version of the PMC. One such example of this is the quadratic theories of gravitation; the so-called $R + R^2$ theories. A proof of the PMC for these theories was given by Maluf [20].

Another interesting result is the connection between positivity and spherical symmetry. The PMC was used to establish a proof that static perfect fluid spacetime is spherically symmetrical. This is also relevant to proving the spherical symmetry of a static stellar model [21].

The PMC further supports the Hamiltonian formulation of GR. If it did not hold then we would suspect that either GR is not the correct theory (or even the correct foundation of some generalized theory) of gravity, or that the GMD action

is incomplete.

There is a general principle in applied mathematics that one should always be able to obtain all the results of a theory within its own framework. More precisely, given a mathematical theory describing some system, one should be able to calculate everything intrinsically without appealing to quantities external to that system. However, many calculations involve much more work when one is restricted to intrinsic notions only. The most obvious example of this principle is that the intrinsic geometry of a manifold requires no knowledge of the space in which it is embedded.

If the dynamics of gravity is to correspond to GMD, even though the actual theory describing gravity might well be much more general than GR, we would expect that all results (outside the quantum regime) determining the dynamics should be obtainable within the framework of classical GR. This is why the proof of Schoen and Yau is so significant, even though it is much more cumbersome than that of Witten. Their proof is purely geometrical and within the confines of the standard formulation of classical GR.

The ideas of Witten are used now when attempting to prove positive energy theorems for more general theories that attempt to quantify gravity and unify it with the other fields. However, to the author's knowledge, thus far the Schoen and Yau results could have been used equally as well. Thus the non-supersymmetric proof of the positive energy theorems further supports the notion that classical GR is correct in so far as describing the universe as a dynamical system in the classical (non-quantum) regime. It will be interesting to see if there will exist realistic theories where the PMC can be obtained with Witten's approach (using supergravity) but not with strictly classical GR.

Regardless of whether or not the classical arguments continue to be sufficient for proving the PMC for future theories of gravity, the notion of positivity of gravitational energy will continue to play a key role in determining whether a given unified

theory is a plausible candidate for describing nature.

References

1. Abbot, L. and Deser, S.: Nucl. Phys. B195, 76-96 (1982)
2. Araki, H.: On weak time-symmetric gravitational waves. Ann. Phys. 7, 156-465 (1959)
3. Arnowitt, R., Deser, S., and Misner, C.: The Dynamics of General Relativity, in: Gravitation: An Introduction to Current Research, ed. Witten, L., Wiley (1962)
4. Baierlein, R., Sharp, D., and Wheeler, J.: Three-Dimensional Geometry as Carrier of Information about Time, Phys. Rev., 126, no. 5, 1864-65 (1962)
5. Batakis, N. A., Gerard, J. and Schucker, T.: Phys. Lett. 147B(4,5), 295-296 (1984)
6. Bizon, P. and Malec, E.: Class. Quant. 3, L123-L128 (1986)
7. Brill, D.: Ann. Phys. 7, 466-483 (1959)
8. Brill, D. and Deser, S.: Ann. Phys. 50, 548-570 (1968).
9. Brill, C. and Pong, S-Y.: The Positive Mass Conjecture, in: General Relativity and Gravitation, ed. Held, A., Plenum (1980)
10. Choquet-Bruhat, Y.: Positive Energy Theorems, in: Relativity Groups and Topology, eds. Dewitt, B. and Stora, R., North-Holland (1984)
11. Choquet-Bruhat, Y. and Marsden, J.: Solution of the Local Mass Problem in General Relativity. Commun. Math. Phys. 51, no. 3, 283 (1976)

12. Choquet-Bruhat, Y., Fischer, A., and Marsden, J.: Maximal Hypersurfaces and Positivity of Mass, in: Isolated Gravitating Systems in General Relativity, ed. Ehlers, J., North Holland (1979)
13. Chrusciel, P.: Class. Quant. 3, L115-L121 (1986)
14. Federer, H.: Geometric Measure Theory, Springer-Verlag (1969)
15. Finn, R.: Comment. Math. Helv. 40, 1-30 (1965)
16. Fischer, A. and Marsden, J.: The Initial Value Problem and the Dynamical Formulation of General Relativity, in General Relativity, eds. Hawking, S. W. and Israel, W.: Cambridge University Press (1979)
17. Fong, C. K.: Variational Methods: A Problem Solving Approach, (not yet published).
18. Gibbons, G., Hawking, S. W. and Horowitz, G.: Commun. Math. Phys. 88,295-308 (1983)
19. Hawking, S. and Ellis, G.: The Large Scale Structure of Spacetime, Cambridge University Press (1973)
20. Maluf, J. W.: Class. Quant. 6, L151-L154 (1989)
21. Masood-ul-Alam, A.K.M.: Class. Quant. 4, 625-633 (1987)
22. Misner, C., Thorne K., and Wheeler J.: Gravitation, Freeman (1973)
23. Moreschi, O. M. and Sparling, G. A.: J. Math. Phys. 27(9),2402-2408 (1986)
24. Morgan, F.: Geometric Measure Theory, Academic Press (1988)
25. Nester, J.: Phys. Lett. 83A(6), 241-243 (1981)
26. O'Murchadha, N. and York, J.: Phys. Rev. D10, 2-345 (1974)

27. Pong, S-Y.: *J. Math. Phys.* 19, 1155 (1978)
28. Regge, T. and Teitelboim, C.: *Role of the Surface Integrals in the Formulation of General Relativity*, *Ann. Phys.* 88, 286-318 (1974)
29. Schoen, R. and Yau, S.-T.: *The Energy and the Linear Momentum of Space-Times in General Relativity*, *Commun. Math. Phys.* 79, 47-51 (1981)
30. Schoen, R. and Yau, S.-T.: *Proof of the Positive Mass Theorem II*, *Commun. Math. Phys.* 79, 231-260 (1981)
31. Schoen, R. and Yau, S.-T.: *On the Proof of the Positive Mass Conjecture*, *Commun. Math. Phys.* 65, 45-76 (1979)
32. Schoen, R. and Yau, S.-T.: *Commun. Math. Phys.* 48, 369-371 (1982)
33. Shiromizu, T. and Sugai, M.: *Phys. Rev. D*, 51(2), 472-476 (1995)
34. Sokolowski, L. M.: *Class. Quant.* 6, 2045-2050 (1989)
35. Wald, R.: *General Relativity*, University of Chicago Press (1984)
36. Witten, E.: *Commun. Math. Phys.* 80, 381 (1981)
37. York, J.: *Energy and Momentum of the Gravitational Field*, in: *Essays in General Relativity*, ed. Tipler, F., Academic Press (1980)

END

1 9 0 6 9 7

FIN