Introduction to Lorentzian Geometry and Einstein Equations in the Large

Piotr T. Chruściel[1]
Département de Mathématiques
Faculté des Sciences
Parc de Grandmont
F-37200 Tours, France

December 8, 2004

[1]Supported in part by the Polish Research Council grant KBN 2 P03B 073 15.
E–mail: chrusciel@univ-tours.fr

Contents

1 Preface **7**
1.1 Various figures . . . 7
1.2 Return to real stuff . . . 7

2 Basic notions **15**
2.1 Conventions . . . 15
2.2 Lorentzian manifolds . . . 15
2.3 The Levi-Civita connection, curvature . . . 17

3 An introduction to the physics of the Einstein equations **21**
3.1 Einstein equations . . . 21
3.2 Energy-momentum tensors . . . 23
3.2.1 Dust, null fluids . . . 23
3.2.2 Perfect fluids . . . 24
3.2.3 Field theoretical models . . . 24
3.3 The divergence identity, equations of motion . . . 25
3.3.1 Geometric optics in the Schwarzschild metric . . . 27
3.4 Weak gravitational fields . . . 28
3.4.1 Small perturbations of Minkowski space-time . . . 28
3.4.2 Coordinate conditions, wave coordinates . . . 29
3.4.3 Linearized Einstein equations in wave coordinates . . . 31
3.4.4 Gravitational radiation? . . . 32
3.4.5 Newton's equations of motion, and why $8\pi G$ is $8\pi G$ . . . 34
3.4.6 Determining the coupling constant: the geodesic equation 36

4 Causality **39**
4.1 Time orientation . . . 39
4.2 Normal coordinates . . . 41
4.3 Causal paths . . . 49
4.4 Futures, pasts . . . 52
4.5 Extendible and inextendible paths . . . 62
4.5.1 Maximally extended geodesics . . . 65
4.6 Limits of curves . . . 66
4.6.1 Achronal causal curves . . . 69
4.7 Causality conditions . . . 70
4.8 Global hyperbolicity . . . 74

4.9 Domains of dependence . . . 77
4.10 Cauchy surfaces . . . 82
4.11 Some applications . . . 86
4.12 The Lorentzian length functional . . . 89
4.13 The Lorentzian distance function . . . 89

5 Splitting theorems 95
5.1 The geometry of C^2 null hypersurfaces . . . 95
5.2 Galloway's null splitting theorem . . . 109
5.3 Timelike splitting theorems . . . 109

6 The Einstein equations 111
6.1 The nature of the Einstein equations . . . 111
6.2 Existence local in time in wave coordinates . . . 114
6.3 The geometry of spacelike submanifolds . . . 117
6.4 Cauchy data, gauge freedom . . . 123
6.5 $3+1$ evolution equations (to be done) . . . 126
6.6 Existence: global in space (to be done) . . . 126
6.7 Constraint equations: the conformal method . . . 126
6.7.1 The vector constraint equation on compact manifolds . . . 129
6.7.2 The scalar constraint equation on compact manifolds . . . 134
6.7.3 The thin sandwich (to be done) . . . 139
6.7.4 The conformal thin sandwich (to be done) . . . 139
6.7.5 Asymptotically hyperboloidal initial data (to be done) . . . 140
6.8 Penrose-Newman-Christodoulou-Klainerman conformal equations (to be done) . . . 140
6.9 Anderson's conformal equations (to be done) . . . 140
6.10 Maximal globally hyperbolic developments (to be done) . . . 140
6.11 Strong cosmic censorship? (to be done) . . . 140

7 Positive energy theorems 141
7.1 The mass of asymptotically Euclidean manifolds . . . 142
7.2 Coordinate independence . . . 152
7.3 The Bartnik-Witten rigidity theorem . . . 160
7.4 Some space-time formulae . . . 161
7.5 Energy in stationary space-times . . . 163
7.6 Moving hypersurfaces in space-time . . . 164
7.7 Spherically symmetric positive energy theorem . . . 165
7.8 Lohkamp's positive energy theorem . . . 166
7.9 Li-Tam's proof of positivity of the Brown-York quasi-local mass . 166
7.10 The Riemannian Penrose inequality . . . 166
7.11 A poor man's positive energy theorem . . . 166
7.12 Small data positive energy theorem . . . 172
7.13 The definition of mass: a Hamiltonian approach . . . 172
7.14 Some background on spinors . . . 173
7.15 (Local) Schrödinger-Lichnerowicz identities . . . 176
7.16 Eigenvalues of certain matrices . . . 189

8 Black holes **191**

A Some elementary facts in differential geometry **195**
A.1 Vector fields . . . 195
A.2 Raising and lowering of indices . . . 196
A.3 Lie derivatives (to be done) . . . 198
A.4 Covariant derivatives . . . 198
A.4.1 Torsion . . . 202
A.5 Curvature . . . 203
A.5.1 Bianchi identities . . . 206
A.5.2 The Levi-Civita connection . . . 208
A.5.3 Further algebraic identities? . . . 211
A.5.4 How to recognise that coordinates are normal . . . 217
A.5.5 Taylor expansions of the metric in normal coordinates . . 219
A.5.6 Further first-order differential identities? . . . 225
A.5.7 Higher differential identities? . . . 230
A.6 Moving frames . . . 232

B Some interesting space-times **237**
B.1 Minkowski space-time . . . 237
B.2 further examples (Schwarzschild, Kerr)... . . . 237
B.3 Taub-NUT space-times . . . 238

C Conformal rescalings of the metric **241**
C.1 The curvature . . . 241
C.2 Non-characteristic hypersurfaces . . . 242

D Sobolev spaces on manifolds **245**
D.1 Weighted Sobolev spaces on manifolds . . . 246

E $W^{k,p}$ manifolds **251**

F A pot-pourri of formulae **253**

Chapter 1

Preface

1.1 Various figures

The following figures are from Bicak gr-qc/0201010

There are some nice embedding figures in Ben-Dov gr-qc 0408066 on Penrose inequality

There are some good Penrose diagrams for Schwarzschild in our encyclopedia paper with Beig, coming from Nicolas in Dissertationes [?] (in fact, they are improved, as compared to the original ones).

1.2 Return to real stuff

This work is based on lectures given at the Levoča Summer School, August 2000, and at the seminar "Géométrie de Lorentz et relativité générale: Examples", Avignon, March 2002, and at Tours University, March-June 2002, together with the lectures given at the "Oberwolfach Seminar on Mathematical General Relativity", November 2002, and at the "Mini-course on Geometric Analysis & Mathematical Aspects of General Relativity" at the National Center for Theoretical Sciences, National Tsing-Hua University, Hsinchu, Taiwan, August 2004. It is intended as a textbook on selected topics in mathematical relativity, designed for graduate students in mathematics which have some familiarity with manifolds and Riemannian geometry. The book contains a fairly extensive survey of several topics, and some of the results here might also be of interest to mathematically minded researchers in general relativity.

Appendix A, the contents of which would usually be the topic of the first lecture in a lecture course addressed to an audience of students of mathematics at a graduate level, is included for the sake of those which are not used to the index manipulations habits of general relativists. It might also be useful to physics students which might be very familiar with index notations, but are unfamiliar with the more abstract approach presented there. Chapter 2

Figure 1.1: The Penrose conformal diagram of an asymptotically flat spacetime. The Cauchy hypersurface and the hyperboloidal hypersurface s are indicated.

Figure 1.2: Two particles uniformly accelerated in opposite directions.

Figure 1.3: The Penrose compactified diagram of a boost-rotation symmetric spacetime. Null infinity can admit smooth sections.

introduces the notion of a Lorentzian metric, establishes some elementary facts, and summarizes the elementary differential geometric facts described in detail in Appendix A. Chapter 3 has an illustrative character, aiming to give the students an idea of a few important applications of general relativity. While it may serve to make the material more attractive, it can be safely skipped when lecturing to a mathematically minded audience. Chapter 4 is the core•1.2.1 of this book, in so far as it gives a new self-contained approach to the causality theory. The expert will note that the definition of timelike and causal curves given here is rather different from the usual ones from [16, 58, 89], and makes the whole theory rather simpler.

•*1.2.1:* **ptc:***this description needs expanding, and the "core" should be thought over*

It should be clear to the reader that most of the book is devoted to the study of the Einstein equations. On the other hand, the title of the book does not mention this. This is not an omission, as we believe that the main object of studying Lorentzian geometry in the large is precisely the study of global solutions of the Einstein equation. •1.2.2

•*1.2.2: to be thought over*

Tours, spring semester 2002:

1. First lecture (2 h): Metrics, connections, curvature (Appendix)
2. Second lecture (1 h): Weak gravitational fields
3. Third lecture (1.75 h): Einstein equations, energy momentum tensors, conservation law for dust and for null fluids, geodesic equations.
4. Fourth lecture (1.5 h): Newtonian equations of motion from Einstein equations. Time orientability, normal coordinates.
5. Fifth lecture (1.5 h): construction and properties of normal coordinates, definitions of futures and pasts, causality in Minkowski space-time.
6. Sixth lecture (1.75 h): local causality on general manifolds. End points of causal curves.
7. Seventh lecture (1.5 h): accumulation curves, causality conditions with examples up to, but not including, global hyperbolicity.
8. Eighth lecture (1.5 h): introduction to global hyperbolicity, closedness of the causal futures.
9. Ninth lecture (2 h): domains of dependence, Cauchy surfaces, beginning of Geroch's theorem (I do not recommend proving in class that interiors of domains of dependence are globally hyperbolic - very long)
10. 10th lecture (1h15): end of Geroch's theorem (without all details), Einstein equations in harmonic coordinates

11. 11th lecture (1h45): Einstein equations in harmonic coordinates continued; geometry of spacelike hypersurfaces, Codazzi-Mainardi equations, the scalar constraint

12. 12th lecture (1h): the vector constraint; the gauge character of the $g'_{0A}s$, K_{ij} and time derivatives

13. 13th lecture (1h40): heuristics for a Riemannian definition of mass, calculation of the divergence identity, mass for conformally flat metrics, in particular for Schwarzschild in isotropic coordinates

14. 14th lecture (1h45): ADM mass for conformally symmetric metrics, for spherically symmetric metrics, thus for Schwarzschild, most of the proof of coordinate-independence

15. 15th lecture (1h30): end of the proof of coordinate independence, introduction to moving frames

16. 16th lecture (1h30): end of moving frames, spherically symmetric positive energy theorem, introduction to spinors

Throughout this work we use indented text, typeset in smaller font, for material which plays secondary — informative or auxiliary — role, and may be skipped without affecting the understanding of the main line of development of the subject.

Acknowledgements I am very grateful to the following persons for their comments concerning mistakes in previous versions of these notes, or for discussions that helped me to better understand the material covered: A. Cabet, E. Dumas, Leipzig student. Of course I am taking full responsibility for the mistakes that remain.

Mathematical Relativity, Oberwolfach

lectures by Robert A. Bartnik and Piotr T. Chruściel

November 10-16, 2002

Monday	9.00-10.30	RAB: Formal aspects of the Initial Value Problem (IVP)
	10.50-12.20	PTC: The conformal method for solving the constraint equations I
	14.00-16.30	Study sessions (with a break of 30 minutes)
Tuesday	9.00-10.30	RAB: Wave equations
	10.50-12.20	PTC: The conformal method for solving the constraint equations II
	14.00-16.30	Study sessions (with a break of 30 minutes)
Wednesday	9.00-10.30	RAB: Local existence for the IVP
	10.50-12.20	PTC: The definition of mass
	afternoon:	hike to Sankt Roman
Thursday	9.00-10.30	RAB: Introduction to spinors and to the Dirac equation
	10.50-12.20	PTC: Introduction to causality theory I
	14.00-16.30	Study sessions (with a break of 30 minutes)
Friday	9.00-10.30	RAB: Witten's proof of the positive mass theorem
	10.50-12.20	PTC: Introduction to causality theory II
	14.00-15.30	PTC: Introduction to causality theory III
	16.00-16.20	T. Jurke (AEI Golm): On future asymptotics of polarised Gowdy space-times
	16.20-16.40	E. Czuchry (Warsaw): Constraint equations for null matter
	16.40-17.00	S. Calogero (AEI Golm): A stellar dynamics model in scalar gravity
	17.00-17.20	K. Schöebel (Iena): Highly accurate calculations of rotating neutron stars
	17.20-17.40	K. Roszkowski (Cracow): Some aspects of wave propagation on Schwarzschild space-time

Tsin-Hua Lectures, August 2004, 1.5 h lectures

1. The definition of mass, the Witten-Bartnik positive energy theorem

2. Introduction to Lorentzian geometry, up to definition of futures and pasts

3. Futures and past, accumulation curves, causality conditions up to the definition of global hyperbolicity

4. Global hyperbolicity, Lorentz distance function, Splitting theorems, Poor man's positive energy theorem (the timelike version)

Chapter 2

Basic notions

2.1 Conventions

All manifolds are Hausdorff and paracompact. The letter n usually denotes the space dimension of the manifold under consideration; thus, to emphasize the distinct character of the space and time variables, space-times will always have dimension $n+1$. Greek indices α, β, *etc.*, correspond to space-time coordinates x^α and take values $0,1,\ldots,n$, while latin indices i,j, *etc.*, take values $1,\ldots,n$ and correspond to space-coordinates x^i. We shall use the summation convention throughout, which means that every pair of indices with one index up and one index down have to be summed over, *e.g.*

$$A_\alpha{}^\beta{}_{\gamma\delta}X^\gamma Y_{\beta\sigma} := \sum_{\gamma,\beta=0}^{n} A_\alpha{}^\beta{}_{\gamma\delta}X^\gamma Y_{\beta\sigma}\ , \qquad A_iX^i := \sum_{i=1}^{n} A_iX^i\ .$$

If f is a function, then we will freely switch between the following notations to denote its derivatives:

$$\frac{\partial f}{\partial x^\mu} = \partial_\mu f = f_{,\mu} = f_{;\mu} = \nabla_\mu f = \nabla_{\partial_\mu} f\ .$$

2.2 Lorentzian manifolds

A couple $(\mathscr{M},g)$ will be called a Lorentzian manifold if $\mathscr{M}$ is an $n+1$ dimensional, differentiable, paracompact, Hausdorff, connected manifold, and g is a non-degenerate symmetric twice covariant continuous tensor field on $\mathscr{M}$ of signature $(-++\ldots+)$. By an abuse of terminology g is often called a metric. We will only consider differentiable manifolds with continuous metrics. By a theorem of Whitney•2.2.1 there is no loss of generality to assume that the manifold •2.2.1: *ref?* is smooth, however such a restriction turns out to be inconvenient for many purposes. For example, we will see below in Chapter **??** that the initial value problem for general relativity leads naturally to manifolds with a Sobolev-type differentiable structure[1]. Unless explicitly indicated otherwise we assume that

[1]Such structures are discussed in detail in [**?**].

the metric is smooth; however, we will often indicate weaker differentiability conditions which are sufficient for the specific problems at hand.

A vector $X \in T_p\mathscr{M}$ is said to be *timelike* if $g(X,X) < 0$; *null* or *lightlike* if $g(X,X) = 0$ and $X \neq 0$; *causal* if $g(X,X) \leq 0$ and $X \neq 0$; *spacelike* if $g(X,X) > 0$. Given any basis of T_pM with the first vector timelike, we can use the standard Gram-Schmidt procedure to construct an *ON-basis* of $T_p\mathscr{M}$, that is, vectors $e_a \in T_p\mathscr{M}$ such that

$$g(e_a, e_b) = \eta_{ab} \ ,$$

where η_{ab} stands for the Minkowski metric diag$(-1, +1, \ldots, +1)$; indices a, b run from 0 to n. If $X = X^a e_a$, $Y = Y^a e_a$ (we use the summation convention, *i.e.*, repeated indices in different positions have to be summed over), then

$$g(X,Y) = -X^0Y^0 + X^1Y^1 + \ldots + X^nY^n \ , \tag{2.2.1}$$

$$g(X,X) = -(X^0)^2 + (X^1)^2 + \ldots + (X^n)^2 \ . \tag{2.2.2}$$

It follows that X is lightlike if and only if

$$X^0 = \pm\sqrt{\sum_i (X^i)^2} \ ,$$

and timelike if and only if

$$|X^0| > \sqrt{\sum_i (X^i)^2} \ .$$

(Throughout lower Latin indices i, j, *etc.* run from 1 to n.) Thus, in every tangent space the sets of timelike, lightlike, *etc.*, vectors are linearly isomorphic to those of Minkowski space-time. In particular the set of timelike vectors is the union of two disjoint open convex cones, the closures of which meet at the origin.

We will say that a causal vector X is *future pointing* if $X^0 > 0$ (compare Section 4.1 below).

We will often use the following elementary facts:

PROPOSITION 2.2.1 *Let X be timelike future pointing, then*

$$g(X,Y) < 0$$

for all future pointing causal Y's.

PROOF: Choose an ON frame with

$$e_0 = X/\sqrt{-g(X,X)} \quad \Longleftrightarrow \quad X = X^0 e_0 \ ,$$

the result is then obvious from (2.2.1). □

The scalar product of causal vectors satisfies the *inverse Cauchy-Schwarz inequality*:

PROPOSITION 2.2.2 *Let X, Y be two causal vectors, then*

$$|g(X,Y)| \geq \sqrt{-g(X,X)}\sqrt{-g(Y,Y)}\ , \tag{2.2.3}$$

with equality holding if and only if X is proportional to Y.

PROOF: For null X's Equation (2.2.3) is trivial, assume thus that X is timelike. By performing a Gram-Schmidt orthonormalisation we can construct an ON basis in which $e^0 = X/\sqrt{-g(X,X)}$, hence $X = \sqrt{-g(X,X)}e^0$. Since Y is causal we have $|Y^0| \geq \sqrt{\sum_i (Y^i)^2}$, hence

$$\begin{aligned} |g(X,Y)| &= |-X^0Y^0| \\ &= \sqrt{-g(X,X)}|Y^0| \\ &\geq \sqrt{-g(X,X)}\sqrt{(Y^0)^2 - \sum_i (Y^i)^2} \\ &= \sqrt{-g(X,X)}\sqrt{-g(Y,Y)}\ , \end{aligned}$$

with equality holding if and only if $\sum_i (Y^i)^2 = 0$, hence Y proportional to X. If Y is null, and equality holds in (2.2.3), then $g(X,Y) = 0$. In this case a Gram-Schmidt orthonormalisation starting from any timelike vector e^0 as a first vector, and X as a second vector, leads to a basis in which $X = X^0(e_0 + e_1)$, so that

$$0 = g(X,Y) = -X^0(Y^0 - Y^1)\ .$$

Since X^0 is non-zero we conclude that $Y^0 = Y^1$. Causality of Y gives

$$|Y^0| \geq \sum_i (Y^i)^2 \quad \Longrightarrow Y = Y^0(e_0 + e_1) \sim X\ .$$

□

As an immediate Corollary of Proposition 2.2.1 we obtain:

COROLLARY 2.2.3 *Let X, Y be two null vectors satisfying*

$$g(X,Y) = 0\ .$$

Then X is proportional to Y.

2.3 The Levi-Civita connection, curvature

In this section we give a short overview of the properties of the Levi-Civita connection for Lorentzian manifolds, and its curvature. The main point here is to codify our notations and conventions for those readers who are already familiar with those notions. Newcomers will find a complete exposition of the results presented here in Appendix A.

Consider a Lorentzian manifold $(\mathscr{M}, g)$; in a way completely analogous to the Riemannian case, to any Lorentzian metric one can associate a connection ∇ which is uniquely defined by the requirements that a) ∇ is g compatible, that is, for all differentiable vector fields X, Y, Z on $\mathscr{M}$ we have

$$X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)\ ,$$

and b) ∇ is torsion free:

$$\nabla_X Y - \nabla_Y X = [X, Y] \ .$$

∇ is called the Levi-Civita connection associated with the metric g. Given any connection ∇, its Riemann tensor is defined as follows:

$$R(X, Y)Z := \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \ . \tag{2.3.1}$$

Let $\partial_\alpha \equiv \partial/\partial x^\alpha$ be a basis of $T\mathscr{M}$ associated with a coordinate system x^α, and let dx^α be the associated dual basis:

$$\langle dx^\alpha, \partial_\beta \rangle = \delta^\alpha_\beta \ , \tag{2.3.2}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing ($\langle \phi, X \rangle := \phi(X)$) and δ^α_β is the Kronecker delta, equal to 1 when $\alpha = \beta$, and zero otherwise. One defines

$$g_{\alpha\beta} := g(\partial_\alpha, \partial_\beta) \iff g = g_{\alpha\beta} dx^\alpha \otimes dx^\beta \ ,$$

$$R^\alpha{}_{\beta\gamma\delta} := \langle dx^\alpha, R(\partial_\gamma, \partial_\delta)\partial_\beta \rangle \ . \tag{2.3.3}$$

It would be logical to use the symbol R for the Riemann tensor, however, that symbol is often used to denote the Ricci scalar defined in (2.3.12) below. For this reason we will write "Riem" instead for the Riemann tensor, so that (2.3.3) can be rewritten as

$$\mathrm{Riem} = R^\alpha{}_{\beta\gamma\delta} \partial_\alpha \otimes dx^\beta \otimes dx^\gamma \otimes dx^\delta \ . \tag{2.3.4}$$

There are actually two ways of understanding objects with indices, such as $R^\alpha{}_{\beta\gamma\delta}$, or $g_{\beta\gamma}$, *etc.* The first refers to a given coordinate system $\{x^\mu\}$, and then (2.3.3) gives a definition of the collection of numbers $\{R^\alpha{}_{\beta\gamma\delta}\}$, associated with the tensor Riem. This provides one practical way of computing things, and then (2.3.4) tells us how to recover Riem out of the $\{R^\alpha{}_{\beta\gamma\delta}\}$'s. The second way, used throughout this work, is to understand the indices in $R^\alpha{}_{\beta\gamma\delta}$ or in $g_{\beta\gamma}$ as markers which indicate the tensor type, independently of any coordinate system. Those markers are very useful when operations on tensors are performed, such as taking contractions (*e.g.*, $B^\alpha{}_\beta \to B^\alpha{}_\alpha$) or traces (*e.g.*, $A^{\alpha\beta} \to A^\alpha{}_\alpha := A^{\alpha\beta} g_{\alpha\beta}$), and is known as *Penrose's abstract index notation* [90]. To summarise, the notation such as $W^\alpha{}_\beta{}^\gamma$, when referring to a tensor field W, provides a short-hand for saying that W is a section of the bundle $TM \otimes T^*M \otimes TM$ (we will write $W \in \Gamma(TM \otimes T^*M \otimes TM)$, or $W \in \Gamma_{C^k}(TM \otimes T^*M \otimes TM)$ to indicate the C^k differentiability class, *etc.*). This short-hand is consistent with the local equations

$$W^\alpha{}_\beta{}^\gamma = W(dx^\alpha, \partial_\beta, dx^\gamma) \quad \iff \quad W = W^\alpha{}_\beta{}^\gamma \partial_\alpha \otimes dx^\beta \otimes \partial_\gamma \ ,$$

when a local coordinate system $\{x^\mu\}$ has been given, but the reader *should not* assume that some coordinate system has been singled out.

Returning to the Riemann tensor, it follows from (2.3.1) that $R^\alpha{}_{\beta\gamma\delta}$ is antisymmetric in its last two indices,

$$R^\alpha{}_{\beta\gamma\delta} = -R^\alpha{}_{\beta\delta\gamma} \ .$$

As in the Riemannian case we have

$$R_{\alpha\beta\gamma\delta} = R_{\gamma\delta\alpha\beta} ,$$

here and throughout we use the metric to raise and lower the indices,

$$R_{\alpha\beta\gamma\delta} := g_{\alpha\lambda} R^{\lambda}{}_{\beta\gamma\delta} .$$

This, and some other identities, are proved in Appendix A. Unless explicitly indicated otherwise, we use the *Einstein summation convention* which requires summing over all matching pairs of sub- and superscripts. We further have the *first Bianchi identity*,

$$R^{\alpha}{}_{\beta\gamma\delta} + R^{\alpha}{}_{\gamma\delta\beta} + R^{\alpha}{}_{\delta\beta\gamma} = 0 . \tag{2.3.5}$$

We use the convention that adding a subscript "$_{;\alpha}$" to a tensor field denotes covariant differentiation in the direction ∂_α, *e.g.*

$$R^{\alpha}{}_{\beta\gamma\delta;\sigma} := \langle dx^\alpha, (\nabla_\sigma R)(\partial_\gamma, \partial_\delta)\partial_\beta \rangle ,$$

where

$$\nabla_\sigma := \nabla_{\partial_\sigma} .$$

In this notation the *second Bianchi identity* reads

$$R^{\alpha}{}_{\beta\gamma\delta;\sigma} + R^{\alpha}{}_{\beta\delta\sigma;\gamma} + R^{\alpha}{}_{\beta\sigma\gamma;\delta} = 0 . \tag{2.3.6}$$

A metric – whether Lorentzian or Riemannian – defines an isomorphism $\flat$: $T\mathscr{M} \to T^*\mathscr{M}$ by the formula

$$X_\flat(Y) = g(X, Y) .$$

If we write X in the coordinate basis ∂_α as $X^\alpha\partial_\alpha$ (hence $X^\alpha = \langle dx^\alpha, X\rangle$), we then have

$$X_\flat = g_{\alpha\beta} X^\alpha dx^\beta =: X_\beta dx^\beta .$$

The operation which takes X^α to X_α is sometimes called "the lowering of indices" in the physics literature, while its inverse is called "the raising of indices".

We can use the map $\flat$ and its inverse $\sharp$ to define a metric $g^\sharp$ on $T^*\mathscr{M}$, by transporting the metric g from $T\mathscr{M}$ to $T^*\mathscr{M}$. If we let $g^{\alpha\beta}$ denote the matrix inverse to $g_{\alpha\beta}$, one easily finds

$$g^\sharp(dx^\alpha, dx^\beta) = g^{\alpha\beta} . \tag{2.3.7}$$

In the physics literature it is customary to use the same symbol g both for g, and $g^\sharp$, as well as for extensions of g to tensor fields of any order, and we shall sometimes do this.

The *Christoffel symbols* $\Gamma^\alpha_{\beta\gamma}$ are defined as

$$\Gamma^\alpha_{\beta\gamma} \equiv \langle dx^\alpha, \nabla_{\partial_\beta}\partial_\gamma \rangle . \tag{2.3.8}$$

It follows that

$$\nabla_\alpha X = (\partial_\alpha X^\beta + \Gamma^\beta_{\alpha\sigma} X^\sigma)\partial_\beta \; .$$

We have the classical formulae•2.3.1

•2.3.1: **ptc:***refer to where it is worked out in more detail*

$$\Gamma^\alpha_{\beta\gamma} = \tfrac{1}{2} g^{\alpha\sigma}(\partial_\beta g_{\sigma\gamma} + \partial_\gamma g_{\sigma\beta} - \partial_\sigma g_{\beta\gamma}) \; , \tag{2.3.9}$$

$$R^\alpha{}_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha_{\beta\delta} - \partial_\delta \Gamma^\alpha_{\beta\gamma} + \Gamma^\alpha_{\sigma\gamma}\Gamma^\sigma_{\beta\delta} - \Gamma^\alpha_{\sigma\delta}\Gamma^\sigma_{\beta\gamma} \; . \tag{2.3.10}$$

The *Ricci tensor* $\mathrm{Ric} = R_{\alpha\beta} dx^\alpha \otimes dx^\beta$ is defined as

$$R_{\alpha\beta} = R^\gamma{}_{\alpha\gamma\beta} \; . \tag{2.3.11}$$

The scalar curvature R defined as

$$R = g^{\alpha\beta} R_{\alpha\beta} \tag{2.3.12}$$

is also called *the Ricci scalar* in the physics literature. The *Einstein tensor* $G = G_{\alpha\beta} dx^\alpha \otimes dx^\beta$ is defined as

$$G_{\alpha\beta} = R_{\alpha\beta} - \frac{1}{2} R g_{\alpha\beta} \; . \tag{2.3.13}$$

The interest of this object stems from the *divergence identity*,

$$G_\alpha{}^\beta{}_{;\beta} = 0 \; , \tag{2.3.14}$$

which is obtained as follows: contracting the second Bianchi identity (3.3.3) over α and σ gives

$$R^\alpha{}_{\beta\gamma\delta;\alpha} - R_{\beta\delta;\gamma} + R_{\beta\gamma;\delta} = 0 \; . \tag{2.3.15}$$

Contracting this equation with $g^{\beta\gamma}$ yields

$$-R^\alpha{}_{\delta;\alpha} - R^\gamma{}_{\delta;\gamma} + R_{;\delta} = 0 \; . \tag{2.3.16}$$

which is, up to some renaming of indices, Equation (2.3.14).

Chapter 3

An introduction to the physics of the Einstein equations

This chapter constitutes a short introduction to the physical aspects of the Einstein equations. It is self-contained, and no material from here is needed in the remaining parts of this work. It can be safely skipped by the mathematically-minded reader.

3.1 Einstein equations

In 1915 Einstein proposed to describe the gravitational field using a Lorentzian metric tensor g. He postulated the following set of equations:

$$\boxed{R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi\frac{G}{c^4}T_{\mu\nu}} \ . \tag{3.1.1}$$

Here $R_{\mu\nu}$ is the Ricci tensor of the metric g, R is its trace, G is Newton's gravitational constant, Λ is a constant called *the cosmological constant*, $T_{\mu\nu}$ is the energy-momentum tensor of matter fields, and c is the speed of light. From now on we shall assume that a system of units has been chosen so that

$$c = 1 \ .$$

We will give examples of energy-momentum tensors in Section 3.2 below. The simplest case is the vacuum one, where $T_{\mu\nu}$ vanishes identically. If one further assumes that Λ vanishes as well, then (3.1.1) reads

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 0 \ . \tag{3.1.2}$$

Multiplying by $g^{\mu\nu}$ and summing over μ and ν one has, in space-time dimension $n+1$,

$$0 = g^{\mu\nu}\Big(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\Big) = \underbrace{g^{\mu\nu}R_{\mu\nu}}_{=R} - \frac{1}{2}R \underbrace{g^{\mu\nu}g_{\mu\nu}}_{=\delta^\mu_\mu = n+1} = -\frac{n-1}{2}R \ .$$

This leads us to the more usual form of the *vacuum Einstein equations*:

$$R_{\mu\nu} = 0 \; . \tag{3.1.3}$$

Here and throughout we assume that the space-dimension n is strictly larger than one: this assumption is motivated by the fact that if $n = 0$ or 1, then (3.1.2) is identically fulfilled for any metric.

One might try to justify the idea that the gravitational field should be described by a metric tensor. The argument that is usually put forward in this context goes as follows: Recall that in Newton's theory of gravitation the key object is a function ϕ satisfying the Laplace equation in a flat space $\mathbb{R}^3$,

$$\Delta\phi = -4\pi G\mu \; , \tag{3.1.4}$$

where μ is the density of mass of the matter fields. Equation (3.1.4) does not seem to be compatible with special relativity in any simple way. Further, the naive generalization, where Δ is replaced by the wave operator

$$\Box_\eta \phi := \eta^{\mu\nu}\partial_\mu\partial_\nu\phi = -4\pi G\mu \; , \tag{3.1.5}$$

does not work: By Einstein's famous formula $E = mc^2$, the mass density can be identified with the energy density carried by the fields:

$$\rho = \mu c^2 \; . \tag{3.1.6}$$

The simplest description of energy density ρ in special relativity is using vector fields: in this approach ρ is not a scalar, but the time-component of the momentum four-vector field p^μ. But then Equations (3.1.5)-(3.1.6) do not have the right covariance under Lorentz-transformations. Now, a correct behaviour could be restored if gravitation were described by a vector field. Einstein tried to do that, and convinced himself that such a theory would not be compatible with experiment. The next simplest possibility is to consider a theory based on a tensor field, call it $g_{\mu\nu}$. Since the Minkowski metric $\eta_{\mu\nu}$ is a tensor field playing a key role in special relativity, it does not seem absurd to suppose that $g_{\mu\nu}$ could be a generalization of the Minkowski metric. In particular it should be non-degenerate, with Lorentzian signature.

Having decided that a description of gravity using a tensor field $g_{\mu\nu}$ could make sense, one needs a field equation. Einstein's idea that *matter curves space* leads one to look for equations on the Riemann curvature tensor, or its contractions. The fact that (3.1.1) is the right combination can be justified *a posteriori* in many ways. Probably the simplest justification is the variational character of the operator appearing there, see Section **??** for details. Another one is the divergence identity, which we are about to derive. (However, we prefer to think of the divergence identity (3.3.2) and its consequences as a consequence of Einstein equations, rather than a justification for those.) Finally, one can adopt the following point of view: we have the elegant geometric equation (3.1.1), let us study the properties of its solutions. This leads to many difficult and fascinating mathematical problems, which provide a sufficient reason to study this equation. As a bonus one obtains various predictions, which have all turned out to be compatible with experiments so far.

3.2 Energy-momentum tensors

3.2.1 Dust, null fluids

The simplest possible energy-momentum tensor is that describing *dust*: one considers a swarm of particles moving along a family of curves $x(s)$, with tangent vector field

$$u^\mu := \frac{dx^\mu}{ds} \; . \qquad (3.2.1)$$

One assumes that the particles fill a region of space-time, and that their space-time trajectories do not cross each-other, so that (3.2.1) defines a vector field u^μ on that region. One of the postulates of Einstein's theory of gravity is that *physical bodies do not move faster than the speed of light.* (This is a requirement independent of the Einstein equations themselves, and one can in principle consider those equations without imposing this restriction.) From a purely mathematical point of view, the speed-of-light limit is encoded in the equation

$$g(u,u) \le 0 \; . \qquad (3.2.2)$$

The limiting case of equality is only allowed for objects which have *zero rest mass*, such as photons. Thus, when modelling a star, or a galaxy, one assumes that

$$g(u,u) < 0 \; . \qquad (3.2.3)$$

If we change the parameterization of the curve $x(s)$ to $x(s(\tau))$ we will obtain

$$\frac{dx}{d\tau} = \frac{ds}{d\tau}\frac{dx}{ds} \qquad \Longrightarrow \qquad g(\frac{dx}{d\tau},\frac{dx}{d\tau}) = \left(\frac{ds}{d\tau}\right)^2 g(\frac{dx}{ds},\frac{dx}{ds}) \; .$$

It is very convenient to choose the new parameterization τ to be a solution of the equation

$$\frac{d\tau}{ds} = \sqrt{\left|g(\frac{dx}{ds},\frac{dx}{ds})\right|} \; .$$

Redefining u as $dx/d\tau$ we will obtain

$$g(u,u) = -1 \; . \qquad (3.2.4)$$

(In the case of a swarm of photons, which have the property that

$$g(u,u) \equiv 0 \; , \qquad (3.2.5)$$

the freedom of changing the parameterization is not taken care of by the above, and there is no canonical way of choosing that normalisation at this stage.) The energy-momentum tensor for such a physical system is postulated to take the form

$$T_{\mu\nu} = \rho u_\mu u_\nu \; . \qquad (3.2.6)$$

We will talk about dust when (3.2.4) holds, and of *null fluid* when (3.2.5) holds. In the case of dust the function ρ is called the *energy density of the particles in their rest frame.* In the null fluid case the interpretation of ρ as an energy density requires some care, because of the freedom left in rescaling u, but we ignore this for the moment.

3.2.2 Perfect fluids

The perfect fluid is the next simplest physical system. The situation is similar to that of the previous section: one considers a swarm of particles moving along space-time trajectories, with a field of tangents u normalized as in (3.2.4). The fluid description requires the addition of a field p which describes the pressure within the fluid. The corresponding energy-momentum tensor is•3.2.1

•3.2.1: **ptc:** *make a comment about enthalpy, refer to Kijowski*

$$T_{\mu\nu} = (\rho + p)u_\mu u_\nu + p g_{\mu\nu} \ , \tag{3.2.7}$$

The simple form (3.2.7) is related to the fact that in a fluid the pressures are the same in all space-directions. Clearly (3.2.6) is a special case of (3.2.7) with $p = 0$. To complete the description one needs to add an *equation of state*, typically this is done by prescribing p as a function of ρ:

$$p = p(\rho) \ .$$

3.2.3 Field theoretical models

Suppose that we want to describe a theory of matter fields, say φ^A, interacting with the gravitational field. The fields φ^A will of course satisfy their own set of equations. In theoretical physics it is usual to consider equations which are derived from a variational principle, with Lagrangean $\mathscr{L}$. In several cases of interest the Lagrangean takes the simple form

$$\mathscr{L} = \mathscr{L}(\varphi^A, \partial_\mu \varphi^A, g) \ .$$

In such situations one defines the energy-momentum tensor of the fields φ^A as

$$\boxed{T_{\mu\nu} := -\frac{2}{\sqrt{|\det g|}}\frac{\partial \mathscr{L}}{\partial g^{\mu\nu}}} \ . \tag{3.2.8}$$

This definition is related to the existence of a joint variational principle for the gravitational field and the φ^A's, see Section **??**. As a simple example, consider a theory of a scalar field φ with Lagrangean

$$\begin{aligned} \mathscr{L} &= -\frac{\sqrt{|\det g|}}{16\pi}\Big(g^{\alpha\beta}\nabla_\alpha\varphi\nabla_\beta\varphi - V(\varphi)\Big) \\ &= -\frac{\sqrt{|\det g|}}{16\pi}\Big(g^{\alpha\beta}\partial_\alpha\varphi\partial_\beta\varphi - V(\varphi)\Big) \ , \end{aligned} \tag{3.2.9}$$

where V is a given function, called the *potential* function. One talks about *a massless scalar field* when $V \equiv 0$ and about a *a massive scalar field* when $V = -m\varphi^2$. Here m is a constant, which is called the *rest mass*, or the *bare rest mass* of the field. Other potentials are often considered, *e.g.* in the so-called *inflationary cosmological models.*

To calculate the corresponding energy-momentum tensor we need to work out

$$\frac{\partial\sqrt{|\det g|}}{\partial g^{\mu\nu}} \ ,$$

this proceeds as follows: Fix an index μ, then the determinant of the matrix $g_{\alpha\beta}$ can be calculated by expanding in the μ'th column:

$$\det g_{\alpha\beta} = \sum_\nu g_{\mu\nu}\Delta^{\mu\nu}$$

(no summation over μ), where $\Delta^{\mu\nu}$ is the matrix of co-factors. Since $\Delta^{\mu\nu}$ does not involve the $g_{\mu\nu}$ entry of the matrix g, we have

$$\frac{\partial(\det g_{\alpha\beta})}{\partial g_{\mu\nu}} = \Delta^{\mu\nu} \ .$$

Now, the matrix of co-factors is related to the matrix $g^{\mu\nu}$, inverse to $g_{\mu\nu}$, by the formula:

$$\Delta^{\mu\nu} = (\det g_{\alpha\beta})g^{\mu\nu} \ ,$$

so that

$$\frac{\partial(\det g_{\alpha\beta})}{\partial g_{\mu\nu}} = (\det g_{\alpha\beta})g^{\mu\nu} \ . \tag{3.2.10}$$

It then follows that

$$\frac{\partial\sqrt{|\det g_{\alpha\beta}|}}{\partial g_{\mu\nu}} = \frac{1}{2}\sqrt{|\det g_{\alpha\beta}|}g^{\mu\nu} \ . \tag{3.2.11}$$

The identity

$$g^{\alpha\beta}g_{\beta\gamma} = \delta^\alpha_\gamma$$

leads to the equation

$$\frac{\partial}{\partial g^{\mu\nu}} = -g_{\alpha\mu}g_{\beta\nu}\frac{\partial}{\partial g_{\alpha\beta}} \ . \tag{3.2.12}$$

From (3.2.11) and (3.2.12) we finally obtain

$$\frac{\partial\sqrt{|\det g_{\alpha\beta}|}}{\partial g^{\mu\nu}} = -\frac{1}{2}\sqrt{|\det g_{\alpha\beta}|}g_{\mu\nu} \ . \tag{3.2.13}$$

(The above argument can be somewhat simplified if one considers $\det g^{\mu\nu}$ to begin with; we have organized the calculations as above because we will need (3.2.11) later on.)

Returning to the calculation of $T_{\mu\nu}$, it immediately follows from Equations (3.2.8), (3.2.9) and (3.2.11) that the energy-momentum tensor for a scalar field takes the form

$$T_{\mu\nu} = \frac{1}{8\pi}\left\{\nabla_\mu\varphi\nabla_\nu\varphi - \frac{1}{2}\Big(g^{\alpha\beta}\nabla_\alpha\varphi\nabla_\beta\varphi - V(\varphi)\Big)g_{\mu\nu}\right\} \ .$$

3.3 The divergence identity, equations of motion

One of the objects appearing in (3.1.1) is the *Einstein tensor* $G_{\mu\nu}$, defined as

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \ . \tag{3.3.1}$$

This tensor satisfies an important identity, sometimes called the *divergence identity*, which reads

$$\boxed{\nabla_\mu G^{\mu\nu} = 0} \, . \tag{3.3.2}$$

In order to establish this, recall the second Bianchi identity•3.3.1

•3.3.1: **ptc:***what about repetitiveness here?*

$$R^\alpha{}_{\beta\gamma\delta;\sigma} + R^\alpha{}_{\beta\delta\sigma;\gamma} + R^\alpha{}_{\beta\sigma\gamma;\delta} = 0 \, . \tag{3.3.3}$$

Contracting (3.3.3) over α and σ gives

$$R^\alpha{}_{\beta\gamma\delta;\alpha} - R_{\beta\delta;\gamma} + R_{\beta\gamma;\delta} = 0 \, . \tag{3.3.4}$$

Contracting this equation with $g^{\beta\gamma}$ yields

$$-R^\alpha{}_{\delta;\alpha} - R^\gamma{}_{\delta;\gamma} + R_{;\delta} = 0 \, . \tag{3.3.5}$$

This is, up to some renaming and raising of indices, identical with Equation (2.3.14).

Using the Einstein tensor we can rewrite the Einstein equation (3.1.1) as

$$G^{\mu\nu} + \Lambda g^{\mu\nu} = 8\pi \frac{G}{c^4} T^{\mu\nu} \tag{3.3.6}$$

(here, as everywhere else unless explicitly specified otherwise, the indices have been raised using the inverse metric tensor $g^{\mu\nu}$). Taking the divergence of both sides we are led to the identity

$$\nabla_\mu T^{\mu\nu} = 0 \, . \tag{3.3.7}$$

This equation necessarily holds for all solutions of (3.3.6).

While (3.3.7) appears as a trivial identity at first sight, it turns out to contain useful information. We shall illustrate this in the case of the energy-momentum tensor (3.2.6),

$$T^{\mu\nu} = \rho u^\mu u^\nu \, , \tag{3.3.8}$$

with

$$u^\mu u_\mu = \text{const} \, . \tag{3.3.9}$$

Indeed, we calculate

$$0 = \nabla_\mu T^{\mu\nu} = \nabla_\mu(\rho u^\mu) u^\nu + \rho u^\mu \nabla_\mu u^\nu \, . \tag{3.3.10}$$

Let $\gamma^\mu(s)$ be the flow lines of u^μ,

$$u^\mu = \dot\gamma^\mu := \frac{d\gamma^\mu}{ds} \, .$$

In regions where $\rho = 0$ the equation (3.3.10) is not interesting, as no matter is present there. However, at points at which ρ does not vanish (3.3.10) says that

$$\boxed{\nabla_{\dot\gamma}\dot\gamma \sim \dot\gamma} \, . \tag{3.3.11}$$

This means, by definition, that γ is a geodesic of the metric g. We have thus obtained the conclusion that *the flow lines of u^μ are geodesics in space-time.* Equivalently, dust particles move on timelike geodesics in space-time; null fluids move on null geodesics in space-time. This is the precise meaning of the statement that *Einstein equations determine the motion of bodies in general relativity* .

When $u_\mu u^\mu = -1$, as considered in (3.2.4), we can actually obtain a stronger conclusion from (3.3.10): first, we note that the normalization condition (3.3.9) implies

$$0 = \nabla_\mu(u^\nu u_\nu) = \nabla_\mu(g^{\alpha\beta}u_\alpha u_\beta) = g^{\alpha\beta}(\nabla_\mu u_\alpha)u_\beta + g^{\alpha\beta}u_\alpha(\nabla_\mu u_\beta) = 2u^\nu\nabla_\mu u_\nu \; . \tag{3.3.12}$$

Multiplying (3.3.10) by u^ν, the second term there vanishes by (3.3.12) leading to

$$\nabla_\mu(\rho u^\mu) \; . \tag{3.3.13}$$

This equation is often called *the continuity equation*, and expresses the conservation of the total energy contained in volumes which are comoving with the fluid.

Inserting (3.3.13) into (3.3.10) we are finally led to

$$\nabla_u u = 0 \; , \tag{3.3.14}$$

which shows that the integral curves of u are *affinely parameterized geodesics.*

3.3.1 Geometric optics in the Schwarzschild metric

In the case of a null fluid the argument leading to (3.3.14) breaks down because $u^\mu u_\mu = 0$, and the multiplication of (3.3.10) by u^μ does not lead to any new information. This does, however, not affect the conclusion that a null fluid moves along null geodesics, because the vanishing of $u^\mu u_\mu$ does not affect the derivation of (3.3.11). Now, the interest of null fluids arises from the fact that *null fluid energy-momentum tensors occur in the geometric optics limit* of the Einstein-Maxwell equations.[1] Thus Equation (3.3.11) shows that, in the geometric optics limit, light propagates on null geodesics. This leads to various spectacular effects which can be already seen in spherical symmetry. Recall that one of the simplest non-trivial solutions of the vacuum Einstein equations is the Schwarzschild metric, defined on $\{(t, r, \theta, \varphi) \in \mathbb{R} \times (2m, \infty) \times S^2\}$ by the formula•3.3.2 •*3.3.2:* **ptc:***refer to where it is derived*

$$g = -(1 - \frac{2m}{r})dt^2 + \frac{1}{1 - \frac{2m}{r}}dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2) \; . \tag{3.3.15}$$

This metric models the external gravitational field of a spherically symmetric body. The parameter m is a constant which has the interpretation of the active gravitational mass of the body (compare Section 7.1). In Figure 3.1 one can see how the curvature of space-time affects the image of a picture of Einstein which has been placed behind, and slightly sideways, of a spherically symmetric star. The curving of geodesics by the geometry results in the existence of multiple images of a single object under certain circumstances. Because of the characteristic arc-shaped form, the resulting images are often called *Einstein arcs.* In Figure 3.2 one observes that

[1] An exact mathematical treatment of some of the issues arising here can be found in the PhD Thesis of Jeanne [**?**].

Figure 3.1: Gravitational lensing by a spherically symmetric star: Einstein arcs. The mass parameter of the metric is $M = 1.5$ in geometrical units $G = c = 1$. The distance of Einstein to the gravitational lens is approximately equal to $25M$; that of the observer is approximately $70M$. The dimensions of the picture are $8M \times 12M$. Figures 3.1 and 3.2 have been obtained by Daniel Weisskopf and Hans Ruder at the University of Tuebingen. The calculation consists of numerically tracing rays along null geodesics in the Schwarzschild geometry.

Figure 3.2: Gravitational lensing by a spherically symmetric star: an Einstein ring. The parameters are the same as in Figure 3.1, except that now the picture of Einstein is directly behing the star.

same image, when placed directly behind the star. If space-time were flat, we would not see an image at all, because the light rays would propagate along straight lines, and the image would have been hidden by the star. In curved space-time we can actually see what happens in a certain region directly behind the star, this results in the formation of *Einstein rings*. Spectacular visualizations of how the image is deformed when the picture is moved from right to left behind the star can be found on `http://www.tat.physik.uni-tuebingen.de/~tmueller`.

The effects observed in Figures 3.1 and 3.2 are an example of a general phenomenon referred to as *gravitational lensing*, an extensive treatment of the subject can be found in [**?**, 92, **?**]. One of the important effects that arise here is a magnification of the image, allowing one to see objects which otherwise would have been much too faint for observations. By now several Einstein arcs have been seen by astrophysicists, Figure 3.3•3.3.3 provide examples of two such observations made by the Hubble space telescope.

•*3.3.3:* **ptc:** *The Hubble image needs improvement*

3.4 Weak gravitational fields

3.4.1 Small perturbations of Minkowski space-time

Consider $\mathbb{R}^{n+1}$ with a metric which in the natural coordinates on $\mathbb{R}^{n+1}$ takes the form

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \ , \tag{3.4.1}$$

and suppose that there exists a small constant ϵ such that we have

$$|h_{\mu\nu}| \ , \ |\partial_\sigma h_{\mu\nu}| \ , \ |\partial_\sigma \partial_\rho h_{\mu\nu}| \le \epsilon \ . \tag{3.4.2}$$

It is then easy to check that

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} + O(\epsilon^2) \ . \tag{3.4.3}$$

Throughout this section we use the metric η to raise and lower indices, *e.g.*,

$$h^\alpha{}_\beta := \eta^{\alpha\mu} h_{\mu\beta} \ , \qquad h^{\alpha\beta} := \eta^{\alpha\mu}\eta^{\beta\nu} h_{\mu\nu} = \eta^{\beta\nu} h^\alpha{}_\nu \ .$$

Figure 3.3: Einstein arcs in the Galaxy Cluster Abbel 2218 (from the STScI Public Archive [?]).

Further,

$$\begin{aligned}
\Gamma^{\alpha}{}_{\beta\gamma} &= \frac{1}{2}g^{\alpha\sigma}\{\partial_\beta g_{\sigma\gamma} + \partial_\gamma g_{\sigma\beta} - \partial_\sigma g_{\beta\gamma}\} \\
&= \frac{1}{2}g^{\alpha\sigma}\{\partial_\beta h_{\sigma\gamma} + \partial_\gamma h_{\sigma\beta} - \partial_\sigma h_{\beta\gamma}\} \\
&= \frac{1}{2}\eta^{\alpha\sigma}\{\partial_\beta h_{\sigma\gamma} + \partial_\gamma h_{\sigma\beta} - \partial_\sigma h_{\beta\gamma}\} + O(\epsilon^2) \\
&= \frac{1}{2}\{\partial_\beta h^{\alpha}{}_{\gamma} + \partial_\gamma h^{\alpha}{}_{\beta} - \partial^\alpha h_{\beta\gamma}\} + O(\epsilon^2) = O(\epsilon)\ . \qquad (3.4.4)
\end{aligned}$$

The Ricci tensor is easily found from (A.5.3)

$$\begin{aligned}
R_{\beta\delta} &= \partial_\alpha \Gamma^{\alpha}{}_{\beta\delta} - \partial_\delta \Gamma^{\alpha}{}_{\beta\alpha} + O(\epsilon^2) \\
&= \frac{1}{2}\left[\partial_\alpha\{\partial_\beta h^{\alpha}{}_{\delta} + \partial_\delta h^{\alpha}{}_{\beta} - \partial^\alpha h_{\beta\delta}\} - \partial_\delta\{\partial_\beta h^{\alpha}{}_{\alpha} + \partial_\alpha h^{\alpha}{}_{\beta} - \partial^\alpha h_{\beta\alpha}\}\right] + O(\epsilon^2) \\
&= \frac{1}{2}\left[\partial_\alpha\{\partial_\beta h^{\alpha}{}_{\delta} + \partial_\delta h^{\alpha}{}_{\beta} - \partial^\alpha h_{\beta\delta}\} - \partial_\delta\partial_\beta h^{\alpha}{}_{\alpha}\right] + O(\epsilon^2)\ . \qquad (3.4.5)
\end{aligned}$$

3.4.2 Coordinate conditions, wave coordinates

The expression (3.4.5) for the Ricci tensor is still more complicated than desired: we will be interested in solving, *e.g.*, the vacuum equation $R_{\mu\nu} = 0$, and it is far from clear how that can be done using (3.4.5). It turns out that one can obtain considerable simplifications if one imposes a set of *coordinate conditions*. Recall that a tensor field g is represented by matrices $g_{\mu\nu}$ in many different ways, depending upon the coordinate system chosen: if a point p has coordinates y^μ in a coordinate system $\{y^\mu\}$, and coordinates x^α in a second coordinate system $\{x^\alpha\}$, then we have

$$\begin{aligned}
g_p = g_{\mu\nu}(y^\sigma)dy^\mu dy^\nu &= g_{\mu\nu}(y^\sigma(x^\alpha))\left(\frac{\partial y^\mu}{\partial x^\beta}dx^\beta\right)\left(\frac{\partial y^\nu}{\partial x^\gamma}dx^\gamma\right) \\
&= g_{\mu\nu}(y^\sigma(x^\alpha))\frac{\partial y^\mu}{\partial x^\beta}\frac{\partial y^\nu}{\partial x^\gamma}dx^\beta dx^\gamma\ , \\
&= g_{\beta\gamma}(x^\alpha)dx^\beta dx^\gamma
\end{aligned}$$

so that

$$\boxed{g_{\beta\gamma}(x^\alpha) = g_{\mu\nu}(y^\sigma(x^\alpha))\frac{\partial y^\mu}{\partial x^\beta}\frac{\partial y^\nu}{\partial x^\gamma}}\ . \qquad (3.4.6)$$

One can make use of this transformation law to obtain a form of the matrix $g_{\mu\nu}$ which is convenient for the problem at hand. (This property is referred to as *"gauge freedom"* in the physics literature.) For example, to show that the Riemann tensor of the Minkowski metric vanishes it is best to use a coordinate system in which all the $g_{\mu\nu}$'s are constants, hence the $\Gamma^\alpha{}_{\beta\gamma}$'s vanish, which obviously implies the result. On the other hand, the same result in spherical coordinates requires a lengthy calculation.

A choice of coordinate system, which is useful for many purposes, is that of *wave coordinates*, sometimes also referred to as *harmonic coordinates*; physicists also talk of the *de Donder gauge* in this context: One requires that the coordinate functions be solutions of the wave equation:

$$\Box_g x^\alpha = 0 \,, \tag{3.4.7}$$

where $\Box_g$ is the wave operator associated with the metric g; in (any) local coordinates:

$$\Box_g f := g^{\mu\nu}\nabla_\mu\nabla_\nu f \,. \tag{3.4.8}$$

For further purposes it is convenient to rewrite (3.4.8) as

$$\Box_g f = \frac{1}{\sqrt{|\det g_{\mu\nu}|}}\partial_\rho\left(\sqrt{|\det g_{\mu\nu}|}g^{\rho\sigma}\partial_\sigma f\right) \,. \tag{3.4.9}$$

In order to show that this formula is indeed correct, we calculate

$$\begin{aligned}
\Box_g f &= g^{\mu\nu}\Big(\partial_\mu\partial_\nu f - \Gamma^\gamma{}_{\mu\nu}\partial_\gamma f\Big) \\
&= g^{\mu\nu}\Big(\partial_\mu\partial_\nu f - \frac{1}{2}g^{\gamma\sigma}(\partial_\mu g_{\sigma\nu} + \partial_\nu g_{\sigma\mu} - \partial_\sigma g_{\mu\nu})\partial_\gamma f\Big) \\
&= g^{\mu\nu}\Big(\partial_\mu\partial_\nu f - \frac{1}{2}g^{\gamma\sigma}(2\partial_\mu g_{\sigma\nu} - \partial_\sigma g_{\mu\nu})\partial_\gamma f\Big) \\
&= g^{\sigma\gamma}\partial_\sigma\partial_\gamma f - \Big(\underbrace{g^{\mu\nu}g^{\gamma\sigma}\partial_\mu g_{\sigma\nu}}_{a} - \frac{1}{2}g^{\gamma\sigma}\underbrace{g^{\mu\nu}\partial_\sigma g_{\mu\nu}}_{b}\Big)\partial_\gamma f \,.
\end{aligned} \tag{3.4.10}$$

Differentiating the identity

$$g^{\gamma\sigma}g_{\sigma\nu} = \delta^\gamma_\nu$$

we obtain

$$g^{\gamma\sigma}\partial_\mu g_{\sigma\nu} = -g_{\sigma\nu}\partial_\mu g^{\gamma\sigma} \,. \tag{3.4.11}$$

It follows that the function a of (3.4.10) equals

$$a = -g^{\mu\nu}g_{\sigma\nu}\partial_\mu g^{\gamma\sigma} = -\partial_\sigma g^{\gamma\sigma} \,. \tag{3.4.12}$$

Further, (3.2.11) gives

$$b = g^{\mu\nu}\partial_\sigma g_{\mu\nu} = \frac{2}{\sqrt{|\det g_{\alpha\beta}|}}\partial_\sigma\Big(\sqrt{|\det g_{\alpha\beta}|}\Big) \,. \tag{3.4.13}$$

Inserting all this into (3.4.10) we obtain

$$\begin{aligned}
\Box_g f &= g^{\sigma\gamma}\partial_\sigma\partial_\gamma f + \Big(\partial_\sigma g^{\gamma\sigma} + g^{\gamma\sigma}\frac{1}{\sqrt{|\det g_{\alpha\beta}|}}\partial_\sigma\Big(\sqrt{|\det g_{\alpha\beta}|}\Big)\Big)\partial_\gamma f \\
&= \frac{1}{\sqrt{|\det g_{\mu\nu}|}}\partial_\rho\left(\sqrt{|\det g_{\mu\nu}|}g^{\rho\sigma}\partial_\sigma f\right) \,,
\end{aligned} \tag{3.4.14}$$

as claimed.

(Local) solutions of (3.4.8) are easily constructed as follows: one chooses any spacelike hypersurface $\mathscr{S} \subset \mathscr{M}$ — by definition, this means that the metric γ induced on $\mathscr{S}$ from g is Riemannian; the explicit formula for γ reads

$$\forall\ X, Y \in T\mathscr{S} \quad \gamma(X,Y) := g(X,Y)\ .$$

As g is Lorentzian, equation (3.4.8) is a second order hyperbolic equation. The standard theory of hyperbolic PDE's (*cf., e.g.* [103]) asserts that for any functions $k, h : \mathscr{S} \to \mathbb{R}$ and any vector field X defined along $\mathscr{S}$ and *transverse* to $\mathscr{S}$ there exists a neighborhood of $\mathscr{S}$ and a unique solution of the wave equation $\Box_g f = 0$ satisfying

$$f|_{\mathscr{S}} = k\ , \qquad X(f)|_{\mathscr{S}} = h\ .$$

So, in order to construct wave coordinates around a point p one chooses any spacelike $\mathscr{S}$ passing through p, together with a coordinate patch $\mathscr{U}$ on $\mathscr{S}$ with coordinates $\{y^i\}$. Replacing $\mathscr{S}$ by $\mathscr{U}$ one can without loss of generality assume that $\mathscr{S} = \mathscr{U}$. Then one solves the Cauchy problem

$$x^0|_{\mathscr{S}} = 0\ , \qquad n(x^0)|_{\mathscr{S}} = 1\ , \tag{3.4.15a}$$

$$x^i|_{\mathscr{S}} = y^i\ , \qquad n(x^i)|_{\mathscr{S}} = 0 \quad , \tag{3.4.15b}$$

where n is the field of unit-normals to $\mathscr{S}$. Let $\mathscr{O} \subset \mathscr{M}$ denote the neigbhourhood of $\mathscr{S}$ on which the solution exists: (3.4.15) shows that for any coordinate system y^α around p the matrix $\partial x^\mu / \partial y^\alpha$ will be non-degenerate on $\mathscr{S}$. It follows that, passing to a subset of $\mathscr{O}$ if necessary, the x^μ's will form a coordinate system on $\mathscr{O}$.

3.4.3 Linearized Einstein equations in wave coordinates

Let us return to metrics of the form (3.4.1) satisfying (3.4.2). As explained in the previous section we can choose a coordinate system in which the coordinate functions will satisfy the wave equation (3.4.7). We wish to show that the expression (3.4.5) for the Ricci tensor simplifies considerably when wave coordinates are chosen. Indeed, it then follows from (3.4.9) that we have

$$\begin{aligned} 0 &= \Box_g x^\alpha \\ &= \frac{1}{\sqrt{|\det g_{\mu\nu}|}} \partial_\rho \Big(\sqrt{|\det g_{\mu\nu}|} g^{\rho\sigma} \underbrace{\partial_\sigma x^\alpha}_{\delta^\alpha_\sigma} \Big) \\ &= \frac{1}{\sqrt{|\det g_{\mu\nu}|}} \partial_\rho \Big(\sqrt{|\det g_{\mu\nu}|} g^{\rho\alpha} \Big)\ . \end{aligned} \tag{3.4.16}$$

We need to calculate this expression up terms of order ϵ^2. In order to do this, we first use (3.2.11) to obtain

$$\begin{aligned} \sqrt{|\det g_{\mu\nu}|} &= \sqrt{|\det g_{\mu\nu}|}\Big|_{g=\eta} + \frac{\partial \sqrt{|\det g_{\mu\nu}|}}{\partial g_{\alpha\beta}}\Big|_{g=\eta} h_{\alpha\beta} + O(\epsilon^2) \\ &= 1 + \frac{1}{2}\eta^{\alpha\beta} h_{\alpha\beta} + O(\epsilon^2) \\ &= 1 + \frac{1}{2} h^\alpha{}_\alpha + O(\epsilon^2)\ . \end{aligned}$$

This, together with (3.4.3) allows us to write

$$\begin{aligned}
0 &= \partial_\rho\Big(\sqrt{|\det g_{\mu\nu}|}g^{\rho\alpha}\Big) \\
&= \partial_\rho\Big((1+\frac{1}{2}h^\beta{}_\beta)(\eta^{\rho\alpha}-h^{\rho\alpha})\Big)+O(\epsilon^2) \\
&= \frac{1}{2}\partial^\alpha h^\beta{}_\beta - \partial_\rho h^{\rho\alpha} + O(\epsilon^2)\ . \qquad (3.4.17)
\end{aligned}$$

Equivalently,

$$\partial_\rho h^\rho{}_\alpha = \frac{1}{2}\partial_\alpha h^\beta{}_\beta + O(\epsilon^2)\ . \qquad (3.4.18)$$

This allows us to rewrite (3.4.5) as

$$\begin{aligned}
R_{\beta\delta} &= \frac{1}{2}\left[\partial_\alpha\{\partial_\beta h^\alpha{}_\delta+\partial_\delta h^\alpha{}_\beta-\partial^\alpha h_{\beta\delta}\}-\partial_\delta\partial_\beta h^\alpha{}_\alpha\right]+O(\epsilon^2) \\
&= \frac{1}{2}\Big[\partial_\beta \underbrace{\partial_\alpha h^\alpha{}_\delta}_{=-\partial_\delta h^\alpha{}_\alpha/2} +\partial_\delta \underbrace{\partial_\alpha h^\alpha{}_\beta}_{=-\partial_\beta h^\alpha{}_\alpha/2} -\partial_\alpha\partial^\alpha h_{\beta\delta}-\partial_\delta\partial_\beta h^\alpha{}_\alpha\Big]+O(\epsilon^2) \\
&= -\frac{1}{2}\partial_\alpha\partial^\alpha h_{\beta\delta}+O(\epsilon^2) \\
&= -\frac{1}{2}\Box_\eta h_{\beta\delta}+O(\epsilon^2)\ . \qquad (3.4.19)
\end{aligned}$$

It follows that — up to higher order terms and a constant multiplicative factor — the Ricci tensor is the Minkowski wave operator acting on h:

$$\boxed{R_{\alpha\beta} = -\tfrac{1}{2}\Box_\eta h_{\alpha\beta}} + O(\epsilon^2)\ . \qquad (3.4.20)$$

3.4.4 Gravitational radiation?

•3.4.1

•*3.4.1:* **ptc:***include some photographs of detectors*

It is well know that solutions of the scalar wave equation on Minkowski space-time carry away energy. For instance, one can show [**?**, 105] (see also Section **??**) that a solution ϕ of the wave equation on Minkowski space-time with smooth initial data which are compactly supported at $t=0$ will behave as, for $r\ge 1$, $t\ge 0$,

$$\phi = \frac{c(t-r,\theta,\varphi)}{r}+O(r^{-2})\ , \qquad (3.4.21)$$

with $O(r^{-2})$ here being understood as follows:

$$f(t,r,\theta,\varphi)=O(r^{-2}) \qquad \Longleftrightarrow \qquad \sup_{r,\theta,\varphi}(1+r)^2|f(t+r,r,\theta,\varphi)|<\infty\ . \qquad (3.4.22)$$

The behaviour (3.4.22) is further preserved under differentiation in the obvious way. We emphasize that the time variable t is shifted by r when increasing r in the right-hand-side of (3.4.22) — this corresponds to moving along outgoing light cones $\{t=u+r, r>0\}$ in Minkowski space-time with u fixed. Standard field-theoretic considerations (*cf., e.g.* [34]) show then that the system radiates away energy along those cones at a rate

$$\frac{dE}{dt}(u) = \frac{1}{8\pi}\int_{S^2}\left(\frac{\partial c}{\partial u}(u,\theta,\varphi)\right)^2\sin\theta d\theta d\varphi\ .$$

Basing on this behaviour, one expects that there should exist a large class of solutions of the vacuum Einstein equation which will be going to zero at large distances as $1/r$ when tending with r to infinity along light cones. Then the small parameter ϵ in (3.4.1) can be thought of as $1/r$, and the terms $O(\epsilon^2)$ in (3.4.19) being thus $O(r^{-2})$. This suggests that the approximation done by neglecting the nonlinear terms in (3.4.22) should be better and better when one moves away from the gravitating system along light cones, so that the approximation of the gravitational field as simply solving

$$\Box_\eta h_{\alpha\beta} = 0$$

should be a very good one at large distances. One then expects that isolated systems radiate energy in the form of gravitational waves in a way roughly analogous to that for scalar fields, or to Maxwell fields.

•3.4.2 One can develop approximation schemes which take into account the quadratic and higher terms in h, obtaining approximate solutions of the Einstein equations describing, *e.g.*, a system of two stars orbiting each other. Such a program has been already undertaken by Einstein, and is still being pursued by several researchers [**?**, **?**]. The mathematical status of the calculations which are done in this context is far from satisfactory. Nevertheless the calculations along those lines of Damour [**?**] have found a beautiful experimental confirmation in the observations by Taylor and Hulse of the millisecond pulsar PSR 1913+16, rewarded by a Nobel price in 1993.[2] •*3.4.2:* **ptc:***side material*

The above heuristic discussion has a lot of ifs attached to it: First, it is easy to construct initial data which satisfy the smallness hypotheses (3.4.2) on compact sets, in which case (3.4.20) is perfectly justified. However, to obtain the wave behavior described by (3.4.21) one needs to take limits $r \to \infty$, and it is not clear whether there will be large classes of solutions which do satisfy (3.4.20) on the relevant regions. Next, to obtain (3.4.27)-(3.4.22) we have assumed that the initial data for the scalar field have compact support. This assumption is not compatible with the relativistic constraint equations, discussed in detail in Chapter 6.7, at least if one assumes that there are no singularities in the initial data set[3]. It is clear that there will be large classes of initial data for the scalar field on the hypersurface $\{t = 0\}$ for which the asymptotic behavior (3.4.21)-(3.4.22) will be correct, but there are no rigorous general results of this kind available in the literature for fields which are not compactly supported. Further, the behavior of light-cones in a curved space-time is different from that of cones in Minkowski space-time. This might imply that the replacement of the wave operator by a flat one, which has been effectively done in (3.4.20), is too drastic to be correct. Finally, one needs to show that the non-linearities do indeed lead to lower order terms from a dynamical point of view, assuming the light-cone problem has been appropriately taken care of. The work of Friedrich [46], of Christodoulou and Klainerman [24], and of Klainerman and Nicolo [71] shows that the naive linearized picture is close to being correct. However, our understanding of all the mathematical issues that arise here is far from being complete, and there is considerable ongoing effort in trying to understand the properties of the gravitational field in the radiation regime, see [35, 47, **?**, **?**, 48] and references therein.

---

[2] An excellent elementary description of the PSR 1913+16 pulsar and of the Taylor-Hulse observations can be found at `http://astrosun.tn.cornell.edu/courses/astro201/psr1913.htm`.

[3] It follows from the positive energy theorem in Chapter 7 that the evolution of non-singular initial data with fall-off faster than $1/r$ will — under the hypotheses of that theorem — be a subset of Minkowski space-time.

3.4.5 Newton's equations of motion, and why $8\pi G$ is $8\pi G$

Einstein's theory is supposed to be a theory of gravitation. We already have one such theory, due to Newton, which works pretty well in several situations. It would thus be desirable if Einstein's theory contained Newton's theory in some limit. This is indeed the case, and can be easily established using the calculations done so far.

There are a few conditions which should obviously hold when trying to recover Newton's theory: since that last theory is a linear one, and Einstein's is not, the gravitational field should be sufficiently weak in order that the non-linearities do not matter. This is taken care of by the parameter ϵ in (3.4.1). Next, the wave operator arising in equation (3.4.21) leads to radiation phenomena when systems with bodies with large relative velocities are considered. On the other hand, Newton's equation

$$\Delta_\delta \phi = -4\pi G \mu \ , \tag{3.4.23}$$

does not exhibit any wave behaviour. In (3.4.23) μ is the matter density, ϕ is the Newtonian potential, Δ_δ is the Euclidean Laplace operator, and G is Newton's constant. This suggests that a regime in which approximate agreement can be obtained is one where time derivatives are small:

$$\partial_t h_{\mu\nu} = O(\epsilon) \ , \quad \partial_t \partial_\alpha h_{\mu\nu} = O(\epsilon) \ . \tag{3.4.24}$$

We consider thus a space-time containing a body made of dust with small energy-density,

$$T_{\mu\nu} = \rho u_\mu u_\nu \ , \quad \rho = O(\epsilon) \ .$$

The body is assumed to be moving slowly,

$$u_\mu dx^\mu = u_0 dt + O(\epsilon) \quad \Longleftrightarrow \quad u_i = O(\epsilon) \ . \tag{3.4.25}$$

We assume Einstein's equation describing the equivalence of *active gravitational mass density* — by definition, this is the function μ which appears in (3.4.23) — and of the energy density:

$$\rho = \mu c^2 = \mu \tag{3.4.26}$$

(recall that we are using units in which the speed of light c equals one). We note that in the calculations below it would suffice that (3.4.26) holds up to terms $O(\epsilon^2)$. Systematically neglecting all ϵ^2's one has, from (3.4.20),

$$\begin{aligned} R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} &= -\frac{1}{2}\Box_\eta \Big(h_{\mu\nu} - \frac{1}{2} h^\alpha{}_\alpha \eta_{\mu\nu} \Big) \\ &= -\frac{1}{2}\Big(\underbrace{-\partial_t^2 + \Delta_\delta}_{\text{reduces to } \Delta_\delta \text{ by (3.4.24)}} \Big)\Big(h_{\mu\nu} - \frac{1}{2} h^\alpha{}_\alpha \eta_{\mu\nu} \Big) \\ &= -\frac{1}{2}\Delta_\delta \Big(h_{\mu\nu} - \frac{1}{2} h^\alpha{}_\alpha \eta_{\mu\nu} \Big) \\ &= \lambda \rho u_\mu u_\nu \ . \end{aligned} \tag{3.4.27}$$

Here we have written the Einstein equations as

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \lambda T_{\mu\nu} \, , \tag{3.4.28}$$

and one of the goals of our calculation will be to determine the constant λ. Multiplying (3.4.27) by $2\eta^{\mu\nu}$ we find (recall that $u^\mu u_\mu = -1 + O(\epsilon)$ and $\eta^{\mu\nu}\eta_{\mu\nu} = \delta^\mu_\mu = 4$)

$$-\eta^{\mu\nu}\Delta_\delta\Big(h_{\mu\nu} - \frac{1}{2}h^\alpha{}_\alpha\eta_{\mu\nu}\Big) = \Delta_\delta h^\alpha{}_\alpha = -2\lambda\rho \, . \tag{3.4.29}$$

If we consider a bounded body, so that ρ is compactly supported, then there exists a unique solution $h^\alpha{}_\alpha$ of that equation which approaches 0 at large distance, which is proportional to the Newtonian potential of ρ:

$$h^\alpha{}_\alpha = \frac{2\lambda}{4\pi G}\phi \, . \tag{3.4.30}$$

Note that for compatibility we should have ϕ of order $O(\epsilon)$.

Next, since $u_i = O(\epsilon)$ we have

$$\Delta_\delta h_{0i} = 0 \, , \qquad \Delta_\delta\Big(h_{ij} - \frac{1}{2}h^\alpha{}_\alpha\delta_{ij}\Big) = 0 \, ,$$

and if we consider only solutions $h_{\mu\nu}$ which approach zero as r goes to infinity, then the maximum principle gives

$$h_{0i} = 0 \, , \qquad h_{ij} = \frac{1}{2}h^\alpha{}_\alpha\delta_{ij} = \frac{\lambda}{4\pi G}\phi\delta_{ij} \, . \tag{3.4.31}$$

Summing over i and j gives

$$\sum_{i=1}^{3} h_{ii} = \frac{3\lambda}{4\pi G}\phi \, ,$$

so that

$$h^\alpha{}_\alpha = -h_{00} + \sum_{i=1}^{3} h_{ii} = -h_{00} + \frac{3\lambda}{4\pi G}\phi \, .$$

Comparing with (3.4.30) leads to

$$h_{00} = \frac{\lambda}{4\pi G}\phi \, , \tag{3.4.32}$$

which together with (3.4.31) completes the solution of our problem. To be consistent, in all the equations above the zeros should be replaced by $O(\epsilon^2)$.

When the configuration of the system is bounded in space and has finite total mass M we have•3.4.3

•*3.4.3:* **ptc:***this looks strange, because it is conformally flat*

$$\phi = \frac{GM}{r} + O(r^{-2})$$

in the vacuum region, which leads to the following asymptotic form of the gravitational field

$$g_{00} = -1 + \frac{\lambda M}{4\pi r} + O\left(\left(\frac{M}{r}\right)^2\right) \;, \quad g_{0i} = O\left(\left(\frac{M}{r}\right)^2\right) \;,$$
$$g_{ij} = \Big(1 - \frac{\lambda M}{4\pi r}\Big)\delta_{ij} + O\left(\left(\frac{M}{r}\right)^2\right) \;. \qquad (3.4.33)$$

At this stage one should redo the whole calculation using (3.4.33) as a starting point, to make sure that the final result is consistent with the remaining calculations and hypotheses — this turns out to be the case.

We note that for static vacuum gravitational fields one can obtain complete asymptotic expansions by a recursive use of the Einstein equations, the reader is referred to [18] for details.

3.4.6 Determining the coupling constant: the geodesic equation

In Section 3.3 we have shown that the integral curves $x^\mu(s)$ of the vector field u^μ appearing in (3.4.25) are affinely parameterized timelike geodesics:

$$\frac{dx^\alpha}{ds} = u^\alpha \;, \qquad \frac{d^2x^\mu}{ds^2} = -\Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} \;. \qquad (3.4.34)$$

We wish to use this fact to derive the equations of motion of the dust particles considered in the last section. In order to do that we calculate

$$\frac{dx^\alpha}{ds} = u^\alpha = g^{\alpha\beta}u_\beta = \eta^{\alpha\beta}u_\beta + O(\epsilon) = \eta^{\alpha 0}u_0 + O(\epsilon) \;,$$

leading to

$$\frac{dx^0}{ds} = 1 + O(\epsilon) \;, \qquad \frac{dx^i}{ds} = O(\epsilon) \;.$$

This implies that

$$\Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} = \Gamma^\mu{}_{00}\frac{dx^0}{ds}\frac{dx^0}{ds} + O(\epsilon) = \Gamma^\mu{}_{00} + O(\epsilon) \;.$$

In order to calculate the space-acceleration d^2x^i/ds^2 of the particles it remains to calculate $\Gamma^i{}_{00}$:

$$\Gamma^i{}_{00} = \frac{1}{2}g^{i\sigma}\left(2\partial_0 g_{\sigma 0} - \partial_\sigma g_{00}\right) = -\frac{1}{2}\partial_i h_{00} + O(\epsilon) \;.$$

From (3.4.32) we thus obtain

$$\Gamma^i{}_{00} = -\frac{\lambda}{8\pi G}\partial_i\phi \;.$$

It follows that

$$\frac{d^2x^i}{ds^2} = \frac{\lambda}{8\pi G}\partial_i\phi \;, \qquad (3.4.35)$$

which is identical with Newton's equations of motion,

$$\frac{d^2x^i}{ds^2} = \partial_i \phi \, , \tag{3.4.36}$$

if and only if

$$\lambda = 8\pi G \, . \tag{3.4.37}$$

We have thus shown that the choice (3.4.37) of the constant appearing in (3.4.28) leads to a theory which reproduces Newtonian's theory of gravitation, in the limit of weak fields, for slowly moving low density bodies made of dust.

It should be borne in mind that all our arguments above have been carried through at a somewhat heuristic level. The problem is that we have considered metrics defined on a neighborhood of the hypersurface $\{t = 0\}$ in $\mathbb{R}^4$. As already pointed out in the gravitational radiation context, hypotheses (3.4.2) on non-compact subsets of space-time need justification. A rigourous treatment would require careful estimates, to show that the terms which we have neglected can indeed be neglected. This can be done in some situations.

We close this section by noting the elegant framework of J. Ehlers [**?**] which geometrizes the Newtonian limit of Einstein's theory. This approach has been used by Heilig [60] to construct general relativistic axially-symmetric stationary star models, using the implicit function theorem in a neighborhood of the corresponding Newtonian solutions of Lichtenstein. See also Rendall [**?**].

Chapter 4

Causality

•4.0.4

•4.0.4: **ptc:**do something about manifolds with boundary????

4.1 Time orientation

•4.1.1 Recall that at each point $p \in \mathscr{M}$ the set of timelike vectors in T_pM has precisely two components. A *time-orientation of* $T_p\mathscr{M}$ is the assignment of the name "future pointing vectors" and "past pointing vectors" to each of those components. The set of future pointing timelike, or causal, vectors, is stable under addition and multiplication by positive numbers; similarly for past pointing ones. (In particular this implies convexity.) In order to see this, suppose that $X = (X^0, \vec{X})$ and $Y = (Y^0, \vec{Y})$ are timelike future pointing, in an ON-frame this is equivalent to

•4.1.1: **ptc:**check Seifert's approach to causality; compare defs; see if he can go along with continuous metrics?

$$|\vec{X}| < X^0 \ , \quad |\vec{Y}| < Y^0 \ ,$$

and the inequality

$$|\vec{X} + \vec{Y}| \le |\vec{X}| + |\vec{Y}| < X^0 + Y^0$$

follows. Two timelike vectors X and Y have the same time orientation if and only if

$$g(X, Y) < 0 \ ; \tag{4.1.1}$$

this follows immediately from (2.2.1) in an ON frame in which X is proportional to e_0.

A time-orientation of $T_p\mathscr{M}$ can always be propagated to a neighborhood of p by choosing any continuous vector field X defined around p which is timelike and future pointing at p. By continuity of the metric and of X, the vector field X will be timelike in a sufficiently small neighborhood $\mathscr{O}_p$ of p, and for $q \in \mathscr{O}_p$ one can define future pointing vectors at q as those lying in the same component of the set of timelike vectors as $X(q)$: for $q \in \mathscr{O}_p$ the vector $Y \in T_q\mathscr{M}$ will be said to be timelike future pointing if and only if $g(Y, X(q)) < 0$. A Lorentzian manifold is said to be *time orientable* if such locally defined time-orientations can be defined globally in a consistent way; that is, we can cover $\mathscr{M}$ by coordinate neighborhoods $\mathscr{O}_p$, each equipped with a vector field $X_{\mathscr{O}_p}$, such that $g(X_{\mathscr{O}_p}, X_{\mathscr{O}_q}) < 0$ on $\mathscr{O}_p \cap \mathscr{O}_q$.

Figure 4.1: The Möbius strip, with the flat metric is $-dt^2 + dx^2$ (so that the light cones are at 45^o) provides an example of a two dimensional Lorentzian manifold which is not time orientable.

Some manifolds will not be time orientable, as is shown by the flat metric[1] on the Möbius strip, *cf.* Figure 4.1. On a time-orientable manifold there are precisely two choices of time-orientation possible, and $(\mathscr{M}, g)$ will be said *time oriented* when such a choice has been done. This leads us to the fundamental definition:

DEFINITION 4.1.1 *A couple $(\mathscr{M}, g)$ will be called a* space-time *if $(\mathscr{M}, g)$ is a time-oriented Lorentzian manifold.*

A Lorentzian manifold which is *not* time orientable has a double cover which is, *cf., e.g.* [16][•4.1.2] for a proof.

•*4.1.2: do the proof*

On any space-time there always exists a globally defined future directed timelike vector field — to show this, consider the locally defined timelike vector fields $X_{\mathscr{O}_p}$ defined on neighborhoods $\mathscr{O}_p$ as described above. One can choose a locally finite covering of $\mathscr{M}$ by such neighborhoods $\mathscr{O}_{p_i}$, $i \in \mathbb{N}$, and construct a globally defined vector field X on $\mathscr{M}$ by setting

$$X = \sum_i \phi_i X_{\mathscr{O}_{p_i}} \ ,$$

where the functions ϕ_i form a partition of unity associated to the covering $\{\mathscr{O}_{p_i}\}_{i \in \mathbb{N}}$. The resulting vector field will be timelike future pointing everywhere, as follows from the fact that the sum of an arbitrary number of future pointing timelike vectors is a future pointing timelike vector.

Now, non-compact manifolds always admit a nowhere vanishing vector field. However, compact manifolds possess a nowhere vanishing vector field if and only if [80][•4.1.3] they have vanishing Euler characteristic χ, which provides a

•*4.1.3:* **ptc:***there are some absurd statements there*

[1] In two dimensions $-g$ is a Lorentzian metric whenever g is, and the operation $g \to -g$ has the effect of interchanging the role of space and of time. The reader will notice that while the Möbius strip with the flat metric g of Figure 4.1 is not time-orientable, it becomes time-orientable when equipped with $-g$.

necessary condition of topological nature for a Lorentzian manifold to be time-orientable. We actually have the following:

PROPOSITION 4.1.2 *A manifold $\mathscr{M}$ admits a space-time structure if and only if there exists a nowhere vanishing vector field on $\mathscr{M}$.*

PROOF: The necessity of the existence of a nowhere vanishing vector field on $\mathscr{M}$ has already been established. Conversely, suppose that such a vector field X exists, and let h be any Riemannian metric on $\mathscr{M}$. Then the formula

$$g(Y,Z) = h(Y,Z) - \frac{h(Y,X)h(Z,X)}{h(X,X)} \tag{4.1.2}$$

defines a Lorentzian metric on $\mathscr{M}$. Finally, the existence of a globally defined timelike vector field X on a Lorentzian manifold $(\mathscr{M},g)$ implies time-orientability of $\mathscr{M}$ in the obvious way – choose $\mathscr{O}_p = \mathscr{M}$ and $X_{\mathscr{O}_p} = X$. $\Box$

We note that simply connected four dimensional manifolds have, by Poincaré duality, Euler characteristic larger than or equal to 2, and therefore do not admit a space-time structure.

4.2 Normal coordinates

For $p \in \mathscr{M}$ the exponential map

$$\exp_p : T_p\mathscr{M} \to \mathscr{M}$$

is defined as follows; if X is a vector in the tangent space $T_p\mathscr{M}$, then $\exp_p(X) \in \mathscr{M}$ is the point reached by following a geodesic with initial point p and initial tangent vector $X \in T_p\mathscr{M}$ for an affine distance one, provided that the geodesic in question can be continued that far. Now an affinely parameterized geodesic solves the equation

$$\nabla_{\dot x}\dot x = 0 \quad \Longleftrightarrow \quad \frac{d^2x^\mu}{ds^2} = -\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}\ , \tag{4.2.1}$$

where the $\Gamma^\mu_{\alpha\beta}$'s are the Christoffel symbols of the metric g, defined as

$$\Gamma^\mu_{\alpha\beta} := \frac{1}{2}g^{\mu\sigma}\left(\frac{\partial g_{\sigma\alpha}}{\partial x^\beta} + \frac{\partial g_{\sigma\beta}}{\partial x^\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x^\sigma}\right)\ , \tag{4.2.2}$$

$$g^{\mu\sigma} := g^\#(dx^\mu, dx^\sigma)\ , \qquad g_{\alpha\beta} := g(\partial_\alpha, \partial_\beta)\ . \tag{4.2.3}$$

We use sometimes use the symbol $g^\#$ to denote the "contravariant metric",•4.2.1 that is, the metric on $T^*\mathscr{M}$ constructed out from g in the canonical way (see Section 6; the matrix $g^{\alpha\beta}$ is thus the matrix inverse to $g_{\alpha\beta}$). However, it is usual in the literature to use the same symbol g for the metric $g^\#$, as well as for all other metrics on Equations (4.2.2)-(4.2.3) show that when the metric is of $C^{1,1}$ differentiability class, then the Christoffel symbols are Lipschitz continuous, which guarantees local existence and uniqueness of solutions of (4.2.1). Due

•4.2.1: **ptc:** *watch out for repetitivity here; also M or $\mathscr{M}$?*

•4.2.2: is this correct? note Woolgar Sorkin ref, and also Keye Martin gr-qc/0408068

•4.2.3: de Sitter? ref? more detail?

to the lack of uniqueness[2] of the Cauchy problem for (4.2.1) for metrics which are not $C^{1,1}$, it appears$^{\bullet 4.2.2}$ to be difficult to do causality theory on manifolds with a metric with less regularity[3] than $C^{1,1}$.

The domain $\mathscr{U}_p$ of $\exp_p$ is always the largest subset of $T_p\mathscr{M}$ on which the exponential map is defined. By construction, and by homogeneity properties of solutions of (4.2.1) under a linear change of parameterization (see (4.2.10)), the set $\mathscr{U}_p$ is star-shaped with respect to the origin (this means that if $X \in \mathscr{U}_p$ then we also have $\lambda X \in \mathscr{U}_p$ for all $\lambda \in [0,1]$). When the metric is $C^{1,1}$, continuity of solutions of ODE's upon initial values shows that $\mathscr{U}_p$ is an open neighborhood of the origin of $T_p\mathscr{M}$.

The exponential map is neither surjective nor injective in general. For example, on $\mathbb{R} \times S^1$ with the flat metric $-dt^2 + dx^2$, the "left-directed" null geodesics $\Gamma_-(s) = (s, -s \mod 2\pi)$ and the "right-directed" null geodesics $\Gamma_+(s) = (s, s \mod 2\pi)$ meet again after going each "half of the way around S^1", and injectivity fails. In anti-de-Sitter$^{\bullet 4.2.3}$ space-time all timelike geodesics meet again at an "antipodal point", which leads to lack of surjectivity of the exponential map.

A Lorentzian manifold is said to be *geodesically complete* if all geodesics can be defined for all real values of affine parameter; this is equivalent to the requirement that for all $p \in \mathscr{M}$ the domain of the exponential map is $T_p\mathscr{M}$. One also talks about *timelike geodesically complete* space-times, *future timelike geodesically complete* space-times, *etc.*, with those notions defined in an obvious way.

The fundamental *Hopf-Rinow* theorem[4] asserts that *compact Riemannian manifolds are geodesically complete.* There is no Lorentzian analogue of this, the standard counter-example proceeds as follows:

EXAMPLE 4.2.1 Consider the following symmetric tensor field on $\mathbb{R}^2$:

$$g = \frac{2dxdy}{x^2+y^2} . \tag{4.2.4}$$

We have

$$g_{\mu\nu} = \frac{1}{x^2+y^2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Longrightarrow \quad \det g_{\mu\nu} = -\frac{1}{(x^2+y^2)^2} , \tag{4.2.5}$$

which shows that g is indeed a Lorentzian metric. Note the for all $\lambda \in \mathbb{R}^*$ the maps

$$\mathbb{R}^2 \ni (x,y) \to \phi_\lambda(x,y) := (\lambda x, \lambda y)$$

[2]Examples of $C^{1,\alpha}$ metrics with non-unique geodesics for $0 < \alpha < 1$ can be found in, *e.g.,* [28, Appendix F]. Here $C^{k,\alpha}$ is the space of k times differentiable functions (or maps, or sections — whichever is the case should be clear from the context), the k'th derivatives of which satisfy, locally, a Hölder condition of order α; no uniformity conditions are imposed unless explicitly indicated otherwise.

[3]We will see in Section **??** that one can construct large classes of solutions to the Cauchy problem for the vacuum Einstein equations which are *not* of $C^{1,1}$ differentiability class; see also [9, 72, 73, 102]. This leads to an unfortunate mismatch in differentiability between the Cauchy problem and causality theory. It would be of interest to find how low one can go with regularity of the metric, while retaining a reasonable theory of causality.

[4]There are various related theorems known under this name [62, 81], this statement is one of several versions thereof.

are isometries of g:

$$\phi_\lambda^* g = \frac{2d(\lambda x)d(\lambda y)}{(\lambda x)^2 + (\lambda y)^2} = \frac{2dxdy}{x^2+y^2} = g \ .$$

It follows that for any $1 \neq \lambda > 0$ the metric g passes to the quotient space

$$\left\{\mathbb{R}^2 \setminus \{0\}\right\} / \phi_\lambda = \{(x,y) \sim (\lambda x, \lambda y)\} \approx S^1 \times S^1 = \mathbb{T}^2 \ .$$

(Clearly the quotient spaces with λ and $1/\lambda$ are the same, so without loss of generality one can assume $\lambda > 1$.) In order to show geodesic incompletess of g we will use the following result:

PROPOSITION 4.2.2 *Let f be a function such that $g(\nabla f, \nabla f)$ is a constant. Then the integral curves of ∇f are affinely parameterized geodesics.*

PROOF: Let $X := \nabla f$, we have

$$\begin{aligned}
(\nabla_X X)^j &= \nabla^i f \nabla_i \nabla^j f = \nabla^i f \nabla^j \nabla_i f \\
&= \frac{1}{2}\nabla^j(\nabla^i f \nabla_i f) = \frac{1}{2}\nabla^j (g(\nabla f, \nabla f)) = 0 \ .
\end{aligned}$$

□

Returning to the metric (4.2.4), let $f = x$, by (4.2.5) we have

$$g^{\mu\nu} = (g^{\mu\nu})^{-1} = (x^2+y^2)\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} ,$$

so that

$$\nabla f = (x^2+y^2)\partial_y \quad \Longrightarrow \quad g(\nabla f, \nabla f) = 0 \ .$$

Proposition 4.2.2 shows that the integral curves of ∇f are null affinely parameterized geodesics. Let $\gamma(s) = (x^\mu(s))$ be any such integral curve, thus

$$\frac{dx^\mu}{ds} = \nabla^\mu f \quad \Longrightarrow \quad \frac{dx}{ds} = 0 \ , \quad \frac{dy}{ds} = (x^2+y^2) \ .$$

It follows that $x(s) = x(0)$ for all s. The equation for y is easily integrated; for our purposes it is sufficient to consider that integral curve which passes through $(0, y_0) \in \mathbb{R}^2 \setminus \{0\}$, $y_0 > 0$ — we then have $x(s) = 0$ for all s and

$$\frac{dy}{ds} = y^2 \quad \Longrightarrow \quad y(s) = \frac{y_0}{1 - y_0 s} \ . \tag{4.2.6}$$

This shows that $y(s)$ runs away to infinity as s approaches

$$s_\infty := \frac{1}{y_0} \ .$$

It follows that γ is indeed incomplete on $\mathbb{R}^2 \setminus \{0\}$. To see that it is also incomplete on the quotient torus $\left\{\mathbb{R}^2 \setminus \{0\}\right\} / \phi_\lambda$, $\lambda > 1$, note that the image of $\gamma(s) = (0, y(s))$ under the equivalence relation $\sim$ is a circle, and there exists a sequence $s_j \to s_\infty$ such that $\gamma(s_j)$ passes again and again through its starting point:

$$y(s_j) = \lambda^j y_0 \quad \Longrightarrow \quad (0, y(s_j)) \sim (0, y_0) \ \text{ in } \ \left\{\mathbb{R}^2 \setminus \{0\}\right\} / \phi_\lambda \ .$$

By (4.2.6) we have

$$\frac{dy}{ds}(s_j) = (y(s_j))^2 = (\lambda_j y_0)^2 \longrightarrow_{s_j \to s_\infty} \infty \ ,$$

which shows that the sequence of tangents $(dy/ds)(s_j)$ at $(0, y_0)$ blows up as j tends to infinity. This clearly implies that γ cannot be extended beyond s_∞ as a C^1 curve.

When the metric is C^2, •4.2.4 the inverse function theorem shows that there exists a neighborhood $\mathscr{V}_p \subset \mathscr{U}_p$ of the origin in $\mathbb{R}^{\dim \mathscr{M}}$ on which the exponential map is a diffeomorphism between $\mathscr{V}_p$ and its image

•4.2.4: can one use Clarke for $C^{1,1}$?

$$\mathscr{O}_p := \exp_p(\mathscr{V}_p) \subset \mathscr{M} \ .$$

This allows one to define *normal coordinates* centred at p:

PROPOSITION 4.2.3 *Let $(\mathscr{M}, g)$ be a C^3 Lorentzian manifold with C^2 metric g. For every $p \in \mathscr{M}$ there exists an open coordinate neighborhood $\mathscr{O}_p$ of p, in which p is mapped to the origin of $\mathbb{R}^{n+1}$, such that the coordinate rays $s \to sx^\mu$ are affinely parameterized geodesics. If the metric g is expressed in the resulting coordinates $(x^\mu) = (x^0, \vec{x}) \in \mathscr{V}_p$, then*

$$g_{\mu\nu}(0) = \eta_{\mu\nu} \ , \quad \partial_\sigma g_{\mu\nu}(0) = 0 \ , \tag{4.2.7}$$

Further, if the function $\sigma_p : \mathscr{O}_p \to \mathbb{R}$ is defined by the formula

$$\boxed{\sigma_p(\exp_p(x^\mu)) := \eta_{\mu\nu} x^\mu x^\nu \equiv -(x^0)^2 + (\vec{x})^2 \ ,} \tag{4.2.8}$$

then

$$\boxed{\nabla\sigma_p \quad \textit{is} \quad \begin{cases} \textit{timelike} & \begin{cases} \textit{past directed} & \textit{on } \{q \,|\, \sigma_p(q) < 0 \ , \ x^0(q) < 0\}, \\ \textit{future directed} & \textit{on } \{q \,|\, \sigma_p(q) < 0 \ , \ x^0(q) > 0\}, \end{cases} \\ \textit{null} & \begin{cases} \textit{past directed} & \textit{on } \{q \,|\, \sigma_p(q) = 0 \ , \ x^0(q) < 0\}, \\ \textit{future directed} & \textit{on } \{q \,|\, \sigma_p(q) = 0 \ , \ x^0(q) > 0\}, \end{cases} \\ \textit{spacelike} & \qquad\qquad\qquad\quad \textit{on } \{q \,|\, \sigma_p(q) > 0\} \ . \end{cases}} \tag{4.2.9}$$

REMARK 4.2.4 The coefficients of a Taylor expansion of $g_{\mu\nu}$ in normal coordinates can be expressed in terms of the Riemann tensor and its covariant derivatives (*cf.*, *e.g.* [84]).•4.2.5

•4.2.5: **ptc:***add ref to appendix*

PROOF: Let us start by justifying that the implicit function theorem can indeed be applied: Let x^μ be any coordinate system around p, and let $e_a = e_a{}^\mu \partial_\mu$ be any ON frame at p. Let

$$X = X^a e_a = X^a e_a{}^\mu \partial_\mu \in T_p M$$

and let $x^\mu(s, X)$ denote the affinely parameterized geodesic passing by p at $s = 0$, with tangent vector

$$\dot{x}^\mu(0, X) := \frac{dx^\mu(s, X)}{ds}\Big|_{s=0} = X^a e_a{}^\mu \ .$$

Homogeneity properties of the ODE (4.2.1) under the change of parameter $s \to \lambda s$ together with uniqueness of solutions of ODE's show that for any constant $a \neq 0$ we have

$$x^\mu(as, X/a) = x^\mu(s, X) \ .$$

This, in turn, implies that there exist functions γ^μ such that

$$x^\mu(s,X) = \gamma^\mu(sX)\ . \tag{4.2.10}$$

From (4.2.1) and (4.2.10) we have

$$x^\mu(s,X) = x_0^\mu + sX^a e_a{}^\mu + O((s|X|)^2)\ .$$

Here x_0^μ are the coordinates of p, $|X|$ denotes the norm of X with respect to some auxiliary Riemannian metric on M, while the $O((s|X|)^2)$ term is justified by (4.2.10). The usual considerations of the proof that solutions of ODE's are differentiable functions of their initial conditions show that

$$\begin{aligned}\frac{\partial x^\mu(s,X)}{\partial X^a} &= \frac{\partial(x_0^\mu + sX^a e_a{}^\mu)}{\partial X^a} + O(s^2)|X| \\ &= s e_a{}^\mu + O(s^2)|X|\ .\end{aligned}$$

At $s=1$ one thus obtains

$$\frac{\partial x^\mu(1,X)}{\partial X^a} = e_a{}^\mu + O(|X|)\ . \tag{4.2.11}$$

This shows that $\partial x^\mu/\partial X^a$ will be bijective at $X=0$ provided that $\det e_a{}^\mu \neq 0$. But this last inequality can be obtained by taking the determinant of the equation

$$g(e_a,e_b) = g_{\mu\nu} e_a{}^\mu e_b^\nu \quad \Longrightarrow \quad -1 = (\det g_{\mu\nu})(\det e_a{}^\mu)^2\ . \tag{4.2.12}$$

This justifies the use of the implicit function theorem to obtain existence of the neighborhood $\mathscr{O}_p$ announced in the statement of the proposition. Clearly $\mathscr{O}_p$ can be chosen to be star-shaped with respect to p. Equation (4.2.11) and the fact that $e^\mu{}_a$ is an ON-frame show that

$$g(\partial_a,\partial_b)\Big|_{X^a=0} = g_{\mu\nu} e_a{}^\mu e_b^\nu\Big|_{X^a=0} = \eta_{ab}\ ,$$

which establishes the first claim in (4.2.7).

By construction the rays

$$s \to \gamma^a(s) := sX^a$$

are affinely parameterized geodesics with tangent $\dot\gamma = X^a\partial_a$, which gives

$$\begin{aligned}0 = (\nabla_{\dot\gamma}\dot\gamma)^a &= \underbrace{\frac{d^2(sX^a)}{ds^2}}_{=0} + \Gamma^a{}_{bc}(sX^d)X^bX^c \\ &= \Gamma^a{}_{bc}(sX^d)X^bX^c\ .\end{aligned}$$

Differentiating this equation twice with respect to X^d and X^e, and setting $X=0$ one obtains

$$\Gamma^a{}_{de}(0) = 0\ .$$

The equation

$$0 = \nabla_a g_{bc} = \partial_a g_{bc} - \Gamma^d{}_{ba} g_{dc} - \Gamma^d{}_{ca} g_{bd}$$

evaluated at $X = 0$ gives the second equality in (4.2.7).

Let us pass now to the proof of the main point here, namely (4.2.9). From now on we will denote by x^μ the normal coordinates obtained so far, and which were denoted by X^a in the arguments just done. For $x \in \mathscr{O}_p$ define

$$f(x) := \eta_{\mu\nu} x^\mu x^\nu \ , \tag{4.2.13}$$

and let $\mathscr{H}_\tau \subset \mathscr{O}_p \setminus \{p\}$ be the level sets of f:

$$\mathscr{H}_\tau := \{x : f(x) = \tau \ , \ x \neq 0\} \ . \tag{4.2.14}$$

We will show that

$$\text{the vector field } x^\mu\partial_\mu \text{ is normal to the } \mathscr{H}_\tau\text{'s.} \tag{4.2.15}$$

Now, $x^\mu\partial_\mu$ is tangent to the geodesic rays $s \to \gamma^\mu(s) := s x^\mu$. As the causal character of the field of tangents to a geodesic[5] is point-independent along the geodesic, we have

$$\begin{array}{rcl} x^\mu\partial_\mu \text{ is timelike at } \gamma(s) & \Longleftrightarrow & f(x) < 0 \ , \\ x^\mu\partial_\mu \text{ is null at } \gamma(s) & \Longleftrightarrow & f(x) = 0 \ , \ x \neq 0 \ , \\ x^\mu\partial_\mu \text{ is spacelike at } \gamma(s) & \Longleftrightarrow & f(x) > 0 \ . \end{array} \tag{4.2.16}$$

This follows from the fact that the right-hand-side is precisely the condition that the geodesic be timelike, spacelike, or null, at $\gamma(0)$. Since ∇f is always normal to the level sets of f, when (4.2.15) holds we will have

$$x^\mu\partial_\mu \text{ is proportional to } \nabla^\mu f. \tag{4.2.17}$$

This shows that (4.2.9) will follow from (4.2.16) when (4.2.15) holds.

It remains to establish (4.2.15). In order to do that, consider any curve $x^\mu(\lambda)$ lying on $\mathscr{H}_\tau$:

$$\eta_{\mu\nu} x^\mu(\lambda) x^\mu(\lambda) = \tau \quad \Longrightarrow \quad \eta_{\mu\nu} x^\mu(\lambda) \partial_\lambda x^\mu(\lambda) = 0 \ . \tag{4.2.18}$$

Let $\gamma^\mu(\lambda, s)$ be the following one-parameter family of geodesic rays:

$$\gamma^\mu(\lambda, s) := s x^\mu(\lambda) \ .$$

For any function f set

$$T(f) = \partial_s \left(f \circ \gamma(s, \lambda)\right) \ , \quad X(f) = \partial_\lambda \left(f \circ \gamma(s, \lambda)\right) \ ,$$

so that

$$T(\lambda, s) := \Big(\partial_s \gamma^\mu(\lambda, s)\Big)\partial_\mu = x^\mu(\lambda, s)\partial_\mu \ , \quad X := \Big(\partial_\lambda \gamma^\mu(\lambda, s)\Big)\partial_\mu \ .$$

[5]Without loss of generality an affine parameterization of a geodesic γ can be chosen, the result follows then from the calculation $d(g(\dot\gamma, \dot\gamma))/ds = 2g(\nabla_{\dot\gamma}\dot\gamma, \dot\gamma) = 0$.

For any fixed value of λ the curves $s \to \gamma^\mu(\lambda, s)$ are geodesics, which shows that

$$\nabla_T T = 0 \ .$$

This gives

$$\frac{d(g(T,T))}{ds} = 2g(\nabla_T T, T) = 0 \ ,$$

hence

$$g(T,T)(s) = g(T,T)(0) = \eta_{\mu\nu} x^\mu(\lambda) x^\nu(\lambda) = \tau$$

by (4.2.18), in particular $g(T,T)$ is λ-independent.

Next, for any twice-differentiable function g we have

$$[T,X](g) := T(X(g)) - X(T(g)) = \partial_s \partial_\lambda (g(\gamma^\mu(s,\lambda)) - \partial_\lambda \partial_s (g(\gamma^\mu(s,\lambda)) = 0 \ ,$$

because of the symmetry of the matrix of second partial derivatives. It follows that

$$[T,X] = \nabla_T X - \nabla_X T = 0 \ .$$

Finally,

$$\begin{aligned} \frac{d(g(T,X))}{ds} &= g(\underbrace{\nabla_T T}_{0}, X) + g(T, \underbrace{\nabla_T X}_{=\nabla_X T}) \\ &= g(T, \nabla_X T) = \frac{1}{2} X(g(T,T)) = \frac{1}{2} \partial_\lambda (g(T,T)) = 0 \ . \end{aligned}$$

This yields

$$g(T,X)(s,\lambda) = g(T,X)(0,\lambda) = \eta_{\mu\nu} x^\mu(\lambda) \partial_\lambda x^\mu(\lambda) = 0$$

again by (4.2.18). Thus T is normal to the level sets of f, which had to be established. □

It should be pointed out that some regularity of the metric is lost when going to normal coordinates; this can be avoided using coordinates which are only approximately normal up to a required order (compare Proposition A.5.2) — this is often sufficient for several purposes.

•4.2.6

•*4.2.6:* **ptc:***add remark about the section about normal coordinates*

It is sometimes useful to have a *geodesic convexity property* at our disposal. This is made precise by the following proposition:

PROPOSITION 4.2.5 *Let $\mathscr{O}$ be the domain of definition of a coordinate system $\{x^\mu\}$. Let $p \in \mathscr{O}$ and let $B_p(r) \subset \mathscr{O}$ denote an open coordinate ball of radius r centred at p. There exists $r_0 > 0$ such that every geodesic segment $\gamma : [a,b] \to \overline{B_p(r_0)} \subset \mathscr{O}$ satisfying*

$$\gamma(a), \gamma(b) \in B_p(r) \ , \quad r < r_0$$

is entirely contained in $B_p(r)$.

PROOF: Let $x^\mu(s)$ be the coordinate representation of γ, set

$$f(s) := \sum_\mu (x^\mu - x_0^\mu)^2 \ ,$$

where x_0^μ is the coordinate representation of p. We have

$$\begin{aligned} \frac{df}{ds} &= 2\sum_\mu (x^\mu - x_0^\mu)\frac{dx^\mu}{ds} \ , \\ \frac{d^2 f}{ds^2} &= 2\sum_\mu \left(\frac{dx^\mu}{ds}\right)^2 + 2\sum_\mu (x^\mu - x_0^\mu)\frac{d^2x^\mu}{ds^2} \\ &= 2\sum_\mu \left(\frac{dx^\mu}{ds}\right)^2 - 2\sum_\mu (x^\mu - x_0^\mu)\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds} \ . \end{aligned}$$

Compactness of $\overline{B_p(r_0)}$ implies that there exists a constant C such that we have

$$\left|\sum_\mu (x^\mu - x_0^\mu)\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}\right| \le Cr_0 \sum_\mu \left(\frac{dx^\mu}{ds}\right)^2 \ .$$

It follows that $d^2f/ds^2 \ge 0$ for r_0 small enough. This shows that f has no interior maximum if r_0 is small enough, whence the result. □

It is convenient to introduce the following notion:

DEFINITION 4.2.6 *An* elementary region *is an open coordinate ball $\mathscr{O}$ within the domain of a normal coordinate neighborhood $\mathscr{U}_p$ such that*

1. *$\mathscr{O}$ has compact closure $\overline{\mathscr{O}}$ in $\mathscr{U}_p$, and*

2. *∇t and ∂_t are timelike on $\overline{\mathscr{U}}$.* •4.2.7

•4.2.7: *add geodesic convexity if needed?*

Note that ∂_t is timelike if and only if

$$g_{tt} = g(\partial_t, \partial_t < 0) \ ,$$

while ∇t is timelike if and only if

$$g^{tt} = g^\#(dt, dt) < 0 \ .$$

Existence of elementary regions containing some point $p \in M$ follows immediately from Proposition 4.2.3: In normal coordinates centred at p one chooses $\mathscr{O}$ be an open coordinate ball of sufficiently small radius. It follows from Proposition 4.2.5 that, by choosing the radius even smaller if necessary, we can also require that every two points in $p, q \in \mathscr{O}$ can be joined by a geodesic Γ_{pq} contained in $\mathscr{O}$. Further, Γ_{pq} may be required to be unique within the class of geodesics entirely contained in $\mathscr{O}$. •4.2.8

•4.2.8: **ptc:***might need elaborating upon*

4.3 Causal paths

•4.3.1

*•4.3.1: **ptc:**All that follows has never been globally proof-read*

Let $(\mathscr{M}, g)$ be a space-time. The basic objects in causality theory are paths. We shall always use *parameterized paths*: by definition, this are continuous maps from some interval to $\mathscr{M}$.[•4.3.2] We will use interchangedly the terms "path", "parameterized path", or "curve".

*•4.3.2: **ptc:**should one ask for injectivity?*

Let $\gamma : I \to \mathscr{M}$ and let $\mathscr{U} \subset \mathscr{M}$, we will write

$$\gamma \subset \mathscr{U}$$

whenever the image $\gamma(I)$ of I by γ is a subset of $\mathscr{U}$.[•4.3.3] If $\gamma(I) \cap \mathscr{U}$ is connected, we will write

*•4.3.3: **ptc:**added, cross-check for repetitiveness later*

$$\gamma \cap \mathscr{U}$$

for the path obtained by removing from I those parameters s for which $\gamma(s) \not\in \mathscr{U}$. If $\gamma(I) \cap \mathscr{U}$ is not connected, then $\gamma \cap \mathscr{U}$ denotes the collection of the connected components of $\gamma(I) \cap \mathscr{U}$.

Some authors define a path in $\mathscr{M}$ as the *image of a parameterized path*; in this approach one forgets about the parameterization of γ, and identifies two paths which have the same image and differ only by a reparameterization. This leads to various difficulties when considering end points of causal paths — *cf.* Section 4.5, or limits of sequences of paths — *cf.* Section 4.6, and therefore we do *not* adopt this approach.

If $\gamma : I \to \mathscr{M}$ where $I = [a, b)$ or $I = [a, b]$, then $\gamma(a)$ is called the *starting point* of the path γ, or of its image $\gamma(I)$. If $I = (a, b]$ or $I = [a, b]$, then $\gamma(b)$ is called the *end point*. (This definition will be extended in Section 4.5, but it is sufficient for the purposes here.) We shall say that $\gamma : [a, b] \to \mathscr{M}$ is a path from p to q if $\gamma(a) = p$ and $\gamma(b) = q$.

In the standard approach [16, 53, 58, 87, 89, 105] to causality theory one defines[•4.3.4] *future directed timelike* paths as those paths γ which are *piecewise differentiable*, with $\dot\gamma$ timelike and future directed wherever defined; at break points one further assumes that both the left-sided and right-sided derivatives are timelike. This definition turns out to be quite inconvenient for several purposes. For instance, when studying the global causal structure of space-times one needs to take limits of timelike curves, obtaining thus — by definition — *causal future directed* paths. Such limits will not be piecewise differentiable most of the time, which leads one to the necessity of considering paths with poorer differentiability properties. One then faces the unhandy situation in which timelike and causal paths have completely different properties. In several theorems separate proofs have then to be given. The approach we present avoids this, leading — we believe — to a considerable simplification of the conceptual structure of the theory.

*•4.3.4: **ptc:**give a more detailed discussion*

It is convenient to choose once and for all some auxiliary Riemannian metric h on $\mathscr{M}$, such that $(\mathscr{M}, h)$ is complete — such a metric always exists[•4.3.5] ; let dist_h denote the associated distance function. A parameterized path $\gamma : I \to \mathscr{M}$ from an interval $I \subset \mathbb{R}$ to M is called *locally Lipschitzian* if for every compact subset K of I there exists a constant $C(K)$ such that

•4.3.5: give construction

$$\forall\ s_1, s_2 \in K \quad \mathrm{dist}_h(\gamma(s_1), \gamma(s_2)) \le C(K)|s_1 - s_2|\ .$$

•4.3.6: **ptc:**_new result and proof, reread_

It is of interest to enquire whether the class of paths so defined depends upon the background metric h:•4.3.6

PROPOSITION 4.3.1 *Let h_1 and h_2 be two complete Riemannian metrics on $\mathscr{M}$. Then a path $\gamma : I \to \mathscr{M}$ is locally Lipschitzian with respect to h_1 if and only if it is locally Lipschitzian with respect to h_2.*

PROOF: Let $K \subset I$ be a compact set, then $\gamma(K)$ is compact. Let L_a, $a = 1, 2$ denote the h_a-length of γ, set

$$\mathscr{K}_a := \cup_{s\in I} B_{h_a}(\gamma(s), L_a) ,$$

where $B_{h_a}(p, r)$ denotes a geodesic ball, with respect to the metric h_a, centred at p, of radius r. Then the $\mathscr{K}_a$'s are compact. Likewise the sets

$$\widehat{\mathscr{K}}_a \subset T\mathscr{M} ,$$

defined as the sets of h_a-unit vectors over $\mathscr{K}_a$, are compact. This implies that there exists a constant C_K such that for all $X \in T_pM$, $p \in \mathscr{K}_a$, we have

$$C_K^{-1} h_1(X, X) \le h_2(X, X) \le C_K h_1(X, X) .$$

Let γ_{a,s_1,s_2} denote any minimising h_a-geodesic between $\gamma(s_1)$ and $\gamma(s_2)$, then

$$\forall\ s_1, s_2 \in I \qquad \gamma_{a,s_1,s_2} \subset \mathscr{K}_a .$$

This implies

$$\begin{aligned}
\mathrm{dist}_{h_2}(\gamma(s_1), \gamma(s_2)) &= \int_{\sigma=\gamma_{2,s_1,s_2}} \sqrt{h_2(\dot\sigma, \dot\sigma)} \\
&\ge C_K^{-1} \int_{\sigma=\gamma_{2,s_1,s_2}} \sqrt{h_1(\dot\sigma, \dot\sigma)} \\
&\ge C_K^{-1} \inf_\sigma \int_\sigma \sqrt{h_1(\dot\sigma, \dot\sigma)} \\
&= C_K^{-1} \int_{\sigma=\gamma_{1,s_1,s_2}} \sqrt{h_1(\dot\sigma, \dot\sigma)} \\
&= C_K^{-1} \mathrm{dist}_{h_1}(\gamma(s_1), \gamma(s_2)) .
\end{aligned}$$

From symmetry with respect to the interchange of h_1 with h_2 we conclude that

$$\forall\, s_1, s_2 \in K \quad C_K^{-1}\mathrm{dist}_{h_1}(\gamma(s_1), \gamma(s_2)) \le \mathrm{dist}_{h_2}(\gamma(s_1), \gamma(s_2)) \le C_K \mathrm{dist}_{h_1}(\gamma(s_1), \gamma(s_2)) ,$$

and the result easily follows. □

More generally, if (N, k) and (M, h) are Riemannian manifolds, then a map ϕ is called *locally Lipschitzian*, or *locally Lipschitz*, if for every compact subset K of N there exists a constant $C(K)$ such that

$$\forall\ p, q \in K \quad \mathrm{dist}_h(\phi(p), \phi(q)) \le C(K)\mathrm{dist}_k(p, q) .$$

A map is called *Lipschitzian* if the constant $C(K)$ above can be chosen independently of K.

The following important theorem of Rademacher will play a key role in our considerations:

THEOREM 4.3.2 (Rademacher) *Let $\phi : M \to N$ be a locally Lipschitz map from a manifold M to a manifold N. Then:*

1. *ϕ is classically differentiable almost everywhere, with "almost everywhere" understood in the sense of the Lebesgue measure in local coordinates on M.*

2. *The distributional derivatives of ϕ are in L^∞_{loc} and are equal to the classical ones almost everywhere.*

3. *Suppose that M is an open subset of $\mathbb{R}$ and N is an open subset of $\mathbb{R}^n$. Then ϕ is the integral of its distributional derivative,*

$$\phi(x) - \phi(y) = \int_y^x \frac{d\phi}{dt} dt \, . \tag{4.3.1}$$

PROOF: Point 1. is actually the classical statement of Rademacher, the proof can be found in [44, Theorem 2, p. 235]. Point 2. is Theorem 5 in [44, p. 131] and Theorem 1 of [44, p. 235]. (In that last theorem the classical differentiability a.e. is actually established for all $W^{1,p}_{\mathrm{loc}}$ functions with $p > n$). Point 3. can be established by approximating ϕ by C^1 functions as in [44, Theorem 1, p. 251], and passing to the limit. □

Point 2. shows that the usual properties of the derivatives of continuously differentiable functions — such as the Leibniz rule, or the chain rule — hold almost everywhere for the derivatives of locally Lipschitzian functions. By point 3. those properties can be used freely whenever integration is involved.
•4.3.7

•*4.3.7:* **ptc:***a comment on integration on non-compact sets commented out*

We will use the symbol

$$\dot\gamma$$

to denote the *classical*•4.3.8 derivative of a path γ, wherever defined. A parameterized path γ will be called *causal future directed* if γ is locally Lipschitzian, with $\dot\gamma$ — causal and future directed almost everywhere[6]. Thus, $\dot\gamma$ is defined almost everywhere; and it is causal future directed almost everywhere on the set on which it is defined. A parameterized path γ will be called *timelike future directed* if γ is locally Lipschitzian, with $\dot\gamma$ — timelike future directed almost everywhere. *Past directed* parameterized paths are defined by changing "future" to "past" in the definitions above.

•*4.3.8:* **ptc:***classical added, but maybe this is a bad idea*

•4.3.9 A useful property of locally Lischitzian paths is that they can be parameterized by h-distance. Let $\gamma : [a, b) \to \mathscr{M}$ be a path, and suppose that $\dot\gamma$ is non-zero almost everywhere — this is certainly the case for causal paths. By Rademacher's theorem the integral

•*4.3.9:* **ptc:***make a formal statement; show that this leads to equivalence of the notion of image and of the notion of parameterized path, so that this is the same as what geometers have been doing, except for a choice of origin on the path*

$$s(t) = \int_a^t |\dot\gamma|_h(u) du$$

[6]Some authors allow constant paths to be causal, in which case the sets $J^\pm(\mathscr{U}; \mathscr{O})$ defined below automatically contain $\mathscr{U}$. This leads to unnecessary discussions when concatenating causal paths, so that we find it convenient not to allow such paths in our definition.

is well defined. Clearly $s(t)$ is a continuous strictly increasing function of t, so that the map $t \to s(t)$ is a bijection from $[a,b)$ to its image. The new path $\hat\gamma := \gamma \circ s^{-1}$ differs from γ only by a reparametrization, so it has the same image in $\mathscr{M}$. The reader will easily check that $|\dot{\hat\gamma}|_h = 1$ almost everywhere.•4.3.10

•4.3.10: **ptc:***prove*

Further, $\hat\gamma$ is Lipschitz continuous with Lipschitz constant smaller than or equal to 1. In order to seee that, let dist_h be the associated distance function, we then have

$$\mathrm{dist}_h(\gamma(s),\gamma(s')) \le |s-s'| \ . \tag{4.3.2}$$

In order to see that, we calculate, for $s > s'$:

$$\begin{aligned} s - s' &= \int_{s'}^{s} dt \\ &= \int_{s'}^{s} \underbrace{\sqrt{h(\dot\gamma,\dot\gamma)}(t)}_{=1}\, dt \\ &\ge \inf_{\tilde\gamma} \int_{\tilde\gamma} \sqrt{h(\dot{\tilde\gamma},\dot{\tilde\gamma})}(t)dt = \mathrm{dist}_h(\gamma(s),\gamma(s')) \ , \end{aligned} \tag{4.3.3}$$

where the infimum is taken over $\tilde\gamma$'s which start at $\gamma(s')$ and finish at $\gamma(s)$.

4.4 Futures, pasts

Let $\mathscr{U} \subset \mathscr{O} \subset \mathscr{M}$. One sets

$$\begin{aligned} I^+(\mathscr{U};\mathscr{O}) &:= \{q \in \mathscr{O} : \text{there exists a timelike future directed path} \\ &\qquad \text{from } \mathscr{U} \text{ to } q \text{ contained in } \mathscr{O}\} \ , \\ J^+(\mathscr{U};\mathscr{O}) &:= \{q \in \mathscr{O} : \text{there exists a causal future directed path} \\ &\qquad \text{from } \mathscr{U} \text{ to } q \text{ contained in } \mathscr{O}\} \cup \mathscr{U} \ . \end{aligned}$$

$I^-(\mathscr{U};\mathscr{O})$ and $J^-(\mathscr{U};\mathscr{O})$ are defined by replacing "future" by "past" in the definitions above. The set $I^+(\mathscr{U};\mathscr{O})$ is called the *timelike future of $\mathscr{U}$ in $\mathscr{O}$*, while $J^+(\mathscr{U};\mathscr{O})$ is called the *causal future of $\mathscr{U}$ in $\mathscr{O}$*, with similar terminology for the timelike past and the causal past. We will write $I^\pm(\mathscr{U})$ for $I^\pm(\mathscr{U};\mathscr{M})$, similarly for $J^\pm(\mathscr{U})$, and one then omits the qualification "in $\mathscr{M}$" when talking about the causal or timelike futures and pasts of $\mathscr{U}$. We will write $I^\pm(p;\mathscr{O})$ for $I^\pm(\{p\};\mathscr{O})$, $I^\pm(p)$ for $I^\pm(\{p\};\mathscr{M})$, *etc.*

Although our definition of causal curves does not coincide with the usual ones [16, 58, 89, 105], it is equivalent to those. It immediately follows that our definition of $J^\pm$ is identical to the standard one. On the other hand, the class of timelike curves as defined here is quite wider then the standard one; nevertheless, the resulting sets $I^\pm$ are again identical to the usual ones (compare Proposition 4.4.11).

It is legitimate to raise the question, why is it interesting to consider sets such as $J^+(\mathscr{O})$. The answer is two-fold: From a mathematical point of view, those sets appear naturally when describing the *finite speed of propagation* property of wave-type equations, such as Einstein's equations, see Section **??**•4.4.1 for details. From a physical point of view, such constructs are related to the

•4.4.1: **ptc:***where?*

fundamental postulate of general relativity, that *no signal can travel faster than the speed of light*. This is equivalent to the statement that the only events of space-times that are influenced by an event $p \in \mathscr{M}$ are those which belong to $J^+(\mathscr{M})$.

EXAMPLE 4.4.1 Let $\mathscr{M} = S^1 \times S^1$ with the flat metric $g = -dt^2 + d\varphi^2$. Geodesics of g through $(0,0)$ are of the form

$$\gamma(s) = (\alpha s \bmod 2\pi, \beta s \bmod 2\pi) \ , \tag{4.4.1}$$

where α and β are constants; the remaining geodesics are obtained by a rigid translation of (4.4.1). Clearly any two points of $\mathscr{M}$ can be joined by a timelike geodesic, which shows that for all $p \in \mathscr{M}$ we have

$$I^+(p) = J^+(p) = \mathscr{M} \ .$$

It is of some interest to point out that for irrational β/α in (4.4.1) the corresponding geodesic is dense in $\mathscr{M}$.

There is an obvious *meta-rule* in the theory of causality that whenever a property involving I^+ or J^+ holds, then an identical property will be true with I^+ replaced by I^-, and with J^+ replaced by J^-, or both. This is proved by changing the time-orientation of the manifold. Thus we will only make formal statements for the futures.

Another useful *meta-rule* is the following: suppose that a property involving $I^+(\mathscr{U})$ holds, and let $\mathscr{V}$ be an open subset of $\mathscr{M}$ containing $\mathscr{U}$. Then one can apply the result to the new space-time $(\mathscr{V}, g|_{\mathscr{V}})$, where $g_{\mathscr{V}}$ is the restriction of g to $\mathscr{V}$, obtaining an identical claim for $I^+(\mathscr{U};\mathscr{V})$.

Example 4.4.1 shows that in causally pathological space-times the notions of futures and pasts need not to carry interesting information. On the other hand those objects are useful tools to study the global structure of those space-times which possess reasonable causal properties.

We start with some elementary properties of futures and pasts:

PROPOSITION 4.4.2

1. $I^+(\mathscr{U}) \subset J^+(\mathscr{U})$.
2. $p \in I^+(q) \quad \Longleftrightarrow \quad q \in I^-(p)$.
3. $\mathscr{V} \subset I^+(\mathscr{U}) \quad \Longrightarrow \quad I^+(\mathscr{V}) \subset I^+(\mathscr{U})$.

Similar properties hold with I^+ replaced by J^+.

PROOF: 1. A timelike curve is a causal curve.

2. If $[0,1] \ni s \to \gamma(s)$ is a future directed causal curve from q to p, then $[0,1] \ni s \to \gamma(1-s)$ is a past directed causal curve from p to q.

3. Let us start by introducing some notation: consider $\gamma_a : [0,1] \to \mathscr{M}$, $a = 1,2$, two causal curves such that $\gamma_1(1) = \gamma_2(0)$. We define the *concatenation operation* $\gamma_1 \cup \gamma_2$ as follows:

$$(\gamma_1 \cup \gamma_2)(s) = \begin{cases} \gamma_1(s) \ , & s \in [0,1] \ , \\ \gamma_2(s-1) \ , & s \in [1,2] \ . \end{cases} \tag{4.4.2}$$

There is an obvious extension of this definition when the ranges of parameters of the γ_a's are not $[0,1]$, or when a finite number $i \geq 3$ of paths is considered, we leave the formal definition to the reader.

Let, now, $r \in I^+(\mathscr{V})$, then there exists $q \in \mathscr{V}$ and a future directed timelike curve γ_2 from q to r. Since $\mathscr{V} \subset I^+(\mathscr{U})$ there exists a future directed timelike curve γ_1 from some point $p \in \mathscr{U}$ to q. Then the curve $\gamma_1 \cup \gamma_2$ is a future directed timelike curve from $\mathscr{U}$ to r. □

We have the following, intuitively obvious, description of futures and pasts of points in Minkowski space-time; in Proposition 4.4.5 below we will shortly prove a similar local result in general space-times, with a considerably more complicated proof.

PROPOSITION 4.4.3 *Let $(\mathscr{M}, g)$ be the $(n+1)$-dimensional Minkowski space-time $\mathbb{R}^{1,n} := (\mathbb{R}^{1+n}, \eta)$, with Minkowskian coordinates $(x^\mu) = (x^0, \vec{x})$ so that*

$$\eta(\partial_\mu, \partial_\nu) = \mathrm{diag}(-1, +1, \ldots, +1) \, .$$

Then

1. *$I^+(0) = \{x^\mu : \eta_{\mu\nu}x^\mu x^\nu < 0, x^0 > 0\}$,*
2. *$J^+(0) = \{x^\mu : \eta_{\mu\nu}x^\mu x^\nu \leq 0, x^0 \geq 0\}$,*
3. *in particular the boundary $\dot{J}^+(0)$ of $J^+(0)$ is the union of $\{0\}$ together with all null future directed geodesics with initial point at the origin.*

PROOF: Let $\gamma(s) = (x^\mu(s))$ be a parameterized causal path in $\mathbb{R}^{1,n}$ with $\gamma(0) = 0$. At points at which γ is differentiable we have

$$\eta(\dot\gamma, \dot\gamma) = -\left(\frac{dx^0}{ds}\right)^2 + \left|\frac{d\vec{x}}{ds}\right|_\delta^2 \leq 0 \, , \quad \frac{dx^0}{ds} \geq \left|\frac{d\vec{x}}{ds}\right|_\delta^2 \geq 0 \, .$$

Now, similarly to a differentiable function, a locally Lipschitzian function is the integral of its distributional derivative (see Theorem 4.3.2) hence

$$\begin{aligned} x^0(s) &= \int_0^s \frac{dx^0}{ds}(u)du && \text{(4.4.3a)} \\ &\geq \int_0^s \left|\frac{d\vec{x}}{ds}\right|_\delta (u)du =: \ell(\gamma_s) \, . && \text{(4.4.3b)} \end{aligned}$$

Here $\ell(\gamma_s)$ is the length, with respect to the flat Riemannian metric δ, of the path γ_s, defined as

$$[0, s] \ni u \to \vec{x}(u) \in \mathbb{R}^n \, .$$

Let dist_δ denote the distance function of the metric δ, thus

$$\mathrm{dist}_\delta(\vec{x}, \vec{y}) = |\vec{x} - \vec{y}|_\delta \, ,$$

it is well known that

$$\ell_s \geq \mathrm{dist}_\delta(\vec{x}(s), \vec{x}(0)) = |\vec{x}(s) - \vec{x}(0)|_\delta = |\vec{x}(s)|_\delta \, .$$

Therefore

$$x^0(s) \geq |\vec{x}(s)|_\delta \ ,$$

and point 2. follows. For timelike curves the same proof applies, with all inequalities becoming strict, establishing point 1. Point 3. is a straightforward consequence of point 2. □

•4.4.2 There is a natural generalisation of Proposition 4.4.3 to the following class of metrics on $\mathbb{R}\times\mathscr{S}$: •4.4.2: **ptc:***added, to be proofread*

$$g = -\varphi dt^2 + h \ , \qquad \partial_t \varphi = \partial_t h = 0 \ , \tag{4.4.4}$$

where h is a Riemannian metric on $\mathscr{S}$. (Such metrics are sometimes called *warped-products*, with *warping function* φ.)

PROPOSITION 4.4.4 *Let $\mathscr{M} = \mathbb{R}\times\mathscr{S}$ with the metric (4.4.4), and let $p\in\mathscr{S}$. Then $J^+((0,p))$ is the graph over $\mathscr{S}$ of the distance function* $\mathrm{dist}_{\hat h}(p,\cdot)$ *of the* optical metric

$$\hat h := \varphi^{-1} h \ ,$$

while $I^+((0,p))$ is the epigraph of $\mathrm{dist}_{\hat h}(p,\cdot)$,

$$I^+((0,p)) = \{(t,q) : t > \mathrm{dist}_{\hat h}(p,q)\} \ .$$

PROOF: Since the causal character of a curve is invariant under conformal transformations, the causal and timelike futures with respect to the metric g coincide with those with respect to the metric

$$\varphi^{-1} g = -dt^2 + \hat h \ .$$

Arguing as in the the proof of Proposition 4.4.3, (4.4.3) becomes

$$x^0(s) = \int_0^s \frac{dx^0}{ds}(u)du \tag{4.4.5a}$$

$$\geq \int_0^s \left|\frac{d\vec{x}}{ds}\right|_{\hat h}(u)du =: \ell_{\hat h}(\gamma_s) \ , \tag{4.4.5b}$$

where $\ell_{\hat h}(\gamma_s)$ denotes the length of γ_s with respect to $\hat h$, and one concludes as before. □

The next result shows that, locally, causal behaviour is identical to that of Minkowski space-time. The proof of this "obvious" fact turns out to be surprisingly involved:

PROPOSITION 4.4.5 *Let $\mathscr{O}_p$ be a domain of normal coordinates x^μ centered at $p\in\mathscr{M}$ as in Proposition 4.2.3. Let*

$$\mathscr{O} \subset \mathscr{O}_p$$

be any normal–coordinate ball such that ∇x^0 is timelike on $\mathscr{O}$. Recall (compare (4.2.8)) that the function $\sigma_p : \mathscr{O}_p \to \mathbb{R}$ has been defined by the formula

$$\sigma_p(\exp_p(x^\mu)) := \eta_{\mu\nu}x^\mu x^\nu \ . \tag{4.4.6}$$

Then

$$\mathscr{O} \ni q = \exp_p(x^\mu) \in \begin{cases} I^+(p;\mathscr{O}) & \Longleftrightarrow \quad \sigma_p(q) < 0\ ,\ x^0 > 0\ , \\ J^+(p;\mathscr{O}) & \Longleftrightarrow \quad \sigma_p(q) \le 0\ ,\ x^0 \ge 0\ , \\ \dot J^+(p;\mathscr{O}) & \Longleftrightarrow \quad \sigma_p(q) = 0\ ,\ x^0 \ge 0\ , \end{cases} \tag{4.4.7}$$

with the obvious analogues for pasts. In particular, a point $q = \exp_p(x^\mu) \in \dot J^+(p;\mathscr{O}_p)$ *if and only if* q *lies on the null geodesic segment* $[0,1] \ni s \to \gamma(s) = \exp_p(sx^\mu) \in \dot J^+(p;\mathscr{O}_p)$.

Remark 4.4.6 Example 4.4.1 shows that $I^\pm(p;\mathscr{O})$, *etc.*, cannot be replaced by $I^\pm(p)$, because causal paths through p can exit $\mathscr{O}_p$ and reenter it; this can actually happen again and again.

Before proving Proposition 4.4.5, we note the following straightforward implication thereof:

Proposition 4.4.7 *Let* $\mathscr{O}$ *be as in Proposition 4.4.5, then* $I^+(p;\mathscr{O})$ *is open.*

□

Proof of Proposition 4.4.5: As the coordinate rays are geodesics, the implications "$\Leftarrow$" in (4.4.7) are obvious. It remains to prove "$\Rightarrow$". We start with a lemma:

•*4.4.3:* **ptc:***do not impose past pointing gradient*

Lemma 4.4.8 *Let* τ *be a* time function, i.e., *a differentiable*•4.4.3 *function with* timelike past-pointing *gradient. For any* τ_0, *a future directed causal path* γ *cannot leave the set* $\{q : \tau(q) > \tau_0\}$*; the same holds for sets of the form* $\{q : \tau(q) \ge \tau_0\}$. *In fact,* τ *is non-decreasing along* γ, *strictly increasing if* γ *is timelike.*

Proof: Let $\gamma : I \to \mathscr{M}$ be a future directed parameterized causal path, then $\tau\circ\gamma$ is a locally Lipschitzian function, hence equals the integral of its derivative on any compact subset of its domain of definition, so that

$$\begin{aligned} \tau(\gamma(s_2)) - \tau(\gamma(s_1)) &= \int_{s_1}^{s_2} \frac{d(\tau\circ\gamma)}{du}(u)du \\ &= \int_{s_1}^{s_2} \langle d\tau, \dot\gamma\rangle(u)du \\ &= \int_{s_1}^{s_2} g(\nabla\tau,\dot\gamma)(u)du \ge 0\ , \end{aligned} \tag{4.4.8}$$

since $\nabla\tau$ is timelike past directed, while $\dot\gamma$ is causal future directed or zero wherever defined. The function $s \to \tau(\gamma(s))$ is strictly increasing when γ is timelike, since then the integrand in (4.4.8) is strictly positive almost everywhere. □

Applying Lemma 4.4.8 to the time function x^0 we obtain the claim about x^0 in (4.4.7). To justify the remaining claims of Proposition 4.4.5, we recall Equation (4.2.9)

$$\nabla\sigma_p \text{ is } \begin{cases} \text{timelike future directed} & \text{on } \{q : \sigma_p(q) < 0\ ,\ x^0(q) > 0\}\ , \\ \text{null future directed} & \text{on } \{q : \sigma_p(q) = 0\ ,\ x^0(q) > 0\}\ . \end{cases} \tag{4.4.9}$$

Let $\gamma = (\gamma^\mu) : I \to \mathscr{O}$ be a parameterized future directed causal path with $\gamma(0) = p$, then $\sigma_p \circ \gamma$ is a locally Lipschitzian function, hence

$$\begin{aligned} \sigma_p \circ \gamma(t) &= \int_0^t \frac{d(\sigma_p \circ \gamma)(s)}{ds} ds \\ &= \int_0^t g(\nabla \sigma_p, \dot\gamma)(s) ds \; . \end{aligned} \tag{4.4.10}$$

We note the following:

LEMMA 4.4.9 *A future directed causal path $\gamma \subset \mathscr{O}_p$ cannot leave the set $\{q : x^0(q) > 0, \sigma_p(q) < 0\}$.*

PROOF: The time function x^0 remains positive along γ by Lemma 4.4.8. If $-\sigma_p$ were also a time function we would be done by the same argument. The problem is that $-\sigma_p$ is a time function only on the set where σ_p is negative, so some care is needed; we proceed as follows: The vector field $\nabla\sigma_p$ is causal future directed on $\{x^0 > 0, \eta_{\mu\nu}x^\mu x^\nu \le 0\}$, while $\dot\gamma$ is causal future directed or zero wherever defined, hence $g(\nabla\sigma_p, \dot\gamma) \le 0$ as long as γ stays in $\{x^0 > 0, \eta_{\mu\nu}x^\mu x^\nu \le 0\}$. By Equation (4.4.9) the function σ_p is non-increasing along γ as long as γ stays in $\{x^0 > 0, \eta_{\mu\nu}x^\mu x^\nu \le 0\}$. Suppose that $\sigma_p(\gamma(s_1)) < 0$ and let

$$s_* = \sup\{u \in I : \sigma_p(\gamma(s)) < 0 \quad \text{on} \quad [s_1, u]\} \; .$$

If $s_* \in I$, then $\sigma_p \circ \gamma(s_*) = 0$ and $\sigma_p \circ \gamma$ is *not* non-increasing on $[s_1, s_*)$, which is not possible since $\gamma(s) \in \{x^0 > 0, \eta_{\mu\nu}x^\mu x^\nu \le 0\}$ for $s \in [s_1, s_*)$. It follows that $\sigma_p \circ \gamma < 0$, as desired. □

Proposition 4.4.5 immediately follows for those future direct causal paths through p which *do enter* the set $\{\eta_{\mu\nu}x^\mu x^\nu < 0\}$. This is the case for γ's such that $\dot\gamma(0) = (\dot\gamma^\mu(0))$ exists and is timelike: We then have

$$\gamma^\mu(s) = s\dot\gamma^\mu(0) + o(s) \; ,$$

hence

$$\eta_{\mu\nu}\gamma^\mu(s)\gamma^\nu(s) = s^2\eta_{\mu\nu}\dot\gamma^\mu(0)\dot\gamma^\nu(0) + o(s^2) < 0$$

for s small enough. It follows that γ enters the set $\{\eta_{\mu\nu}x^\mu x^\nu < 0\} \equiv \{q : \sigma_p(q) < 0\}$, and remains there for $|s|$ small enough. We conclude using Lemma 4.4.9.

We continue with arbitrary parameterized future directed *timelike* paths $\gamma : [0, b) \to \mathscr{M}$, with $\gamma(0) = p$, thus $\dot\gamma$ exists and is timelike future directed for almost all $s \in [0, b)$. In particular there exists a sequence $s_i \to_{i\to\infty} 0$ such that $\dot\gamma(s_i)$ exists and is timelike.

Standard properties of solutions of ODE's show that for each $q \in \mathscr{O}_p$ there exists a neighborhood $\mathscr{W}_{p,q}$ of p such that the function

$$\mathscr{W}_{p,q} \ni r \to \sigma_r(q)$$

is defined, continuous in r. For i large enough we will have $\gamma(s_i) \in \mathscr{W}_{p,\gamma(s)}$; for such i's we have just shown that

$$\sigma_{\gamma(s_i)}(\gamma(s)) < 0 \; .$$

Passing to the limit $i \to \infty$, by continuity one obtains

$$\sigma_p(\gamma(s)) \leq 0 \ , \tag{4.4.11}$$

thus

$$\gamma \subset \{x^0 \geq 0, \eta_{\mu\nu}x^\mu x^\nu \leq 0\} \ . \tag{4.4.12}$$

Since $\dot\gamma$ is timelike future directed wherever defined, and $\nabla\sigma_p$ is causal future directed on $\{x^0 > 0, \eta_{\mu\nu}x^\mu x^\nu \leq 0\}$, Equations (4.4.10) and (4.4.12) show that the inequality in (4.4.11) must be strict.

To finish the proof, we reduce the general case to the last one by considering perturbed metrics, as follows: let e_0 be any unit timelike vector field on $\mathscr{O}$ (e_0 can, *e.g.*, be chosen as $\nabla x^0/\sqrt{-g(\nabla x^0, \nabla x^0)}$), for $\epsilon > 0$ define a family of Lorentzian metrics g_ϵ on $\mathscr{O}$ by the formula

$$g_\epsilon(X,Y) = g(X,Y) - \epsilon g(e_0,X)g(e_0,Y) \ .$$

Consider any vector X which is causal for g, then

$$\begin{aligned} g_\epsilon(X,X) &= g(X,X) - \epsilon(g(e_0,X))^2 \\ &\leq -\epsilon(g(e_0,X))^2 < 0 \ , \end{aligned}$$

so that X is timelike for g_ϵ. Let $\sigma(g_\epsilon)_p$ be the associated functions defined as in Equation (4.4.6), where the exponential map there is the one associated to the metric g_ϵ. Standard properties of solutions of ODE's imply that for any compact subset K of $\mathscr{O}_p$ there exists an $\epsilon_K > 0$ and a neighborhood $\mathscr{O}_{p,K}$ of K such that for all $\epsilon \in [0, \epsilon_K]$ the functions

$$\mathscr{O}_{p,K} \ni q \to \sigma(g_\epsilon)_p(q)$$

are defined, and depend continuously upon ϵ. We take K to be $\gamma([0,s])$, where s is such that $[0,s] \subset I$, and consider any ϵ in $(0, \epsilon_{\gamma([0,s])})$. Since γ is timelike for g_ϵ, the results already established show that we have

$$\sigma(g_\epsilon)_p(\gamma(s)) < 0 \ .$$

Continuity in ϵ implies

$$\sigma_p(\gamma(s)) \leq 0 \ .$$

Since s is arbitrary in I, Proposition 4.4.5 is established. □

For certain considerations it is useful to have the following:

COROLLARY 4.4.10 *Under the hypotheses of Proposition 4.4.5, let $\gamma \subset \mathscr{O}$ be a causal curve from p to*

$$q = \exp_p(x^\mu) \in \dot J^+(p;\mathscr{O}) \ .$$

Then there exists a reparameterization $s \to r(s)$ of γ so that

$$[0,1] \ni s \to \gamma(r(s)) = \exp_p(sx^\mu) \ .$$

PROOF: Proposition 4.4.5 shows that $\sigma_p(q) = 0$. It follows that

$$0 = \sigma_p \circ \gamma(t) = \int_0^t g(\nabla\sigma_p, \dot\gamma)(s)ds \; . \tag{4.4.13}$$

Since $\nabla\sigma_p$ and $\dot\gamma$ are causal oppositely directed we have $g(\nabla\sigma_p, \dot\gamma) \geq 0$ almost everywhere. It thus follows from (4.4.13) that

$$g(\nabla\sigma_p, \dot\gamma) = 0$$

almost everywhere. This is only possible if

$$\nabla\sigma_p \sim \dot\gamma$$

almost everywhere (see Proposition 2.2.2), which gives the result.•4.4.4 □

•4.4.4: **ptc:***needs a better justification*

Penrose's approach [89]•4.4.5 to the theory of causality is based on the notion of *timelike or causal trips*: by definition, a causal trip is a piecewise broken causal geodesic. The following result can be used to show equivalence of the definitions of I^+, *etc.*, given here, to those of Penrose:

•4.4.5: **ptc:***move near the definition of timelike curve; make a summary of all definitions: Wald, Seifert (look it up), HE, GerochDoD,BeemEhrlich*

COROLLARY 4.4.11 *If $q \in I^+(p)$, then there exists a future directed piecewise broken future directed timelike geodesic from p to q. Similarly, if $q \in J^+(p)$, then there exists a future directed piecewise broken future directed causal geodesic from p to q.*

PROOF: Let $\gamma : [0,1] \to \mathscr{M}$ be a parameterized future directed causal path with $\gamma(0) = p$ and $\gamma(1) = q$. Continuity of γ implies that for every $s \in [0,1]$ there exists $\epsilon_s > 0$ such that

$$\gamma(u) \in \mathscr{O}_{\gamma(s)}$$

for all

$$u \in (s - 2\epsilon_2, s + 2\epsilon_2) \cap [0,1] = [\max(0, s - 2\epsilon_s), \min(1, s + 2\epsilon_s)] \; ,$$

where $\mathscr{O}_r$ is a normal-coordinates ball centred at r, and satisfying the requirements of Proposition 4.4.5. Compactness of $[0,1]$ implies that from the covering $\{(s - \epsilon_s, s + \epsilon_s)\}_{s\in[0,1]}$ a finite covering $\{(s_i - \epsilon_{s_i}, s_i + \epsilon_{s_i})\}_{i=0,\dots,N}$ can be extracted, with $s_0 = 0$, $s_N = 1$. Reordering the s_i's if necessary we may assume that $s_i < s_i + 1$. By definition we have

$$\gamma|_{[s_i, s_{i+1}]} \subset \mathscr{O}_{\gamma(s_i)} \; ,$$

and by Proposition 4.4.5 there exists a causal future directed geodesic segment from $\gamma(s_i)$ to $\gamma(s_{i+1})$: if $\gamma(s_{i+1}) = \exp_{\gamma(s_i)}(x^\mu)$, then the required geodesic segment is given by

$$[0,1] \ni s \to \exp_{\gamma(s_i)}(sx^\mu) \; .$$

If γ is timelike, then all the segments are timelike. Concatenating the segments together provides the claimed piecewise broken geodesic. □

Proposition 4.4.3 shows that the sets $I^\pm(p)$ are open in Minkowski spacetime. Similarly it follows from Proposition 4.4.5 that the sets $I^\pm(p;\mathscr{O}_p)$ are open. This turns out to be true in general:

PROPOSITION 4.4.12 *For all $\mathscr{U} \subset \mathscr{M}$ the sets $I^\pm(\mathscr{U})$ are open.*

PROOF: Let $q \in I^+(\mathscr{U})$, and let, as in the proof of Corollary 4.4.11, s_{N-1} be such that $q \in \mathscr{O}_{\gamma(s_{N-1})}$. Then

$$\mathscr{O}_{\gamma(s_{N-1})} \cap I^+(\gamma(s_{N-1}); \mathscr{O})$$

is an open neighborhood of q by Corollary 4.4.7. Clearly

$$I^+(\gamma(s_{N-1}); \mathscr{O}) \subset I^+(\gamma(s_{N-1})) .$$

Since $\gamma(s_{N-1}) \in I^+(\mathscr{U})$ we have

$$I^+(\gamma(s_{N-1})) \subset I^+(\mathscr{U})$$

(see point 3. of Proposition 4.4.2). It follows that

$$\mathscr{O}_{\gamma(s_{N-1})} \cap I^+(\gamma(s_{N-1}); \mathscr{O}) \subset I^+(\mathscr{U}) ,$$

which implies our claim. $\square$

In Minkowski space-time the sets $J^\pm(p)$ are closed, with

$$\overline{I^\pm(p)} = J^\pm(p) . \tag{4.4.14}$$

We will show below (see Corollary 4.4.17) that we always have

$$\overline{I^\pm(p)} \supset J^\pm(p) , \tag{4.4.15}$$

but this requires some work. Before proving (4.4.15), let us point out that (4.4.14) does not need to be true in general:

EXAMPLE 4.4.13 Let $(\mathscr{M}, g)$ be the two-dimensional Minkowski space-time $\mathbb{R}^{1,1}$ from which the set $\{x^0 = 1, x^1 \le -1\}$ has been removed. Then

$$J^+(0; \mathscr{M}) = J^+(0, \mathbb{R}^{1,1}) \setminus \{x^0 = -x^1 \ , x^1 \in (-\infty, 1]\} ,$$

cf. Figure 4.2, hence $J^+(0; \mathscr{M})$ is neither open nor closed, and Equation (4.4.14) does not hold.

We have the following:

LEMMA 4.4.14 ("Push-up Lemma 1") *For any $\Omega \subset \mathscr{M}$ we have*

$$I^+(J^+(\Omega)) = I^+(\Omega) . \tag{4.4.16}$$

PROOF: The obvious property

$$\mathscr{U} \subset \mathscr{V} \quad \Longrightarrow \quad I^+(\mathscr{U}) \subset I^+(\mathscr{V})$$

provides inclusion of the right-hand-side of (4.4.16) into the left-hand-side. It remains to prove that

$$I^+(J^+(\Omega)) \subset I^+(\Omega) .$$

Figure 4.2: $J^+(p)$ is not closed unless some global causal regularity conditions are fulfilled by $(\mathscr{M}, g)$.

Let $r \in I^+(J^+(\Omega))$, thus there exists a past-directed timelike curve γ_0 from r to a point $q \in J^+(\Omega)$. Since $q \in J^+(\Omega)$, then either $q \in \Omega$, and there is nothing to prove, or there exists a past-directed causal curve $\gamma : I \to \mathscr{M}$ from q to some point $p \in \Omega$. We want to show that there exists a past-directed timelike curve $\hat{\gamma}$ starting at r and ending at p. The curve $\hat{\gamma}$ can be obtained by "pushing-up" γ slightly, to make it timelike, the construction proceeds as follows: Using compactness, we cover γ by a finite collection $\mathscr{U}_i$, $i = 0, \cdots, N$, of elementary regions $\mathscr{U}_i$ centered at $p_i \in \gamma(I)$, with

$$p_0 = q \;, \quad p_i \in \mathscr{U}_i \cap \mathscr{U}_{i+1} \;, \quad p_{i+1} \subset J^-(p_i) \;, \quad p_N = p \;.$$

Let $\gamma_0 : [0, s_0] \to \mathscr{M}$ be the already mentioned causal curve from r to $q \in \mathscr{U}_1$; let s_1 be close enough to s_0 so that $\gamma_0(s_1) \in \mathscr{U}_1$. By Proposition 4.2.3 together with the definition of elementary regions there exists a past directed timelike curve $\gamma_1 : [0, 1] \to \mathscr{U}_1$ from $\gamma_0(s_1)$ to $p_1 \in \mathscr{U}_1 \cap \mathscr{U}_2$. For s close enough to 1 the curve γ_1 enters $\mathscr{U}_2$, choose an s_2 such that $\gamma_1(s_2) \in \mathscr{U}_2$, again by Proposition 4.2.3 there exists a a past directed timelike curve $\gamma_2; [0, 1] \to \mathscr{U}_2$ from $\gamma_1(s_2)$ to p_2. One repeats that construction iteratively obtaining a (finite) sequence of past-directed timelike curves $\gamma_i \subset I^+(\gamma) \cap \mathscr{O}$ such that the end point $\gamma_i(s_{i+1})$ of $\gamma_i|_{[0,s_{i+1}]}$ coincides with the starting point of γ_{i+1}. Concatenating those curves together gives the desired path $\hat{\gamma}$. □

We have the following, slightly stronger, version of Lemma 4.4.14, which gives a sufficient condition to be able to deform a *causal* curve to a *timelike* one, keeping the deformation as small as desired:

COROLLARY 4.4.15 *Consider a <u>causal</u> future directed curve $\gamma : [0, 1] \to \mathscr{M}$ from p to q. If there exist $s_1 < s_2 \in [0, 1]$ such that $\gamma|_{[s_1,s_2]}$ is <u>timelike</u>, then in any neighborhood $\mathscr{O}$ of γ there exists a <u>timelike</u> future directed curve $\hat{\gamma}$ from p to q.*

REMARK 4.4.16 The so-called *maximizing* causal geodesics can *not* be deformed as above to timelike curves, whether locally or globally. For example, any causal geodesic in Minkowski space-time has this property.

PROOF: : If $s_2 = 1$, then Corollary 4.4.15 is essentially a special case of Lemma 4.4.14: the only difference is the statement about the neighborhood $\mathscr{O}$. This last requirement can be satisfied by choosing the sets $\mathscr{U}_i$ in the proof of Lemma 4.4.14 so that $\mathscr{U}_i \subset \mathscr{O}$. If $s_1 = 0$ (and regardless of the value of s_2) the result is obtained by changing time-orientation, applying the result already established to the path $\gamma'(s) = \gamma(1-s)$, and changing-time orientation again. The general case is reduced to the ones already covered by first deforming the curve $\gamma|_{[0,s_2]}$ to a new timelike curve $\tilde{\gamma}$ from p to $\gamma(s_2)$, and then applying the result again to the curve $\tilde{\gamma} \cup \gamma|_{[s_2,1]}$. □

As another straightforward corollary of Lemma 4.4.14 one obtains:

COROLLARY 4.4.17 *For any $p \in \mathscr{M}$ we have*

$$J^+(p) \subset \overline{I^+(p)} \, .$$

PROOF: Let $q \in J^+(p)$, and let $r_i \in I^+(q)$ be any sequence of points accumulating at q, then $r_i \in I^+(p)$ by Lemma 4.4.14, hence $q \in \overline{I^+(p)}$. □

4.5 Extendible and inextendible paths

A useful concept, when studying causality, is that of a causal path with cannot be extended any further. Recall that, from a physical point of view, the image in space-time of a timelike path is supposed to represent the history of some observer, and it is sometimes useful to have at hand idealised observers which do never stop to exist. Here it is important to have in mind the geometrical picture in mind, where all that matters is the image in space-time of the path, independently of any paramaterisation: if that image "stops", then one can sometimes continue the path by concatenating with a further one; continuing in this way one hopes to be able to obtain paths which are inextendible.

In order to make things precise, let $\gamma : [a,b) \to \mathscr{M}$, be a parameterized, causal, future directed path. A point p is called a *future end point* of γ if $\lim_{s\to b} \gamma(s) = p$. Past end points are defined in the obvious analogous way. An *end point* is a point which is either a past end point or a future end point.

Given γ as above, together with an end point p, one is tempted to extend γ to a new path $\hat{\gamma} : [a,b] \to \mathscr{M}$ defined as

$$\hat{\gamma}(s) = \begin{cases} \gamma(s), & s \in [a,b) \, , \\ p, & s = b \, . \end{cases} \tag{4.5.1}$$

The first problem with this procedure is that the resulting curve might fail to be locally Lipschitz in general. An example is given by the timelike future-directed path

$$[0,1) \ni \to \gamma_1(s) = (-(1-s)^{1/2}, 0) \in \mathbb{R}^{1,1} \, ,$$

which is locally Lipschitzian on $[0,1)$, but is *not* on $[0,1]$. (This follows from the fact that the difference quotient $(f(s) - f(s'))/(s-s')$ blows up as s and s' tend to one when $f(s) = (1-s)^{1/2}$). Recall that in our definition of a causal curve γ, a prerequisite condition is the locally Lipschitz character, so that the extension $\hat{\gamma}_1$ fails to be causal even though γ_1 is.

The problem is even worse if $b = \infty$: consider the timelike future-directed curve

$$[1, \infty) \ni s \to \gamma_3(s) = (-1/s, 0) \in \mathbb{R}^{1,1} .$$

Here there is no way to extend the curve to the future, as an application from a subset of $\mathbb{R}$ to $\mathscr{M}$, because the range of parameters already covers all $s \geq 1$. Now, the image of both γ_1 and γ_2 is simply the interval $[-1, 0) \times \mathbb{R}$, which can be extended to a longer causal curve in $\mathbb{R}^{1,1}$ in many ways if one thinks in terms of images rather than of maps.

Both problems above can be taken care of by requiring that the parameter s be the proper distance parameter of some auxiliary Riemannian metric h.•4.5.1 (At this stage h is not required to be complete). This might require reparameterizing the path. From the point of view of our definition this means that we are passing to a different path, but the image in space-time of the new path coincides with the previous one. If one thinks of timelike paths as describing observers, the new observer will thus have experienced identical events, even though he will be experiencing those events at different times on his time-measuring device. We note, moreover, that (locally Lipschitz) reparameterizations do not change the timelike or causal character of paths.

•*4.5.1:* **ptc:***wrong, since you can move the parameterization by shifting it*

We have already shown in Section 4.3 that a locally Lipschitzian path can always be reparameterized by h-distance, leading to a uniformly Lipschitzian path, with Lipschitz constant one. It should be clear from the examples given above, as well as from the examples discussed at the beginning of Section 4.6, that it is sensible to use such a parameterization, and it is tempting to build this requirement into the definition of a causal path. One reason for not doing that is the existence of affine parameterization for geodesics, which is geometrically significant, and which is convenient for several purposes. Another reason is the arbitrariness related to the choice of h. Last but not least, a limit curve for a sequence of dist_h-parameterized curves does not have to be dist_h-parameterized. Therefore we will not assume *a priori* an h-distance parameterization, but such a reparameterization will often be used in the proofs.•4.5.2

•*4.5.2:* **ptc:***added VIII.2004, to reread*

Returning to (4.5.1), we want to show that $\hat{\gamma}$ will be uniformly Lipschitz if dist_h-parameterization is used for γ. More generally, suppose that γ is uniformly Lipschitz with Lipschitz with constant L,

$$\mathrm{dist}_h(\gamma(s), \gamma(s')) \leq L|s - s'| . \tag{4.5.2}$$

Passing with s' to b in that equation we obtain

$$\mathrm{dist}_h(\gamma(s), p) \leq L|s - b| ,$$

and the Lipschitzian character of $\hat{\gamma}$ easily follows. We have therefore proved:

LEMMA 4.5.1 *Let $\gamma : [a, b) \to \mathscr{M}$, $b < \infty$, be a uniformly Lipschitzian path with an end point p. Then γ can be extended to a uniformly Lipschitzian path $\hat{\gamma} : [a, b] \to \mathscr{M}$, with $\hat{\gamma}(b) = p$.*

Let $\gamma : [a, b) \to \mathscr{M}$, $b \in \mathbb{R} \cup \{\infty\}$ be a path, then p is said to be an *ω-limit point* of γ if there exists a sequence $s_k \to b$ such that $\gamma(s_k) \to p$. An end point is always an ω-limit point, but the inverse does not need to be true in general (consider $\gamma(s) = \exp(is) \in \mathbb{C}$, then every point $\exp(ix) \in S^1 \subset \mathbb{C}^1$ is an ω-limit point of γ — set $s_k = x + 2\pi k$). For $b < \infty$ and for uniformly Lipschitz paths the notions of end point and of ω-limit point coincide:

LEMMA 4.5.2 *Let $\gamma : [a, b) \to \mathscr{M}$, $b < \infty$, be a uniformly Lipschitzian path. Then every ω-limit point of γ is an end point of γ. In particular, γ has at most one ω-limit point.*

PROOF: By (4.5.2) we have

$$\mathrm{dist}_h(\gamma(s_i), \gamma(s)) \leq L|s_i - s| ,$$

and since dist_h is a continuous function of its arguments we obtain, passing to the limit $i \to \infty$

$$\mathrm{dist}_h(p, \gamma(s)) \leq L|b - s| .$$

Thus p is an end point of γ. Since there can be at most one end point, the result follows. □

A future directed causal curve $\gamma : [a, b) \to \mathscr{M}$ will be said to be *future extendible* if there exists $b < c \in \mathbb{R} \cup \{\infty\}$ and a causal curve $\tilde{\gamma} : [a, c) \to \mathscr{M}$ such that

$$\tilde{\gamma}|_{[a,b)} = \gamma . \tag{4.5.3}$$

The path $\tilde{\gamma}$ is then said to be an *extension* of γ. The curve γ will be said *future inextendible* if it is not future extendible. The notions of *past extendibility*, or of *extendibility*, are defined in the obvious way.

Extendibility in the class of causal paths forces a causal $\gamma : [a, b) \to \mathscr{M}$ to be uniformly Lipschitzian: This follows from the fact that $[a, b]$ is a compact subset of the domain of definition of any extension $\tilde{\gamma}$, so that $\tilde{\gamma}|_{[a,b]}$ is uniformly Lipschitzian there. But then $\tilde{\gamma}|_{[a,b)}$ is also uniformly Lipschitzian, and the result follows from (4.5.3).

Whenever a uniformly Lipschitzian path can be extended by adding an end point, it can also be extended as a strictly longer path:

LEMMA 4.5.3 *A uniformly Lipschitzian causal path $\gamma : [a, b) \to \mathscr{M}$, $b < \infty$ is extendible if and only if it has an end point.*

PROOF: Let $\hat{\gamma}$ be given by Proposition 4.5.1, and let $\tilde{\gamma} : [0, d)$ be any maximally extended to the future, future directed causal geodesic starting at p, for an appropriate $d \in (0, \infty)$. Then $\hat{\gamma} \cup \tilde{\gamma}$ is an extension of γ. □

It turns out that the paths considered in Lemma 4.5.3 are always extendible:

THEOREM 4.5.4 *Let $\gamma : [a, b) \to \mathscr{M}$, $b \in \mathbb{R} \cup \{\infty\}$, be a future directed causal path parameterized by h-distance, where h is any complete auxiliary Riemannian metric. Then γ is future inextendible if and only if $b = \infty$.*

PROOF: Suppose that $b < \infty$. Let $B_h(p,r)$ denote the open h-distance ball, with respect to the metric h, of radius r, centred at p. Since γ is parameterized by h-distance we have, by (4.3.2),

$$\gamma([a,b)) \subset B_h(\gamma(a), b-a) \; .$$

The Hopf-Rinow theorem[4] asserts that $\overline{B_h(\gamma(a), b-a)}$ is compact, therefore there exists $p \in \overline{B_h(\gamma(a), b-a)}$ and a sequence s_i such that

$$[a,b) \ni s_i \to_{i\to\infty} b \quad \text{and} \quad \gamma(s_i) \to p \; .$$

Thus p is an ω-limit point of γ. Clearly γ is uniformly Lipschitzian (with Lipschitz modulus one), and Lemma 4.5.2 shows that p is an end point of γ. The result follows now from Lemma 4.5.3. □

4.5.1 Maximally extended geodesics

Consider the Cauchy problem for an affinely-parameterized geodesic γ:

$$\nabla_{\dot\gamma}\dot\gamma = 0 \; , \quad \gamma(0) = p \; , \quad \dot\gamma(0) = X \; . \tag{4.5.4}$$

This is a second-order ODE which, by the standard theory [?], for $C^{1,1}$ metrics, has unique solutions defined on a maximal interval $I = I(p,X) \ni 0$. I is maximal in the sense that if I' is another interval containing 0 on which a solution of (4.5.4) is defined, then $I' \subset I$. When I is maximal the geodesic will be called *maximally extended.* Now, it is not immediately obvious that a *maximally extended* geodesic is *inextendible* in the sense just defined: To start with, the notion of inextendibility involves only the pointwise properties of a path, while the notion of maximally extended geodesic involves the ODE (4.5.4), which involves both the first and second derivatives of γ. Next, the inextendibility criteria given above have been formulated in terms of uniformly Lipschitzian parameterizations. While an affinely parameterized geodesic is certainly locally Lipschitzian, there is no *a priori* reason why it should be uniformly so, when maximally extended. All these issues turn out to be irrelevant, and we have the following:

PROPOSITION 4.5.5 *A geodesic $\gamma : I \to \mathscr{M}$ is maximally extended as a geodesic if and only if γ is inextendible as a causal path.*

PROOF: Suppose, for contradiction, that γ is a maximally extended geodesic which is extendible as a path, thus γ can be extended to a path $\hat\gamma$ by adding its end point p as in (4.5.1). Working in a normal coordinate neighborhood $\mathscr{O}_p$ around p, $\hat\gamma \cap \mathscr{O}_p$ has a last component which is a geodesic segment which ends at p. By construction of normal coordinates the component of $\hat\gamma$ in question is simply a half-ray through the origin, which can be clearly be continued through p as a geodesic. This contradicts maximality of γ as a geodesic. It follows that a maximally extended geodesic is inextendible. Now, if γ is inextendible as a path, then γ can clearly not be extended as a geodesic, which establishes the reverse implication. □

A result often used in causality theory is the following:

THEOREM 4.5.6 *Let γ, be a future directed causal, respectively timelike, path. Then there exists an inextendible causal, respectively timelike, extension of γ.*

PROOF: If $\gamma : [a, b) \to \mathscr{M}$ is inextendible there is nothing to prove; otherwise the path $\hat{\gamma} \cup \tilde{\gamma}$, where $\hat{\gamma}$ is given by Proposition 4.5.1, and $\tilde{\gamma}$ is any maximally extended future directed causal geodesic as in the proof of Proposition 4.5.3, provides an extension. This extension is inextendible by Proposition 4.5.5. □
•4.5.3

•*4.5.3:* **ptc:***Note that this does not work for rough metrics, so another argument has to be given*

4.6 Limits of curves

A key tool in the analysis of global properties of space-times is the analysis of sequences of curves. One typically wants to obtain a limiting curve, and study its properties. The object of this section is to establish the existence of such limiting curves.

We wish, first, to find the ingredients needed for a useful notion of a limit of curves. It is enlightening to start with several examples. The first question that arises is whether to consider a sequence of curves γ_n defined on a common interval I, or whether one should allow different domains I_n for each γ_n. To illustrate that this last option is very unpractical, consider the family of curves

$$(-1/n, 1/n) \ni s \to \gamma_n(s) = (s, 0) \in \mathbb{R}^{1,1} \ . \tag{4.6.1}$$

The only sensible geometric object to which the $\gamma_n(s)$ converge is the constant map

$$\{0\} \ni s \to \gamma_\infty(s) = 0 \in \mathbb{R}^{1,1} \ , \tag{4.6.2}$$

which is quite reasonable, except that it takes us away from the class of causal curves. To avoid such behavior we will therefore assume that all the curves γ_n have a common domain of definition I.

Next, there are various reasons why a sequence of curves might fail to have an "accumulation curve". First, the whole sequence might simply run to infinity. (Consider, for example, the sequence

$$\mathbb{R} \ni s \to \gamma_n(s) = (s, n) \in \mathbb{R}^{1,1} \ .)$$

This is avoided when one considers curves such that $\gamma_n(0)$ converges to some point $p \in \mathscr{M}$.

Further, there might be a problem with the way the curves are parameterized. As an example, let γ_n be defined as

$$(-1, 1) \ni s \to \gamma_n(s) = (s/n, 0) \in \mathbb{R}^{1,1} \ .$$

As in (4.6.1), the $\gamma_n(s)$ converge to the constant map

$$(-1, 1) \ni s \to \gamma_\infty(s) = 0 \in \mathbb{R}^{1,1} \ , \tag{4.6.3}$$

again not a causal curve. Another example of pathological parameterizations is given by the family of curves

$$\mathbb{R} \ni s \to \gamma_n(s) = (ns, 0) \in \mathbb{R}^{1,1} \ .$$

In this case one is tempted to say that the γ_n's accumulate at the path, say γ_1, if parameterization is not taken into account. However, such a convergence is extremely awkward to deal with when attempting to actually prove something. This last behavior can be avoided by assuming that all the curves are uniformly Lipschitz continuous, with the same Lipschitz constant. The simplest way of ensuring this is to parameterize all the curves by a length parameter with respect to our auxiliary complete Riemannian metric h.

Yet another problem arises when considering the family of Euclidean-distance-parameterized causal curves

$$\mathbb{R} \ni s \to \gamma_n(s) = (s+n, 0) \in \mathbb{R}^{1,1} \ .$$

This can be gotten rid of by shifting the distance parameter so that the sequence $\gamma_n(s_0)$ stays in a compact set, or converges, for some s_0 in the domain I.

The above discussion motivates the hypotheses of the following result:

PROPOSITION 4.6.1 *Let $\gamma_n : I \to \mathscr{M}$ be a sequence of uniformly Lipschitz future directed causal curves, and suppose that there exist $p \in \mathscr{M}$ such that*

$$\gamma_n(0) \to p \ . \tag{4.6.4}$$

Then there exists a future directed causal curve $\gamma : I \to \mathscr{M}$ and a subsequence γ_{n_i} converging to γ in the topology of uniform convergence of compact subsets of I.

Proposition 4.6.1 provides the justification for the following definition:

DEFINITION 4.6.2 *Let $\gamma_n : I \to \mathscr{M}$ be a sequence of paths. We shall say that $\gamma : I \to \mathscr{M}$ is an <u>accumulation curve</u> of the γ_n's, or that the γ_n's accumulate at γ, if there exists a subsequence γ_{n_i} that converges to γ, uniformly on compact subsets of I.*•4.6.1

•4.6.1: **ptc:***definition changed, should be checked for consistency*

In their treatment of causal theory, Hawking and Ellis [58] introduce a notion of *limit curve* for paths, regardless of parameterization, which we find very awkward to work with. A related but slightly more convenient notion of *cluster curve* is considered in [75], where the name of "limit curve" is used for yet another notion of convergence. As discussed in [16, 75], those definitions lead to pathological behavior in some situations. We have found the above notion of "accumulation curve" the most convenient to work with from several points of view.

A sensible terminology, in the context of Definition 4.6.2, could be "C^0_{loc}-limits of curves", but we prefer not to use the term "limit" in this context, as limits are usually unique, while Definition 4.6.2 allows for more than one accumulation curve.

PROOF OF PROPOSITION 4.6.1: The hypothesis that all the γ_n's are uniformly Lipschitz reads

$$\mathrm{dist}_h(\gamma_n(s), \gamma_n(s')) \le L|s - s'| \ , \tag{4.6.5}$$

for some constant L. This shows that the family $\{\gamma_n\}$ is equicontinuous, and (4.6.4) together with the Arzela-Ascoli theorem implies that for every compact set $K \subset I$ there exists a curve $\gamma_K : K \to \mathscr{M}$ and a subsequence γ_{n_i} which converges uniformly to γ_K on K. One can obtain a K-independent curve γ by a so-called *diagonalisation* procedure.

The diagonalisation procedure goes as follows: For ease of notation we consider $I = \mathbb{R}$, the same argument applies on any interval with obvious modifications. Let $\gamma_{n(i,1)}$ be the sequence which converges to $\gamma_{[-1,1]}$; applying Arzela-Ascoli to this sequence one can extract a subsequence $\gamma_{n(i,2)}$ of $\gamma_{n(i,1)}$ which converges uniformly to some curve $\gamma_{[-2,2]}$ on $[-2,2]$. Since $\gamma_{n(i,2)}$ is a subsequence of $\gamma_{n(i,1)}$, and since $\gamma_{n(i,1)}$ converges to $\gamma_{[-1,1]}$ on $[-1,1]$, one finds that $\gamma_{[-2,2]}$ restricted to $[-1,1]$ equals $\gamma_{[-1,1]}$. One continues iteratively: suppose that $\{\gamma_{n(i,k)}\}_{i\in\mathbb{N}}$ has been defined for some k, and converges to a curve $\gamma_{[-k,k]}$ on $[-k,k]$, then the sequence $\{\gamma_{n(i,k+1)}\}_{i\in\mathbb{N}}$ is defined as a subsequence of $\{\gamma_{n(i,k)}\}_{i\in\mathbb{N}}$ which converges to some curve $\gamma_{[-(k+1),k+1]}$ on $[-(k+1),k+1]$. The curve γ is finally defined as

$$\gamma(s) = \gamma_{[-k,k]}(s) \ ,$$

where k is any number such that $s \le k$. The construction guarantees that $\gamma_{[-k,k]}(s)$ does not depend upon k as long as $s \le k$.

It remains to show that γ is causal. Passing to the limit $n \to \infty$ in (4.6.5) one finds

$$\mathrm{dist}_h(\gamma(s),\gamma(s')) \le L|s-s'| \ . \tag{4.6.6}$$

For $q \in \mathscr{M}$ let $\mathscr{O}_q$ be an elementary neighborhood of q as in Proposition 4.4.5, and let σ_q be the associated function defined by (4.4.6). Let $s \in \mathbb{R}$ and consider any point $\gamma(s) \in \mathscr{M}$. Now, the size of the sets $\mathscr{O}_q$ can be controlled uniformly when q varies over compact subsets of $\mathscr{M}$. It follows that for all s' close enough to s and for all n large enough we have $\gamma_n(s') \in \mathscr{O}_{\gamma_n(s)}$. Since the γ_n's are causal, Proposition 4.4.5 shows that we have

$$\sigma_{\gamma_n(s)}(\gamma_n(s')) \le 0 \ . \tag{4.6.7}$$

Since $\sigma_q(p)$ depends continuously upon q and p, passing to the limit in (4.6.7) gives

$$\sigma_{\gamma_n(s)}(\gamma(s')) \le 0 \ . \tag{4.6.8}$$

This is only possible if γ is causal, which can be seen as follows: Suppose that γ is differentiable at s. In normal coordinates on $\mathscr{O}_{\gamma(s)}$ we have, by definition of a derivative,

$$\gamma^\mu(s') = \underbrace{\gamma^\mu(s)}_{=0} + \dot\gamma^\mu(s)(s'-s) + o(s'-s) \ ,$$

hence

$$0 \ge \sigma_{\gamma_n(s)}(s') = \eta_{\mu\nu}\dot\gamma^\mu(s)\dot\gamma^\nu(s)(s'-s)^2 + o((s'-s)^2) \ .$$

For $s'-s$ small enough this is only possible if

$$\eta_{\mu\nu}\dot\gamma^\mu(s)\dot\gamma^\nu(s) \le 0 \ ,$$

and $\dot\gamma$ is causal, as we desired to show. □

Let us address now the question of inextendibility of accumulation curves. We note the following lemma:

LEMMA 4.6.3 *Let γ_n be a sequence of* dist_h*-parameterized inextendible causal curves converging to γ uniformly on compact subsets of $\mathbb{R}$, then γ is inextendible.*•4.6.2

•4.6.2: **ptc:***new result, to be proved, this is claimed both in Beem Ehrlich page 76 and Kriele and HE*

PROOF: Note that the parameter range of γ is $\mathbb{R}$, and the result would follow from Theorem 4.5.4 if γ were dist_h-parameterized, but this might fail to be the case (consider a sequence of null geodesics in $\mathbb{R}^{1,1} = (\mathbb{R}^2, g = -dt^2 + dx^2)$ threading back and forth a space-distance $1/n$ around the $\{x = 0\}$ axis, with $h = dt^2 + dx^2$ the limit curve is $\gamma(s) = (s/\sqrt{2}, 0)$ which is not dist_h-parameterized). •4.6.3

•4.6.3: **ptc:***I am not sure that this is true, and how to state it not to make it trivial, and whether this is needed at all*

□

In summary, it follows from Lemmata 4.5.3 and 4.6.3 together with Proposition 4.6.1 that:

THEOREM 4.6.4 *Every sequence of future directed, inextendible, causal curves which converges at one point accumulates at some future directed, inextendible, causal curve.* □

One is sometimes interested in sequences of maximally extended geodesics:

PROPOSITION 4.6.5 *Let γ_n be a sequence of maximally extended geodesics accumulating at γ. Then γ is a maximally extended geodesic.*•4.6.4

•4.6.4: **ptc:***new result, to be crosschecked*

PROOF: If we use a dist_h-parameterization of the γ_n's such that $\gamma_n(0) \to \gamma(0)$, then by Proposition 4.6.1 (passing to a subsequence if necessary) the γ_n's converge to γ, uniformly on compact subsets of $\mathbb{R}$. Let $\mathscr{K}$ be a compact neighborhood of $\gamma(0)$, compactness of $\cup_{p\in\mathscr{K}} U_p\mathscr{M}$, where $U_p\mathscr{M} \subset T_p\mathscr{M}$ is the set of h-unit vectors tangent to $\mathscr{M}$, implies that there exists a subsequence such that $\dot\gamma_n(0)$ converges to some vector $X \in U_{\gamma(0)}\mathscr{M} \subset T_{\gamma(0)}\mathscr{M}$. Let $\sigma : (a,b) \to \mathscr{M}$, $a \in \mathbb{R} \cup \{-\infty\}$, $b \in \mathbb{R} \cup \{\infty\}$, be an affinely parameterised maximally extended geodesic through $\gamma(0)$ with initial tangent vector X. By continuous dependence of ODE's upon initial values it follows that 1) for any $a < \alpha < \beta < b$ all the γ_n's, except perhaps for a finite number, are defined on $[\alpha, \beta]$ when affinely parameterized, and 2) they converge to $\sigma|_{[\alpha,\beta]}$ in uniform $C^1([\alpha,\beta], \mathscr{M})$ topology. Thus $\dot\gamma_n(s) \to \dot\sigma(s)$ uniformly on compact subsets of (a,b), which implies that a dist_h-parameterization is preserved under taking limits. Hence the γ_n's, when dist_h-parameterized, converge uniformly to a dist_h-reparameterization of σ on compact subsets of $\mathbb{R}$, call it μ. It follows that $\gamma = \mu$, and γ is a maximally extended geodesic. □

4.6.1 Achronal causal curves

•4.6.5 A curve $\gamma : I \to \mathscr{M}$ is called *achronal* if

•4.6.5: **ptc:***new section, needs a global reread*

$$\forall\ s, s' \qquad \gamma(s) \notin I^+(\gamma(s'))\ .$$

Any spacelike geodesic in Minkowski space-time is achronal. More interestingly, it follows from Proposition 4.4.3 that this is also true for null geodesics. However, null geodesics do not have to be achronal in general: consider, *e.g.*, the two-dimensional space-time $\mathbb{R} \times S^1$ with the flat metric $-dt^2 + dx^2$, where x is an angle-type coordinate along S^1 with periodicity, say, 2π. Then the points $(0,0)$ and $(2\pi, 0)$ both lie on the null geodesic

$$s \to (s, s \bmod 2\pi) \ ,$$

and are clearly timelike related to each other.

In this section we will be interested in causal curves that are achronal. We start with the following:

PROPOSITION 4.6.6 *Let γ be an* achronal causal *curve, then γ is a null geodesic.*

PROOF: Let $\mathscr{O}$ be any elementary neighborhood, then any connected component of $\gamma \cap \mathscr{O}$ is a null geodesic by Corollary 4.4.10. □

THEOREM 4.6.7 *Let $\gamma_n : I \to \mathscr{M}$ be a sequence of achronal causal curves accumulating at γ, then γ is achronal.*

REMARK 4.6.8 It follows from Propositions 4.6.6 and 4.6.5 that γ is inextendible if the γ_n's are.

PROOF: Changing time-orientation if necessary we can without loss of generality assume that all the γ_n's are future directed, for n large enough. It follows from Proposition 4.6.1 that passing to a subsequence and reparameterizing if necessary, the γ_n's converge to γ uniformly on compact subsets of $\mathbb{R}$. Suppose γ is not achronal, then there exist $s_1, s_2 \in I$ such that $\gamma(s_2) \in I^+(\gamma(s_1))$, thus there exists a *timelike* curve $\hat{\gamma} : [s_1, s_2] \to \mathscr{M}$ from $\gamma(s_1)$ to $\gamma(s_2)$. Choose some $\hat{s} \in (s_1, s_2)$. We have $\gamma(s_2) \in I^+(\hat{\gamma}(\hat{s}))$, and since $I^+(\hat{\gamma}(\hat{s}))$ is open there exists an open neighborhood $\mathscr{O}_2$ of $\gamma(s_2)$ such that $\mathscr{O}_2 \subset I^+(\hat{\gamma}(\hat{s}))$. Similarly there exists an open neighborhood $\mathscr{O}_1$ of $\gamma(s_1)$ such that $\mathscr{O}_1 \subset I^-(\hat{\gamma}(\hat{s}))$. This shows that any point $p_2 \in \mathscr{O}_2$ lies in the timelike future of any point $p_1 \in \mathscr{O}_1$: indeed, one can go from p_1 along some timelike path to $\hat{\gamma}(\hat{s})$, and continue along another timelike path from $\hat{\gamma}(\hat{s})$ to p_2.

By Proposition 4.6.1, passing to a subsequence if necessary, $\gamma_n(s_1)$ converges to $\gamma(s_1)$ and $\gamma_n(s_2)$ converges to $\gamma(s_2)$. Then $\gamma_n(s_1) \in \mathscr{O}_1$ and $\gamma_n(s_2) \in \mathscr{O}_2$ for n large enough, leading to $\gamma_n(s_2) \in I^+(\gamma_n(s_1))$, contradicting achronality of γ_n. □

•4.6.6

•4.6.6: **ptc:***prove that γ is achronal if the γ_n's were maximising timelike; show that arc length is upper semicontinuous and that lorentzian distance is lower semicontinuous*

4.7 Causality conditions

Space-times can exhibit various causal pathologies, most of which are undesirable from a physical point of view. The simplest example of unwanted causal behaviour is the existence of closed timelike curves. A space-time is said to be *chronological* if no such curves exist. An example of a space-time which is not chronological is provided by $S^1 \times \mathbb{R}$ with the flat metric $-dt^2 + dx^2$, where t is a local coordinate defined modulo 2π on S^1. Then every circle $x = \text{const}$ is a closed timelike curve.

The class of compact manifolds is a very convenient one from the point of view of Riemannian geometry. The following result of Geroch shows that such manifolds are always pathological from a Lorentzian perspective:

PROPOSITION 4.7.1 (Geroch [?]) *Every compact space-time $(\mathscr{M}, g)$ contains a closed timelike curve.*

PROOF: Consider the covering of $\mathscr{M}$ by the collection of open sets $\{I^-(p)\}_{p\in\mathscr{M}}$, by compactness a finite covering $\{I^-(p_i)\}_{i=1,\dots,I}$ can be chosen. The possibility $p_1 \in I^-(p_1)$ yields immediately a closed timelike curve through p_1, otherwise there exists $p_{i(1)}$ such that $p_1 \in I^-(p_{i(1)})$. Again if $p_{i(1)} \in I^-(p_{i(1)})$ we are done, otherwise there exists $p_{i(2)}$ such that $p_{i(1)} \in I^-(p_{i(2)})$. Continuing in this way we obtain a — finite or infinite — sequence of points $p_{i(j)}$ such that

$$p_{i(j)} \in I^-(p_{i(j+1)}) \,. \tag{4.7.1}$$

If the sequence is finite we are done. Now, we have only a finite number of p_i's at our disposal, therefore if the sequence is finite it has to contain repetitions:

$$p_{i(j+\ell)} = p_{i(j)}$$

for some j, and some $\ell > 0$. It should be clear from (4.7.1) that there exists a closed timelike curve through $p_{i(j)}$. □

REMARK 4.7.2 Galloway [?] has shown that in compact space-times $(\mathscr{M}, g)$ there exist closed timelike curves through any two points p and q, under the supplementary condition that the Ricci tensor Ric satisfies the following energy condition:

$$\mathrm{Ric}(X, X) > 0 \quad \text{for all causal vectors } X. \tag{4.7.2}$$

The chronology condition excludes closed <u>timelike</u> curves, but it just fails to exclude the possibility of occurrence of <u>closed</u> causal curves. A space-time is said to be *causal* if no such curves can be found. The existence of space-times which are *chronological* but not *causal* requires a little work:

EXAMPLE 4.7.3 Let $\mathscr{M} = \mathbb{R} \times S^1$ with the metric

$$g = 2dt\, dx + f(t)dx^2 \,,$$

where f is any function satisfying

$$f \geq 0 \,, \quad \text{with } f(t) = 0 \text{ iff } t = 0 \,.$$

(The function $f(t) = t^2$ will do.) Here t runs over the $\mathbb{R}$ factor of $\mathscr{M}$, while x is a coordinate defined modulo 2π on S^1. In matrix notation we have

$$[g_{\mu\nu}] = \begin{bmatrix} 0 & 1 \\ 1 & f \end{bmatrix} ,$$

which leads to the following inverse metric

$$[g^{\mu\nu}] = \begin{bmatrix} -f & 1 \\ 1 & 0 \end{bmatrix} .$$

It follows that

$$g(\nabla t, \nabla t) = -f \leq 0 \text{ , with } g(\nabla t, \nabla t) = 0 \text{ iff } t = 0 \; . \tag{4.7.3}$$

Recall, now, that a function τ is called a *time function* if $\nabla\tau$ is timelike, past-pointing. Equation (4.7.3) shows that t is a time function on the set $\{t \neq 0\}$. Since a time-function is strictly increasing on any causal curve (see Lemma 4.4.8), one easily concludes that no closed causal curve in $\mathscr{M}$ can intersect the set $\{t \neq 0\}$. In other words, closed causal curves — if they do exist — must be entirely contained in the set $\{t = 0\}$. Now, any curve γ contained in this last set is of the form

$$\gamma(s) = (0, x(s)) \; ,$$

with tangent vector

$$\dot\gamma = \dot x \partial_x \quad \Longrightarrow \quad g(\dot\gamma, \dot\gamma) = (\dot x)^2 g(\partial_x, \partial_x) = (\dot x)^2 g_{xx} = 0 \; .$$

This shows in particular that

- $\mathscr{M}$ does contain closed causal curves: an example is given by $x(s) = s \bmod 2\pi$.
- All closed causal curves are null.

It follows that $(\mathscr{M}, g)$ is indeed chronological, but not causal, as claimed.

It is desirable to have a condition of causality which is stable under small changes of the metric. By way of example, consider a space-time which contains a family of causal curves γ_n with both $\gamma_n(0)$ and $\gamma_n(1)$ converging to p. Such curves can be thought of as being "almost closed". Further, it is clear that one can produce an arbitrarily small deformation of the metric which will allow one to obtain a closed causal curve in the deformed space-time. The object of our next causality condition is to exclude this behaviour. A space-time will be said to be *strongly causal* if every neighborhood $\mathscr{O}$ of a point $p \in \mathscr{M}$ contains a neighborhood $\mathscr{U}$ such that for every causal curve $\gamma : I \to \mathscr{M}$ the set

$$\{s \in I : \gamma(s) \in \mathscr{U}\} \subset I$$

is a connected subset of I. In other words, γ does not re-enter $\mathscr{U}$ once it has left it.

Clearly, a strongly causal space-time is necessarily causal. However, the inverse does not always hold. An example is given in Figure 4.7

The definition of strong causality appears, at first sight, somewhat unwieldy to verify, so simpler conditions are desirable. The following provides a useful criterion: A space-time $(\mathscr{M}, g)$ is said to be *stably causal* if there exists a time function t globally defined on $\mathscr{M}$. Recall — see Lemma 4.4.8 — that time functions are strictly increasing on causal curves. It then easily follows that *stable causality* implies *strong causality*:

PROPOSITION 4.7.4 *If $(\mathscr{M}, g)$ is stably causal, then it is strongly causal.*

Figure 4.3: A causal space-time which is not strongly causal. Here the metric is the flat one $-dt^2 + dx^2$, so that the light cones are at 45^o. It should be clear that no matter how small the neighborhood $\mathscr{U}$ of p is, there will exist a causal curve as drawn in the figure which will intersect this neighborhood twice. In order to show that $(\mathscr{M}, g)$ is causal one can proceed as follows: suppose that γ is any closed causal curve in $\mathscr{M}$, then γ has to intersect the hypersurfaces $\{t = \pm 1\}$ at some points $x_\pm$, with $x_- > 1$ and $x_+ < -1$. If we parameterize γ so that $\gamma(s) = (s, x(s))$ we obtain $-2 > x_+ - x_- = \int_{-1}^{1} \frac{dx}{ds} ds$, hence there must exist $s_* \in [-1, 1]$ such that $dx/ds < -1$, contradicting causality of γ.

Proof: Let $\mathscr{O}$ be a connected open neighborhood of $p \in \mathscr{M}$, and let φ be a nonegative smooth function such that $\varphi(p) \neq 0$ and such that the support $\mathrm{supp}\varphi$ of φ is a compact set contained in $\mathscr{O}$. Let τ be a time function on $\mathscr{M}$, for $a \in \mathbb{R}$ set

$$\tau_a := \tau + a\varphi \ .$$

As $\nabla\tau$ is timelike, the function $g(\nabla\tau, \nabla\tau)$ is bounded away from zero on the compact set $\mathrm{supp}\varphi$, which implies that there exists $\epsilon > 0$ small enough so that $\tau_{\pm\epsilon}$ are time functions on $\mathrm{supp}\varphi$. Now, $\tau_{\pm\epsilon}$ coincides with τ away from $\mathrm{supp}\varphi$, so that the $\tau_{\pm\epsilon}$'s are actually time functions on $\mathscr{M}$ as well. We set

$$\mathscr{U} := \{q : \tau_{-\epsilon}(q) < \tau(p) < \tau_{+\epsilon}(q)\} \ .$$

We have:

- $p \in \mathscr{U}$, therefore $\mathscr{U}$ is not empty;
- $\mathscr{U}$ is open because the τ_a's are continuous;
- $\mathscr{U} \subset \mathscr{O}$ because φ vanishes outside of $\mathscr{O}$.

Consider any causal curve γ the image of which intersects $\mathscr{U}$, γ can enter or leave $\mathscr{U}$ only through

$$\partial\mathscr{U} \subset \{q : \tau_{-\epsilon}(q) = \tau(p)\} \cup \{q : \tau(p) = \tau_{+\epsilon}(q)\} \ . \tag{4.7.4}$$

At a point s_- at which $\gamma(s_-) \in \{q : \tau_{-\epsilon}(q) = \tau(p)\}$ we have

$$\tau(p) = \tau_{-\epsilon}(\gamma(s_-)) = \tau(\gamma(s_-)) - \epsilon\varphi(\gamma(s_-)) \quad \Longrightarrow \quad \tau(\gamma(s_-)) > \tau(p) \; .$$

Similarly at a point s_+ at which $\gamma(s_+) \in \{q : \tau_{+\epsilon}(q) = \tau(p)\}$ we have

$$\tau(p) = \tau_{+\epsilon}(\gamma(s_+)) = \tau(\gamma(s_+)) + \epsilon\varphi(\gamma(s_+)) \quad \Longrightarrow \quad \tau(\gamma(s_+)) < \tau(p) \; .$$

As τ is increasing along γ, we conclude that γ can enter $\mathscr{U}$ only through $\{q : \tau_{+\epsilon}(q) = \tau(p)\}$, and leave $\mathscr{U}$ only through $\{q : \tau_{-\epsilon}(q) = \tau(p)\}$. Lemma 4.4.8 shows that γ can intersect each of the two sets at the right-hand-side of (4.7.4) at most once. Those facts obviously imply connectedness of the intersection of (the image of) γ with $\mathscr{U}$. □

The strongest causality condition is that of *global hyperbolicity*, considered in the next section.

4.8 Global hyperbolicity

A space-time $(\mathscr{M}, g)$ said to be *globally hyperbolic* if it is *strongly causal*, and if for every $p, q \in M$ the sets $J^+(p) \cap J^-(q)$ are compact.

It is sometimes useful to consider space-times with boundary — *i.e.*, $\mathscr{M}$ is a manifold with boundary, with g extending differentiably to $\partial\mathscr{M}$. We will allow such space-times throughout this section.•4.8.1

•*4.8.1:* **ptc:***this should be said already well before here, and all results cross-checked against this hypothesis*

It is not too difficult to show that Minkowski space-time $\mathbb{R}^{1,n}$ is globally hyperbolic: first, the Minkowski time x^0 provides a time-function on $\mathbb{R}^{n,1}$ — this implies strong causality. Compactness of $J^+(p) \cap J^-(q)$ for all p's and q's is easily checked by drawing pictures; it is also easy to write a formal proof using Proposition 4.4.3, this is left as an exercise to the reader.

The notion of globally hyperbolicity provides excellent control over causal properties of $(\mathscr{M}, g)$. This will be made clear at several other places in this work. Anticipating, let us list a few of those:

1. Let $(\mathscr{M}, g)$ be globally hyperbolic. If $J^+(p) \cap J^-(q) \neq \emptyset$, then there exists a causal geodesic from p to q. Similarly if $I^+(p) \cap I^-(q) \neq \emptyset$, then there exists a timelike geodesic from p to q.

2. The Cauchy problem for linear wave equations is globally solvable on globally hyperbolic space-times.

3. A key theorem of Choquet-Bruhat and Geroch asserts that *maximal globally hyperbolic* solutions of the Cauchy problem for Einstein's equations are unique up to diffeomorphism.

We start our study of globally hyperbolic space-times with the following property:

Proposition 4.8.1 *Let $(\mathscr{M}, g)$ be globally hyperbolic, and let γ_n be a family of causal curves accumulating both at p and q. Then there exists a causal curve γ, accumulation curve of the (perhaps reparameterized) γ_n's, which passes both through p and q.*

REMARK 4.8.2 The stably-causal space-time in Figure **??** shows that the result is wrong without the hypothesis of global hyperbolicity.

PROOF: Extending the γ_n's to inextendible curves, and reparameterizing if necessary, we can assume that the γ_n's are dist_h-parameterized, with common domain of definition $I = \mathbb{R}$, and with $\gamma_n(0)$ converging to p. If $p = q$ the result has already been established in Proposition 4.6.1, so we assume that $p \neq q$. Consider the compact set

$$\mathscr{K} := \big(J^+(p) \cap J^-(q)\big) \cup \big(J^+(q) \cap J^-(p)\big) \tag{4.8.1}$$

(since a globally hyperbolic space-time is causal, one of those sets is, of course, necessarily empty). $\mathscr{K}$ can be covered by a finite number of elementary domains $\mathscr{U}_i$, $i = 1, \cdots, N$. Strong causality allows us to choose the $\mathscr{U}_i$'s small enough so that for every n the image of γ_n is a connected subset in $\mathscr{U}_i$. We can choose a parameterization of the γ_n's by h–length so that, passing to a subsequence of the γ_n's if necessary, we have $\gamma_n(0) \to p$. Extending the γ_n's if necessary we can assume that all the γ_n's are defined on $\mathbb{R}$. We note the following:

LEMMA 4.8.3 *Let $\mathscr{U}$ be an elementary neighborhood, as defined in Definition 4.2.6. There exists a constant ℓ such that for any causal curve $\gamma : I \to \mathscr{U}$ the h–length $|\gamma|_h$ of γ is bounded by ℓ.*

To prove Lemma 4.8.3 we need the following variation of the inverse Cauchy-Schwarz inequality (compare Proposition 2.2.2):

LEMMA 4.8.4 *Let K be a compact set and let X be a continuous timelike vector field defined there, then there exists a strictly positive constant C such that for all $q \in K$ and for all causal vectors $Y \in T_q\mathscr{M}$ we have*

$$|g(X,Y)| \geq C|Y|_h \ . \tag{4.8.2}$$

PROOF: By homogeneity it is sufficient to establish (4.8.2) for causal $Y \in T_q\mathscr{M}$ such that $|Y|_h = 1$; let us denote by $U(h)_q$ this last set. The result follows then by continuity of the strictly positive function

$$\cup_{q\in K} U(h)_q \ni Y \to |g(X,Y)|$$

on the compact set $\cup_{q\in K} U(h)_q$. □

Returning to the proof of Lemma 4.8.3, let x^0 be the local time coordinate on $\mathscr{U}$, since ∇x^0 is timelike we can use Lemma 4.8.4 with $K = \overline{\mathscr{U}}$ to conclude that there exists a constant C such that for any causal curve $\gamma \subset \mathscr{U}$ we have

$$|g(X,\dot\gamma)| \geq C > 0$$

at all points at which γ is differentiable. This implies, for $s_2 \geq s_1$,

$$\begin{aligned} |x^0(s_2) - x^0(s_1)| &\geq \int_{s_1}^{s_2} |g(\nabla x^0, \dot\gamma)| ds \\ &\geq C \int_{s_1}^{s_2} ds = C|s_2 - s_1| \ . \end{aligned}$$

It follows that

$$|\gamma|_h \le \ell := \frac{2}{C} \sup_{\mathscr{U}} |x^0| < \infty \;, \tag{4.8.3}$$

as desired. □

Returning to the proof of Proposition 4.8.1, Lemma 4.8.3 shows that there exists a constant L_i — independent of n — such that the h–length $|\gamma_n \cap \mathscr{U}_i|_h$ is bounded by L_i. Consequently the h–length $|\gamma_n \cap \mathscr{K}|_h$, with $\mathscr{K}$ as in (4.8.1), is bounded by

$$|\gamma_n \cap \mathscr{K}|_h \le L := L_1 + L_2 + \ldots + L_I \;. \tag{4.8.4}$$

By hypothesis the γ_n's accumulate at q, therefore there exists a sequence s_n (passing again to a subsequence if necessary) such that

$$\gamma_n(s_n) \to q \;.$$

Equation (4.8.4) shows that the sequence s_n is bounded, hence — perhaps passing to a subsequence — we have $s_n \to s_*$ for some $s_* \in \mathbb{R}$.

At this stage we could use Proposition 4.6.1, but one might as well argue directly: by our choice of parametrization we have

$$\mathrm{dist}_h(\gamma_n(s), \gamma_n(s')) \le |s - s'| \tag{4.8.5}$$

(see (4.3.2)-(4.3.3)). This shows that the family $\{\gamma_n\}$ is equicontinuous, and (4.8.5) together with the Arzela-Ascoli theorem (on the compact set $[-L, L]$) implies existence of a curve $\gamma : [-L, L] \to \mathscr{M}$ and a subsequence γ_{n_i} which converges uniformly to γ on $[-L, L]$. As $\gamma_{n_i}(s_{n_i})$ converges both to $\gamma(s_*)$ and to q we have

$$\gamma(s_*) = q \;.$$

This shows that γ is the desired causal curve joining p with q. □

As a straightforward corollary of Proposition 4.8.1 we obtain:

COROLLARY 4.8.5 *Let $(\mathscr{M}, g)$ be globally hyperbolic, then*

$$\overline{I^\pm(p)} = J^\pm(p) \;.$$

PROOF: Let $q_n \in I^+(p)$ be a sequence of points accumulating at q, thus there exists a sequence γ_n of causal curves from p to q, then $q \in J^+(p)$ by Proposition 4.8.1. □

As already mentioned, global hyperbolicity gives us control over causal geodesics (see Section 4.12 for a proof):

THEOREM 4.8.6 *Let $(\mathscr{M}, g)$ be globally hyperbolic, if $q \in I^+(p)$, respectively $q \in J^+(p)$, then there exists a timelike, respectively causal, future directed geodesic from p to q.*•4.8.2

•4.8.2: **ptc:***watch out if with boundary*

4.9 Domains of dependence

A set $\mathscr{U} \subset \mathscr{M}$ is said to be *achronal* if

$$I^+(\mathscr{U}) \cap I^-(\mathscr{U}) = \emptyset .$$

There is an obvious analogous definition of an *acausal* set

$$J^+(\mathscr{U}) \cap J^-(\mathscr{U}) = \emptyset .$$

Let $\mathscr{S}$ be an achronal topological hypersurface in a space-time $(\mathscr{M}, g)$. (By a hypersurface we mean an embedded submanifold of codimension one.) The *future domain of dependence* $\mathscr{D}^+(\mathscr{S})$ of $\mathscr{S}$ is defined as the set of points $p \in \mathscr{M}$ with the property that *every past-directed past-inextendible timelike curve starting at p meets $\mathscr{S}$ precisely once.* The *past domain of dependence* $\mathscr{D}^-(\mathscr{S})$ is defined by changing *past-directed past-inextendible* to *future-directed future-inextendible* above. Finally one sets

$$\mathscr{D}(\mathscr{S}) := \mathscr{D}^+(\mathscr{S}) \cup \mathscr{D}^-(\mathscr{S}) . \tag{4.9.1}$$

The "precisely" in "precisely once" above follows of course already from achronality of $\mathscr{S}$ — the repetitiveness in our definition is deliberate, to emphasize this property. We always have

$$\mathscr{S} \subset \mathscr{D}^\pm(\mathscr{S}) .$$

We have found it useful to build in the fact that $\mathscr{S}$ is a topological hypersurface in the definition of $\mathscr{D}^+(\mathscr{S})$. Some authors do not impose this restriction [53], which can lead to various pathologies. From the point of view of differential equations the only interesting case is that of a hypersurface anyway.

Hawking and Ellis [58] define the domain of dependence using causal curves instead of timelike ones (on the other hand timelike curves are used by Geroch [53] and by Penrose [89]). The definition with causal curves has the advantage that the resulting $\mathscr{D}(\mathscr{S})$ is an open set when $\mathscr{S}$ is an acausal topological hypersurface. However, this excludes piecewise null hypersurfaces as Cauchy surfaces, and this is the reason why we prefer the definition above.

The following examples are instructive; they are left as exercices to the reader:

Example 4.9.1 Let $\mathscr{S} = \{x^0 = 0\}$ in Minkowski space-time $\mathbb{R}^{1,n}$, where x^0 is the usual time coordinate on $\mathbb{R}^{1,n}$. Then $\mathscr{D}(\mathscr{S}) = \mathbb{R}^{1,n}$. Thus both $\mathscr{D}^+(\mathscr{S})$ and $\mathscr{D}^-(\mathscr{S})$ are non-trivial, and their union covers the whole space-time.

Example 4.9.2 Let $\mathscr{S} = \{x^0 - x^1 = 0\}$ in Minkowski space-time, where the x^μ's are the usual Minkowskian coordinates on $\mathbb{R}^{1,n}$. Then $\mathscr{D}^+(\mathscr{S}) = \mathscr{D}^-(\mathscr{S}) = \mathscr{D}(\mathscr{S}) = \mathscr{S}$, which makes the objects uninteresting.

Example 4.9.3 Let $\mathscr{S} = \{x^0 = |x^1|\}$ in Minkowski space-time. Then $\mathscr{D}^-(\mathscr{S}) = \mathscr{S}$, $\mathscr{D}^+(\mathscr{S}) = \{x^0 \geq |x^1|\}$. Thus $\mathscr{D}^-(\mathscr{S}) = \mathscr{S}$, which is not very useful. On the other hand $\mathscr{D}^+(\mathscr{S})$ covers the whole future of $\mathscr{S}$.

EXAMPLE 4.9.4 Let $\mathscr{S} = \{\eta_{\mu\nu}x^\mu x^\nu = -1 \ , x^0 > 0\}$ be the upper connected component of the unit spacelike hyperboloid in Minkowski space-time. Then $\mathscr{D}(\mathscr{S}) = J^+(0)$. Thus both $\mathscr{D}^-(\mathscr{S})$ and $\mathscr{D}^+(\mathscr{S})$ are non-trivial, however $\mathscr{D}^-(\mathscr{S})$ does not cover the whole past of $\mathscr{S}$.

As a warm-up, let us prove the following elementary property of domains of dependence:

PROPOSITION 4.9.5 *Let $p \in \mathscr{D}^+(\mathscr{S})$, then*

$$I^-(p) \cap J^+(\mathscr{S}) \subset \mathscr{D}^+(\mathscr{S}) \ .$$

PROOF: Let $q \in I^-(p) \cap J^+(\mathscr{S})$, thus there exists a past-directed timelike curve γ_0 from p to q. Let γ_1 be a past-inextendible timelike curve γ_1 starting at q. The curve $\gamma := \gamma_0 \cup \gamma_1$ is a past-inextendible past-directed timelike curve starting at p, thus it meets $\mathscr{S}$ precisely once at some point $r \in \mathscr{S}$. Suppose that γ passes through r before passing through q, as $q \in J^+(\mathscr{S})$ Lemma 4.4.14 shows that $r \in I^+(\mathscr{S})$, contradicting achronality of $\mathscr{S}$. This shows that γ must meet $\mathscr{S}$ after passing through q, hence γ_1 meets $\mathscr{S}$ precisely once. □

Let $\mathscr{S}$ be achronal, we shall say that a set $\mathscr{O}$ forms a one-sided future neighborhood of $p \in \mathscr{S}$ if there exists an open set $\mathscr{U} \subset \mathscr{M}$ such that $\mathscr{U}$ contains p and

$$\mathscr{U} \cap J^+(\mathscr{S}) \subset \mathscr{O} \ .$$

As $I^-(p)$ is open, Proposition 4.9.5 immediately implies:

COROLLARY 4.9.6 *Suppose that $\mathscr{D}^+(\mathscr{S}) \neq \mathscr{S}$, consider any point $p \in \mathscr{D}^+(\mathscr{S}) \setminus \mathscr{S}$. For any $q \in \mathscr{S} \cap I^-(p)$ the set $\mathscr{D}^+(\mathscr{S})$ forms a one-sided future neighborhood of q.* □

Local coordinate considerations near $\mathscr{S}$ should make it clear that $\mathscr{D}(\mathscr{S})$ forms a neighborhood of $\mathscr{S}$, for $\mathscr{S}$'s — achronal, smooth, <u>spacelike</u> hypersurfaces.•4.9.1 Example 4.9.2 shows that this will not be the case for more general $\mathscr{S}$'s. Let us show that things are well behaved for <u>acausal</u> topological hypersurfaces $\mathscr{S}$, regardless of their differentiability properties:

•*4.9.1:* **ptc:***this should be done*

PROPOSITION 4.9.7 *If $\mathscr{S}$ is an <u>acausal</u> topological hypersurface, then $\mathscr{D}(\mathscr{S})$ forms a neighborhood of $\mathscr{S}$.*

PROOF: •4.9.2 □

•*4.9.2:* **Warning:** *I am not sure this is correct; proof missing anyway*

The next theorem shows that achronal topological hypersurfaces can be used to produce globally hyperbolic space-times:

THEOREM 4.9.8 *Let $\mathscr{S}$ be an achronal hypersurface in $(\mathscr{M}, g)$, and suppose that the interior $\mathring{\mathscr{D}}(\mathscr{S})$ of the domain of dependence $\mathscr{D}(\mathscr{S})$ of $\mathscr{S}$ is not empty. Then $\mathring{\mathscr{D}}(\mathscr{S})$ equipped with the metric obtained by restriction from g is globally hyperbolic.*

PROOF: We need first to show that a causal curve can be pushed-up by an amount as small as desired to yield a timelike curve:

•4.9.3: **ptc:**q *here is probably the end point of gamma??????*

LEMMA 4.9.9 ("Push-up lemma") *Let γ be a past-inextendible past-directed causal curve starting at p,•4.9.3 and let $\mathscr{O}$ be a neighborhood of (the image of) γ. Then for every $r \in I^+(q) \cap \mathscr{O}$ there exists a past-inextendible past-directed timelike curve $\hat\gamma$ starting at r such that*

$$\hat\gamma \subset I^+(\gamma) \cap \mathscr{O} \ .$$

PROOF: The construction is essentially identical to that of the proof of Lemma 4.4.14, except that we will have to deal with a countable collection of curves, rather than a finite one. One also needs to make sure that the final curve is inextendible. As always, we parameterize γ by h–distance as measured from p. Using an exhaustion of $[0,\infty)$ by compact intervals $[m, m+1]$ we cover γ by a countable collection $\mathscr{U}_i \subset \mathscr{O}$, $i \in \mathbb{N}$ of elementary regions $\mathscr{U}_i$ centered at

$$p_i = \gamma(r_i)$$

with

$$p_1 = p \ , \quad p_i \in \mathscr{U}_i \cap \mathscr{U}_{i+1} \ , \quad p_{i+1} \subset J^-(p_i) \ .$$

We can further require impose the following condition on the $\mathscr{U}_i$'s: if $r_i \in [j, j+1)$, then the corresponding $\mathscr{U}_i$ is contained in a h-distance ball $B_h(p_i, 1/(j+1))$. Let $\gamma_0 : [0, s_0] \to \mathscr{M}$ be a past directed causal curve from r to $p \in \mathscr{U}_1 \cap \mathscr{U}_2$; let s_1 be close enough to s_0 so that

$$\gamma_0(s_1) \in \mathscr{U}_2 \ .$$

Proposition 4.2.3 together with the definition of elementary regions shows that there exists a past directed timelike curve $\gamma_1 : [0,1] \to \mathscr{U}_1 \subset \mathscr{O}$ from q to p_2. (In particular $\gamma_1 \setminus \{p\} \subset I^+(p) \subset I^+(\gamma)$). Similarly, for any $s \in [0,1]$ there exists a a past directed timelike curve $\gamma_{2,s} : [0,1] \to \mathscr{U}_2 \subset \mathscr{O}$ from $\gamma_1(s)$ to p_2. We choose $s =: s_2$ small enough so that

$$\gamma_1(s_2) \in \mathscr{U}_3 \ .$$

One repeats that construction iteratively, obtaining a sequence of past-directed timelike curves $\gamma_i \subset I^+(\gamma) \cap \mathscr{U}_i \subset I^+(\gamma) \cap \mathscr{O}$ such that the end point of γ_i lies in $\mathscr{U}_{i+1}$ and coincides with the starting point of γ_{i+1}. Concatenating those curves together gives the desired path $\hat\gamma$. Since every path γ_i lies in $I^+(\gamma) \cap \mathscr{O}$, so does their union.

Since $\gamma_i \subset \mathscr{U}_i \subset B_h(p_i, 1/(j+1))$ when $r_i \in [j, j+1)$ we obtain, for $r \in [j, j+1)$,

$$\mathrm{dist}_h(\gamma(r), \hat\gamma) \le \mathrm{dist}_h(\gamma(r), \gamma(r_i)) + \mathrm{dist}_h(\gamma(r_i), \gamma_i) \le \frac{2}{j+1} \ ,$$

where we have ensured that $\mathrm{dist}_h(\gamma(r), \gamma(r_i)) < 1/(j+1)$ by choosing r_i appropriately. It follows that

$$\mathrm{dist}_h(\gamma(r), \hat\gamma) \le \frac{2}{r} \ . \tag{4.9.2}$$

To finish the proof, suppose that $\hat\gamma : [0, s_*) \to \mathscr{M}$ is extendible, call $\hat p$ be the end point of $\hat\gamma$. By (4.9.2)

$$\lim_{r\to\infty} \mathrm{dist}_h(\gamma(r), \hat p) = 0 \;.$$

Thus $\hat p$ is an end point of γ, which together withTheorem 4.5.4, contradicts inextendibility of γ. □

By the definition of domains of dependence, inextendible *timelike* curves through $p \in \mathscr{D}^+(\mathscr{S})$ intersect all the sets $\mathscr{S}$, $I^+(\mathscr{S})$, and $I^-(\mathscr{S})$. This is wrong in general for inextendible *causal* curves through points in $\mathscr{D}^+(\mathscr{S}) \setminus \mathring{\mathscr{D}}^+(\mathscr{S})$, as shown on Figure **??**.•4.9.4 Nevertheless we have:

•4.9.4: *Figure missing*

Lemma 4.9.10 *If $p \in \mathring{\mathscr{D}}(\mathscr{S})$, then every inextendible causal curve γ through p intersects $\mathscr{S}$, $I^-(\mathscr{S})$ and $I^+(\mathscr{S})$.*

Remark 4.9.11 In contradistinction with the timelike case, for causal curves the intersection of γ with $\mathscr{S}$ does not have to be a point. An example is given by the hypersurface $\mathscr{S}$ of Figure **??**.

Proof: Changing time-orientation if necessary we may suppose that $p \in \mathscr{D}^+(\mathscr{S})$. Let $\gamma : I \to \mathscr{M}$ be any past-directed inextendible causal curve through p. Since p is an interior point of $\mathscr{D}^+(\mathscr{S})$ there exists $q \in I^+(p) \cap \mathscr{D}^+(\mathscr{S})$. By the push-up Lemma 4.9.9 (with $\mathscr{O} = \mathscr{M}$) there exists a past-inextendible past-directed timelike curve $\hat\gamma$ starting at q which lies to the future of γ. Since $q \in \mathscr{D}^+(\mathscr{S})$ the curve $\hat\gamma$ enters $I^-(\mathscr{S})$, and since γ lies to the past of $\hat\gamma$ it must enter $I^-(\mathscr{S})$ as well.

Suppose, first, that $p \in \mathscr{S}$, then we can repeat the argument above with the time-orientation changed, showing that γ enters $I^+(\mathscr{S})$ as well, and we are done.

Consider, finally, the possibility that $p \notin \mathscr{S}$, then p is necessarily in $I^+(\mathscr{S})$, hence γ meets $I^+(\mathscr{S})$. Now, each of the two disjoint sets

$$I_\pm := \{s \in I : \gamma(s) \in I^\pm(\mathscr{S})\} \subset \mathbb{R}$$

is open in the connected interval I. They cover I if γ does not meet $\mathscr{S}$, which implies that either I_+ or I_- must be empty when $\gamma \cap \mathscr{S} = \emptyset$. But we have shown that both I_+ and I_- are not empty, which implies that γ has to meet $\mathscr{S}$, as desired. □

Returning to the proof of Theorem 4.9.8, suppose that $\mathring{\mathscr{D}}(\mathscr{S})$ is not strongly causal. Then there exists $p \in \mathring{\mathscr{D}}(\mathscr{S})$ and a sequence $\gamma_n : \mathbb{R} \to \mathring{\mathscr{D}}(\mathscr{S})$ of inextendible past directed causal curves which exit the h-distance geodesic ball $B_h(p, 1/n)$ (centred at p and of radius $1/n$) and renter $B_h(p, 1/n)$ again. (As usual we assume that the γ_n's are parameterized by h–distance, with $\gamma_n(0) \in B_h(p, 1/n)$). Thus there exists a strictly increasing sequence $s_n > 0$ such that $\gamma_n(s_n) \in B_h(p, 1/n)$, with each point $\gamma_n(s_n)$ lying on a different connected component of $\gamma \cap B_h(p, \epsilon)$ for some $\epsilon > 0$.•4.9.5 Changing time-orientation if necessary, without loss of generality we may assume that $p \in \mathring{\mathscr{D}}^+(\mathscr{S}) \cup \mathscr{S}$.

•4.9.5: **Warning:** *one should add the (easy) justification that the s_n's are strictly separated away from zero*

Let γ be an accumulation curve of the γ_n's passing through p. The curve γ is causal and p is either in the interior of $\mathscr{D}^+(\mathscr{S})$ or on $\mathscr{S}$, we can therefore use Lemma 4.9.10 to conclude that γ enters $I^-(\mathscr{S})$. Since $I^-(\mathscr{S})$ is open, and since the γ_n's accumulate at γ, the γ_n's have to enter $I^-(\mathscr{S})$ for n large enough (passing again to a subsequence if necessary).

Suppose that s_n is bounded, then (passing to a subsequence if necessary) we have $s_n \to s_*$, $\gamma_n(s_*) \to p$, which shows that γ is a closed causal curve through p when restricted to $[0, s_*]$. We have $s_* > 0$ so that γ is non-trivial. We can obtain an inextendible (periodic) causal curve by circling $\gamma|_{[0,s_*]}$ over and over again. By Lemma 4.9.10 the curve so obtained meets periodically $I^+(\mathscr{S})$ and $I^-(\mathscr{S})$, so there exist points $q_\pm \in \gamma \cap I^\pm(\mathscr{S})$. We then have $q_+ \in J^-(p)$, with $I^-(q_+) \cap \mathscr{S} \neq \emptyset$, and Lemma 4.4.14 implies that $I^-(p) \cap \mathscr{S} \neq \emptyset$. Following γ backwards we find $q_- \in J^+(p)$, with $I^-(q_-) \cap \mathscr{S} \neq \emptyset$, and Lemma 4.4.14 implies that $I^+(p) \cap \mathscr{S} \neq \emptyset$. This contradicts achronality of $\mathscr{S}$.

It follows that $s_n \to \infty$. Let s_* be such that $\gamma(s_*) \in I^-(\mathscr{S})$, since the γ_n's accumulate at γ we will have that $\gamma_n(s_*) \in I^-(\mathscr{S})$ for n large enough. Since $s_n \to \infty$ for n large enough we will have $s_n > s_*$, which shows that γ_n has visited $I^-(\mathscr{S})$ at $s = s_*$ before meeting $B_h(p, 1/n) \subset \mathring{\mathscr{D}}^+(\mathscr{S})$ at $s = s_n$. If $p \in I^+(\mathscr{S})$, this contradicts again achronality of $\mathscr{S}$.

The only case left therefore is that in which $p \in \mathscr{S}$ and $s_n \to \infty$. In that case we will obtain a violation of strong causality at $\gamma(s_*) \in I^-(\mathscr{S})$ by following γ_n from 0 to s_n, and then following a causal curve which remains close to γ ...
•4.9.6 ... This establishes strong causality of $\mathring{\mathscr{D}}^+(\mathscr{S})$.

•*4.9.6:* **Warning:** *this isn't complete, there is a gap in the proof here, to be filled in*

To finish the proof we need to prove compactness of all the sets of the form

$$J^+(p) \cap J^-(q) \;, \quad p, q \in \mathring{\mathscr{D}}(\mathscr{S}) \;.$$

If p and q are such that this is set is empty there is nothing to prove; otherwise, consider a sequence $r_n \in J^+(p) \cap J^-(q)$, thus there exists a future directed causal curve $\hat{\gamma}_n$ from p to q which passes through r_n,

$$\gamma_n(s_n) = r_n \;. \tag{4.9.3}$$

Changing time-orientation and passing to a subsequence if necessary we may without loss of generality assume that $p \in I^-(\mathscr{S}) \cup \mathscr{S}$. Replacing p by q and passing again to a subsequence if necessary we may further assume that $r_n \in I^-(\mathscr{S}) \cup \mathscr{S}$. Let γ_n be any dist_h-parameterized, inextendible causal curve extending $\hat{\gamma}_n$, with $\gamma_n(0) = p$. Let γ be an inextendible accumulation curve of the γ_n's, then γ is a future inextendible causal curve through $p \in \mathscr{D}^-(\mathscr{S}) \cup \mathscr{S}$, so that by Lemma 4.9.10 there exists s_+ such that $\gamma(s_+) \in I^+(\mathscr{S})$. Passing to a subsequence, the γ_n's converge uniformly to γ on $[0, s_+]$, which implies that for n large enough the γ_n's do have to enter $I^+(\mathscr{S})$ at some time $s < s_+$. This, together with achronality of $\mathscr{S}$, shows that the sequence s_n defined by (4.9.3) is bounded; eventually passing to a subsequence we thus have $s_n \to s_\infty$. This implies

$$r_n \to \gamma(s_\infty) \in J^+(p) \cap J^-(q) \;,$$

which had to be established. □

We have the following characterisation of *interiors* of domains of dependence:

THEOREM 4.9.12 *A point $p \in \mathscr{M}$ is in $\mathring{\mathscr{D}}^+(\mathscr{S})$ if and only if*

$$\text{the set } \overline{I^-(p) \cap \mathscr{S}} \text{ is non-empty, and compact as a subset of } \mathscr{S}. \quad (4.9.4)$$

•*4.9.7*

•*4.9.7:* **Warning:** *I have not done that in class; the proof is incomplete and needs finishing*

REMARK 4.9.13 The reader is warned that condition (4.9.4) cannot be replaced by the requirement that $J^-(p)$ ....

PROOF: For $p \in \mathring{\mathscr{D}}^+(\mathscr{S})$ compactness of $\overline{I^-(p) \cap \mathscr{S}}$ can be established by an argument very similar to that given in the last part of the proof of Theorem 4.9.8, the details are left to the reader.

In order to prove the reverse implication assume that (4.9.4) holds, then there exists a future directed causal curve $\gamma : [0,1] \to \mathscr{M}$ from some point $q \in \mathscr{S}$ to p. Set

$$I := \{t \in [0,1] : \gamma(s) \in \mathring{\mathscr{D}}^+(\mathscr{S}) \ \text{ for all } s \le t\} \subset [0,1] \, .$$

Then I is not empty, as the fact that $\mathscr{S}$ is a topological hypersurface implies that $\mathring{\mathscr{D}}^+(\mathscr{S})$ contains a neighborhood of $\mathscr{S}$. Clearly I is open in $[0,1]$, in order to show that it equals $[0,1]$ set

$$t_* := \sup I \, .$$

Consider any past-inextendible past-directed causal curve $\hat\gamma$ starting at $\gamma(t_*)$. For $t < t_*$ let $\hat\gamma_t$ be a family of past-inextendible causal push-downs of $\hat\gamma$ which start at $\gamma(t)$, and which have the property that

$$\mathrm{dist}_h(\gamma_t(s), \gamma(s) \le 1/(t-t_*) for 0 \le s \le 1/(t-t_*) \, .$$

Then $\hat\gamma_t$ intersects $\mathscr{S}$ at some point $q_t \in J^-(p)$. Compactness of $J^-(p) \cap \mathscr{S}$ implies that the curve $t \to q_t \in \mathscr{S}$ accumulates at some point $q_* \in \mathscr{S}$, which clearly is the point of intersection of γ with $\mathscr{S}$. This shows that every causal curve γ through $\gamma(t_*)$ meets $\mathscr{S}$, in particular $\gamma(t_*) \in \mathscr{D}^+(\mathscr{S})$. □

4.10 Cauchy surfaces

A topological hypersurface $\mathscr{S}$ is said to be a *Cauchy surface* if

$$\mathscr{D}^+(\mathscr{S}) = \mathscr{M} \, .$$

Theorem 4.9.8 shows that a necessary condition for this equality is that $\mathscr{M}$ be globally hyperbolic. A celebrated theorem due independently to Geroch and Seifert shows that this condition is also sufficient:

THEOREM 4.10.1 (Geroch [53], Seifert [97]) *A space-time $(\mathscr{M}, g)$ is globally hyperbolic if and only if there exists on $\mathscr{M}$ a time function τ with the property that all its level sets are Cauchy surfaces. τ can be chosen to be smooth if g is smooth.*•4.10.1

•4.10.1: **ptc:** *What is the differentiability threshold for τ here?*

PROOF: The proof uses volume functions, defined as follows: let φ_i, $i \in \mathbb{N}$, be any partition of unity on $\mathscr{M}$, set

$$V_i := \int_{\mathscr{M}} \varphi_i d\mu \ ,$$

where $d\mu$ is, say, the Riemannian measure associated to the auxiliary Riemannian metric h on $\mathscr{M}$. Define

$$\nu := \sum_{i\in\mathbb{N}} \frac{1}{2^i V_i} \varphi_i \ .$$

Then ν is smooth, positive, nowhere vanishing, with

$$\int_M \nu \, d\mu = 1 \ .$$

Following Geroch, we define

$$V_\pm(p) := \int_{J^\pm(p)} \nu \, d\mu \ .$$

We clearly have

$$\forall p \in \mathscr{M} \quad 0 < V_\pm(p) < 1 \ .$$

The functions $V_\pm$ may fail to be continuous in general, an example is given in Figure **??**. It turns out that such behavior cannot occur under the current conditions:

LEMMA 4.10.2 *On globally hyperbolic space-times the functions $V_\pm$ are continuous.*

PROOF: Let p_i be any sequence converging to p, and let the symbol φ_Ω denote the characteristic function of a set Ω. Let q be any point such that $q \in I^-(p) \Leftrightarrow p \in I^+(q)$, since $I^+(q)$ forms a neighborhood of p we have $p_i \in I^+(q) \Leftrightarrow q \in I^-(p_i)$ for i large enough. Equivalently,

$$\forall i \geq i_0 \qquad \varphi_{I^-(p_i)}(q) = 1 = \varphi_{I^-(p)}(q) \ . \tag{4.10.1}$$

Since the right-hand-side of (4.10.1) is zero for $q \notin I^-(p)$ we obtain

$$\forall q \qquad \liminf_{i\to\infty} \varphi_{I^-(p_i)}(q) \geq \varphi_{I^-(p)}(q) \ . \tag{4.10.2}$$

By Corollary 4.8.5 $J^-(p)$ differs from $I^-(p)$ by a topological hypersurface•4.10.2 , so that

•4.10.2: **ptc:** *this needs justification*

$$\liminf_{i\to\infty} \varphi_{J^-(p_i)} \geq \varphi_{J^-(p)} \quad \text{a.e.} \tag{4.10.3}$$

To obtain the inverse inequality, let q be such that

$$\limsup_{i\to\infty} \varphi_{J^-(p_i)}(q) = 1 \; ,$$

hence there exists a sequence γ_j of future directed, dist_h-parameterised•4.10.3 causal curves from q to p_{i_j}. By Proposition 4.8.1 there exists a future directed accumulation curve of the γ_j's from q to p. We have thus shown the implication

•*4.10.3:* **ptc:***added because of change of definitions, check*

$$\limsup_{i\to\infty} \varphi_{J^-(p_i)}(q) = 1 \quad \Longrightarrow \quad \varphi_{J^-(p)}(q) = 1 \; .$$

Since the function appearing at the left-hand-side of the implication above can only take values zero or one, it follows that

$$\limsup_{i\to\infty} \varphi_{J^-(p_i)} \le \varphi_{J^-(p)} \; . \tag{4.10.4}$$

Equations (4.10.1)-(4.10.4) show that

$$\lim_{i\to\infty} \varphi_{J^-(p_i)} \;\; \text{exists a.e., and equals} \;\; \varphi_{J^-(p)} \;\; \text{a.e.}$$

Since

$$0 \le \varphi_{J^-(p)} \le 1 \in \mathscr{L}^1(\nu\, d\mu) \; ,$$

the Lebesgue dominated convergence theorem gives

$$V_-(p) = \int_{\mathscr{M}} \varphi_{J^-(p)} \nu\, d\mu = \lim_{i\to\infty} \int_{\mathscr{M}} \varphi_{J^-(p_i)} \nu\, d\mu = \lim_{i\to\infty} V_-(p_i) \; .$$

Changing time orientation one also obtains continuity of V_+. □

We continue with the following observation:

Lemma 4.10.3 *V_- tends to zero along any past-inextendible causal curve $\gamma : [a,b) \to \mathscr{M}$.*

Proof: Let X_i be any partition of $\mathscr{M}$ by sets with compact closure, the dominated convergence theorem shows that

$$\lim_{k\to\infty} \sum_{i\ge k} \int_{X_i} \nu\, d\mu = 0 \; . \tag{4.10.5}$$

Suppose that there exists $k < \infty$ such that

$$\forall s \quad J^-(\gamma(s)) \cap \left(\cup_{i=1}^k X_i \right) \ne \emptyset \; .$$

Equivalently, there exists a sequence $s_i \to b$ such that

$$\gamma(s_i) \in K := \overline{\cup_{i=1}^k X_i} \; .$$

Compactness of K implies that there exists (passing to a subsequence if necessary) a point $q_\infty \in K$ such that $\gamma(s_i) \to q_\infty$. Strong causality of $\mathscr{M}$ implies that there exists an elementary neighborhood $\mathscr{O}$ of q_∞ such that $\gamma \cap \mathscr{O}$ is connected, and Lemma 4.8.3 shows that $\gamma\cap\mathscr{O}$ has finite h-length, which contradicts inextendibility of γ (compare Theorem 4.5.4). This implies that for any k we have

$$J^-(\gamma(s)) \cap \left(\cup_{i=1}^k X_i\right) = \emptyset$$

for s large enough, say $s \ge s_k$. In particular

$$s \ge s_k \quad \Longrightarrow \quad \int_{J^-(\gamma(s))\cap\left(\cup_{i=1}^k X_i\right)} \nu \, d\mu = 0 \; .$$

This implies

$$\forall s \ge s_k \quad V_-(\gamma(s)) = \int_{J^-(\gamma(s))\cap\left(\cup_{i=k+1}^\infty X_i\right)} \nu \, d\mu \le \sum_{i\ge k+1} \int_{X_i} \nu \, d\mu \; ,$$

which, in view of (4.10.5), can be made as small as desired by choosing k sufficiently large. □

We are ready now to pass to the proof of Theorem 4.10.1. Set

$$\tau := \frac{V_-}{V_+} \; .$$

Then τ is continuous by Lemma 4.10.2. Let $\gamma : (a,b) \to \mathscr{M}$ be any inextendible future-directed causal curve. By Lemma 4.10.3

$$\lim_{s\to b} \tau(\gamma(s)) = \infty \; , \qquad \lim_{s\to a} \tau(\gamma(s)) = 0 \; .$$

Thus τ runs from 0 to ∞ on any such curves, in particular γ intersects every level set of τ at least once. •4.10.4 From the definition of the measure $\nu \, d\mu$ it should be clear that τ is actually strictly increasing on any causal curve, hence the level sets of τ are met by causal curves precisely once. •*4.10.4:* **ptc:***needs better justification*

The differentiability properties of τ constructed above are not completely clear. It thus remains to show that for smooth metrics τ can be modified, if necessary, so that it is actually smooth. This is done by a localisation-convolution procedure, as follows:•4.10.5 □ •*4.10.5:* **Warning:** *to be finished*

An important corollary of Theorem 4.10.1 is:

Corollary 4.10.4 *A globally hyperbolic space-time is necessarily diffeomorphic to $\mathbb{R}\times\mathscr{S}$, with the coordinate along the $\mathbb{R}$ factor having timelike gradient.*

Proof: Let X by any smooth timelike vector field on X — if the time function τ of Theorem 4.10.1 is smooth then $\nabla\tau$ will do, but any other choice works equally well. Choose any number τ_0 in the range of τ. Define a bijection $\varphi : \mathscr{M} \to \mathbb{R}\mathscr{S}$ as follows: for $p \in M$ let $q(p)$ be the point on the level set $\mathscr{S}_0 = \{r \in \mathscr{M} : \tau(r) = \tau_0\}$ which lies on the integral curve of X through p. Such a point exists because any inextendible timelike curve in $\mathscr{M}$ meets $\mathscr{S}_0$; it is unique by achronality of $\mathscr{S}_0$. The map φ is continuous by continuous dependence of ODE's upon initial values. If τ is merely continuous, one can invoque the invariance of domain theorem•4.10.6 to prove that φ is a homeomorphism; if τ is differentiable, its level sets are differentiable, X meets those level sets transversely, and the fact that φ is a diffeomorphism follows from the implicit function theorem. □

•4.10.6: **ptc:***give ref*

It is not easy to decide whether or not a hypersurface $\mathscr{S}$ is a Cauchy hypersurface, except in *spatially compact* space-times:

THEOREM 4.10.5 (Budič *et al.* [**?**, **?**] Galloway [52]) *Let* $(\mathscr{M}, g)$ *be globally hyperbolic and suppose that* $\mathscr{M}$ *contains a smooth, compact, connected spacelike hypersurface* $\mathscr{S}$*. Then* $\mathscr{S}$ *is a Cauchy surface for* $\mathscr{M}$*.*•*4.10.7*

•*4.10.7:* **ptc:***give proof*

REMARK 4.10.6 Some further results concerning Cauchy surface criteria can be found in [52, 57].

•4.10.8

•*4.10.8:* **ptc:***add the new guys reference, with their Cauchy surface criterion, and smoothing argument*

THEOREM 4.10.7 *Let* $\mathscr{S}$ *be a spacelike hypersurface in a space-time* $(\mathscr{M}, g)$*. Then the Cauchy problem for the wave equation has a unique globally defined solution for all smooth initial data if and only if* $\mathscr{S}$ *is a Cauchy surface for* $\mathscr{S}$*.*•*4.10.9*

•*4.10.9:* **ptc:***This result certainly needs a proof, if true*

4.11 Some applications

Any formalism is only useful to something if it leads to interesting applications. In this section we will list some of those.
•4.11.1

•*4.11.1:* **ptc:***start with Choquet-Bruhat Geroch*

We say that $(\mathscr{M}, g)$ satisfies the *timelike focussing condition*, or *timelike convergence condition*, if the Ricci tensor satisfies

$$R_{\mu\nu}n^\mu n^\nu \geq 0 \ \text{ for all } \underline{\text{timelike}} \text{ vectors } \ n^\mu \ . \tag{4.11.1}$$

By continuity, the inequality in (4.11.1) will also hold for causal vectors. Condition (4.11.1) can of course be rewritten as a condition on the matter fields using the Einstein equation, and is satisfied in many cases of interest, including vacuum general relativity, or the Einstein-Maxwell theory, or the Einstein-Yang-Mills theory. This last two examples actually have the property that the corresponding energy-momentum tensor is trace-free; whenever this happens, (4.11.1) is simply the requirement that the energy density of the matter fields is non-negative for all observers:

$$8\pi T_{\mu\nu}n^\mu n^\nu = (R_{\mu\nu} - \frac{1}{2}\underbrace{R}_{=0 \text{ if } g^{\alpha\beta}T_{\alpha\beta}=0} g_{\mu\nu})n^\mu n^\nu = R_{\mu\nu}n^\mu n^\nu \ . \tag{4.11.2}$$

We say that $(\mathscr{M}, g)$ satisfies the *null energy condition* if

$$R_{\mu\nu}n^\mu n^\nu \geq 0 \ \text{ for all } \underline{\text{null}} \text{ vectors } \ n^\mu \ . \tag{4.11.3}$$

Clearly, the timelike focussing condition implies the null energy condition. Because $g_{\mu\nu}n^\mu n^\nu = 0$ for all null vectors, the R term in the calculation (4.11.2) drops out regardless of whether or not $T_{\mu\nu}$ is traceless, so the null energy condition is equivalent to positivity of energy density of matter fields without any provisos.

The simplest *geodesic incompleteness theorem* is:•4.11.2

•*4.11.2:* **ptc:***Hawking has a claim on this theorem too?*

THEOREM 4.11.1 (Geroch's geodesic incompleteness theorem [?]) *Let $(\mathscr{M}, g)$ be a globally hyperbolic satisfying the timelike focussing condition, and suppose that $\mathscr{M}$ contains a Cauchy surface $\mathscr{S}$ with strictly negative mean curvature:*

$$\mathrm{tr}_h K < 0 \, ,$$

where (h, K) are the usual Cauchy data induced on $\mathscr{S}$ by g. Then $(\mathscr{M}, g)$ is future timelike geodesically incomplete.

Let $\mathscr{S}$ be a spacelike hypersurface in $(\mathscr{M}, g)$, and a consider a surface $S \subset \mathscr{S}$. We shall also assume that S is two-sided in $\mathscr{S}$, this means that there exists a globally defined field m of unit normals to S within $\mathscr{S}$. There are actually two such fields, m and $-m$, we arbitrarily choose one and way call it *outer pointing*. In situations where S does actually bound a compact region, the outer-pointing one should of course be chosen to point away from the compact region. We let H denote the mean extrinsic curvature of S within $\mathscr{S}$:

$$H := D_i m^i \, , \tag{4.11.4}$$

where D is the covariant extrinsic of the metric h induced on $\mathscr{S}$. We say that S is *outer-future-trapped* if•4.11.3

•*4.11.3:* **ptc:***crosscheck sign*

$$\theta_+ := H + K_{AB} m^A m^B \le 0 \, , \tag{4.11.5}$$

with an obvious symmetric definition for *inner-future-trapped*:

$$\theta_- := -H + K_{AB} m^A m^B \ge 0 \, , \tag{4.11.6}$$

(One also has the obvious *past* version thereof, where the sign in front of the K term should be changed.) A celebrated theorem of Penrose[7] reads:

THEOREM 4.11.2 (Penrose's geodesic incompleteness theorem [?]) *Let $(\mathscr{M}, g)$ be a globally hyperbolic space-time satisfying the* null energy condition*, and suppose that $\mathscr{M}$ contains a* non-compact *Cauchy surface $\mathscr{S}$. If there exists a* compact *trapped surface within S which is both* inner-future-trapped *and* outer-future-trapped*, then $(\mathscr{M}, g)$ is geodesically incomplete.*

Future-trapped surfaces signal the existence of black holes — a formal statement requires the introduction of the notion of a black hole, as well as several global regularity conditions, and will therefore not be given here.•4.11.4

•*4.11.4:* **ptc:***should be done in the black holes section*

Another example of application of the causality theory developed so far is the *area theorem*:

[7] Penrose's theorem is slightly more general; this requires a definition of $\theta_\pm$ which involves a discussion of null geometry which we prefer to avoid here. This is the reason why we have stated this theorem in the current form.

THEOREM 4.11.3 ([30, 58]) *Let $\mathscr{E}$ be a future geodesically complete acausal null hypersurface, and let $\mathscr{S}_1$, $\mathscr{S}_2$ be two spacelike acausal hypersurfaces. If*

$$\mathscr{E} \cap \mathscr{S}_1 \subset J^-(\mathscr{E} \cap \mathscr{S}_2) ,$$

then

$$Area(\mathscr{E} \cap \mathscr{S}_1) \leq Area(\mathscr{E} \cap \mathscr{S}_2) .$$

•4.11.5

•*4.11.5:* **ptc:***show existence of maximising geodesics in globally hyperbolic space-times*

4.12 The Lorentzian length functional

Let γ be a causal path, we define the *Lorentzian length* of γ with respect to the metric g as

$$\ell_g(\gamma) = \int_\gamma \sqrt{-g(\dot\gamma,\dot\gamma)}\ . \tag{4.12.1}$$

We will write $\ell(\gamma)$ when no ambiguities concerning the metric can arise.

A key property is the *upper semi-continuity* of ℓ with respect to uniform convergence on compact sets:•4.12.1

•4.12.1: **ptc:***I probably don't need that, do I? remove the coco environment if eventually needed*

PROPOSITION 4.12.1 (Penrose [89], Eschenburg and Galloway [43]) *Let a sequence of causal curves $\gamma_n : I \to \mathscr{M}$ converge to $\gamma : I \to \mathscr{M}$, uniformly on compact subsets of I. Then*

$$\ell(\gamma) \ge \limsup \ell(\gamma_n)\ .$$

PROOF: Suppose, first, that $(\mathscr{M}, g)$ is strongly causal, and that I is compact. Then..... •4.12.2

•4.12.2: **ptc:***to be proved, this is in Penrose's notes*

In general, we can partition I using a countable sequence $t_i \in I$, $t_i < t_{i+1}$, $i \in \mathbb{Z}$, such that $\gamma|_{[t_i,t_{i+1}]}$ is contained in an elementary neighborhood $\mathscr{O}_i$. Each of the $\mathscr{O}_i$'s with the induced metric g is strongly causal (since x^0 is a time function there). By uniform convergence we have $\gamma_n|_{[t_i,t_{i+1}]} \subset \mathscr{O}_i$ for i large enough, so that we can use the result already established to show that

$$\ell(\gamma|_{[t_i,t_{i+1}]}) \ge \limsup \ell(\gamma_n|_{[t_i,t_{i+1}]})\ .$$

Summing over i proves the proposition. □

4.13 The Lorentzian distance function

Let $\Omega(p,q)$ be the set of all future directed causal curves from p to q. We define the *Lorentzian distance function* $d : M \times M \to [0,\infty]$ as follows:

$$d(p,q) = \begin{cases} \sup\{\ell(\gamma) : \gamma \in \Omega(p,q)\}, & q \in J^+(p), \\ 0, & \text{otherwise}, \end{cases} \tag{4.13.1}$$

where sup is understood in $\mathbb{R} \cup \{\infty\}$. We shall sometimes write d_g for d when the need to indicate explicitly the metric arises.

It is tempting to define $d(p,q) = -\infty$ for $q \notin J^+(p)$, but this leads to a function d which is never continuous (compare Proposition 4.13.12), which is the reason for using (4.13.1).

It is legitimate to ask the question, what would happen if one took the infimum rather than the supremum in (4.13.1). The result is not very interesting, the reader should easily convince himself that one can approach a C^0 curve γ as close as desired by threading back and forth near γ along null geodesics, each of which has zero Lorentzian length. So taking the infimum in (4.13.1) always gives zero.

In any case the calculation of σ in Minkowski space-time, which we are about to do, should make it clear that the right thing to do is to take the supremum:

EXAMPLE 4.13.1 Let $q = (x^\mu) = (x^0, \vec{x}) \in \mathbb{R}^{1,n}$ be in the timelike future of the origin, thus $x^0 > |\vec{x}|_\delta$ by Proposition 4.4.3. Let us make a Gram-Schmidt orthonormalization starting from (x^μ), viewed as a vector tangent to $\mathbb{R}^{1,n}$ at 0, thus $x^\mu e_\mu = \sqrt{(x^0)^2 - |\vec{x}|^2_\delta} e_0$. Using normal coordinates based on the new basis e_μ one obtains a Minkowskian coordinate system in which $q = (t, \vec{0})$, with

$$t = \sqrt{(x^0)^2 - |\vec{x}|^2_\delta} \,. \tag{4.13.2}$$

Consider, now any causal curve $\gamma(s) = (\gamma^0(s), \vec{\gamma}(s))$, from the origin to q, causality gives

$$|\dot\gamma^0| \geq |\dot{\vec{\gamma}}|_\delta \,,$$

with $|\dot\gamma^0|$ without zeros (since the inequality is strict for timelike vectors, and if it is an equality than neither side can vanish, otherwise $\dot\gamma$ would vanish, which is not allowed for causal vectors). This shows that we can reparameterize γ so that

$$\gamma^0(s) = s \,, \quad s \in [0, t] \,.$$

In this parameterization we have

$$\ell(\gamma) = \int_0^t \sqrt{1 - |\dot{\vec{\gamma}}|^2_\delta} \,,$$

and this formula makes it clear that the supremum is attained on the path

$$\gamma(s) = (s, \vec{0}) \quad \Longrightarrow \quad \ell(\gamma) = t \,.$$

It follows that

$$d_\eta(p,q) = \begin{cases} \sqrt{(x^0)^2 - \sum_i (x^i)^2} \,, & q \in J^+(p), \\ 0, & \text{otherwise.} \end{cases} \tag{4.13.3}$$

We note that (4.13.3) follows also from Proposition 4.13.9 below.

EXAMPLE 4.13.2 The two-dimensional space-time $\mathscr{M} = S^1 \times S^1$ with the flat metric $-dt^2 + dx^2$, where t is a $\mod 2\pi$–coordinate on the first S^1 factor, and x is a similar coordinate on the second factor, provides an example where $d(p,q) = \infty$ for all $p, q \in \mathscr{M}$. This is seen by noting that if γ is a timelike curve from $p = (t_0, x_0)$ to q, then one can obtain a causal curve of length $2\pi n + \ell(\gamma)$ be going n times around the timelike circle $x = x_0$, and then following γ from p to q.

It follows immediately from its definition that the Lorentzian distance function obeys the reverse triangle inequality: for $p \in J^-(q)$ and $q \in J^-(r)$ it holds that

$$d(p,r) \geq d(p,q) + d(q,r) \,. \tag{4.13.4}$$

Indeed, if $p \in J^-(q)$ and $q \in J^-(r)$, then the class $\Omega(p,r)$ of future directed causal curves from p to r contains the class $\Omega(p,q) \cup \Omega(q,r)$, where the union $\cup$ is understood as the concatenation operation on paths. It follows that the sup

Figure 4.4: The space-time of Example 4.13.3.

of ℓ over $\Omega(p,q)$ is greater than or equal to the sup of ℓ over $\Omega(p,q)\cup\Omega(p,r)$, which implies (4.13.4)

The function d needs not to be continuous in general, as seen in the following example:

EXAMPLE 4.13.3 ([16, p. 141]) Let $\mathscr{M}$ be

$$\{(t,x):0\le t\le 2\}\setminus\{(1,x):-1\le x\le 1\}\ ,$$

with the identifications $(x,0)\sim(x,2)$ (see Figure 4.4)•4.13.1 , equipped with the flat metric $g=-dt^2+dx^2$. Let p be the origin and let $q=(0,1/2)$, set $p_n=(0,1/n)\to p$. Clearly $p_n\in I^+(p_n)$ which implies $d(p_n,p_n)=\infty$, and also $d(p_n,q)=\infty$ for $n>2$. By Proposition 4.13.9 we have $d(p,q)=1/2$, so that the function $d(\cdot,q)$ is not continuous at p.

•4.13.1: **ptc:**_make figure_

Let us show that d is *lower semi-continuous*:

PROPOSITION 4.13.4 *If $p_n\to p$ and $q_n\to q$ then*

$$d(p,q)\le\liminf d(p_n,q_n)\ . \tag{4.13.5}$$

(This implies in particular that d is continuous on $d^{-1}(\{\infty\})$.)

PROOF: If $d(p,q)=0$ there is nothing to prove. Otherwise let $\gamma:[0,1]\to\mathscr{M}$ be any causal curve from p to q with non-zero Lorentzian length. Let us start by showing that we can always deform γ to a new causal curve $\tilde\gamma$, which is *timelike near p and q*, reducing the length by less than $\epsilon/2$ (if at all). Indeed, let $s_*>0$ be the last point on γ such that $\gamma(s)\not\in I^+(p)$, then $s<1$, and by Proposition 4.6.6 $\gamma|_{[0,s_*]}$ is a null geodesic, in particular $\ell(\gamma|_{[0,s_*]})=0$. By definition of s_* there exists a sequence $s_i\searrow s_*$ such that $\gamma(s_i)\in I^+(\gamma(s_*))$. Since

$$\ell(\gamma)=\ell(\gamma|_{[s_*,1]})\ ,$$

by choosing i large enough we will have $\ell(\gamma|_{[s_i,1]})\ge\ell(\gamma)-\epsilon/4$. One can use the path constructed in the proof of Lemma 4.4.14 to replace $\gamma|_{[0,s_i]}$ by a timelike curve from p to $\gamma(s_i)$, leading to a new curve from p to q which is timelike near p and which has length not less than $\ell(\gamma)-\epsilon/4$. A similar construction near q leads to a causal path $\tilde\gamma:[0,1]\to\mathscr{M}$ from p to q, timelike near p and q, such that

$$\ell(\tilde\gamma)\ge\ell(\gamma)-\epsilon/2\ .$$

Let, now, $s_->0$ be small enough so that

$$\ell(\tilde\gamma|_{[0,s_-]})<\epsilon/4\ .$$

Since $\tilde\gamma$ is timelike near p the set $I^-(\gamma(s_-))$ is an open neighborhood of p and therefore, for n large enough, there exists a timelike curve from p_n to $\gamma(s_-)$. Similarly we can find s_+ close to q so that

$$\ell(\tilde\gamma|_{[s_+,1]})<\epsilon/4\ ,$$

and there exists a timelike curve from $\gamma(s_+)$ to q_n for n large enough. Concatenating those curves results in a causal curve from p_n to q_n with length not less than $\ell(\gamma)-\epsilon$. Taking a supremum over γ's proves (4.13.5) except for $-\epsilon$ at the left-hand-side. However, as ϵ is arbitrary, (4.13.5) follows.

If $d(p,q) = \infty$, then (4.13.5) gives $\liminf d(p_n,q_n) = \infty$, which clearly implies $\lim d(p_n,q_n)=\infty$, as desired. $\Box$

We shall say that a future directed causal path $\gamma : I \to \mathscr{M}$ is *maximising* if

$$\forall\ s,s' \in I\ ,\ s<s' \qquad d(\gamma(s),\gamma(s')) = \ell(\gamma|_{[s,s']})\ . \tag{4.13.6}$$

We have the following simple observation:

Proposition 4.13.5 *$d(p,q)$ is attained on $\gamma : [0,1] \to \mathscr{M}$ if and only if γ is maximising.*

Proof: The triangle inequality gives

$$d(p,q) \geq d(p,\gamma(s_1)) + d(\gamma(s_1),\gamma(s_2)) + d(\gamma(s_2),q)\ . \tag{4.13.7}$$

Since $d(p,q)$ is attained on γ we have

$$\begin{aligned} d(p,q) &= \int_0^1 \sqrt{g(\dot\gamma,\dot\gamma)}(s)ds \\ &= \int_0^{s_1} \sqrt{g(\dot\gamma,\dot\gamma)}(s)ds + \underbrace{\int_{s_1}^{s_2} \sqrt{g(\dot\gamma,\dot\gamma)}(s)ds}_{=\ell(\gamma|_{[s_1,s_2]})} + \int_{s_2}^1 \sqrt{g(\dot\gamma,\dot\gamma)}(s)ds\ . \end{aligned}$$

If γ were not maximising, then there would exist $s_1,s_2 \in [0,1]$ and a causal curve σ from $\gamma(s_1)$ to $\gamma(s_2)$ with Lorentzian length larger than $\ell(\gamma|_{[s_1,s_2]})$. One would then obtain a causal path longer than γ by following γ from p to $\gamma(s_1)$, then following σ, and then following γ from $\gamma(s_2)$ to q. This contradicts the fact that $d(p,q)$ is attained on γ. It follows that $\ell(\gamma|_{[s_1,s_2]}) = d(\gamma(s_1),\gamma(s_2))$ for all $s_1,s_2\in[0,1]$. $\Box$

We also note:

Proposition 4.13.6 *A null geodesic γ is maximising if and only if it is achronal.*

Proof: $\Longrightarrow$: From the definition of d we have $d(\gamma(s_1),\gamma(s_2)) > 0$ whenever $\gamma(s_2) \in I^+(\gamma(s_1))$, therefore all the terms at the right-hand-side of (4.13.7) are zero if $d(p,q)$ vanishes. It follows that γ is achronal, and the fact that γ is a null geodesic follows from Proposition 4.6.6.

$\Longleftarrow$: Since γ is achronal the inequality in (4.13.7) is an inequality, with all terms vanishing. $\Box$

A Jacobi field along a geodesic γ is a solution of *Jacobi equation*:

$$\frac{D^2Z}{ds^2}(s) = R(\dot\gamma,Z)\dot\gamma\ . \tag{4.13.8}$$

The point $\gamma(s_2)$ is said to be *conjugate* to $\gamma(s_1)$ if there exists a non-trivial solution of (4.13.8) which vanishes at both points.

A key result concerning maximising curves is:

THEOREM 4.13.7 *If γ is maximising, then γ is a geodesic without conjugate points.* •4.13.2

•4.13.2: **ptc:***the proof of the theorem to be finished*

PROOF: Proposition 4.13.5 reduces the problem to showing that if $d(p,q)$ is attained on γ, then γ is a geodesic without conjugate points. This is proved by variational arguments, as follows. We start with a Lemma:•4.13.3

•4.13.3: **ptc:***watch out for differentiability of the metric, here we assume that it is $C^{1,1}$, say it:*

LEMMA 4.13.8 *Let $q \in I^+(p)$, and let $\Omega_{C^{1,1}}(p,q)$ denote the class of $C^{1,1}$ timelike paths from p to q. Then*

$$\sup_{\gamma\in\Omega(p,q)} \ell(\gamma) = \sup_{\gamma\in\Omega_{C^{1,1}}(p,q)} \ell(\gamma) \ .$$

□

It follows from Proposition A.5.14 in Appendix A.5.5 that the exponential map ceases to be a local diffeomorphism at conjugate points. This implies, by definition, that there are no points conjugate to the origin in elementary neighborhoods, leading to:•4.13.4

•4.13.4: **ptc:***this does not need the conjugate points story and can be proved using Gauss coordinates on a normal neighborhood so it could be used in the proof of the previous theorem*

PROPOSITION 4.13.9 *Let $\mathscr{O}$ be an elementary neighborhood centred at p. For $q \in \mathscr{O}$ the geodesic segment from p to q is the longest causal curve from p to q entirely contained in $\mathscr{O}$. Equivalently, the distance function $\sigma(p,\cdot)$ within the space-time $(\mathscr{O}, g|_{\mathscr{O}})$ coincides with $\sqrt{-\sigma_p}$ on $J^+(p,\mathscr{O})$, and vanishes elsewhere.*
□

Let us show that accumulation curves of maximising paths are maximising:

PROPOSITION 4.13.10 *Let γ_n be a sequence of maximising curves accumulating at γ, then γ is a maximising geodesic.*

PROOF: By Theorem 4.13.7 the γ_n's are geodesics, so that γ is a geodesic by Proposition 4.6.5. Reparameterizing if necessary, we can assume that the γ_n's are affinely parameterised and defined on a common compact interval I. Passing to a subsequence if necessary, the γ_n's converge to γ in $C^1(I,\mathscr{M})$, so that $\ell(\gamma_n) \to \ell(\gamma)$. Suppose that γ is not maximising between $p = \lim \gamma_n(s_1)$ and $q = \lim \gamma_n(s_2)$, then there exists a causal curve $\tilde\gamma$ from p to q with

$$r := \ell(\tilde\gamma) - \ell(\gamma|_{[s_1,s_2]}) > 0 \ ,$$

in particular $q \in I^+(p)$. Since $\ell(\gamma_n) \to \ell(\gamma)$ we have $\ell(\gamma_n) < \ell(\gamma) - r/2$ for n large enough. However, the construction in the proof of Proposition 4.13.4 provides, for n large enough, a causal curve from $\gamma_n(s_1)$ to $\gamma_n(s_2)$ strictly longer than $\ell(\gamma) - r/2$ if we choose ϵ there to be smaller than $r/2$, contradicting the maximising property of γ_n. □

We are almost ready to show that d is continuous on globally hyperbolic space-times. We start with the proof of a slightly stronger version of Theorem 4.8.6:

THEOREM 4.13.11 *Let $(\mathscr{M}, g)$ be globally hyperbolic, if $q \in I^+(p)$, respectively $q \in J^+(p)$, then there exists a timelike, respectively causal, future directed, maximising geodesic from p to q, on which $d(p,q)$ is attained.*•4.13.5

•*4.13.5:* **ptc:***watch out if with boundary*

PROOF: Let γ_n be any sequence of dist_h-parameterized causal paths from p to q such that $\ell(\gamma_n) \to \mathrm{dist}_g(p,q)$. By Proposition 4.8.1 there exists a causal curve γ through p and q which is an accumulation curve of the γ_n's. With some work one shows that γ has to be a geodesic — this uses the fact that γ is the longest causal curve from p to q. If $q \in I^+(p)$ then $\mathrm{dist}_g(p,q) > 0$ and the resulting geodesic has to be timelike.•4.13.6 □

•*4.13.6:* **ptc:***the details of this should be filled in*

PROPOSITION 4.13.12 *On globally hyperbolic space-times d is continuous, everywhere finite.*

PROOF: Let $q_n \in I^+(p_n)$ and suppose that (p_n, q_n) converge to (p,q). Let γ_n be any maximising geodesic from p_n to q_n, then (passing to a subsequence if necessary), by Proposition 4.13.10, γ_n converges to a maximising geodesic from p to q in C^2 topology. This implies that $\ell(\gamma_n) \to \ell(\gamma)$, proving the result. □

Chapter 5

Splitting theorems

5.1 The geometry of C^2 null hypersurfaces

Null hypersurfaces, which we will define shortly, provide an important tool in the study of Lorentzian manifolds. Generic null hypersurfaces arising naturally in general relativity will typically not be C^2; however, it is convenient to start the analysis with C^k surfaces, with $k \geq 2$. Our exposition follows that in [50], as expanded in [30].

It is convenient to start with some elementary algebra. Let W be a linear space equipped with a Lorentzian scalar product g. Recall that a hyperplane V in W is a vector subspace of codimension one. Let

$$V^0 = \{\alpha \in W^* : \alpha(v) = 0 \ \text{ for all } \ v \in V\} \subset W^* \ .$$

Then V^0 is one-dimensional, choose some $0 \neq \alpha \in V^0$. Recall that W^* is equipped with a natural Lorentzian metric $g^\flat$; V is called *spacelike* if α is timelike, *timelike* if α is spacelike, and *null* if α is null. In the first two cases the restriction of g is non-degenerate (with signature $(+\ldots+)$ for spacelike hyperplanes and $(-+\ldots+)$ for the timelike ones). On the other hand, for null hyperplanes the scalar product restricted to V is necessarily degenerate: indeed, let $K = g^\flat(\alpha, \cdot) \in (W^*)^* = W$, then $K \in V$ since $\alpha(K) = g^\flat(\alpha,\alpha) = 0$. Next, K is isotropic since $g^\flat(\alpha,\alpha) = g(K,K) = 0$. Thus, $g|_V$ has at least one isotropic direction, and it follows from non-degeneracy of g that there is only one such direction. It should be clear that the condition of degeneracy of $g|_V$ is equivalent to that of α being isotropic. •5.1.1

•5.1.1: **ptc:***to be cleaned-up*

Let (M,g) be a spacetime of dimension $n+1 \geq 3$. We denote the Lorentzian metric on M by g or $\langle\,,\rangle$. A C^2 *null hypersurface* in M is a C^2 embedded submanifold $\mathcal{H}$ of M, of co-dimension one, such that the pullback of the metric g to $\mathcal{H}$ is degenerate. The signature of the restriction of g to $T\mathcal{H}$ is thus $(0,1,\cdots,1)$, and by standard results on quadratic forms for every $p \in \mathcal{H}$ there exists a unique one dimensional subspace $\mathcal{K}_p \subset T_p\mathcal{H}$ such that

$$g(K,X) = 0 \ \text{ for each } K \in \mathcal{K}_p \text{ and } X \in T_p\mathcal{H}.$$

Clearly we can choose a vector $K(p) \in \mathcal{K}_p$ in a way as differentiable as the hypersurface allows (C^{k-1} if $\mathcal{H}$ is C^k), which shows that a C^2 hypersurface $\mathcal{H}$

admits a C^1 non-vanishing future directed null vector field $K \in \Gamma T\mathcal{H}$. Clearly, K is unique up to a positive scale factor. K has the property that the normal space of K at a point $p \in \mathcal{H}$ coincides with the tangent space of $\mathcal{H}$ at p, *i.e.*,

$$K_p^\perp = T_p\mathcal{H} \text{ for all } p \in \mathcal{H} . \tag{5.1.1}$$

As the signature of the restriction of g to $T\mathcal{H}$ is $(0, 1, \cdots, 1)$, those tangent vectors to $\mathcal{H}$ which are not parallel to K are spacelike.

We have:

Proposition 5.1.1 *The integral curves of K are null geodesics.*

Proof: Let φ be any defining function for $\mathcal{H}$, *i.e.*, $\mathcal{H}$ is the zero level set of φ, with $d\varphi$ nowhere vanishing on $\mathcal{H}$. Then $d\varphi$ annihilates $T\mathcal{H}$, so does $K^\flat = g(K, \cdot)$, and since $T\mathcal{H}$ has codimension one we obtain that $K_\flat$ is proportional to $d\varphi$. Since $g(K, K) = 0$ we have $g(\nabla\varphi, \nabla\varphi) = 0$ on $\mathcal{H}$, and Proposition 4.2.2 shows that the integral curves of φ intersecting $\mathcal{H}$ are null geodesics[1]. But the integral curves of K differ from those of $\nabla\varphi$ only by a reparametrization[2], whence the result. □

These integral curves are called the *null geodesic generators* of $\mathcal{H}$.

As K is orthogonal to $\mathcal{H}$, one can introduce the null Weingarten map and null second fundamental form of $\mathcal{H}$ with respect K in a manner roughly analogous to what is done for spacelike hypersurfaces (see Section **??**), as follows: We start by introducing an equivalence relation on tangent vectors: for $X, X' \in T_p\mathcal{H}$,

$$X' = X \bmod K \text{ if and only if } X' - X = \lambda K$$

for some $\lambda \in \mathbb{R}$. Let $\overline{X}$ denote the equivalence class of X. We have

Proposition 5.1.2 *Let $X, Y \in T_p\mathcal{H}$. If $X' = X$ mod K and $Y' = Y$ mod K then*

1. $\langle X', Y' \rangle = \langle X, Y \rangle$, *and*
2. $\langle \nabla_{X'} K, Y' \rangle = \langle \nabla_X K, Y \rangle$.

Proof: 1. $\langle X + \lambda K, Y + \mu K \rangle = \langle X, Y \rangle + \mu \underbrace{\langle X, K \rangle}_{=0} + \lambda \underbrace{\langle K, Y + \mu K \rangle}_{=0}$.

2. Since the integral curves of K are null geodesics we have $\nabla_K K = \gamma K$ for some function γ. This gives

$$\langle \nabla_{X+\lambda K} K, Y + \mu K \rangle = \langle \nabla_X K, Y \rangle + \lambda \underbrace{\langle \underbrace{\nabla_K K}_{=\gamma K}, Y + \mu K \rangle}_{=0} + \mu \langle \nabla_X K, K \rangle .$$

[1]Strictly speaking, one should argue as follows: since K is tangent to $\mathcal{H}$, so is $\nabla\varphi$, hence the integral curves of $\nabla\varphi$ are tangent to $\mathcal{H}$. Thus $g(\nabla\varphi, \nabla\varphi) = 0$ along those integral curves, and the argument of the proof of Proposition 4.2.2 applies.

[2]Let $x(s)$ satisfy $dx/ds = K(x(s))$, let φ be any positive function, and let $ds/ds' = \varphi(x(s))$. Then $x(s(s'))$ satisfies the equation $dx/ds' = ds/ds' \cdot dx/ds = \varphi(x(s))K(x(s))$. It follows indeed that the integral curves of φK coincide, as point sets, with those of K.

Now, since X is tangent to $\mathcal{H}$,

$$\langle \nabla_X K, K\rangle = \frac{1}{2} X(\underbrace{\langle K, K\rangle}_{=0}) = 0 \ .$$

□

Proposition 5.1.2 shows that components along K are irrelevant for some objects of interest. For this reason it is convenient to work with the tangent space of $\mathcal{H}$ modded out by K:

$$T_p\mathcal{H}/K = \{\overline{X} \mid X \in T_p\mathcal{H}\} \ \text{ and } \ T\mathcal{H}/K = \cup_{p\in\mathcal{H}} T_p\mathcal{H}/K \ .$$

When the space-time M has dimension $n+1$, then $T\mathcal{H}/K$ is a rank $n-1$ vector bundle over $\mathcal{H}$. This vector bundle does not depend on the particular choice of null vector field K. There is a natural positive definite metric h in $T\mathcal{H}/K$ induced from g: For each $p \in \mathcal{H}$, define $h : T_p\mathcal{H}/K \times T_p\mathcal{H}/K \to \mathbb{R}$ by

$$h(\overline{X}, \overline{Y}) = \langle X, Y\rangle \ . \tag{5.1.2}$$

Point 1 of Proposition 5.1.2 shows that h is indeed well-defined.

The *null Weingarten map* $b = b_K$ of $\mathcal{H}$ with respect to K is, for each point $p \in \mathcal{H}$, a linear map $b : T_p\mathcal{H}/K \to T_p\mathcal{H}/K$ defined by

$$b(\overline{X}) = \overline{\nabla_X K} \ . \tag{5.1.3}$$

Point 2. of Proposition 5.1.2 shows that b is well-defined. As b involves taking a derivative of K, which is C^1, the tensor b will be C^0, but no more regularity can be expected in general. Note that if

$$\widetilde{K} = fK \ , \tag{5.1.4}$$

with $f \in C^1(\mathcal{H})$, is any other future directed null vector field tangent to $\mathcal{H}$, then

$$\nabla_X \widetilde{K} = f\nabla_X K \bmod K \ .$$

This shows that

$$b_{fK} = f b_K \ . \tag{5.1.5}$$

It follows that the Weingarten map b of $\mathcal{H}$ is unique up to a scale factor φ, with φ positive once a time-orientation of K has been chosen.

Let us show that b is self-adjoint with respect to h:

$$h(b(\overline{X}), \overline{Y}) = h(\overline{X}, b(\overline{Y})) \ . \tag{5.1.6}$$

Indeed,

$$h(b(\overline{X}), \overline{Y}) = \langle \nabla_X K, Y\rangle = X(\underbrace{\langle K, Y\rangle}_{=0}) - \langle K, \nabla_X Y\rangle \ .$$

Now, since the torsion vanishes we have

$$\begin{aligned}\langle K, \nabla_X Y\rangle &= \langle K, \nabla_Y X + \underbrace{[X,Y]}_{\in K^\perp}\rangle = \langle K, \nabla_Y X\rangle = Y(\underbrace{\langle K, X\rangle}_{=0}) - \langle \nabla_Y K, x\rangle \\ &= -h(\overline{X}, b(\overline{Y})) \ ,\end{aligned}$$

whence (5.1.6).

The *null second fundamental form* $B = B_K$ of $\mathcal{H}$ with respect to K is the bilinear form associated to b via h: For each $p \in \mathcal{H}$,

$$B : T_p\mathcal{H}/K \times T_p\mathcal{H}/K \to \mathbb{R}$$

is defined by

$$B(\overline{X}, \overline{Y}) = h(b(\overline{X}), \overline{Y}) = \langle \nabla_X K, Y \rangle \; . \tag{5.1.7}$$

Since b is self-adjoint, B is symmetric,

$$B(\overline{X}, \overline{Y}) = B(\overline{Y}, \overline{X}) \; . \tag{5.1.8}$$

Recall that a submanifold N is called *totally geodesic* if any geodesic initially tangent to N remains entirely in N. In a manner analogous to the second fundamental form for non-characteristic hypersurfaces, B measure deviation from geodeticity:

PROPOSITION 5.1.3 ([76, Theorem 30]) *A null hypersurface $\mathcal{H}$ is totally geodesic if and only if B vanishes identically.*

PROOF: The following simple proof has been shown to us by J. Jezierski: Let X and Y be vector fields tangent to $\mathcal{H}$, by (5.1.1) the vector field $\nabla_X Y$ will be tangent to $\mathcal{H}$ if and only if the scalar product $\langle \nabla_X Y, K \rangle$ vanishes. Now,

$$\langle \nabla_X Y, K \rangle = X(\underbrace{\langle Y, K \rangle}_{=0}) - \langle Y, \nabla_X K \rangle = -B(\overline{X}, \overline{Y}) \; ,$$

which shows that ∇ defines a connection on $\mathcal{H}$ in a natural way when B vanishes. Let us denote by $\tilde{\nabla}$ this connection, and let γ be a geodesic of $\tilde{\nabla}$, thus γ is a curve lying on $\mathcal{H}$ such that $\tilde{\nabla}_{\dot\gamma}\dot\gamma$ is proportional to γ. Since $\nabla_{\dot\gamma}\dot\gamma = \tilde{\nabla}_{\dot\gamma}\dot\gamma$, the vector $\nabla_{\dot\gamma}\dot\gamma$ is also proportional to γ, thus γ is also a geodesic of ∇. Uniqueness of the Cauchy problem for the geodesic equation yields the result. □

The *null mean curvature* of $\mathcal{H}$ with respect to K is the continuous scalar field $\theta \in C^0(\mathcal{H})$ defined by

$$\theta = \operatorname{tr} b \; ; \tag{5.1.9}$$

in the general relativity literature θ is often referred to as the *convergence* or *divergence* of the horizon. Let $e_1, e_2, \ldots, e_{n-1}$ be $n-1$ orthonormal spacelike vectors (with respect to $\langle \, , \rangle$) tangent to $\mathcal{H}$ at p. Then $\{\overline{e}_1, \overline{e}_2, \ldots, \overline{e}_{n-1}\}$ is an orthonormal basis (with respect to h) of $T_p\mathcal{H}/K$. Hence at p,

$$\begin{aligned} \theta &= \operatorname{tr} b = \textstyle\sum_{i=1}^{n-1} h(b(\overline{e}_i), \overline{e}_i) \\ &= \textstyle\sum_{i=1}^{n-1} \langle \nabla_{e_i} K, e_i \rangle. \end{aligned} \tag{5.1.10}$$

REMARK 5.1.4 A more direct, equivalent way of introducing θ is as follows: Let K be any vector field along $\mathcal{H}$ tangent to the null generators of $\mathcal{H}$, and let, as above, $e_1, e_2, \ldots, e_{n-1}$ be $n-1$ orthonormal spacelike vectors (with respect to $\langle \, , \rangle$) tangent to $\mathcal{H}$ at p. Then one can use the second line in (5.1.10) to define θ. The calculations above show that the definition is independent of the choice of the e_i's.

Let Σ be the intersection, transverse to K, of a hypersurface in M with $\mathcal{H}$. Then Σ is a C^2 $(n-1)$ dimensional spacelike submanifold of M contained in $\mathcal{H}$ which meets K orthogonally. From Equation (5.1.10),$^{\bullet 5.1.2}$

•5.1.2: **ptc:***to jest jakas bzura bo dywergencja to jest dla wektorow stycznych do sigma i pochodnej kowariantnej innej wiec albo to trzeba wytlumaczyc albo wyrzucic*

$$\theta|_\Sigma = \mathrm{div}_\Sigma K$$

and hence the null mean curvature gives a measure of the divergence of the null generators of $\mathcal{H}$. Note that

$$\boxed{\text{if } \widetilde{K} = fK \text{ then } \widetilde{\theta} = f\theta.}$$

Thus the null mean curvature inequalities $\theta \geq 0$, $\theta \leq 0$, are invariant under positive scaling of K. In Minkowski space, a future null cone $\mathcal{H} = \partial I^+(p) - \{p\}$ (respectively, past null cone $\mathcal{H} = \partial I^-(p) - \{p\}$) has positive null mean curvature, $\theta > 0$ (respectively, negative null mean curvature, $\theta < 0$); this is most easily seen using (5.1.27) below.

The null second fundamental form of a null hypersurface obeys a well-defined comparison theory roughly similar to the comparison theory satisfied by the second fundamental forms of a family of parallel spacelike hypersurfaces (*cf.* Eschenburg [41]).

Let $\eta\,(a,b) \to M$, $s \to \eta(s)$, be a future directed affinely parameterized null geodesic generator of $\mathcal{H}$. For each $s \in (a,b)$, let

$$b(s) = b_{\eta'(s)} : T_{\eta(s)}\mathcal{H}/\eta'(s) \to T_{\eta(s)}\mathcal{H}/\eta'(s)$$

be the Weingarten map based at $\eta(s)$ with respect to the null vector $K = \eta'(s)$. We wish to derive the equation satisfied by b, this needs an appropriate notion of covariant derivative. Let $s \to V(s)$ be a $T\mathcal{H}/K$–vector field along η, *i.e.*, for each s, $V(s)$ is an element of $T_{\eta(s)}\mathcal{H}/K$. Say $s \to V(s)$ is smooth if (at least locally) there is a smooth — in the usual sense — vector field $s \to Y(s)$ along η such that $V(s) = \bar{Y}(s)$ for each s. Then the covariant derivative of $s \to V(s)$ along η can be defined by setting

$$V'(s) = \overline{Y'(s)}$$

where Y' is the usual covariant derivative. Since $\nabla_K K$ is proportional to K, and $\dot\eta$ is also proportional to K, we have

$$\nabla_{\dot\eta}(Y + \mu K) = \nabla_{\dot\eta} Y \mod K\ ,$$

so that that V' so defined is independent of the choice of Y. This definition applies in particular to the equivalence class of vector fields $b(\bar{X})$.

We want to show that the null Weingarten map b of a *smooth*[3] null hypersurface $\mathcal{H}$ satisfies a Ricatti equation (*cf.* [16, p. 431]):

$$\boxed{b' + b^2 + R = 0\ .} \tag{5.1.11}$$

[3]In Proposition 5.1.5 below we will show that Equation (5.1.11) still holds for C^2 hypersurfaces.

Here $'$ denotes covariant differentiation in the direction $\eta'(s)$, with η – an affinely parameterized null geodesic generator of $\mathcal{H}$; more precisely, if $X = X(s)$ is a vector field along η tangent to $\mathcal{H}$, then, as required by Leibniz's rule,

$$b'(\overline{X}) = b(\overline{X})' - b(\overline{X'}). \tag{5.1.12}$$

Finally

$$R : T_{\eta(s)}\mathcal{H}/\eta'(s) \to T_{\eta(s)}\mathcal{H}/\eta'(s)$$

is the curvature endomorphism defined by

$$R(\overline{X}) = \overline{R(X,\eta'(s))\eta'(s)}$$

where $(X,Y,Z) \to R(X,Y)Z$ is the Riemann curvature tensor of M (recall, in our conventions, $R(X,Y)Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z$).

We pass now to the proof of Equation (5.1.11). Fix a point $p = \eta(s_0)$, $s_0 \in (a,b)$, on η. On a neighborhood U of p in $\mathcal{H}$ we can scale the null vector field K so that K is a geodesic vector field, $\nabla_K K = 0$, and so that K, restricted to η, is the velocity vector field to η, *i.e.*, for each s near s_0, $K_{\eta(s)} = \eta'(s)$. Let $X \in T_pM$. Shrinking U if necessary, we can extend X to a smooth vector field on U so that

$$[X,K] = \nabla_X K - \nabla_K X = 0 \ .$$

Then,

$$R(X,K)K = \nabla_X\nabla_K K - \nabla_K\nabla_X K - \nabla_{[X,K]}K = -\nabla_K\nabla_K X \ .$$

Hence along η we have,

$$X'' = -R(X,\eta')\eta'$$

(which implies that X, restricted to η, is a Jacobi field along η, compare Section **??**). Thus, from Equation (5.1.12), at the point p we have,

$$\begin{aligned} b'(\overline{X}) &= \overline{\nabla_X K}\,' - b(\overline{\nabla_K X}) = \overline{\nabla_K X}\,' - b(\overline{\nabla_X K}) \\ &= \overline{X''} - b(b(\overline{X})) = -\overline{R(X,\eta')\eta'} - b^2(\overline{X}) \\ &= -R(\overline{X}) - b^2(\overline{X}), \end{aligned} \tag{5.1.13}$$

which establishes Equation (5.1.11).

Equation (5.1.11) leads to the so-called *Raychaudhuri equation*: by taking the trace of (5.1.11) we obtain the following formula for the derivative of the null mean curvature $\theta = \theta(s)$ along η,

$$\boxed{\theta' = -\mathrm{Ric}(\eta',\eta') - \sigma^2 - \tfrac{1}{n-1}\theta^2 \ ,} \tag{5.1.14}$$

where σ, the shear scalar, is the trace of the square of the trace free part of b. This equation shows how the Ricci curvature of spacetime influences the null mean curvature of a null hypersurface.

Some readers might find the above derivation of the Raychaudhuri equation too abstract. A direct way, without invoking quotient spaces, proceeds as follows: Let e_i be as in Remark 5.1.4. The space $\mathrm{Vect}\{e_i\}^\perp$ is a two dimensional Lorentzian space, let f_0 and f_1 be an ON basis there. Replacing f_1 by $-f_1$ if necessary, the null vector K can be written as

$$K = a(f_0 + f_1)\ ,$$

for some $a > 0$. Set

$$L = a^{-1}(f_0 - f_1)\ ,$$

then

$$g(K,K) = g(L,L) = g(K,e_i) = g(L,e_i) = 0\ , \qquad g(K,L) = -2\ . \tag{5.1.15}$$

Whenever needed, we will extend K and L to any vector fields defined in a neighborhood of $\mathcal{H}$, and we will check what objects in the final formulae are independent of the extensions chosen.

We also note the following formula for the contravariant metric $g^\sharp$:

$$\begin{aligned} g^\sharp &= -f_0\otimes f_0 + f_1\otimes f_1 + \sum_{i=1}^{n-1} e_i\otimes e_i \\ &= -\frac{1}{4}\Big(a^{-1}K + aL\Big)\otimes\Big(a^{-1}K + aL\Big) + \frac{1}{4}\Big(a^{-1}K - aL\Big)\otimes\Big(a^{-1}K - aL\Big) + \sum_{i=1}^{n-1} e_i\otimes e_i \\ &= -\frac{1}{2}K\otimes L - \frac{1}{2}L\otimes K + \sum_{i=1}^{n-1} e_i\otimes e_i\ . \end{aligned} \tag{5.1.16}$$

Next, suppose that we use the normalisation $K_\mu = \nabla_\mu\varphi$, where φ is a defining function for $\mathcal{H}$ (compare the proof of Proposition 5.1.1). We then have

$$\nabla_\mu K_\nu = \nabla_\mu\nabla_\nu\varphi = \nabla_\nu\nabla_\mu\varphi = \nabla_\nu K_\mu\ ,$$

so that the tensor of first covariant derivatives of K is symmetric. In general we will have $K_\mu = f\nabla_\mu\varphi$, for some function f, so that

$$\begin{aligned} \nabla_\mu K_\nu &= \nabla_\mu(f\nabla_\nu\varphi) = f\nabla_\mu\nabla_\nu\varphi + \nabla_\mu f\nabla_\nu\varphi = f\nabla_\nu\nabla_\mu\varphi + \nabla_\mu f\nabla_\nu\varphi \\ &= \nabla_\nu(f\nabla_\mu\varphi) - \nabla_\nu f\nabla_\mu\varphi + \nabla_\mu f\nabla_\nu\varphi \\ &= f\nabla_\nu K_\mu + 2\nabla_{[\mu}f\,\nabla_{\nu]}\varphi\ . \end{aligned}$$

Equivalently,

$$\nabla_\mu K_\nu = \nabla_\nu K_\mu + 2b_{[\mu}K_{\nu]} \tag{5.1.17}$$

with $b_\mu = \nabla_\mu \ln f$. (This equation is equivalent to symmetry of ∇K modulo K, that is, symmetry of the null second fundamental form B.) In what follows we will use the geodesic gauge,

$$K^\nu\nabla_\nu K_\mu = 0\ . \tag{5.1.18}$$

This implies, in view of (5.1.17),

$$0 = K^\nu\nabla_\nu K_\mu = \underbrace{K^\nu\nabla_\mu K_\nu}_{=\frac{1}{2}\nabla_\mu(K^\nu K_\nu)=0} -2K^\mu b_{[\mu}K_{\nu]} = -K^\mu b_\mu K_\nu\ ,$$

showing that

$$K^\mu b_\mu = 0\ . \tag{5.1.19}$$

Set

$$a = L^\mu L^\nu \nabla_\mu K_\nu = L^\mu L^\nu \nabla_{(\mu} K_{\nu)} \ , \ a_\nu = L^\mu \nabla_{(\mu} K_{\nu)} + \frac{1}{2} a K_\nu \ , \tag{5.1.20}$$

which implies

$$a_\mu K^\mu = a_\mu L^\mu = 0 \ , \quad L^\mu \nabla_{(\mu} K_{\nu)} = a_\nu - \frac{1}{2} a K_\nu \ . \tag{5.1.21}$$

Finally we define

$$B_{\mu\nu} := \nabla_{(\mu} K_{\nu)} - \frac{1}{4} a K_\mu K_\nu + a_{(\mu} K_{\nu)} \ . \tag{5.1.22}$$

We claim that

$$B_{\mu\nu} K^\mu = B_{\mu\nu} L^\nu = 0 \ . \tag{5.1.23}$$

Indeed, all contractions of $B_{\mu\nu}$ with K vanish because of (5.1.18) and (5.1.19). Further,

$$L^\mu B_{\mu\nu} \quad = \quad \underbrace{L^\mu \nabla_{(\mu} K_{\nu)}}_{= a_\nu - \frac{1}{2} a K_\nu} - \frac{1}{4} a \underbrace{L^\mu K_\mu}_{=-2} K_\nu + \frac{1}{2} \underbrace{L^\mu a_\mu}_{=0} K_\nu + \frac{1}{2} \underbrace{L^\mu K_\mu}_{=-2} a_\nu = 0 \ ,$$

as claimed. $B_{\mu\nu}$ is of course closely related to the abstract Weingarten tensor B previously defined.

From (5.1.17) and (5.1.22) one is led to

$$\nabla_\mu K_\nu = b_{[\mu} K_{\nu]} + B_{\mu\nu} + \frac{1}{4} a K_\mu K_\nu - a_{(\mu} K_{\nu)} \ , \tag{5.1.24}$$

with (see (5.1.19), (5.1.21) and (5.1.23))

$$B_{[\mu\nu]} = B_{\mu\nu} K^\mu = B_{\mu\nu} L^\nu = b_\mu K^\mu = a_\mu K^\mu = a_\mu L^\mu = 0 \ . \tag{5.1.25}$$

All vector fields that we have been using so far, except L^μ, are tangent to $\mathcal{H}$. Thus, the only part of $\nabla_\mu K_\nu$ in the decomposition above which depends upon the extension of K off $\mathcal{H}$ is $L^\mu \nabla_\mu K_\nu$. From (5.1.25) and (5.1.24) we find

$$L^\mu \nabla_\mu K_\nu = L^\mu b_{[\mu} K_{\nu]} - \frac{1}{2} a K_\nu \ ,$$

which shows that neither $B_{\mu\nu}$ nor a_μ depend upon the extension of K chosen.

Let us relate the calculations here to the quotient space description. By definition of the metric h we have

$$h(\bar{e}_i, \bar{e}_j) = g(e_i, e_j) = \delta_{ij} \ .$$

Consider, next, the Weingarten map B defined in (5.1.7): we have

$$B(\bar{e}_i, \bar{e}_j) = g(\nabla_i K, e_j) = B_{ij} \ ,$$

with $B_{\mu\nu}$ as in (5.1.22), because all the remaining factors in (5.1.24) give zero contribution when contracted with two vectors tangent to $\mathcal{H}$. This shows consistency of notation between (5.1.7) and (5.1.22).

It is often convenient to decompose B into a trace-free part $\sigma_{\mu\nu}$ and into its trace θ, using h:

$$\sigma_{\mu\nu} := B_{\mu\nu} - \frac{\theta}{n-1} h_{\mu\nu} \ , \tag{5.1.26}$$

where $h_{\mu\nu}$ is the space-time counterpart of the metric h defined in (5.1.2)

$$h_{\mu\nu} := g_{\mu\nu} + \frac{1}{2}K_\mu L_\nu + \frac{1}{2}K_\nu L_\nu \ .$$

(Note that $h_{\mu\nu}K^\mu = h_{\mu\nu}L^\mu = 0$ by (5.1.15), similarly $h_{\mu\nu}e_i^\mu e_j^\nu = g_{\mu\nu}e_i^\mu e_j^\nu = g(e_i, e_j) = \delta_{ij}$, which shows that $h_{\mu\nu}$ is indeed closely related to h.) The tensor field $\sigma_{\mu\nu}$ is sometimes called the *shear tensor of $\mathcal{H}$*.

Now, it follows from (5.1.24) that

$$\nabla_\mu K^\mu = g^{\mu\nu}B_{\mu\nu}$$

by (5.1.25). Further, by (5.1.16),

$$g^{\mu\nu}B_{\mu\nu} = -\underbrace{K^\mu L^\nu B_{\mu\nu}}_{=0} + \underbrace{\sum_i B_{ii}}_{=\theta} \ ,$$

which gives an alternative formula for θ, in any[4] affine gauge:

$$\theta = \nabla_\mu K^\mu \ . \tag{5.1.27}$$

This equation provides a convenient starting point for the derivation of the Raychaudhuri equation:

$$\begin{aligned}
\frac{d\theta}{ds} &= K^\nu \nabla_\nu \nabla_\mu K^\mu \\
&= K^\nu(\nabla_\mu \nabla_\nu K^\mu + \underbrace{R^\mu{}_{\sigma\nu\mu}}_{=-R_{\sigma\nu}} K^\sigma) \\
&= \nabla_\mu(\underbrace{K^\nu \nabla_\nu K^\mu}_{=0}) - \nabla_\mu K^\nu \nabla_\nu K^\mu - R_{\sigma\nu}K^\nu K^\sigma \ .
\end{aligned}$$

From (5.1.24) and (5.1.26)

$$\nabla_\mu K^\nu \nabla_\nu K^\mu = B_{\mu\nu}B^{\mu\nu} = \sigma_{\mu\nu}\sigma^{\mu\nu} + \frac{\theta^2}{n-1} = |\sigma|_h^2 + \frac{\theta^2}{n-1} \ ,$$

recovering (5.1.14).

---

[4]It should be pointed out that for a timelike or spacelike geodesic γ there is a natural affine parameterization arising from the normalisation $g(\dot\gamma, \dot\gamma) \in \{\pm 1\}$. For null geodesics there is no natural way of getting rid of the freedom of replacing an affine parameter s by αs, where α is a non-zero constant. Therefore there is a one-parameter family of affine gauges on each generator of $\mathcal{H}$. This leads to a family of rescalings $K \to \alpha K$, where α is any function on $\mathcal{H}$ satisfying $K^\mu \nabla_\mu \alpha = 0$, and therefore a family of affine gauges parameterized by such functions.

The above calculations were done assuming that $\mathcal{H}$ is smooth. Now, when $\mathcal{H}$ is only C^2, all we know is that b is a C^0 tensor field so that there is no reason *a priori* that the derivative b' should exist. We shall show now that this derivative does exist, and that b satisfies the expected differential equation. As the function $s \mapsto R_{\eta(s)}$ is C^∞ then the Riccati equation implies that actually the dependence of $b_{\eta(s)}$ on s is C^∞. This will be clear from the proof below for other reasons:•5.1.3

•5.1.3: **ptc:***not clear whether this should be kept; from now on neither proofread nor adapted to the current notation in any case; on the other hand, Proposition 5.1.8 gives a geometrical interpretation of θ*

PROPOSITION 5.1.5 *Let $\mathcal{H}$ be a C^2 null hypersurface in the $(n+1)$ dimensional spacetime (M, g) and let b be the one parameter family of Weingarten maps along an affine parameterized null generator η. Then the covariant derivative b' defined by Equation (5.1.12) exists and satisfies Equation (5.1.11). Similarly, (5.1.14) holds.*

PROOF: Let $\eta\ (a, b) \to \mathcal{H}$ be an affinely parameterized null generator of $\mathcal{H}$. To simplify notation we assume that $0 \in (a, b)$ and choose a C^∞ spacelike hypersurface Σ of M that passes through $p = \eta(0)$ and let $N = \mathcal{H} \cap \Sigma$. Then N is a C^2 hypersurface in Σ. Now let $\tilde{N}$ be a C^∞ hypersurface in Σ so that $\tilde{N}$ has second order contact with N at p. Let $\tilde{K}$ be a smooth null normal vector field along $\tilde{N}$ such that at p, $\tilde{K} = \eta'(0)$. Consider the hypersurface $\tilde{\mathcal{H}}$ obtained by exponentiating normally along $\tilde{N}$ in the direction $\tilde{K}$; by Lemma **??** there are no focal points along η as long as η stays on $\mathcal{H}$. Passing to a subset of $\tilde{N}$ if necessary to avoid cut points, $\tilde{\mathcal{H}}$ will then be a C^∞ null hypersurface in a neighborhood of η. Let $B(s)$ and $\tilde{B}(s)$ be the null second fundamental forms of $\mathcal{H}$ and $\tilde{\mathcal{H}}$, respectively, at $\eta(s)$ in the direction $\eta'(s)$. We claim that $\tilde{B}(s) = B(s)$ for all $s \in (a, b)$. Since the null Weingarten maps $\tilde{b} = \tilde{b}(s)$ associated to $\tilde{B} = \tilde{B}(s)$ satisfy Equation (5.1.11), this is sufficient to establish the lemma.

We first show that $\tilde{B}(s) = B(s)$ for all $s \in [0, c]$ for some $c \in (0, b)$. By restricting to a suitable neighborhood of p we can assume without loss of generality that M is globally hyperbolic. Let $X \in T_p\Sigma$ be the projection of $\eta'(0) \in T_pM$ onto $T_p\Sigma$. By an arbitrarily small second order deformation of $\tilde{N} \subset \Sigma$ (depending on a parameter ϵ in a fashion similar to Equation (**??**)) we obtain a C^∞ hypersurface $\tilde{N}^+_\epsilon$ in Σ which meets N only in the point p and lies to the side of N into which X points. Similarly, we obtain a C^∞ hypersurface $\tilde{N}^-_\epsilon$ in Σ which meets N only in the point p and lies to the side of N into which $-X$ points. Let $\tilde{K}^\pm_\epsilon$ be a smooth null normal vector field along $\tilde{N}^\pm_\epsilon$ which agrees with $\eta'(0)$ at p. By exponentiating normally along $\tilde{N}^\pm_\epsilon$ in the direction $\tilde{K}^\pm_\epsilon$ we obtain, as before, in a neighborhood of $\eta|_{[0,c]}$ a C^∞ null hypersurface $\tilde{\mathcal{H}}^\pm_\epsilon$, for some $c \in (0, b)$. Let $\tilde{B}^\pm_\epsilon(s)$ be the null second fundamental form of $\tilde{\mathcal{H}}^\pm_\epsilon$ at $\eta(s)$ in the direction $\eta'(s)$.

By restricting the size of Σ if necessary we find open sets W, $W^\pm_\epsilon$ in Σ, with $W^-_\epsilon \subset W \subset W^+_\epsilon$, such that $N \subset \partial_\Sigma W$ and $\tilde{N}^\pm_\epsilon \subset \partial_\Sigma W^\pm_\epsilon$. Restricting to a sufficiently small neighborhood of $\eta|_{[0,c]}$, we have $\mathcal{H} \cap J^+(\Sigma) \subset \partial J^+(W)$ and $\tilde{\mathcal{H}}^\pm_\epsilon \cap J^+(\Sigma) \subset \partial J^+(W^\pm_\epsilon)$. Since $J^+(\overline{W^-_\epsilon}) \subset J^+(\overline{W}) \subset J^+(\overline{W^+_\epsilon})$, it follows that $\tilde{\mathcal{H}}^-_\epsilon$ is to the future of $\mathcal{H}$ near $\eta(s)$ and $\mathcal{H}$ is to the future of $\tilde{\mathcal{H}}^+_\epsilon$ near $\eta(s)$, $s \in [0, c]$. Now if two null hypersurfaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are tangent at a point p, and $\mathcal{H}_2$ is to the future of $\mathcal{H}_1$, then the difference of the null second fundamental forms

$B_2 - B_1$ is positive semidefinite at p. We thus obtain $\tilde{B}^-_\epsilon(s) \geq B(s) \geq \tilde{B}^+_\epsilon(s)$. Letting $\epsilon \to 0$, (i.e., letting the deformations go to zero), we obtain $\tilde{B}(s) = B(s)$ for all $s \in [0, c]$. A straightforward continuation argument implies, in fact, that $\tilde{B}(s) = B(s)$ for all $s \in [0, b)$. A similar argument establishes equality for $s \in (a, 0]$. □

In the last result above the hypersurface $\mathcal{H}$ had to be of at least C^2 differentiability class. Now, in our applications we have to consider hypersurfaces $\mathcal{H}$ obtained as a collection of null geodesics normal to a C^2 surface. A naive inspection of the problem at hand shows that such $\mathcal{H}$'s could in principle be of C^1 differentiability only. Let us show that one does indeed have C^2 differentiability of the resulting hypersurface:

PROPOSITION 5.1.6 *Consider a C^{k+1} spacelike submanifold $N \subset M$ of co–dimension two in an $(n + 1)$ dimensional spacetime (M, g), with $k \geq 1$. Let $\mathbf{k}$ be a non-vanishing C^k null vector field along N, and let $\mathcal{U} \subseteq \mathbb{R} \times N \to M$ be the set of points where the function*

$$f(t, p) := \exp_p(t\mathbf{k}(p))$$

is defined. If $f_{(t_0,p_0)*}$ is injective then there is an open neighborhood $\mathcal{O}$ of (t_0, p_0) so that the image $f[\mathcal{O}]$ is a C^{k+1} embedded hypersurface in M.

REMARK 5.1.7 In our application we only need the case $k = 1$. This result is somewhat surprising as the function $p \mapsto \mathbf{k}(p)$ used in the definition of f is only C^k. We emphasize that we are *not* assuming that f is injective. We note that f will not be of C^{k+1} differentiability class in general, which can be seen as follows: Let $t \to r(t)$ be a C^{k+1} curve in the x-y plane of Minkowski 3-space which is *not* of C^{k+2} differentiability class. Let $t \to \mathbf{n}(t)$ be the spacelike unit normal field along the curve in the x-y plane, then $t \to \mathbf{n}(t)$ is C^k and is *not* C^{k+1}. Let $T = (0, 0, 1)$ be the unit normal to the x-y plane. Then $K(t) = \mathbf{n}(t)+T$ is a C^k normal null field along $t \to r(t)$. The normal exponential map $f\ R^2 \to \mathbb{R}^3$ in the direction K is given by $f(s, t) = r(t) + s[\mathbf{n}(t) + T]$, and hence $df/dt = r'(t) + sn'(t)$, showing explicitly that the regularity of f can be no greater than the regularity of $\mathbf{n}(t)$, and hence no greater than the regularity of $r'(t)$.

PROOF: This result is local in N about p_0 so there is no loss of generality, by possibly replacing N by a neighborhood of p_0 in N, in assuming that N is a embedded submanifold of M. The map f is of class C^k and the derivative $f_{(t_0,p_0)*}$ is injective so the implicit function theorem implies $f[\mathcal{U}]$ is a C^k hypersurface near $f(t_0, p_0)$. Let η be any nonzero timelike C^∞ vector field on M defined near p_0 (some restrictions to be put on η shortly) and let Φ_s be the flow of η. Then for sufficiently small ε the map $\tilde{f}\ (-\varepsilon, \varepsilon) \times N \to M$ given by

$$\tilde{f}(s, p) := \Phi_s(p)$$

is injective and of class C^{k+1}. Extend $\mathbf{k}$ to any C^k vector field $\tilde{\mathbf{k}}$ along $\tilde{f}$. (It is not assumed that the extension $\tilde{k}$ is null.) That is $\tilde{\mathbf{k}}\ (-\varepsilon, \varepsilon) \times N \to TM$ is a C^k map and $\tilde{\mathbf{k}}(s,p) \in T_{\tilde{f}(s,p)}M$. Note that we can choose $\tilde{\mathbf{k}}(s,p)$ so that the covariant derivative $\frac{\nabla \tilde{\mathbf{k}}}{\partial s}(0,p_0)$ has any value we wish at the one point $(0,p_0)$. Define a map $F\ (t_0-\varepsilon, t_0+\varepsilon)\times(-\varepsilon,\varepsilon)\times N \to M$ by

$$F(t,s,p) = \exp(t\tilde{\mathbf{k}}(s,p)).$$

We now show that F can be chosen to be a local diffeomorphism near $(t_0,0,p_0)$. Note that $F(t,0,p) = f(t,p)$ and by assumption $f_{*(t_0,p_0)}$ is injective. Therefore the restriction of $F_{*(t_0,0,p_0)}$ to $T_{(t_0,p_0)}(\mathbb{R}\times N) \subset T_{(t_0,0,p_0)}(\mathbb{R}\times\mathbb{R}\times N)$ is injective. Thus by the inverse function theorem it is enough to show that $F_{*(t_0,0,p_0)}(\partial/\partial s)$ is linearly independent of the subspace $F_{*(t_0,0,p_0)}[T_{(t_0,p_0)}(\mathbb{R}\times N)]$. Let

$$V(t) = \frac{\partial F}{\partial s}(t,s,p_0)\bigg|_{s=0}.$$

Then $V(t_0) = F_{*(t_0,0,p_0)}(\partial/\partial s)$ and our claim that F is a local diffeomorphism follows if $V(t_0) \notin F_{*(t_0,0,p_0)}[T_{(t_0,p_0)}(\mathbb{R}\times N)]$. For each s,p the map $t \mapsto F(s,t,p)$ is a geodesic and therefore V is a Jacobi field along $t \mapsto F(0,t,p_0)$. (Those geodesics might change type as s is varied at fixed p_0, but this is irrelevant for our purposes.) The initial conditions of this geodesic are

$$V(0) = \frac{\partial}{\partial s}F(0,s,p_0)\bigg|_{s=0} = \frac{\partial}{\partial s}\Phi_s(p_0)\bigg|_{s=0} = \eta(p_0)$$

and

$$\frac{\nabla V}{\partial t}(0) = \frac{\nabla}{\partial t}\frac{\nabla}{\partial s}F(t,s,p_0)\bigg|_{s=0,t=0} = \frac{\nabla}{\partial s}\frac{\nabla}{\partial t}F(t,s,p_0)\bigg|_{s=0,t=0} = \frac{\nabla\tilde{\mathbf{k}}}{\partial s}(0,p_0)\ .$$

From our set up we can choose $\eta(p_0)$ to be any timelike vector and $\frac{\nabla \tilde{\mathbf{k}}}{\partial s}(0,p_0)$ to be any vector. As the linear map from $T_{p_0}M \times T_{p_0}M \to T_{f(t_0,p_0)}M$ which maps the initial conditions $V(0)$, $\frac{\nabla V}{\partial t}(0)$ of a Jacobi field V to its value $V(t_0)$ is surjective[5] it is an open map. Therefore we can choose $\eta(p_0)$ and $\frac{\nabla \tilde{\mathbf{k}}}{\partial s}(0,p_0)$ so that $V(t_0)$ is not in the nowhere dense set $F_{*(t_0,0,p_0)}[T_{(t_0,p_0)}(\mathbb{R}\times N)]$. Thus we can assume F is a local C^k diffeomorphism on some small neighborhood $\mathcal{A}$ of $(t_0,0,p_0)$ onto a small neighborhood $\mathcal{B} := F[\mathcal{A}]$ of $F(t_0,0,p_0)$ as claimed.

Consider the vector field $F_*(\partial/\partial t) = \partial F/\partial t$ along F. Then the integral curves of this vector field are the geodesics $t \mapsto F(t,s,p) = \exp(t\tilde{\mathbf{k}}(s,p))$. (This is true even when F is not injective on its entire domain.) These geodesics and their velocity vectors depend smoothly on the initial data. In the case at hand the initial data is C^k so $\partial F/\partial t$ is a C^k vector field along F. Therefore the one form α defined by $\alpha(X) := \langle X, \partial F/\partial t\rangle$ on the neighborhood $\mathcal{B}$ of q_0 is C^k. The definition of F implies that $f(t,p) = F(t,0,p)$ and therefore the vector field $\partial F/\partial t$ is tangent to $f[\mathcal{O}]$ and the null geodesics $t \mapsto f(t,p) = F(t,0,p)$ rule $f[\mathcal{O}]$ so that $f[\mathcal{O}]$ is a null hypersurface. Therefore for any vector X tangent

[5] If $v \in T_{f(t_0,p_0)}N$ there is a Jacobi field with $V(t_0) = v$ and $\frac{\nabla V}{\partial t}(t_0) = 0$, which implies subjectivity.

to $f[\mathcal{O}]$ we have $\alpha(X) = \langle X, \partial F/\partial t\rangle = 0$. Thus $f[\mathcal{O}]$ is an integral submanifold for the distribution $\{X \mid \alpha(X) = 0\}$ defined by α. But, as is easily seen by writing out the definitions in local coordinates, an integral submanifold of a C^k distribution is a C^{k+1} submanifold. (Note that in general there is no reason to believe that the distribution defined by α is integrable. However, we have shown directly that $f[\mathcal{O}]$ is an integral submanifold of that distribution.) □

We shall close this appendix with a calculation, needed in the main body of the paper, concerning Jacobians. Let us start by recalling the definition of the Jacobian needed in our context. Let $\phi\ M \to N$ be a C^1 map between Riemannian manifolds, with $\dim M \le \dim N$. Let $n = \dim M$ and let $e_1, \dots, e_n$ be an orthonormal basic of T_pM then the Jacobian of ϕ at p is $J(\phi)(p) = \|\phi_{*p}e_1 \wedge \phi_{*p}e_2 \wedge \cdots \wedge \phi_{*p}e_n\|$. When $\dim M = \dim N$ and both M and N are oriented with ω_M being the volume form on M, and ω_N being the volume form on N, then $J(\phi)$ can also be described as the positive scalar satisfying: $\phi^*(\omega_N) = \pm J(\phi)\,\omega_M$.

Let S be a C^2 co-dimension two acausal spacelike submanifold of a smooth spacetime M, and let K be a past directed C^1 null vector field along S. Consider the normal exponential map in the direction K, $\Phi\ \mathbb{R} \times S \to M$, defined by $\Phi(s, x) = \exp_x sK$. (Φ need not be defined on all of $\mathbb{R} \times S$.) Suppose the null geodesic $\eta\ s \to \Phi(s, p)$ meets a given acausal spacelike hypersurface Σ at $\eta(1)$. Then there is a neighborhood W of p in S such that each geodesic $s \to \Phi(s, x)$, $x \in W$ meets Σ, and so determines a C^1 map $\phi\, W \to \Sigma$, which is the projection into Σ along these geodesics. Let $J(\phi)$ denote the Jacobian determinant of ϕ at p. $J(\phi)$ may be computed as follows. Let $\{X_1, X_2, \dots, X_k\}$ be an orthonormal basis for the tangent space T_pS. Then ,

$$J(\phi) = \|\phi_{*p}X_1 \wedge \phi_{*p}X_2 \wedge \cdots \wedge \phi_{*p}X_k\|.$$

Suppose there are no focal points to S along $\eta|_{[0,1]}$. Then by shrinking W and rescaling K if necessary, $\Phi\,[0, 1] \times W \to M$ is a C^1 embedded null hypersurface N such that $\Phi(\{1\} \times W) \subset \Sigma$. Extend K to be the C^1 past directed null vector field, $K = \Phi_*(\frac{\partial}{\partial s})$ on N. Let $\theta = \theta(s)$ be the null mean curvature of N with respect to $-K$ along η. For completeness let us give a proof of the following, well known result:

Proposition 5.1.8 *With $\theta = \theta(s)$ as described above,*

1. *If there are no focal points to S along $\eta|_{[0,1]}$, then*

$$J(\phi) = \exp\left(-\int_0^1 \theta(s)ds\right) . \tag{5.1.28}$$

2. *If $\eta(1)$ is the first focal point to S along $\eta|_{[0,1]}$, then*

$$J(\phi) = 0 .$$

Remark 5.1.9 In particular, if N has nonnegative null mean curvature with respect to the future pointing null normal, *i.e.*, if $\theta \ge 0$, we obtain that $J(\phi) \le 1$.

REMARK 5.1.10 Recall that θ was only defined when a normalization of K has been chosen. We stress that in (5.1.28) that normalization is so that K is tangent to an affinely parameterized geodesic, with s being an affine distance along η, and with p corresponding to $s = 0$ and $\phi(p)$ corresponding to $s = 1$.

PROOF: 1. To relate $J(\phi)$ to the null mean curvature of N, extend the orthonormal basis $\{X_1, X_2, \dots, X_k\}$ to Lie parallel vector fields $s \to X_i(s)$, $i = 1, \dots, k$, along η, $\mathcal{L}_K X_i = 0$ along η. Then by a standard computation,

$$\begin{aligned} J(\phi) &= \|\phi_{*p}X_1 \wedge \phi_{*p}X_2 \wedge \cdots \wedge \phi_{*p}X_k\| \\ &= \|X_1(1) \wedge X_2(1) \wedge \cdots \wedge X_k(1)\| \\ &= \sqrt{g}\,\Big|_{s=1}\,, \end{aligned}$$

where $g = \det[g_{ij}]$, and $g_{ij} = g_{ij}(s) = \langle X_i(s), X_j(s)\rangle$. We claim that along η,

$$\theta = -\frac{1}{\sqrt{g}}\frac{d}{ds}\sqrt{g}\,.$$

The computation is standard. Set $b_{ij} = B(\overline{X}_i, \overline{X}_j)$, where B is the null second fundamental form of N with respect to $-K$, $h_{ij} = h(\overline{X}_i, \overline{X}_j) = g_{ij}$, and let g^{ij} be the i, jth entry of the inverse matrix $[g_{ij}]^{-1}$. Then $\theta = g^{ij}b_{ij}$. Differentiating g_{ij} along η we obtain,

$$\begin{aligned} \frac{d}{ds}g_{ij} = K\langle X_i, X_j\rangle &= \langle \nabla_K X_i, X_j\rangle + \langle X_i, \nabla_K X_j\rangle \\ &= \langle \nabla_{X_i} K, X_j\rangle + \langle X_i, \nabla_{X_j} K\rangle \\ &= -(b_{ij} + b_{ji}) = -2b_{ij}\,. \end{aligned}$$

Thus,

$$\theta = g^{ij}b_{ij} = -\frac{1}{2}g^{ij}\frac{d}{ds}g_{ij} = -\frac{1}{2}\,\frac{1}{g}\frac{dg}{ds} = -\frac{1}{\sqrt{g}}\frac{d}{ds}\sqrt{g}\,,$$

as claimed. Integrating along η from $s = 0$ to $s = 1$ we obtain,

$$J(\phi) = \sqrt{g}\,\Big|_{s=1} = \sqrt{g}\,\Big|_{s=0} \cdot \exp\left(-\int_0^1 \theta\, ds\right) = \exp\left(-\int_0^1 \theta\, ds\right).$$

2. Suppose now that $\eta(1)$ is a focal point to S along η, but that there are no focal points to S along η prior to that. Then we can still construct the C^1 map $\Phi\ [0,1] \times W \to M$, with $\Phi(\{1\} \times W) \subset \Sigma$, such that Φ is an embedding when restricted to a sufficiently small open set in $[0,1] \times W$ containing $[0,1) \times \{p\}$. The vector fields $s \to X_i(s)$, $s \in [0,1)$, $i = 1, .., k$, may be constructed as above, and are Jacobi fields along $\eta|_{[0,1)}$, which extend smoothly to $\eta(1)$. Since $\eta(1)$ is a focal point, the vectors $\phi_* X_1 = X_1(1)$, $\dots$, $\phi_* X_k = X_k(1)$ must be linearly dependent, which implies that $J(\phi) = 0$. $\square$

5.2 Galloway's null splitting theorem

5.3 Timelike splitting theorems

In Riemannian geometry *a line* is defined to be a complete geodesic which is minimising between each pair of its points. A milestone theorem of Cheeger and Gromoll [21] asserts that *a complete Riemannian manifold* (M, g)*, with non-negative Ricci curvature, which contains a line splits as a metric product*

$$M = \mathbb{R} \times N \ , \qquad g = dx^2 + h \ , \tag{5.3.1}$$

where x is a coordinate along the $\mathbb{R}$ factor, while h is an (x-independent) complete metric on N. This result is known under the name of *Cheeger-Gromoll splitting theorem*.

It turns out that there is a corresponding result in Lorentzian geometry, with obvious modifications: First, a line is defined by changing "minimising" to "maximising" in the definition above. Next, in the definition of "splitting" one replaces (5.3.1) with

$$M = \mathbb{R} \times N \ , \qquad g = -dt^2 + h \ , \tag{5.3.2}$$

where we use now t to denote the coordinate along the $\mathbb{R}$ factor. One has the following:

THEOREM 5.3.1 *Let* $(\mathscr{M}, g)$ *be a space-time satisfying the* timelike focusing condition*,*

$$\mathrm{Ric}(X, X) \geq 0 \quad \forall\ X \ \textit{timelike}\ . \tag{5.3.3}$$

Suppose that $\mathscr{M}$ *contains a line and either*

1. $(\mathscr{M}, g)$ *is globally hyperbolic, or*

2. $(\mathscr{M}, g)$ *is timelike geodesically complete.*

Then $(\mathscr{M}, g)$ *splits as in (5.3.2), for some complete metric* h *on* N*.*

REMARK 5.3.2 The "geodesically complete version" of Theorem 5.3.1 was known as *Yau's splitting conjecture* before its proof by Newman [86]. The globally hyperbolic version was proved by Galloway [51]. The result assuming both timelike geodesic completeness and global hyperbolicity had previously been established by Eschenburg [42]. The proof•5.3.1 below is a simplified version pointed out to us by Galloway, based on the analysis in [**?**] and the results in [5]. •*5.3.1:* **ptc:***proof or outline of the proof?*

PROOF: •5.3.2 •*5.3.2:* **ptc:***give proof*

Chapter 6

The Einstein equations

6.1 The nature of the Einstein equations

The *vacuum Einstein equations with cosmological constant* Λ read

$$G_{\alpha\beta} + \Lambda g_{\alpha\beta} = 0 \,, \tag{6.1.1}$$

where $G_{\alpha\beta}$ is the Einstein tensor defined in (2.3.13). We will sometimes refer to those equations as *the vacuum Einstein equations*, regardless of whether or not the cosmological constant vanishes, in situations in which the non-vanishing of Λ is irrelevant for the discussion at hand.•6.1.1 Taking the trace of (6.1.1) one obtains •6.1.1: cos tu jest bez sensu

$$R = \frac{2(n+1)}{n-1}\Lambda \,, \tag{6.1.2}$$

where, as elsewhere, $n+1$ is the dimension of space-time. This leads to the following equivalent version of (6.1.1):

$$\mathrm{Ric} = \frac{2\Lambda}{n-1} g \,. \tag{6.1.3}$$

Thus the Ricci tensor of the metric is proportional to the metric. Pseudo-Lorentzian manifolds the metric of which satisfies Equation (6.1.3) are called *Einstein manifolds* in the mathematical literature.

Given a manifold $\mathscr{M}$, Equation (6.1.1) or, equivalently, Equation (6.1.3) forms a system of partial differential equations for the metric; more precisely, it follows from Equations (2.3.11) and (2.3.1) that the Ricci tensor is an object built out of the Christoffel symbols and their first derivatives, while Equation (2.3.8) shows that the Christoffel symbols are built out of the metric and its first derivatives. The same equations show that the Ricci tensor is linear in the second derivatives of the metric, with coefficients which are rational functions of the $g_{\alpha\beta}$'s, and quadratic in the first derivatives of g, again with coefficients rational in g. Equations linear in the highest order derivatives are called *quasi-linear*, hence the vacuum Einstein equations constitute a second order system of quasi-linear partial differential equations for the metric g.

In the discussion above we have assumed that the manifold $\mathscr{M}$ has been given. Such a point of view might seem to be too restrictive, and sometimes it is argued that the Einstein equations should be interpreted as equations both for the metric and the manifold. The sense of such a statement is far from being clear, one possibility of understanding that is that the manifold arises as a result of the evolution of the metric g. We are going to discuss in detail the evolution point of view below, let us, however, anticipate somewhat and mention the following: vacuum space-times constructed by evolution of initial data all have topology $\mathbb{R}\times\mathscr{S}$, where $\mathscr{S}$ is the n-dimensional manifold on which the initial data have been prescribed. Thus, from the Cauchy problem point of view the resulting space-times (as defined precisely by Theorem **??** below) have topology and differential structure which are determined by the initial data. As will be discussed in more detail in Section 6.11, the space-times obtained by evolution of the data are sometimes extendible; now, the author is not aware of any conditions which would guarantee uniqueness of the extensions which arise in this way, and therefore it does not seem useful to consider the Einstein equations as equations determining the manifold in those cases. We conclude that in the evolutionary point of view the manifold can be also thought as being given *a priori* — namely $\mathscr{M} = \mathbb{R}\times\mathscr{S}$ — even though there is no *a priori* known natural time coordinate which can be constructed by evolutionary methods, and which leads to the decomposition $\mathscr{M} = \mathbb{R}\times\mathscr{S}$.

Let us pass now to the derivation of a somewhat more explicit form of the Einstein equations. In index notation Equation (2.3.1) takes the form

$$\nabla_\mu\nabla_\nu X^\alpha - \nabla_\nu\nabla_\mu X^\alpha = R^\alpha{}_{\beta\mu\nu}X^\beta \ . \tag{6.1.4}$$

A contraction over α and μ gives

$$\nabla_\alpha\nabla_\nu X^\alpha - \nabla_\nu\nabla_\alpha X^\alpha = R_{\beta\nu}X^\beta \ . \tag{6.1.5}$$

Suppose that X is the gradient of a function ϕ, $X = \nabla\phi$, then we have

$$\nabla_\alpha X^\beta = \nabla_\alpha\nabla^\beta\phi = \nabla^\beta\nabla_\alpha\phi \ ,$$

because of the symmetry of second partial derivatives. Further

$$\nabla_\alpha X^\alpha = \Box_g\phi \ ,$$

where we use the symbol

$$\Box_k \equiv \nabla_\mu\nabla^\mu$$

to denote the wave operator associated with a Lorentzian metric k; *e.g.*, for a scalar field we have

$$\Box_k\phi = \nabla_\mu\nabla^\mu\phi = \frac{1}{\sqrt{-\det g_{\alpha\beta}}}\partial_\mu(\sqrt{-\det g_{\rho\sigma}}g^{\mu\nu}\partial_\nu\phi) \ . \tag{6.1.6}$$

For such vector fields Equation (6.1.5) can be rewritten as

$$\nabla_\alpha\nabla^\alpha\nabla_\nu\phi - \nabla_\nu\nabla_\alpha\nabla^\alpha\phi = R_{\beta\nu}\nabla^\beta\phi \ ,$$

or, equivalently,

$$\Box_g d\phi - d(\Box_g\phi) = \mathrm{Ric}(\nabla\phi,\cdot) \ , \tag{6.1.7}$$

where d denotes exterior differentiation. Consider Equation (6.1.7) with ϕ replaced by y^A, where y^A is any collection of functions,

$$\Box_g dy^A = d\lambda^A + \mathrm{Ric}(\nabla y^A, \cdot) \,, \tag{6.1.8}$$

$$\lambda^A \equiv \Box_g y^A \,. \tag{6.1.9}$$

•6.1.2 Set

$$g^{AB} \equiv g(dy^A, dy^B) \,; \tag{6.1.10}$$

•6.1.2: nie wiem dlaczego sie nie zgadza znak; zdecydowac sie jak pisac metryke odwrotna

this is consistent with Equation (2.3.7) except that we haven't assumed yet that the y^A's form a coordinate system. By the chain rule we have

$$\begin{aligned}
\Box_g g^{AB} &= \nabla_\mu \nabla^\mu (g(dy^A, dy^B)) \\
&= \nabla_\mu (g(\nabla^\mu dy^A, dy^B) + g(dy^A, \nabla^\mu dy^B)) \\
&= g(\Box_g dy^A, dy^B) + g(dy^A, \Box_g dy^B) + 2g(\nabla_\mu dy^A, \nabla^\mu dy^B) \\
&= g(d\lambda^A, dy^B) + g(dy^A, d\lambda^B) + 2g(\nabla_\mu dy^A, \nabla^\mu dy^B) \\
&\quad +2\mathrm{Ric}(\nabla y^A, \nabla y^B) \,.
\end{aligned} \tag{6.1.11}$$

Let us suppose that the functions y^A solve the homogeneous wave equation:

$$\lambda^A = \Box_g y^A = 0 \,. \tag{6.1.12}$$

The Einstein equation (2.3.13) implies then

$$E^{AB} \equiv \Box_g g^{AB} - 2g(\nabla_\mu dy^A, \nabla^\mu dy^B) - \frac{4\Lambda}{n-1} g^{AB} \tag{6.1.13a}$$

$$= 0 \,. \tag{6.1.13b}$$

Now,

$$\begin{aligned}
\nabla_\mu (dy_A) &= \nabla_\mu (\partial_\nu y^A \, dx^\nu) \\
&= (\partial_\mu \partial_\nu y^A - \Gamma^\sigma{}_{\mu\nu} \partial_\sigma y^A) dx^\nu \,.
\end{aligned} \tag{6.1.14}$$

Suppose that the $d\phi^A$'s are linearly independent and form a basis of $T^*\mathscr{M}$, then Equation (6.1.13b) is *equivalent* to (2.3.13). Further we can choose the y^A's as coordinates, at least on some open subset of $\mathscr{M}$; in this case we have

$$\partial_A y^B = \delta^B_A \,, \qquad \partial_A \partial_C y^B = 0 \,,$$

so that Equation (6.1.14) reads

$$\nabla_B dy^A = -\Gamma^A{}_{BC} dy^C \,.$$

This, together with (6.1.13b), leads to

$$\Box_g g^{AB} - 2g^{CD} g^{EF} \Gamma^A_{CE} \Gamma^B_{DF} - \frac{4\Lambda}{n-1} g^{AB} = 0 \,. \tag{6.1.15}$$

Here the Γ^A_{BC}'s should be calculated in terms of the g_{AB}'s and their derivatives as in Equation (2.3.8), and the wave operator $\Box_g$ is understood as acting on scalars. We have thus shown that *in "wave coordinates" the Einstein equation forms a second-order quasi-linear wave-type system of equations* (6.1.15) for the metric functions g^{AB}. This establishes the *hyperbolic*, evolutionary character of the Einstein equations.

It turns out that (6.1.13b) allows one also to *construct* solutions of Einstein equations [45], this will be done in the following sections.

Before analyzing the existence question, it is natural to ask the following question: given a solution of the Einstein equations, can one always find local coordinate systems y^A satisfying the wave condition (6.1.12)? The answer is yes, the standard way of obtaining such functions proceeds as follows: Let $\mathscr{S}$ be any spacelike hypersurface in $\mathscr{M}$; by definition, the restriction of the metric g to $T\mathscr{S}$ is positive non-degenerate. Let $\mathscr{O} \subset \mathscr{S}$ be any open subset of $\mathscr{S}$, and let X be any smooth vector field on $\mathscr{M}$, defined along $\mathscr{O}$, which is transverse to $\mathscr{S}$; by definition, this means that for each $p \in \mathscr{O}$ the tangent space $T_p\mathscr{M}$ is the direct sum of $T_p\mathscr{S}$ and of the linear space $\mathbb{R}X(p)$ spanned by $X(p)$. (Any timelike vector X would do — *e.g.*, the unit normal to $\mathscr{S}$ — but transversality is sufficient for our purposes here.) The following result is standard:•6.1.3

•*6.1.3: cos z tym trzeba zrobic, albo np dac referencje?*

THEOREM 6.1.1 *For any smooth functions f, g on $\mathscr{O} \subset \mathscr{S}$ there exists a unique smooth solution ϕ defined on $\mathscr{D}(\mathscr{O})$ of the problem*

$$\Box_g \phi = 0 \,, \qquad \phi|_{\mathscr{O}} = f \,, \quad X(\phi)|_{\mathscr{O}} = g \,.$$

Once a hypersurface $\mathscr{S}$ has been chosen, *local wave coordinates adapted to $\mathscr{S}$* may be constructed as follows: Let $\mathscr{O}$ be any coordinate patch on $\mathscr{S}$ with coordinate functions x^i, $i = 1, \ldots, n$, and let e^0 be the field of unit future pointing normals to $\mathscr{O}$. On $\mathscr{D}(\mathscr{O})$ define the y^A's to be the unique solutions of the problem

$$\Box_g y^A = 0 \,,$$
$$y^0|_{\mathscr{O}} = 0 \,, \quad e^0(y^0)|_{\mathscr{O}} = 1 \,, \tag{6.1.16}$$
$$y^i|_{\mathscr{O}} = x^i \,, \quad e^0(y^i)|_{\mathscr{O}} = 0 \,, \quad i = 1, \ldots, n \,. \tag{6.1.17}$$

We note that there is a considerable freedom in the construction of the y^i's (because of the freedom of choice of the x^i's), but the function y^0 is defined uniquely by $\mathscr{S}$. Since the x^i's form a coordinate system on $\mathscr{O}$, a simple application of the implicit function theorem shows that there exists a neighborhood $\mathscr{U} \subset \mathscr{D}(\mathscr{O})$ of $\mathscr{O}$ which is coordinatized by the y^A's.•6.1.4

•*6.1.4:* **ptc:***all this needs cleaning up*

6.2 Existence local in time in wave coordinates

Let us return to (6.1.11). Assume again that the y^A's form a local coordinate system, but do not assume for the moment that the y^A's solve the wave equation. In that case (6.1.11) together with the definition (6.1.13a) of E^{AB} lead to•6.2.1

•*6.2.1:* **ptc:***is there not a sign wrong in front of the E piece?*

$$R^{AB} = \frac{1}{2}E^{AB} - \nabla^A \lambda^B - \nabla^B \lambda^A + \frac{2\Lambda}{n-1} g^{AB} \,. \tag{6.2.1}$$

The idea, due to Yvonne Choquet-Bruhat [45], is to use the hyperbolic character of the equation $E^{AB} = 0$ to construct a metric g. If we manage to make sure that λ^A vanishes as well, it will then follow from (6.2.1) then g will also solve the Einstein equation. The following result is again standard:•6.2.2

•*6.2.2:* **ptc:***give ref*

THEOREM 6.2.1 *For any initial data*

$$g^{AB}(y^i,0)\in H^{k+1}\ ,\quad \partial_0 g^{AB}(y^i,0)\in H^k\ ,\qquad k>n/2\ , \tag{6.2.2}$$

prescribed on an open subset $\mathscr{O}\subset\{0\}\times\mathbb{R}^n\subset\mathbb{R}\times\mathbb{R}^n$ there exists a unique solution g^{AB} defined on an open neighborhood $\mathscr{U}\subset\mathbb{R}\times\mathbb{R}^n$ of $\mathscr{O}$, satisfying

$$E^{AB}=0\ . \tag{6.2.3}$$

The set $\mathscr{U}$ can be chosen so that g^{AB} defines a Lorentzian metric, with $(\mathscr{U},g)$ — globally hyperbolic with Cauchy surface $\mathscr{O}$.

REMARK 6.2.2 The recent results of [**?**, 102] allow one to reduce the differentiability threshold above.•6.2.3

•6.2.3: **ptc:***do something about it*

It remains to find out how to ensure the conditions (6.1.12). The key observation of Yvonne Choquet-Bruhat is that (6.2.3) and the Bianchi identities imply a wave equation for λ^A's. In order to see that, recall that it follows from the Bianchi identities that the Ricci tensor of the metric g necessarily satisfies a divergence identity:

$$\nabla_A\Big(R^{AB}-\frac{R}{2}g^{AB}\Big)=0\ .$$

Assuming that (6.2.3) holds, (6.2.1) implies then

$$\begin{aligned}0 &= -\nabla_A\Big(\nabla^A\lambda^B+\nabla^B\lambda^A-\nabla_C\lambda^C g^{AB}\Big)\\ &= -\Big(\Box\lambda^B+\nabla_A\nabla^B\lambda^A-\nabla^B\nabla_C\lambda^C\Big)\\ &= -\Big(\Box\lambda^B+R^B{}_A\lambda^A\Big)\ .\end{aligned}$$

This shows that λ^A necessarily satisfies the second order hyperbolic system of equations

$$\Box\lambda^B+R^B{}_A\lambda^A=0\ .$$

Now, it is a standard fact in the theory of hyperbolic equations that we will have

$$\lambda^A\equiv 0$$

on the domain of dependence $\mathscr{D}(\mathscr{O})$ provided that both λ^A and its derivatives vanish at $\mathscr{O}$.

From now on we shall assume that y^0 is the coordinate along the $\mathbb{R}$ factor of $\mathbb{R}\times\mathbb{R}^n$, so that the initial data surface $\{0\}\times\mathscr{O}$ is given by the equation $y^0=0$. We have

$$\begin{aligned}\Box y^A &= \frac{1}{\sqrt{|\det g|}}\partial_B\Big(\sqrt{|\det g|}g^{BC}\partial_C y^A\Big)\\ &= \frac{1}{\sqrt{|\det g|}}\partial_B\Big(\sqrt{|\det g|}g^{BA}\Big)\ .\end{aligned}$$

Clearly a necessary condition for the vanishing of $\Box y^A$ is that it vanishes at $y^0 = 0$, and this allows us to calculate some time derivatives of the metric in terms of space ones:

$$\partial_0\Big(\sqrt{|\det g|}g^{0A}\Big) = -\partial_i\Big(\sqrt{|\det g|}g^{iA}\Big) . \tag{6.2.4}$$

This implies that the initial data (6.2.2) for the equation (6.2.3) cannot be chosen arbitrarily if we want both (6.2.3) and the Einstein equation to be simultaneously satisfied. It turns out that further constraints arise from the requirement of the vanishing of the derivatives of λ. Supposing that (6.2.4) holds at $y^0 = 0$ — equivalently, supposing that λ vanishes on $\{y^0 = 0\}$, we then have

$$\partial_i \Lambda^A = 0$$

on $\{y^0 = 0\}$, where the index i is used to denote tangential derivatives. In order that all derivatives vanish initially it remains to ensure that some transverse derivative does. A convenient transverse direction is provided by the field n of unit timelike normals to $\{y^0 = 0\}$, and the vanishing of $\nabla_n\lambda$ is conveniently expressed as

$$\Big(G_{\mu\nu} + \Lambda g_{\mu\nu}\Big)n^\mu = 0 . \tag{6.2.5}$$

In order to see that it is most convenient to use an ON frame e_μ, with $e_0 = n$.
•6.2.4: **ptc:***crosscheck* It follows from the equation $E_{AB} = 0$ and (6.2.1) that•6.2.4

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = -\Big(\nabla_\mu\lambda_\nu + \nabla_\nu\lambda_\mu - \nabla^\alpha\lambda_\alpha g_{\mu\nu}\Big) ,$$

which gives

$$\begin{aligned}
-\Big(G_{\mu\nu} + \Lambda g_{\mu\nu}\Big)n^\mu n^\nu &= 2\nabla_0\lambda_0 - \nabla^\alpha\lambda_\alpha \underbrace{g_{00}}_{=-1} \\
&= 2\nabla_0\lambda_0 + (-\nabla_0\lambda_0 + \underbrace{\nabla_i\lambda_i}_{=0}) \\
&= \nabla_0\lambda_0 ,
\end{aligned} \tag{6.2.6}$$

which shows that the vanishing of $\nabla_0\lambda_0$ is equivalent to the vanishing of the $\mu = 0$ component in (6.2.5). Finally

$$\begin{aligned}
-\Big(G_{i0} + \Lambda g_{i0}\Big) &= \underbrace{\nabla_i\lambda_0}_{=0} + \nabla_0\lambda_i - \nabla^\alpha\lambda_\alpha \underbrace{g_{i0}}_{=0} \\
&= \nabla_0\lambda_i ,
\end{aligned} \tag{6.2.7}$$

as desired. Equations (6.2.5) are called the *general relativistic constraint equations*. We will shortly see that (6.2.4) has quite a different character from (6.2.5); the former will be referred to as a *gauge equation*.

In conclusion, in the wave gauge $\lambda^A = 0$ the Cauchy data for the vacuum Einstein equations consist of

1. An open subset $\mathscr{O}$ of $\mathbb{R}^n$,

2. together with matrix-valued functions g^{AB}, $\partial_0 g^{AB}$ prescribed there, so that g^{AB} is symmetric with signature $(-,+,\cdots,+)$ at each point.

3. The constraint equations (6.2.5) hold, and

4. the algebraic gauge equation (6.2.4) holds.

6.3 The geometry of spacelike submanifolds

Let $\mathscr{S}$ be a hypersurface in a space-time $(\mathscr{M},g)$, we want to analyze the geometry of such hypersurfaces. Set

$$h := g|_{T\mathscr{S}} \; . \tag{6.3.1}$$

More precisely,

$$\forall\; X,Y \in T\mathscr{S} \qquad h(X,Y) := g(X,Y) \; .$$

The tensor field h is called *the first fundamental form of* $\mathscr{S}$; when non-degenerate, it is also called *the metric induced by* g *on* h. If $\mathscr{S}$ is considered as an abstract manifold with embedding $i : \mathscr{S} \to \mathscr{M}$, then h is simply the pull-back i^*g.

A hypersurface $\mathscr{S}$ will be said to be *spacelike* at $p \in \mathscr{S}$ if h is Riemannian at p, *timelike* at p if h is Lorentzian at p, and finally *null* or *isotropic* or *lightlike* at p if h is degenerate at p. $\mathscr{S}$ will be called *spacelike* if it spacelike at all $p \in \mathscr{S}$, *etc.* An example of null hypersurface is given by $\dot J(p)$ for any $p \in \mathscr{M}$, wherever $\dot J(p)$ is differentiable.

When g is Riemannian, then h is always a Riemannian metric on $\mathscr{S}$, and then $T\mathscr{S}$ is in direct sum with $(T\mathscr{S})^\perp$. Whatever the signature of g, in this section we will always assume that this is the case:

$$T\mathscr{S} \cap (T\mathscr{S})^\perp = \{0\} \quad \Longrightarrow \quad T\mathscr{M} = T\mathscr{S} \oplus (T\mathscr{S})^\perp \; . \tag{6.3.2}$$

Recall that (6.3.2) fails precisely at those points $p \in \mathscr{S}$ at which h is degenerate. Hence, in this section we consider hypersurfaces which are either timelike throughout, or spacelike throughout. Depending upon the character of $\mathscr{S}$ we will then have

$$\epsilon := g(n,n) = \pm 1 \; , \tag{6.3.3}$$

where n is the field of unit normals to $\mathscr{S}$.

For $p \in T\mathscr{S}$ let $P : T_p\mathscr{M} \to T_p\mathscr{M}$ be defined as

$$T_p\mathscr{M} \ni X \to P(X) = X - \epsilon g(X,n)n \; . \tag{6.3.4}$$

Remark 6.3.1 The formalism developed in this section applies with minor changes to the following setup [**?**]: Let K be a Killing vector field on $\mathscr{M}$, at those points at which $g(K,K)$ does not vanish one then sets

$$n = \frac{K}{\sqrt{|g(K,K)|}} \; .$$

Let P be given again by (6.3.4), and suppose that K *is transverse to* $T\mathscr{S}$. One can then define a scalar product on $T\mathscr{S}$ by the formula

$$T\mathscr{S} \ni X, Y \ \to \ h(X,Y) = g(P(X), P(Y)) \ . \tag{6.3.5}$$

•6.3.1: **ptc:***fill in name*

If n is orthogonal to $\mathscr{S}$ then we recover the definition (6.3.1). The modifications needed to cover this situation will be indicated as we go along.[1]•6.3.1 The metric h is called *the orbit-space metric*. Since $\mathscr{S}$ is transverse to the orbits of X, the space of those orbits of X which meet $\mathscr{S}$ can often be locally identified with $\mathscr{S}$ (passing to a subset of $\mathscr{S}$ and $\mathscr{M}$ if necessary). When this is the case, then h coincides with the natural *metric on the space of orbits* at those points at which the space of orbits is a smooth manifold.

We note the following properties of P:

- P annihilates n:

$$P(n) = P(n - \epsilon g(n,n)n) = P(n) - \epsilon^2 P(n) = 0 \ .$$

- P is a projection operator:

$$\begin{aligned} P(P(X)) &= P(X - \epsilon g(X,n)n) \\ &= P(X) - \epsilon g(X,n)P(n) = P(X) \ . \end{aligned}$$

- P restricted to $n^\perp$ is the identity:

$$g(X,n) = 0 \quad \Longrightarrow \quad P(X) = X \ .$$

- P is symmetric:

$$g(P(X),Y) = g(X,Y) - \epsilon g(X,n)g(Y,n) = g(X,P(Y)) \ .$$

The *Weingarten map* $B : T\mathscr{S} \to T\mathscr{S}$ is defined by the equation

$$T\mathscr{S} \ni X \quad \to \quad B(X) := P(\nabla_X n) \in T\mathscr{S} \subset T\mathscr{M} \ . \tag{6.3.6}$$

Here, and in other formulae involving differentiation, one should in principle choose an extension of n off $\mathscr{S}$; however, (6.3.6) involves only derivatives in directions tangent to $\mathscr{S}$, so that the result will not depend upon that extension.

Remark 6.3.1 continued: In the set-up of remark 6.3.1 there is a natural identification between $T\mathscr{S}$ and the space H of vector fields defined in a neighborhood of $\mathscr{S}$ and satisfying the conditions

$$H := \{\mathscr{L}_K X = 0 \ , \ g(K,n) = 0\} \ . \tag{6.3.7}$$

This follows from the fact that n is transverse to $T_p\mathscr{S}$, which in turn implies that the map

$$T_p\mathscr{S} \in X \mapsto P(X) \in T_p\mathscr{M}$$

is bijective. Equation (6.3.6) should then be understood as

$$T\mathscr{S} \approx H \ni X \quad \to \quad B(X) := P(\nabla_X n) \in H \approx T\mathscr{S} \ . \tag{6.3.8}$$

This applies to all further equations in this section.

[1] I am grateful to xxx for pointing out errors in a previous version of the notes about that issue.

The map B is closely related to the *second fundamental form* K of $\mathscr{S}$, also called the *extrinsic curvature tensor* in the physics literature:

$$T\mathscr{S} \ni X, Y \quad \to \quad K(X,Y) := g(P(\nabla_X n), Y) \tag{6.3.9a}$$
$$= g(B(X), Y) \; . \tag{6.3.9b}$$

It is often convenient to have at our disposal index formulae, for this purpose let us consider a local ON frame $\{e_\mu\}$ such that $e_0 = n$ along $\mathscr{S}$. We then have•6.3.2

$$g_{\mu\nu} = \mathrm{diag}(-1, +1, \dots, +1)$$

•6.3.2: **ptc:***this is not coherent, there should be an ϵ for g_{00}, and the signature should be allowed to be Riemannian*

REMARK 6.3.1 CONTINUED: Here the e_μ's should be chosen so that $e_0 = n$, while $e_i \in H$.

Using the properties of P listed above,

$$\begin{aligned} K_{ij} &:= K(e_i, e_j) = g(P(\nabla_{e_i} n), e_j) = g(P(P(\nabla_{e_i} n)), e_j) \\ &= g(P(\nabla_{e_i} n), P(e_j)) = h(P(\nabla_{e_i} n), e_j) = h(B^k{}_i e_k, e_j) \\ &= h_{kj} B^k{}_i \; , \end{aligned} \tag{6.3.10}$$

$$B^k{}_i := \varphi^k(B(e_i)) \; , \tag{6.3.11}$$

where $\{\varphi^k\}$ is a basis of $T^*\mathscr{S}$ dual to the basis $\{P(e_i)\}$ of $T\mathscr{S}$.

Let us show that K is symmetric: first,

$$\begin{aligned} K(X,Y) &= g(\nabla_X n, Y) \\ &= X(\underbrace{g(n,Y)}_{=0}) - g(n, \nabla_X Y) \; . \end{aligned} \tag{6.3.12}$$

Now, ∇ has no torsion, which implies

$$\nabla_X Y = \nabla_Y X - [X,Y] \; .$$

Further, the commutator of vector fields tangent to $\mathscr{S}$ is a vector field tangent to $\mathscr{S}$, which implies

$$\forall \; X, Y \in T\mathscr{S} \qquad g(n, [X,Y]) = 0 \; .$$

Returning to (6.3.12), it follows that

$$K(X,Y) = -g(n, \nabla_Y X - [X,Y]) = -g(n, \nabla_Y X) \; ,$$

and the equation

$$K(X,Y) = K(Y,X)$$

immediately follows from (6.3.12).

REMARK 6.3.1 CONTINUED: The only essential difference in this section arises here: K will not necessarily be symmetric, instead we will have

$$K(X,Y) = K(Y,X) + g([X,Y], n) \; .$$

To continue, for X, Y — sections of $T\mathscr{S}$ we set

$$D_X Y := P(\nabla_X Y)\ . \tag{6.3.13}$$

First, we claim that D is a connection: Linearity with respect to addition in all variables, and with respect to multiplication of X by a function, is straightforward. It remains to check the Leibniz rule:

$$\begin{aligned}
D_X(\alpha Y) &= P(\nabla_X(\alpha Y)) \\
&= P(X(\alpha)Y + \alpha\nabla_X Y) \\
&= X(\alpha)P(Y) + \alpha P(\nabla_X Y) \\
&= X(\alpha)Y + \alpha D_X Y\ .
\end{aligned}$$

It follows that all the axioms of a covariant derivative on vector fields are fulfilled, as desired. It turns out that D is actually the Levi-Civita connection of the metric h. Recall that the Levi-Civita connection is determined uniquely by the requirement of vanishing torsion, and that of metric-compatibility. Both results are straightforward:

$$D_X Y - D_Y X = P(\nabla_X Y - \nabla_Y X) = P([X,Y]) = [X,Y]\ ;$$

in the last step we have again used the fact that the commutator of two vector fields tangent to $\mathscr{S}$ is a vector field tangent to $\mathscr{S}$.•6.3.3 In order to establish metric-compatibility, we calculate for all vector fields X, Y, Z tangent to $\mathscr{S}$:

•6.3.3: **ptc:**ca ne marche plus en general? il faut corriger avec ?

$$\begin{aligned}
X(h(Y,Z)) &= X(g(Y,Z)) \\
&= g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \\
&= g(\nabla_X Y, \underbrace{P(Z)}_{=Z}) + g(\underbrace{P(Y)}_{=Y}, \nabla_X Z) \\
&= \underbrace{g(P(\nabla_X Y), Z) + g(Y, P(\nabla_X Z))}_{P \text{ is symmetric}} \\
&= g(D_X Y, Z) + g(Y, D_X Z) \\
&= h(D_X Y, Z) + h(Y, D_X Z)\ .
\end{aligned}$$

Equation (6.3.13) turns out to be very convenient when trying to express the curvature of h in terms of that of g. To distinguish between both curvatures let us use the symbol ρ for the curvature tensor of h; by definition, for all vector fields tangential to $\mathscr{S}$,

$$\begin{aligned}
\rho(X,Y)Z &= D_X D_Y Z - D_Y D_X Z - D_{[X,Y]}Z \\
&= P\Big(\nabla_X(P(\nabla_Y Z)) - \nabla_Y(P(\nabla_X Z)) - \nabla_{[X,Y]}Z\Big)\ .
\end{aligned}$$

Now, for any vector field W (not necessarily tangent to $\mathscr{S}$) we have

$$\begin{aligned}
P\Big(\nabla_X(P(W))\Big) &= P\Big(\nabla_X(W - \epsilon g(n,W)n)\Big) \\
&= P\Big(\nabla_X W - \underbrace{\epsilon X(g(n,W))n}_{P(n)=0} - \epsilon g(n,W)\nabla_X n\Big)
\end{aligned}$$

$$\begin{aligned} &= P\Big(\nabla_X W\Big) - \epsilon g(n,W) P\Big(\nabla_X n\Big) \\ &= P\Big(\nabla_X W\Big) - \epsilon g(n,W) B(X) \; . \end{aligned}$$

Applying this equation to $W = \nabla_Y Z$ we obtain

$$\begin{aligned} P\Big(\nabla_X(P(\nabla_Y Z))\Big) &= P(\nabla_X\nabla_Y Z) - \epsilon g(n,\nabla_Y Z)B(X) \\ &= P(\nabla_X\nabla_Y Z) + \epsilon K(Y,Z)B(X) \; , \end{aligned}$$

and in the last step we have used (6.3.12). It now immediately follows that

$$\rho(X,Y)Z = P(R(X,Y)Z) + \epsilon\Big(K(Y,Z)B(X) - K(X,Z)B(Y)\Big) \; . \tag{6.3.14}$$

In an adapted ON frame as discussed above this reads

$$\boxed{\rho^i{}_{jk\ell} = R^i{}_{jk\ell} + \epsilon(K^i{}_k K_{j\ell} - K^i{}_\ell K_{jk})} \; . \tag{6.3.15}$$

Here $K^i{}_k$ is the tensor field K_{ij} with an index raised using the contravariant form $h^\#$ of the metric h, compare (6.3.10).

We are ready now to derive the *general relativistic scalar constraint equation*: Let ρ_{ij} denote the Ricci tensor of the metric h, we then have

$$\begin{aligned} \rho_{j\ell} &:= \rho^i{}_{ji\ell} \\ &= \underbrace{R^i{}_{ji\ell}}_{=R^\mu{}_{j\mu\ell} - R^0{}_{j0\ell}} + \epsilon(K^i{}_i K_{j\ell} - K^i{}_\ell K_{ji}) \\ &= R_{j\ell} - R^0{}_{j0\ell} + \epsilon(\mathrm{tr}_h K K_{j\ell} - K^i{}_\ell K_{ji}) \; . \end{aligned}$$

Defining $R(h)$ to be the scalar curvature of h, it follows that[•6.3.4]

•6.3.4: **ptc:** *this is only correct for lorentzian signature at this stage*

$$\begin{aligned} R(h) &= \rho^j{}_j \\ &= \underbrace{R^j j}_{=R^\mu{}_\mu - R^0{}_0} - \underbrace{R^{0j}{}_{0j}}_{=R^{0\mu}{}_{0\mu}} + \epsilon(\mathrm{tr}_h K K^j{}_j - K^{ij} K_{ji}) \\ &= R(g) - 2\underbrace{R^0{}_0}_{=-R_{00}} + \epsilon\Big((\mathrm{tr}_h K)^2 - |K|^2_h\Big) \\ &= 16\pi T_{00} + 2\Lambda + \epsilon\Big((\mathrm{tr}_h K)^2 - |K|^2_h\Big) \; , \end{aligned}$$

and we have used the Einstein equation

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi \frac{G}{c^4} T_{\mu\nu} \tag{6.3.16}$$

with $G = c = 1$. Assuming that $\epsilon = -1$ we obtain the desired scalar constraint:

$$\boxed{R(h) = 16\pi T_{\mu\nu} n^\mu n^\nu + 2\Lambda + |K|^2_h - (\mathrm{tr}_h K)^2} \; . \tag{6.3.17}$$

In particular in vacuum, with $\Lambda = 0$, one obtains

$$R(h) = |K|^2_h - (\mathrm{tr}_h K)^2 \; . \tag{6.3.18}$$

The *vector constraint equation* carries the remaining information contained in the equation $G_{\mu\nu}n^\mu = 0$. In order to understand that equation let Y be tangent to $\mathscr{S}$, we then have

$$\begin{aligned} G_{\mu\nu}n^\mu Y^\nu &= \Big(R_{\mu\nu} - \frac{1}{2}R(g)g_{\mu\nu}\Big)n^\mu Y^\nu \\ &= \mathrm{Ric}(n,Y) - \frac{1}{2}R(g)g(n,Y) \\ &= \mathrm{Ric}(n,Y) \; . \end{aligned}$$

We calculate

$$\begin{aligned} \mathrm{Ric}(Y,n) &= g^{\mu\nu}g(e_\mu, R(Y,e_\nu)n) \\ &= g^{\mu\nu}g(e_\mu, \nabla_Y\nabla_{e_\nu}n - \nabla_{e_\nu}\nabla_Y n - \nabla_{[Y,e_\nu]}n) \\ &= -g(n, \nabla_Y\nabla_n n - \nabla_n\nabla_Y n - \nabla_{[Y,n]}n) \\ &\quad + g^{ij}g(e_i, \nabla_Y\nabla_{e_j}n - \nabla_{e_j}\nabla_Y n - \nabla_{[Y,e_j]}n) \; . \end{aligned} \qquad (6.3.19)$$

Since $g(n,n)$ is constant (recall that we have extended n to have norm equal to plus or minus one in a neighborhood of $\mathscr{S}$) we have, for any vector field Z,

$$g(n, \nabla_Z n) = \frac{1}{2}Z(g(n,n)) = 0 \; . \qquad (6.3.20)$$

It follows that

$$\begin{aligned} g(n,\nabla_Y\nabla_n n) &= Y(\underbrace{g(n,\nabla_n n)}_{=0}) - g(\nabla_Y n, \nabla_n n) \; , \\ g(n,\nabla_n\nabla_Y n) &= n(\underbrace{g(n,\nabla_Y n)}_{=0}) - g(\nabla_Y n, \nabla_n n) \; , \\ g(n, \nabla_{[Y,n]}) &= 0 \; . \end{aligned}$$

Those three equations show that the first term in the before-last line in (6.3.19) cancels the second one there, while the third one gives no contribution. Equation (6.3.20) further implies that for any vector field Z the vector field $\nabla_Z n$ is tangent to $\mathscr{S}$:

$$P(\nabla_Z n) = \nabla_Z n \; .$$

Using $P(e_i) = e_i$ together with symmetry of P we can thus rewrite (6.3.19) as

$$\begin{aligned} \mathrm{Ric}(Y,n) &= g^{ij}g\Big(e_i, P(\nabla_Y\nabla_{e_j}n - \nabla_{e_j}\nabla_Y n - \nabla_{[Y,e_j]}n)\Big) \\ &= g^{ij}g\Big(e_i, P(\nabla_Y(B(e_j)) - \nabla_{e_j}(B(Y))) - B([Y,e_j])\Big) \\ &= g^{ij}g\Big(e_i, D_Y(B(e_j)) - D_{e_j}(B(Y)) - B(D_Y e_j - D_{e_j}Y)\Big) \; . \end{aligned} \qquad (6.3.21)$$

By the Leibniz rule under pairings we have

$$(D_X B)(Y) = D_X\Big(B(Y)\Big) - B(D_X Y) \; ,$$

which allows us to rewrite (6.3.21) in the form

$$\mathrm{Ric}(Y,n) = g^{ij} g\Big(e_i, (D_Y B)(e_j) - (D_{e_j} B)(Y)\Big) . \tag{6.3.22}$$

•6.3.5: **ptc:***is the sign right here?*

In index notation this can be rewritten as•6.3.5

$$\begin{aligned} G_{k\mu} n^\mu &= g^{ij} g\Big(e_i, D_k K^\ell{}_j e_\ell - D_j K^\ell{}_k e_\ell\Big) \\ &= \underbrace{g^{ij} g_{i\ell}}_{=\delta^j_\ell} (D_k K^\ell{}_j - D_j K^\ell{}_k) && (6.3.23) \\ &= D_k K^j{}_j - D_j K^j{}_k . && (6.3.24) \end{aligned}$$

Using the Einstein equation (6.3.16) we obtain the *vector constraint equation*:

$$\boxed{D_k K^j{}_j - D_j K^j{}_k = 8\pi T_{\mu\nu} n^\mu h^\nu{}_k .} \tag{6.3.25}$$

•6.3.6

•6.3.6: **ptc:***for completeness one should also give the other embedding equations here*

6.4 Cauchy data, gauge freedom

Let us return to the discussion of the end of Section 6.1. We shall adopt a slightly general point of view than that presented there, where we assumed that the initial data were given on an open subset $\mathscr{O}$ of the zero-level set of the function y^0. A correct geometric picture here is to start with an n dimensional hypersurface $\mathscr{S}$, and prescribe initial data there; the case where $\mathscr{S}$ is $\mathscr{O}$ is thus a special case of this construction. At this stage there are two attitudes one may wish to adopt: the first is that $\mathscr{S}$ is a subset of the space-time $\mathscr{M}$ — this is essentially what we assumed in Section 6.3. Another way of looking at this is to consider $\mathscr{S}$ as a hypersurface of its own, equipped with an embedding

$$i : \mathscr{S} \to \mathscr{M} .$$

The most convenient approach is to go back and forth between those points of view, and this is the strategy that we will follow.

As made clear by the results in Section 6.3, the metric h is uniquely defined by the space-time metric g once that $\mathscr{S} \subset \mathscr{M}$ (or $i(\mathscr{S}) \subset \mathscr{M}$) has been prescribed; the same applies to the extrinsic curvature tensor K. A *vacuum initial data set* $(\mathscr{S}, h, K)$ is a triple where $\mathscr{S}$ is an n–dimensional manifold, h is a Riemannian metric on $\mathscr{S}$, and K is a symmetric two-covariant tensor field on $\mathscr{S}$. Further (h,K) are supposed to satisfy the vacuum constraint equations (**??**), perhaps (but not necessarily so) with a non-zero cosmological constant Λ.

Let us start by showing that specyfing K is equivalent to prescribing the time-derivatives of the space-part g_{ij} of the resulting space-time metric g. Suppose, indeed, that a space-time (M,g) has been constructed (not necessarily vacuum) such that K is the extrinsic curvature tensor of $\mathscr{S}$ in $(\mathscr{M}, g)$. Consider any domain of coordinates $\mathscr{O} \subset \mathscr{S}$ and construct coordinates y^μ in some $\mathscr{M}$–neighborhood of $\mathscr{U}$ such that $\mathscr{S} \cap \mathscr{U} = \mathscr{O}$; those coordinates could be wave-coordinates, as described at the end of Section 6.1, but this is not necessary at this stage. Since y^0 is constant on $\mathscr{S}$ the one-form dy^0 annihilates $T\mathscr{S}$, so does the one form $g(n,\cdot)$, and since $\mathscr{S}$ has codimension one it follows that dy^0 must be proportional to $g(n,\cdot)$:

$$n_A dy^A = n_0 dy^0$$

on $\mathscr{O}$. The normalisation

$$-1 = g(n,n) = g^{\mu\nu} n_\mu n_\nu = g^{00}(n_0)^2$$

gives

$$n_A dy^A = \frac{1}{\sqrt{|\det g|}} dy^0 \ .$$

We then have, using (**??**),

$$\begin{aligned} K_{ij} &:= g(\nabla_i n, \partial_j) = \nabla_i n_j \\ &= \partial_i n_j - \Gamma^\mu{}_{ji} n_\mu \\ &= -\Gamma^0_{ji} n_0 \\ &= -\frac{1}{2} g^{0\sigma}\Big(\partial_j g_{\sigma i} + \partial_i g_{\sigma j} - \partial_\sigma g_{ij}\Big) n_0 \ . \end{aligned} \tag{6.4.1}$$

This shows that the knowledge of $g_{\mu\nu}$ *and* $\partial_0 g_{ij}$ at $y^0 = 0$ allows one to calculate K_{ij}. Reciprocally, (6.4.1) can be rewritten as

$$\partial_0 g_{ij} = \frac{2}{g^{00} n_0} K_{ij} + \text{ terms determined by the } g_{\mu\nu}\text{'s and their space–derivatives} \ ,$$

so that the knowledge of the $g_{\mu\nu}$'s and of the K_{ij}'s at $y^0 = 0$ allows one to calculate $\partial_0 g_{ij}$. Thus, K_{ij} is the geometric counterpart of the $\partial_0 g_{ij}$'s.

To continue, we wish to show the *gauge character* of the g_{0A}'s. By this it is usually meant that the objects under consideration do not have any intrinsic meaning, and their values can be changed using the action of some family of transformations, relevant to the problem at hand, without changing the geometric, or physical, information carried by those objects. In our case the relevant transformations are the coordinate ones, and things are made precise by the following proposition:

Proposition 6.4.1 *Let g_{AB}, $\tilde g_{AB}$ be two metrics such that*

$$g_{ij}|_{\{y^0=0\}} = \tilde g_{ij}|_{\{y^0=0\}} \ , \qquad K_{ij}|_{\{y^0=0\}} = \tilde K_{ij}|_{\{y^0=0\}} \ . \tag{6.4.2}$$

Then there exists a coordinate transformation ϕ defined in a neighborhood of $\{y^0 = 0\}$ which preserves (6.4.2) such that

$$g_{0A}|_{\{y^0=0\}} = (\phi^* \tilde g)_{0A}|_{\{y^0=0\}} \ . \tag{6.4.3}$$

In fact, for any metric g there exist local coordinate systems $\{\bar y^\mu\}$ such that $\{y^0 = 0\} = \{\bar y^0 = 0\}$ and, if we write $g = \bar g_{AB} d\bar y^A d\bar y^B$ etc. *in the barred coordinate system, then*

$$\begin{aligned} g_{ij}|_{\{y^0=0\}} = \bar g_{ij}|_{\{\bar y^0=0\}} \ , &\quad K_{ij}|_{\{y^0=0\}} = \bar K_{ij}|_{\{\bar y^0=0\}} \ , \\ \bar g_{00}|_{\{y^0=0\}} = -1 \ , &\quad \bar g_{0i}|_{\{y^0=0\}} = 0 \ . \end{aligned} \tag{6.4.4}$$

REMARK 6.4.2 We can actually always achieve $\bar g_{00} = -1$, $\bar g_{0i} = 0$ in a whole neighborhood of $\mathscr{S}$: this is done by shooting geodesics normally to $\mathscr{S}$, choosing y^0 to be the affine parameter along those geodesics, and by transporting the coordinates y^i from $\mathscr{S}$ by requiring them to be constant along the normal geodesics. The coordinate system will break down wherever the normal geodesics start intersecting, but the implicit function theorem guarantees that there will exist a neighborhood of $\mathscr{S}$ on which this does not happen. The resulting coordinates are called *Gauss coordinates*. While those coordinates are geometrically natural, in this coordinate system the Einstein equations do not have any good properties from the PDE point of view.

PROOF: It suffices to prove the second claim: for if $\bar\phi$ is the transformation that brings g to the form (6.4.4), and $\tilde\phi$ is the corresponding transformation for $\tilde g$, then $\phi := \tilde\phi \circ \bar\phi^{-1}$ will satisfy (6.4.3).

Let us start by calculating the change of the metric coefficients under a transformation of the form

$$y^0 = \varphi \bar y^0 \;, \qquad y^i = \bar y^i + \psi^i \bar y^0 \;. \tag{6.4.5}$$

If $\varphi > 0$ then clearly

$$\{y^0 = 0\} = \{\bar y^0 = 0\} \;.$$

Further, one has

$$\begin{aligned}
g\Big|_{\{y^0=0\}} &= \Big(g_{00}(dy^0)^2 + 2g_{0i}dy^0dy^i + g_{ij}dy^idy^j\Big)\Big|_{\{y^0=0\}} \\
&= \Big(g_{00}(\bar y^0 d\varphi + \varphi d\bar y^0)^2 + 2g_{0i}(\bar y^0 d\varphi + \varphi d\bar y^0)(d\bar y^i + \bar y^0 d\psi^i + \psi^i d\bar y^0) \\
&\quad + g_{ij}(d\bar y^i + \bar y^0 d\psi^i + \psi^i d\bar y^0)(d\bar y^j + \bar y^0 d\psi^j + \psi^j d\bar y^0)\Big)\Big|_{\{y^0=0\}} \\
&= \Big(g_{00}(\varphi d\bar y^0)^2 + 2g_{0i}\varphi d\bar y^0(d\bar y^i + \psi^i d\bar y^0) \\
&\quad + g_{ij}(d\bar y^i + \psi^i d\bar y^0)(d\bar y^j + \psi^j d\bar y^0)\Big)\Big|_{\{y^0=0\}} \\
&= \Big((g_{00}\varphi^2 + 2g_{0i}\psi^i + g_{ij}\psi^i\psi^j)(d\bar y^0)^2 \\
&\quad + 2(g_{0i}\varphi + g_{ij}\psi^j)d\bar y^0 d\bar y^i + g_{ij}d\bar y^i d\bar y^j\Big)\Big|_{\{y^0=0\}} \\
&=: \bar g_{\mu\nu} d\bar y^\mu d\bar y^\nu \;.
\end{aligned}$$

We shall apply the above transformation twice: first we choose $\varphi = 1$ and

$$\psi^i = h^{ij} g_{0j} \;,$$

where h^{ij} is the matrix inverse to g_{ij}; this leads to a metric with $\bar g_{0i} = 0$. We then apply a second transformation of the form (6.4.5) to the new metric, now with the new $\psi^i = 0$, and with a φ chosen so that the final g_{00} equals minus one. □

In Proposition 6.4.1 we have *not* assumed that the vacuum Einstein equations hold, or that the metrics g and $\tilde g$ are isometric away from $\mathscr{S}$. Now, our discussion in Section 6.1 shows that we can always find a solution of the vacuum Einstein equations for any prescribed values $g_{0A}|_{\{y^0=0\}}$. Let us show that two solutions differing only by the values $g_{0A}|_{\{y^0=0\}}$ are (locally) isometric:•6.4.1

•6.4.1: **ptc:**_needs finishing_

6.5 $3+1$ evolution equations (to be done)

6.6 Existence: global in space (to be done)

6.7 Constraint equations: the conformal method

A set (M,g,K), where (M,g) is a Riemannian manifold, and K is a symmetric tensor field on M, will be called a *vacuum initial data set* if the vacuum constraint equations (6.3.17), (6.3.25) hold:

$$D_j K^j{}_k = D_k K^j{}_j \ , \tag{6.7.1a}$$

$$R(g) = 2\Lambda + |K|_g^2 - (\mathrm{tr}_g K)^2 \ . \tag{6.7.1b}$$

Here, as before, Λ is a constant. The object of this section is to present the *conformal method* for constructing solutions of (6.7.1). This method requires $\mathrm{tr}_g K$ to be constant over M:

$$\partial_i(\mathrm{tr}_g K) = 0 \ . \tag{6.7.2}$$

(We shall see shortly that (6.7.2) leads to a decoupling of the equations (6.7.1), in a sense which will be made precise below.) Hypersurfaces M in a space-times $\mathscr{M}$ satisfying (6.7.2) are known as *constant mean curvature (CMC) surfaces*. Equation (6.7.2) is sometimes viewed as a "gauge condition", in the following sense: if we require that (6.7.2) be satisfied by all hypersurfaces M_τ within a family of hypersurfaces in the space-time, then this condition restricts the freedom of choice of the associated time function t which labels those hypersurfaces. Unfortunately there exist space-times in which no CMC hypersurfaces exist [12, 67]. Now, the conformal method is the only method known which produces *all* solutions satisfying a reasonably mild "gauge condition", it is therefore unfortunate that condition (6.7.2) is a restrictive one.

The conformal method seems to go back to Lichnerowicz [78], except that Lichnerowicz proposes a different treatment of the vector constraint there.•6.7.1 The associated analytical aspects have been implemented in various contexts: asymptotically flat [25, ?], asymptotically hyperbolic [2–4], or spatially compact [64]; see also [23, 65, 109, ?]. There exist a few other methods for constructing solutions of the constraint equations which do not require constant mean curvature: the "thin sandwich approach" of Baierlain, Sharp and Wheeler [?], further studied in [?, 15, ?]; the gluing approach of Corvino and Schoen [29, 37, 38]; the conformal gluing technique of Joyce [70], as extended by Isenberg, Mazzeo and Pollack [66, 67]; the quasi-spherical construction of Bartnik [13, ?] and its extension due to Smith and Weinstein [101]. One can also use the implicit function theorem, or variations thereof [22, 68, 69], to construct solutions of the constraint equations for which (6.7.2) does not necessarily hold.•6.7.2

•6.7.1: **ptc:**_Ask Yvonne about that?_

•6.7.2: **ptc:**_mention the thin sandwich here, or describe it elsewhere, or both_

•6.7.3

•6.7.3: **ptc:**look up Tam and Shi [98] on positivity of York Brown mass of convex bodies

As is made clear by the name, the method exploits the properties of (6.7.1) under conformal transformations: consider a metric $\widetilde{g}$ related to g by a conformal rescaling:

$$\widetilde{g}_{ij} = \phi^\ell g_{ij} \quad \Longleftrightarrow \quad \widetilde{g}^{ij} = \phi^{-\ell} g^{ij} . \tag{6.7.3}$$

It is straightforward to derive

$$\begin{aligned}
\widetilde{\Gamma}^i{}_{jk} &= \frac{1}{2}\widetilde{g}^{im}(\partial_j \widetilde{g}_{km} + \partial_k \widetilde{g}_{jm} - \partial_m \widetilde{g}_{jk}) \\
&= \frac{1}{2}\phi^{-\ell} g^{im}(\partial_j(\phi^\ell \widetilde{g}_{km}) + \partial_k(\phi^\ell \widetilde{g}_{jm}) - \partial_m(\phi^\ell g_{jk})) \\
&= \Gamma^i{}_{jk} + \frac{\ell}{2\phi}(\delta^i_k \partial_j \phi + \delta^i_j \partial_k \phi - g_{jk} D^i \phi) ,
\end{aligned} \tag{6.7.4}$$

where, as before, D denotes the covariant derivative of g. Let us start by analysing what happens with (6.7.1a). Let $\widetilde{D}$ denote the covariant derivative operator of the metric $\widetilde{g}$, and consider any trace-free symmetric tensor field $\widetilde{L}^{ij}$, we have

$$\begin{aligned}
\widetilde{D}_i \widetilde{L}^{ij} &= \partial_i \widetilde{L}^{ij} + \widetilde{\Gamma}^i{}_{ik} \widetilde{L}^{kj} + \widetilde{\Gamma}^j{}_{ik} \widetilde{L}^{ik} \\
&= D_i \widetilde{L}^{ij} + (\widetilde{\Gamma}^i{}_{ik} - \Gamma^i{}_{ik}) \widetilde{L}^{kj} + (\widetilde{\Gamma}^j{}_{ik} - \Gamma^j{}_{ik}) \widetilde{L}^{ik} .
\end{aligned}$$

Now, from (6.7.4) we obtain

$$\begin{aligned}
\widetilde{\Gamma}^i{}_{ik} &= \Gamma^i{}_{ik} + \frac{\ell}{2\phi}(\delta^i_k \partial_i \phi + \delta^i_i \partial_k \phi - g_{ik} D^i \phi) \\
&= \Gamma^i{}_{ik} + \frac{n\ell}{2\phi}\partial_k \phi ,
\end{aligned} \tag{6.7.5}$$

and we are assuming that we are in dimension n. As $\widetilde{L}$ is traceless we obtain

$$\begin{aligned}
\widetilde{D}_i \widetilde{L}^{ij} &= D_i \widetilde{L}^{ij} + \frac{n\ell}{2\phi}\partial_k \phi \widetilde{L}^{kj} + \frac{\ell}{2\phi}(\delta^j_k \partial_i \phi + \delta^j_i \partial_k \phi - \underbrace{g_{ik} D^j \phi) \widetilde{L}^{ik}}_{\sim g_{ik}\widetilde{L}^{ik}=0} \\
&= D_i \widetilde{L}^{ij} + \frac{(n+2)\ell}{2\phi}\partial_k \phi \widetilde{L}^{kj} \\
&= \phi^{-(n+2)\ell/2} D_i(\phi^{(n+2)\ell/2} \widetilde{L}^{ij}) .
\end{aligned} \tag{6.7.6}$$

It follows that

$$\widetilde{D}_i \widetilde{L}^{ij} = 0 \quad \Longleftrightarrow \quad D_i(\phi^{(n+2)\ell/2} \widetilde{L}^{ij}) = 0 . \tag{6.7.7}$$

This observation leads to the following: suppose that the CMC condition (6.7.2) holds, set

$$L^{ij} := K^{ij} - \frac{\mathrm{tr}_g K}{n} g^{ij} . \tag{6.7.8}$$

Then L^{ij} is symmetric and trace-free whenever K^{ij} satisfies the vector constraint equation (6.7.1a). Reciprocally, let τ be any constant, and let $\widetilde{L}^{ij}$ be symmetric, trace-free, and $\widetilde{g}$*–divergence free*: by definition, this means that

$$\widetilde{D}_i \widetilde{L}^{ij} = 0 .$$

Set

$$L^{ij} := \phi^{(n+2)\ell/2}\widetilde{L}^{ij} \tag{6.7.9a}$$

$$K^{ij} := L^{ij} + \frac{\tau}{n}g^{ij} , \tag{6.7.9b}$$

then K^{ij} satisfies (6.7.1a).

To analyse the scalar constraint equation (6.7.1b) we shall use the following formula, derived in Appendix C,•6.7.4

•6.7.4: **ptc:***to be done*

$$R(g)\phi^{-\ell} = \widetilde{R} + \frac{(n-1)\ell}{\phi}\Delta_{\widetilde{g}}\phi - \frac{(n-1)\ell\{(n-2)\ell+4\}}{4\phi^2}|d\phi|^2_{\widetilde{g}} , \tag{6.7.10}$$

where $\widetilde{R}$ is the scalar curvature of $\widetilde{g}$. Clearly it is convenient to choose

$$\ell = -\frac{4}{n-2} , \tag{6.7.11}$$

as then the last term in (6.7.4) drops out. In order to continue we use (6.7.9) to calculate

$$\begin{aligned} |K|^2_g - (\mathrm{tr}_g K)^2 &= g_{ik}g_{jl}K^{ij}K^{kl} - \tau^2 \\ &= g_{ik}g_{jl}(L^{ij} + \frac{\tau}{n}g^{ij})(L^{kl} + \frac{\tau}{n}g^{kl}) - \tau^2 \\ &= \underbrace{g_{ik}}_{=\phi^{-\ell}g_{ik}} \; g_{jl} \; \underbrace{L^{ij}}_{=\phi^{(n/2+1)\ell}\widetilde{L}^{ij}} \; L^{kl} - \tau^2(1-\frac{1}{n}) \\ &= \phi^{n\ell}\widetilde{g}_{ik}\widetilde{g}_{jl}\widetilde{L}^{ij}\widetilde{L}^{kl} - \tau^2(1-\frac{1}{n}) , \end{aligned}$$

giving thus

$$|K|^2_g - (\mathrm{tr}_g K)^2 = \phi^{n\ell}|\widetilde{L}|^2_{\widetilde{g}} - \frac{n-1}{n}\tau^2 . \tag{6.7.12}$$

Equations (6.7.1b), (6.7.4) and (6.7.12) with ℓ given by (6.7.11) finally yield

$$\Delta_{\widetilde{g}}\phi - \frac{n-2}{4(n-1)}\widetilde{R}\phi = -\sigma^2\phi^{(2-3n)/(n-2)} + \beta\phi^{(n+2)/(n-2)} , \tag{6.7.13}$$

where•6.7.5

•6.7.5: **ptc:***Jim says that the coefficient of sigma square should be 1/8 in dim 3, which does not agree here*

$$\sigma^2 := \frac{n-2}{2(n-1)}|\widetilde{L}|^2_{\widetilde{g}} , \quad \beta := \left[\frac{n-2}{4n}\tau^2 - \frac{n-2}{2(n-1)}\Lambda\right] . \tag{6.7.14}$$

In dimension $n=3$ this equation is known as the *Lichnerowicz equation*:

$$\boxed{\Delta_{\widetilde{g}}\phi - \tfrac{\widetilde{R}}{8}\phi = -\sigma^2\phi^{-7} + \beta\phi^5 .} \tag{6.7.15}$$

We note that σ^2 is positive, as the notation suggests, while β is a constant, non-negative if $\Lambda = 0$, or in fact if $\Lambda \le 0$.

The strategy is now the following: let $\widetilde{g}$ be a given Riemannian metric on M, and let $\widetilde{L}^{ij}$ be any symmetric transverse $\widetilde{g}$-divergence free tensor field. We then solve (if possible) (6.7.13) for ϕ, and obtain a vacuum initial data set by calculating g using (6.7.3), and by calculating K using (6.7.9).

6.7.1 The vector constraint equation on compact manifolds

The starting point of the conformal method is the tensor field $\widetilde{L}$, and to obtain a complete prescription we have to provide such tensors. A prescription how to do that has been given by York: here one starts with an arbitrary symmetric traceless tensor field B^{ij}, which will be referred to as the *seed field*. One then writes•6.7.6

$$\widetilde{L}^{ij} = B^{ij} + C(Y)^{ij} \,, \tag{6.7.16}$$

•6.7.6: **ptc:***it could be that all such $\widetilde{L}$'s are zero, one should talk about the York splitting somewhere here; the fact that the space of tt tensors is infinite dimensional in dimension larger than two is claimed in [19]*

where $C(Y)$ is the *conformal Killing operator*:

$$C(Y)^{ij} := \widetilde{D}^i Y^j + \widetilde{D}^j Y^i - \frac{2}{n}\widetilde{D}_k Y^k \widetilde{g}^{ij} \,. \tag{6.7.17}$$

The requirement that $\widetilde{L}^{ij}$ be divergence free becomes then an equation for the vector field Y:

$$\widetilde{D}_i \widetilde{L}^{ij} = 0 \quad \Longleftrightarrow \quad L(Y)^j := \widetilde{D}_i(\widetilde{D}^i Y^j + \widetilde{D}^j Y^i - \frac{2}{n}\widetilde{D}_k Y^k \widetilde{g}^{ij}) = -\widetilde{D}_i B^{ij} \,. \tag{6.7.18}$$

While (6.7.18) looks complicated at first sight, it is rather natural: we want to produce transverse traceless tensors by solving an elliptic differential equation. Since the condition of being divergence-free is already a first order equation, and it is not elliptic, then the lowest possible order of such an equation will be two. Now, the divergence operation produces vector fields, so the most straightforward way of ensuring ellipticity is to seek an equation for a vector field. The simplest object that we obtain by differentiating a vector field is the tensor field $\widetilde{D}^i Y^j$; in order to achieve the correct symmetries we need to symmetrise and remove the trace, which leads to the conformal Killing operator (6.7.17).

The operator L defined in (6.7.18) is known as the *conformal vector Laplacian*. It is a second order linear partial differential equation for Y, the solvability of which can be easily analysed. In this section we shall consider spatially compact manifolds M. We will give an existence proof for (6.7.18), this requires a few analytical ingredients. The main ingredients of the existence proof which we will present shortly are the following:

1. *Function spaces*: one uses the spaces H_k, $k \in \mathbb{N}$, defined as the completion of the space of smooth tensor fields on M with respect to the norm

$$\|u\|_k := \sqrt{\sum_{0\le \ell \le k} \int_M |D^\ell u|^2 d\mu} \,, \tag{6.7.19}$$

where $D^\ell u$ is the tensor of ℓ-th covariant derivatives of u with respect to some covariant derivative operator D. For compact manifolds[2] this space is identical with that of fields in L^2 such that their distributional derivatives of order less than or equal to k are also in L^2. Again for compact manifolds, different choices of measure $d\mu$ (as long as it remains absolutely continuous with respect to the coordinate one), of the tensor norm $|\cdot|$, or of the connection D, lead to the same space, with equivalent norm.

[2] For non-compact manifolds this is not always the case, compare [7].

Recall that if $u \in L^2$ then $\partial_i u = \rho_i$ in a distributional sense if for every smooth compactly supported vector field we have

$$\int_M X^i \rho_i = -\int_M D_i X^i u \ .$$

More generally, let A be a linear differential operator of order m and let A^t be its *formal* L^2 adjoint, which is the operator obtained by differentiating by parts:

$$\int_M \langle u, L^\dagger v\rangle := \int_M \langle Lu, v\rangle \ , \quad u, v \in C^m_c ;$$

the above formula defines $L^\dagger$ uniquely if it holds for all u, v in the space C^m_c of C^m compactly supported fields. (Incidentally, the reader will note by comparing the last two equations that the formal adjoint of the derivative operator is minus the divergence operator.) Then, for $u \in L^1_{\mathrm{loc}}$ (this is the space of measurable fields u which are Lebesgue-integrable on any compact subset of the manifold), the distributional equation $Lu = \rho$ is said to hold if for all smooth compactly supported v's we have

$$\int_M \langle u, L^\dagger v\rangle = \int_M \langle \rho, v\rangle \ .$$

One sometimes talks about *weak solutions* rather than distributional ones.

The spaces H_k are Hilbert spaces with the obvious scalar product:

$$\langle u, v\rangle_k = \sum_{0\le \ell\le k} \int_M \langle D^\ell u, D^\ell v\rangle d\mu \ .$$

The Sobolev embedding theorem [8] asserts that H_k functions are, locally, of $C^{k'}$ differentiability class, where k' is the largest integer satisfying

$$k' < k - n/2 \ . \tag{6.7.20}$$

On a compact manifold the result is true globally,

$$H_k \subset C^{k'} \ , \tag{6.7.21}$$

with the inclusion map being continuous:

$$\|u\|_{C^{k'}} \le C\|u\|_{H_k} \ . \tag{6.7.22}$$

2. *Orthogonal complements in Hilbert spaces:* Let H be a Hilbert space, and let E be a *closed linear subspace* of H. Then (see, *e.g.*, [106]) we have the direct sum

$$H = E \oplus E^\perp \ . \tag{6.7.23}$$

   This result is sometimes called *the projection theorem*.

3. *Rellich-Kondrashov compactness:* we have the obvious inclusion

$$H_k \subset H_{k'} \ \text{ if } \ k \ge k' \ .$$

The *Rellich-Kondrashov theorem* (see, *e.g.*, $[1, 8, 56, 74]$) asserts that this inclusion is *compact*. Equivalently,[3] if u_n is any sequence satisfying $\|u_n\|_k \leq C$, and if $k' < k$, then there exists a subsequence u_{n_i} and $u_\infty \in H_k$ such that u_{n_i} converges to u_∞ in $H_{k'}$ topology as i tends to infinity.

4. *Elliptic regularity:* If $Y \in L^2$ satisfies $LY \in H_k$ in a distributional sense, with L — an elliptic operator of order m with smooth coefficients, then $Y \in H_{k+m}$, and Y satisfies the equation in the classical sense. Further for every k there exists a constant C_k such that

$$\|Y\|_{k+m} \leq C_k(\|LY\|_k + \|Y\|_0) . \tag{6.7.24}$$

Our aim is to show that solvability of (6.7.18) can be easily studied using the above basic facts. We start by verifying ellipticity of L. Recall that the *symbol* σ of a linear partial differential operator L of the form

$$L = \sum_{0 \leq \ell \leq m} a^{i_1 \ldots i_\ell} D_{i_1} \ldots D_{i_\ell} ,$$

where the $a^{i_1 \ldots i_\ell}$'s are linear maps from fibers of a bundle E to fibers of a bundle F, is defined as the map

$$T^*M \ni p \mapsto \sigma(p) := a^{i_1 \ldots i_m} p_{i_1} \ldots p_{i_m} .$$

Thus, every derivative D_i is replaced by p_i, and all terms other than the top order ones are ignored. An operator is said to be *elliptic* if the symbol is an isomorphism of fibers for all $p \neq 0$. In our case (6.7.18) the operator L acts on vector fields and produces vector fields, with

$$TM \ni Y \to \sigma(p)(Y) = p_i(p^i Y^j + p^j Y^i - \frac{2}{n} p_k Y^k \tilde{g}^{ij})\partial_j \in TM . \tag{6.7.25}$$

(The indices on p^i have been raised with the metric $\tilde{g}$.) To prove bijectivity of $\sigma(p)$, $p \neq 0$, it suffices to verify that $\sigma(p)$ has trivial kernel. Assuming $\sigma(p)(Y) = 0$, a contraction with p_j gives

$$p_j p_i(p^i Y^j + p^j Y^i - \frac{2}{n} p_k Y^k \tilde{g}^{ij}) = |p|^2 p_j Y^j (2 - \frac{2}{n}) = 0 ,$$

hence $p_j Y^j = 0$ for $n > 1$ since $p \neq 0$. Contracting instead with Y_j and using the last equality we obtain

$$Y_j p_i(p^i Y^j + p^j Y^i - \frac{2}{n} p_k Y^k \tilde{g}^{ij}) = |p|^2 |Y|^2 = 0 ,$$

and $\sigma(p)$ has no kernel, as desired.

[3] In this statement we have also made use of the *Tichonov-Alaoglu* theorem, which asserts that bounded sets in Hilbert spaces are weakly compact; *cf., e.g.* [106].

To gain some more insight into the conformal vector Laplacian L let us calculate its formal L^2–adjoint: let thus X and Y be smooth, or C^2, we write

$$\begin{aligned}
\int_M X_i L(Y)^i d\mu_{\widetilde g} &= \int_M X_i \widetilde D_j(\widetilde D^i Y^j + \widetilde D^j Y^i - \frac{2}{n}\widetilde g^{ij}\widetilde D_k Y^k)\, d\mu_{\widetilde g} \\
&= -\int_M \widetilde D_j X_i(\widetilde D^i Y^j + \widetilde D^j Y^i - \frac{2}{n}\widetilde g^{ij}\widetilde D_k Y^k)\, d\mu_{\widetilde g} \\
&= -\frac{1}{2}\int_M (\widetilde D_j X_i + \widetilde D_j X_i)\underbrace{(\widetilde D^i Y^j + \widetilde D^j Y^i - \frac{2}{n}\widetilde g^{ij}\widetilde D_k Y^k)}_{\text{symmetric in } i \text{ and } j}\, d\mu_{\widetilde g} \\
&= -\frac{1}{2}\int_M (\widetilde D_j X_i + \widetilde D_j X_i - \frac{2}{n}\widetilde D_k X^k \widetilde g_{ij})\underbrace{(\widetilde D^i Y^j + \widetilde D^j Y^i - \frac{2}{n}\widetilde g^{ij}\widetilde D_k Y^k)}_{\text{trace free}}\, d\mu_{\widetilde g} \\
&= -\int_M (\widetilde D_j X_i + \widetilde D_j X_i - \frac{2}{n}\widetilde D_k X^k \widetilde g_{ij})\widetilde D^i Y^j\, d\mu_{\widetilde g} \\
&= \int_M \widetilde D^i(\widetilde D_j X_i + \widetilde D_j X_i - \frac{2}{n}\widetilde D_k X^k \widetilde g_{ij}) Y^j\, d\mu_{\widetilde g} \\
&= \int_M L(X)^j Y_j\, d\mu_{\widetilde g}\ .
\end{aligned} \tag{6.7.26}$$

Recall that the *formal adjoint* $L^\dagger$ of L is defined by integration by parts:

$$\int \langle u, Lv\rangle = \int \langle L^\dagger u, v\rangle$$

for all smooth compactly supported fields u, v. (Recall that the definition of a self-adjoint operator further requires an equality of domains, an issue which is, fortunately, completely ignored in the formal definition.) We have thus shown that the conformal vector Laplacian is *formally self adjoint*:

$$L^\dagger = L\ . \tag{6.7.27}$$

We further note that the fourth line in (6.7.26) implies

$$\int_M Y_i L(Y)^i = -\frac{1}{2}\int_M |C(Y)|^2\ , \tag{6.7.28}$$

in particular if Y is C^2 then

$$L(Y) = 0 \quad \Longleftrightarrow \quad C(Y) = 0\ . \tag{6.7.29}$$

This implies that manifolds for which L has a non-trivial kernel are very special.

REMARK 6.7.1 In fact, solutions of the equation $C(Y) = 0$ are called *conformal Killing vectors*. The existence of non-trivial conformal Killing vectors implies the existence of conformal isometries of (M, g). A famous theorem of Obata [77]•6.7.7 implies that either there is a conformal rescaling such that Y is a Killing vector, or (M, g) is conformally isometric to S^n with a round metric. In the former case (M, g) has a non-trivial isometry group, which imposes restrictions on the topology of M, and forces g to be very special. For instance, the existence of non-trivial

•6.7.7: **ptc:***find a ref to Obata*

•6.7.8: **ptc:**_give refs_

Lie group of isometries implies that M admits an S^1 action, which is a serious topological restriction, and in fact is not possible for "most" topologies•6.7.8 . It is also true that even if M admits S^1 actions, then there exists an open and dense set of metrics, in a C^3 topology, or in a H^3 topology, for which no nontrivial solutions of the over-determined system of equations $C(Y) = 0$ exist.

In order to continue we shall need a somewhat stronger version of (6.7.24):•6.7.9

•6.7.9: **ptc:**_second order not needed here, and in the next theorem_

PROPOSITION 6.7.2 *Let L be a second order elliptic operator and suppose that there are no non-trivial smooth solutions of the equation $L(Y) = 0$. Then (6.7.24) can be strengthened to*

$$\|Y\|_{k+2} \leq C'_k \|L(Y)\|_k \; . \tag{6.7.30}$$

REMARK 6.7.3 Equation (6.7.30) implies that L has trivial kernel, which shows that the condition on the kernel is necessary.

PROOF: Suppose that the result does not hold, then for every $n \in \mathbb{N}$ there exists $Y_n \in H_{k+2}$ such that

$$\|Y_n\|_{k+2} \geq n \|L(Y_n)\|_k \; . \tag{6.7.31}$$

Multiplying Y_n by an appropriate constant if necessary we can suppose that

$$\|Y\|_{L^2} = 1 \; . \tag{6.7.32}$$

The basic elliptic inequality (6.7.24) gives

$$\|Y_n\|_{k+2} \leq C_2(\|LY_n\|_k + \|Y_n\|_0) \leq \frac{C_2}{n}\|Y_n\|_{k+2} + C_2 \; ,$$

so that for n such that $C_2/n \leq 1/2$ we obtain

$$\|Y_n\|_{k+2} \leq 2C_2 \; .$$

It follows that Y_n is bounded in H_{k+2}; further (6.7.31) gives

$$\|L(Y_n)\|_k \leq \frac{2C_2}{n} \; . \tag{6.7.33}$$

By the Rellich-Kondrashov compactness we can extract a subsequence, still denoted by Y_n, such that Y_n converges in L^2 to $Y_* \in H_{k+2}$. Continuity of the norm together with L^2 convergence implies that

$$\|Y_*\|_{L^2} = 1 \; , \tag{6.7.34}$$

so that $Y_* \neq 0$. One would like to conclude from (6.7.33) that $L(Y_*) = 0$, but that is not completely clear because we do not know whether or not

$$LY_* = \lim_{n\to\infty} LY_n \; .$$

Instead we write the distributional equation: for every smooth X we have

$$\int_M \langle L(Y_n), X\rangle = \int_M \langle Y_n, L^\dagger(X)\rangle \; .$$

Now, $L(Y_n)$ tends to zero in L^2 by (6.7.33), and Y_n tends to Y_* in L^2, so that passing to the limit we obtain

$$0 = \int_M \langle Y_*, L^\dagger(X)\rangle \ .$$

It follows that Y_* satisfies $L(Y_*) = 0$ in a distributional sense. Elliptic regularity implies that Y_* is a smooth solution of $LY_* = 0$, it is non-trivial by (6.7.34), a contradiction. □

We are ready to prove now:

THEOREM 6.7.4 *Let L be any elliptic partial differential operator on a compact manifold and suppose that the equations $Lu = 0$, $L^\dagger v = 0$ have no non-trivial smooth solutions, where $L^\dagger$ is the formal adjoint of L. Then for any $k \geq 0$ the map*

$$L : H_{k+2} \to H_k$$

is an isomorphism.

PROOF: An element of the kernel is necessarily smooth by elliptic regularity, it remains thus to show surjectivity. We start by showing that the image of L is closed: let Z_n be a Cauchy sequence in $\operatorname{Im} L$, then there exists $Z_\infty \in L^2$ and $Y_n \in H_{k+2}$ such that

$$LY_n = Z_n \stackrel{L^2}{\to} Z_\infty \ .$$

Applying (6.7.30) to $Y_n - Y_m$ we find that Y_n is Cauchy in H_{k+2}, therefore converges in H_{k+2} to some element $Y_\infty \in H_{k+2}$. By continuity of L we have that LY_n converges to LY_∞ in L^2, hence $Z_\infty = LY_\infty$, as desired.

Consider, first, the case $k = 0$. By the orthogonal decomposition theorem we have now

$$L^2 = \operatorname{Im} L \oplus (\operatorname{Im} L)^\perp \ ,$$

and if we show that $(\operatorname{Im} L)^\perp = \{0\}$ we are done. Let, thus, $Z \in (\operatorname{Im} L)^\perp$, this means that

$$\int_M \langle Z, L(Y)\rangle = 0 \tag{6.7.35}$$

for all $Y \in H_{m+2}$. In particular (6.7.35) holds for all smooth Y, which implies that $L^\dagger(Z) = 0$ in a distributional sense. Now, the symbol of $L^\dagger$ is the transpose of the symbol of L, which shows that $L^\dagger$ is also elliptic. We can thus use elliptic regularity to conclude that Z is smooth, and $Z = 0$ follows.

The result in L^2 together with elliptic regularity immediately imply the result in H_k. □

•6.7.10

•6.7.10: **ptc:***one should add a discussion of the bERGER EBIN SPLITTING HERE*

6.7.2 The scalar constraint equation on compact manifolds

Theorem 6.7.4, together with Equation (6.7.29) and Remark 6.7.1, gives a reasonably complete description of the solvability of (6.7.18). We simply note that if B^{ij} there is smooth, then the associated solution will be smooth by elliptic regularity. To finish the presentation of the conformal method we need to address the question of existence of solutions of the Lichnerowicz equation (6.7.13). We will assume that $\Lambda = 0$ so that we have[•6.7.11]

*•6.7.11: **ptc:**it does not make sense to assume vanishing lambda, rather replace τ by τ_Λ in the equation, where τ_Λ is redefined as in (6.7.14), except that of course Jim's analysis only allows positive tau square*

$$\Delta_{\widetilde{g}}\phi - \frac{n-2}{4(n-1)}\widetilde{R}\phi = -\sigma^2\phi^{(2-3n)/(n-2)} + \frac{n-2}{4n}\tau^2\phi^{(n+2)/(n-2)} \ . \tag{6.7.36}$$

We note that the case $\sigma = 0$ corresponds to the so-called *Yamabe* equation; its solutions produce metrics with constant scalar curvature $(n-1)\tau^2$. We shall shortly give a complete answer to the question of solvability of (6.7.36), due to Isenberg [64], based on the known results about the Yamabe problem. Before doing that let us show that (6.7.36) cannot always be solved: this is seen by multiplying (6.7.36) by ϕ and integrating by parts

$$\begin{aligned}\int_M \phi(\Delta_{\widetilde{g}}\phi - \frac{n-2}{4(n-1)}\widetilde{R}\phi) &= -\int_M |\widetilde{D}\phi|^2_{\widetilde{g}} + \frac{n-2}{4(n-1)}\widetilde{R}\phi^2 \\ &= \int_M -\sigma^2\phi^{(2-3n)/(n-2)+1} + \frac{n-2}{4n}\tau^2\phi^{(n+2)/(n-2)+1}) \ .\end{aligned}$$

It then immediately follows:

PROPOSITION 6.7.5 *If $n \geq 3$, $\widetilde{R} \geq 0$ and $\sigma \equiv 0$, then (6.7.36) has no positive solutions unless $\widetilde{R} = \tau = 0$, in which case all solutions are constants.*

We emphasize that this is *not* a failure of the conformal method to produce solutions, but a *no-go* result; we will return to this issue shortly.[•6.7.12] On the positive side, we shall construct solutions of (6.7.36) under several circumstances, using the *monotone iteration scheme*, which we are going to describe now. For completeness we start by proving a simple version of the *maximum principle*:

*•6.7.12: **ptc:**is this true? think it over*

PROPOSITION 6.7.6 *Let (M, g) be compact, suppose that $c < 0$ and let $u \in C^2(M)$. If*

$$\Delta u + cu \geq 0 \ , \tag{6.7.37}$$

then $u \leq 0$. If equality in (6.7.37) holds then $u \equiv 0$.

PROOF: Suppose that u has a strictly positive maximum at p. In local coordinates around p we then have

$$g^{ij}\partial_i\partial_j u - g^{ij}\Gamma^k{}_{ij}\partial_k u \geq -cu \ .$$

The second term on the left-hand-side vanishes at p because ∂u vanishes at p, the first term is non-positive because at a maximum the matrix of second partial derivatives is non-positive definite. On the other hand the right-hand-side is strictly positive, which gives a contradiction. If equality holds in (6.7.37) then both u and minus u are non-positive, hence the result. □

Consider, now, the operator

$$L = \Delta_{\widetilde{g}} + c$$

for some $c < 0$. The symbol of L reads

$$\sigma_L(p) = g^{ij} p_i p_j \neq 0 \quad \text{if} \quad p \neq 0 ,$$

which shows that L is elliptic. It is well-known that $\Delta_{\tilde{g}}$ is formally self-adjoint (with respect to the measure $d\mu_{\tilde{g}}$), and Proposition 6.7.6 allows us to apply Theorem 6.7.4 to conclude existence of H_{k+2} solutions of the equation

$$Lu = \rho$$

for any $\rho \in H_k$, u smooth if ρ is smooth.

Returning to the Lichnerowicz equation (**??**), let us rewrite this equation in the form

$$\Delta_{\tilde{g}} \phi = F(\phi, x) . \tag{6.7.38}$$

A C^2 function ϕ_+ is called a *super-solution* of (6.7.38) if

$$\Delta_{\tilde{g}} \phi_+ \leq F(\phi_+, x) . \tag{6.7.39}$$

Similarly a C^2 function ϕ_- is called a *sub-solution* of (6.7.38) if

$$\Delta_{\tilde{g}} \phi_- \geq F(\phi_-, x) . \tag{6.7.40}$$

A solution is both a sub-solution and a super-solution. This shows that a necessary condition for existence of solutions is the existence of sub- and super-solutions. It turns out that this condition is also sufficient, modulo an obvious inequality between ϕ_- and ϕ_+:

THEOREM 6.7.7 *Suppose that (6.7.38) admits a sub-solution ϕ_- and a super-solution ϕ_+ satisfying*

$$\phi_- \leq \phi_+ .$$

If F is differentiable in ϕ, then there exists a C^2 solution ϕ of (6.7.38) such that

$$\phi_- \leq \phi \leq \phi_+ .$$

(ϕ is smooth if F is.)

PROOF: The argument is known as the *monotone iteration scheme*, or the *method of sub- and super-solutions.* We set

$$\phi_0 = \phi_+ ,$$

and our aim is to construct a sequence of functions such that

$$\phi_- \leq \phi_n \leq \phi_+ , \tag{6.7.41a}$$

$$\phi_{n+1} \leq \phi_n . \tag{6.7.41b}$$

We start by chosing c to be a positive constant large enough so that the function

$$\phi \to F_c(\phi, x) := F(\phi, x) - c\phi$$

is monotone decreasing for $\phi_- \leq \phi \leq \phi_+$. This can clearly be done on a compact manifold. By what has been said we can solve the equation

$$(\Delta_{\tilde{g}} - c)\phi_{n+1} = F_c(\phi_n, x) \ .$$

Clearly (6.7.41a) holds with $n = 0$. Suppose that (6.7.41a) holds for some n, then

$$\begin{aligned}(\Delta_{\tilde{g}} - c)(\phi_{n+1} - \phi_+) &= F_c(\phi_n, x) - \underbrace{\Delta_{\tilde{g}}\phi_+}_{\leq F(\phi_+, x)} - c\phi_+ \\ &\geq F_c(\phi_n, x) - F_c(\phi_+, x) \geq 0 \ ,\end{aligned}$$

by monotonicity of F_c. The maximum principle gives

$$\phi_{n+1} \leq \phi_+ \ ,$$

and induction establishes the second inequality in (6.7.41a). Similarly we have

$$\begin{aligned}(\Delta_{\tilde{g}} - c)(\phi_- - \phi_{n+1}) &= \underbrace{\Delta_{\tilde{g}}\phi_-}_{\geq F(\phi_-, x)} - c\phi_- - F_c(\phi_n, x) \\ &\geq F_c(\phi_-, x) - F_c(\phi_n, x) \geq 0 \ ,\end{aligned}$$

and (6.7.41a) is established. Next, we note that (6.7.41a) implies (6.7.41b) with $n = 0$. To continue the induction, suppose that (6.7.41b) holds for some $n \geq 0$, then

$$(\Delta_{\tilde{g}} - c)(\phi_{n+2} - \phi_{n+1}) = F_c(\phi_{n+1}, x) - -F_c(\phi_n, x) \geq 0 \ ,$$

again by monotonicity of F_c, and (6.7.41b) is proved.

Since ϕ_n is monotone decreasing and bounded there exists ϕ such that ϕ_n tends pointwise to ϕ as n tends to infinity. Continuity of F gives

$$F_n := F(\phi_n, x) \to F_\infty = F(\phi, x) \ ,$$

again pointwise. By the Lebesgue dominated theorem F_n converges to F_∞ in L^2, and the elliptic inequality (6.7.24) gives

$$\|\phi_n - \phi_m\|_{H^2} \leq C_2(\|(\Delta_{\tilde{g}} - c)(\phi_n - \phi_m)\|_{L^2} + \|\phi_n - \phi_m\|_{L^2} \ .$$

Completeness of H^2 implies that there exists $\phi_\infty \in H^2$ such that $\phi_n \to \phi_\infty$ in H^2. Recall that from any sequence converging in L^2 we can extract a subsequence converging pointwise almost everywhere, which shows that $\phi = \phi_\infty$ almost everywhere, hence $\phi \in H^2$. Continuity of $\Delta_{\tilde{g}} + c$ on H^2 shows that

$$(\Delta_{\tilde{g}} - c)\phi = \lim_{n\to\infty}(\Delta_{\tilde{g}} - c)\phi_n = F_c(\phi, x) = F(\phi, x) - c\phi \ ,$$

so that ϕ satisfies the equation, as desired. The remaining claims follow from the elliptic regularity theory. □

In order to apply Theorem 6.7.7 to the Lichnerowicz equation (**??**) we need appropriate sub- and super-solutions. We will seek those in the form of constants. Setting $\phi_- = \epsilon$, we need to show that

$$0 = \Delta_{\tilde g}\epsilon \geq F(\epsilon, x) \equiv \frac{n-2}{4(n-1)}\widetilde{R}\epsilon - \sigma^2\epsilon^{(2-3n)/(n-2)} + \frac{n-2}{4n}\tau^2\epsilon^{(n+2)/(n-2)} \tag{6.7.42}$$

for ϵ small enough. Since $2-3n$ is negative this inequality will clearly hold if $\widetilde{R}$ is strictly negative and ϵ is sufficiently small. Next, we set $\phi_+ = M$, with M again a constant, and we need to check that

$$0 \leq \frac{n-2}{4(n-1)}\widetilde{R}M - \sigma^2 M^{(2-3n)/(n-2)} + \frac{n-2}{4n}\tau^2 M^{(n+2)/(n-2)}\ . \tag{6.7.43}$$

Here we see that a sufficiently large constant will do as soon as $\tau \neq 0$. It follows that:

PROPOSITION 6.7.8 *The Lichnerowicz equation can always be solved if $\widetilde{R}$ is strictly negative and $\tau \neq 0$.*

A complete analysis of the Lichnerowicz equation can be carried through using the celebrated solution of the *Yamabe* problem. Recall that this is the problem of conformally rescaling a given metric so that the resulting new metric has constant scalar curvature. It should be recognised that making use of the solution of this problem sweeps the real difficulties under the carpet: as already mentioned, the Lichnerowicz equation with $\sigma = 0$ is nothing else but the Yamabe equation. Nevertheless, there remains some analysis to do even after the Yamabe part of the problem has been solved. •6.7.13

•*6.7.13:* **ptc:***Is Bacri's argument sufficiently simple to be included here?*

Before proceeding further let us classify the metrics on M as follows: we shall say that $g \in \mathscr{Y}^+$ if g can be conformally rescaled to achieve positive scalar curvature. We shall say that $g \in \mathscr{Y}^0$ if g can be conformally rescaled to achieve zero scalar curvature•6.7.14 but $g \notin \mathscr{Y}^+$. Finally, we let $\mathscr{Y}^-$ be the collection of the remaining metrics. It is easily seen that a manifold which is *not* in the positive Yamabe class can not have•6.7.15

•*6.7.14:* **ptc:***??*

•*6.7.15:* **ptc:***something fishy here; talk about it with Robert; mention obstruction to maximal surfaces; discuss the topology and Yamabe types*

One then has the following result of Isenberg [64]:

THEOREM 6.7.9 *The following table summarizes whether or not the Lichnerowicz equation (**??**) can be solved:*

	$\sigma^2 \equiv 0,\ \tau = 0$	$\sigma^2 \equiv 0,\ \tau \neq 0$	$\sigma^2 \not\equiv 0,\ \tau = 0$	$\sigma^2 \not\equiv 0,\ \tau \neq 0$
$\tilde g \in \mathscr{Y}^+$	*no*	*no*	*yes*	*yes*
$\tilde g \in \mathscr{Y}^0$	*yes*	*no*	*no*	*yes*
$\tilde g \in \mathscr{Y}^-$	*no*	*yes*	*no*	*yes*

For initial data in the class $(\mathscr{Y}^0, \sigma \equiv 0, \tau = 0)$ *all solutions are constants, and any positive constant is a solution. In all other cases the solutions are unique.*

As a Corollary of this theorem one obtains:

THEOREM 6.7.10 *Any compact manifold* M *carries some vacuum initial data set.*

PROOF: Let g be any metric on M which does not have any non-trivial discrete conformal isometries, then we can construct non-trivial solutions of the vector constraint equation using the method of Section 6.7.1. Choosing any $\tau \neq 0$ we can then solve the Lichnerowicz equation whatever the Yamabe type of g by Theorem 6.7.9. □

As already pointed out, we have the following:

PROPOSITION 6.7.11 *All CMC solutions of the vacuum constraint equation can be constructed by the conformal method.*

PROOF: A trivial, but not very enlightening proof goes as follows: if (M, g, K) be a vacuum initial data set, then the result is established by setting $Y = 0$, $\phi = 1$, $\widetilde{L}^{ij} = K^{ij}$.

The following proof seems to convey more information: fix a manifold M, and a conformal class of metric $[g]$. Using the method of Section 6.7.1 we obtain all symmetric, g–transverse, traceless tensor fields by varying the seed field B^{ij}. The conformal covariance of the vector constraint equation with $d\mathrm{tr}_g K = 0$ shows that we will have also obtained all tensors which are symmetric, $\widetilde{g}$–transverse, traceless, for any metric $\widetilde{g}$ which is in the conformal class of g. •6.7.16 □

•*6.7.16:* **ptc:***needs a thought, and finishing: there is an issue with the kernel here*

Proposition 6.7.11 highlights the importance of Isenberg's theorem 6.7.4.

6.7.3 The thin sandwich (to be done)

•6.7.17

•*6.7.17:* **ptc:***this can be more or less cut and pasted from Isenberg Bartnik*

6.7.4 The conformal thin sandwich (to be done)

•6.7.18

•*6.7.18:* **ptc:***this can be more or less cut and pasted from Isenberg Bartnik*

6.7.5 Asymptotically hyperboloidal initial data (to be done)

6.8 Penrose-Newman-Christodoulou-Klainerman conformal equations (to be done)

6.9 Anderson's conformal equations (to be done)

6.10 Maximal globally hyperbolic developments (to be done)

6.11 Strong cosmic censorship? (to be done)

Chapter 7

Positive energy theorems

7.1 The mass of asymptotically Euclidean manifolds

•7.1.1

•7.1.1: **ptc:**_add the Li Tam theorem, and the riemann curvature estimates of Finster and friends_

There exist various approaches to the definition of mass in general relativity, the first one being due to Einstein [40] himself. In Section 7.13 below we will outline two geometric Hamiltonian approaches to that question. However, those approaches require some background knowledge in symplectic field theory, and it appears useful to present an elementary approach which quickly leads to the correct definition for asymptotically flat Riemannian manifolds without any prerequisites.

In the remainder in this chapter we will restrict ourselves to dimensions greater than or equal to three, as the situation turns out to be completely different in dimension two: Indeed, it should be clear from the considerations below that the mass is an object which is related to the integral of the scalar curvature over the manifold. Now, in dimension two, that integral is a topological invariant for compact manifolds, while it is related to a *"deficit angle"* in the non-compact case. This angle, which will be discussed in detail in Remark 7.1.3 below, appears thus to be the natural two-dimensional equivalent of the notion of mass.

The Newtonian approximation provides the simplest situation in which it is natural to assign a mass to a Riemannian metric: recall that in this case the space-part of the metric takes the form

$$g_{ij} = (1+2\phi)\delta_{ij} \;, \tag{7.1.1}$$

where ϕ is the Newtonian potential,

$$\Delta_\delta \phi = -4\pi\mu \;, \tag{7.1.2}$$

with μ – the energy density; compare Section 3.4.5. When μ has compact support $\operatorname{supp}\mu \subset B(0,R)$ we have

$$\phi = \frac{M}{r} + O(r^{-2}) \;, \tag{7.1.3}$$

where M is the total Newtonian mass of the sources:

$$\begin{aligned} M &= \int_{\mathbb{R}^3} \mu \, d^3x \\ &= -\frac{1}{4\pi}\int_{\mathbb{R}^3} \Delta_\delta\phi \\ &= -\frac{1}{4\pi}\int_{B(0,R)} \Delta_\delta\phi \\ &= -\frac{1}{4\pi}\int_{S(0,R)} \nabla^i\phi \, dS_i \\ &= -\lim_{R\to\infty}\frac{1}{4\pi}\int_{S(0,R)} \nabla^i\phi \, dS_i \;. \end{aligned} \tag{7.1.4}$$

Here dS_i denotes the usual coordinate surface element,

$$dS_i = \partial_i \rfloor dx \wedge dy \wedge dz \;, \tag{7.1.5}$$

with $\rfloor$ denoting contraction. Then the number M appearing in (7.1.3) or, equivalently, given by (7.1.4), will be called the mass of the metric (7.1.1).

In Newtonian theory it is natural to suppose that $\mu \geq 0$. We then obtain the simplest possible version of the *positive mass theorem*:

THEOREM 7.1.1 (Conformally flat positive mass theorem) *Consider a C^2 metric on $\mathbb{R}^3$ of the form (7.1.1) with a strictly positive function $1+2\phi$ satisfying*

$$-4\pi\mu := \Delta_\delta \phi \leq 0 \;, \qquad \phi \to_{r\to\infty} 0 \;.$$

Then

$$0 \leq m := -\lim_{R\to\infty} \frac{1}{4\pi} \int_{S(0,R)} \nabla^i \phi \, dS_i \leq \infty \;,$$

with m vanishing if and only if g_{ij} is flat.

PROOF: The result follows from (7.1.4); we simply note that m will be finite if and only if μ is in $L^1(\mathbb{R}^3)$. $\square$

Somewhat more generally, suppose that

$$g_{ij} = \psi\delta_{ij} + o(r^{-1}) \;, \quad \partial_k(g_{ij} - \psi\delta_{ij}) = o(r^{-2}) \;, \tag{7.1.6}$$

with ψ tending to 1 as r tends to infinity. Then a natural generalisation of (7.1.4) is

$$m := -\lim_{R\to\infty} \frac{1}{8\pi} \int_{S(0,R)} \nabla^i \psi \, dS_i \;, \tag{7.1.7}$$

provided that the limit exists.

Let us see whether Definition (7.1.7) can be applied to the Schwarzschild metric:

$$^4g = -(1-2m/r)dt^2 + \frac{dr^2}{1-2m/r} + r^2 d\Omega^2 \;, \tag{7.1.8}$$

where

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\varphi^2 \;. \tag{7.1.9}$$

Here we have decorated ^4g with a subscript four, emphasising its four dimensional character, and we shall be using the symbol g for its three dimensional space-part. Now, every spherically symmetric metric is conformally flat, so that the space-part of the Schwarzschild metric can be brought to the form (7.1.6) *without the error term*, as follows: We want to find ρ such that

$$g := \frac{dr^2}{1-2m/r} + r^2 d\Omega^2 = \psi\left(d\rho^2 + \rho^2 d\Omega^2\right) \;. \tag{7.1.10}$$

Let us check that the answer is

$$\psi = \left(1 + \frac{m}{2\rho}\right)^4 \;.$$

Comparing the coefficients in front of $d\theta^2$, or in front of $d\varphi^2$, in (7.1.10) yields

$$r = \left(1 + \frac{m}{2\rho}\right)^2 \rho \;. \tag{7.1.11}$$

To finish verifying (7.1.10) it suffices to check the g_{rr} term. Differentiating we have

$$dr = \left(1+\frac{m}{2\rho}\right)\left(2\times\left(-\frac{m}{2\rho^2}\right)\times\rho+1+\frac{m}{2\rho}\right)d\rho = \left(1+\frac{m}{2\rho}\right)\left(1-\frac{m}{2\rho}\right)d\rho\ , \tag{7.1.12}$$

while

$$\begin{aligned} 1-\frac{2m}{r} &= 1-\frac{2m}{(1+\frac{m}{2\rho})^2\rho} \\ &= \frac{(1+\frac{m}{2\rho})^2-\frac{2m}{\rho}}{(1+\frac{m}{2\rho})^2} \\ &= \frac{1+\frac{m}{\rho}+\left(\frac{m}{2\rho}\right)^2-\frac{2m}{\rho}}{(1+\frac{m}{2\rho})^2} \\ &= \frac{(1-\frac{m}{2\rho})^2}{(1+\frac{m}{2\rho})^2}\ , \end{aligned}$$

and (7.1.10) readily follows. Hence

$$g = \left(1+\frac{m}{2|\vec{y}|}\right)^4\delta\ , \tag{7.1.13}$$

where δ denotes the flat Euclidean metric in the coordinate system (y^i). From the asymptotic development

$$\left(1+\frac{m}{2|\vec{y}|}\right)^4 = 1+\frac{2m}{|\vec{y}|}+O(|\vec{y}|^{-2})$$

we find that the space-part of the Schwarzschild metric has mass m, as desired. More precisely, one finds a mass m in the coordinate system in which g takes the form (7.1.13). This raises immediately the question, whether the number so obtained does, or does not, depend upon the coordinate system chosen to calculate it. We will shortly see that m is coordinate-independent, and indeed a geometric invariant.

For further reference we note that we have also obtained

$${}^4g = -\frac{(1-\frac{m}{2\rho})^2}{(1+\frac{m}{2\rho})^2}dt^2+\left(1+\frac{m}{2\rho}\right)^4\left(d\rho^2+\rho^2(d\theta^2+\sin^2\theta d\varphi^2)\right)\ . \tag{7.1.14}$$

Actually, (7.1.12) shows that the map which to ρ assigns r is not a diffeomorphism, since $dr/d\rho$ vanishes at

$$r=2m \Longleftrightarrow \rho = m/2\ .$$

A closer inspection of (7.1.11)-(7.1.12) shows that the manifold $\mathbb{R}_t\times\{\rho>0\}\times S^2$ contains the Schwarzschild manifold $\mathbb{R}_t\times\{r>2m\}\times S^2$ twice, once for $\rho>m/2$, and one more copy for $\rho<m/2$; we will return to this later, in Section **??**.

A hint how to proceed in general is given by the conformally flat positive energy theorem 7.1.1, where we have used positivity properties of the "mass density $\mu := \Delta_\delta\phi/(-4\pi)$" to obtain information about the asymptotic behavior of the metric. Recall that the general relativistic correspondent of the mass density μ is the energy density ρ, see (3.4.26). Thus, we need an equation which involves ρ. A candidate here is the *scalar constraint equation* (6.3.17),

$$R(g) = 8\pi\rho + |K|^2 - (\mathrm{tr}_g K)^2\ , \qquad \rho := T_{\mu\nu}n^\mu n^\nu. \tag{7.1.15}$$

(Recall that we are working in the asymptotically flat context here, which requires $\Lambda = 0$.) Here n^μ is the field of unit normals to the spacelike initial data hypersurface $\mathscr{S} \subset M$, with space metric g induced from the space-time metric ${}^4g$. Further, $T_{\mu\nu}$ is the energy-momentum tensor, so that ρ has the interpretation of energy-per-unit-volume of matter fields on $\mathscr{S}$.

Now, R contains a linear combination of second derivatives of g, which is vaguely reminiscent of (7.1.2), however there are also terms which are quadratic in the Christoffel symbols, and it is not completely clear that this is the right equation. We shall, however, hope for the best, manipulate the equation involving $R(g)$, and see what comes out of that. Thus, we isolate all the second derivatives terms in $R(g)$ and we reexpress them as the divergence of a certain object:

$$\begin{aligned} R(g) &= g^{ij}\mathrm{Ric}_{ij} \\ &= g^{ij}R^k{}_{ikj} \\ &= g^{ij}\left(\partial_k\Gamma^k{}_{ij} - \partial_j\Gamma^k{}_{ik} + q\right)\ , \end{aligned}$$

where q denotes an object which is quadratic in the first derivatives of g_{ij} with coefficients which are rational functions of g_{kl}. Now,

$$\Gamma^k{}_{ij} = \frac{1}{2}g^{k\ell}\left(\partial_j g_{\ell i} + \partial_i g_{\ell j} - \partial_\ell g_{ij}\right)\ ,$$

hence

$$\Gamma^k{}_{ik} = \frac{1}{2}g^{k\ell}\left(\partial_k g_{\ell i} + \partial_i g_{\ell k} - \partial_\ell g_{ik}\right) = \frac{1}{2}g^{k\ell}\partial_i g_{\ell k}\ .$$

It follows that

$$\begin{aligned} R(g) &= \frac{1}{2}g^{ij}g^{k\ell}\left(\partial_k\partial_j g_{\ell i} + \partial_k\partial_i g_{\ell j} - \partial_k\partial_\ell g_{ij} - \partial_j\partial_i g_{\ell k} + q\right) \\ &= g^{ij}g^{k\ell}\left(\partial_k\partial_j g_{\ell i} - \partial_j\partial_i g_{\ell k}\right) + \frac{q}{2} \\ &= \partial_j\left(g^{ij}g^{k\ell}\left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right)\right) + q'\ , \end{aligned}$$

with a different quadratic remainder term. We will need to integrate this expression, so we multiply everything by $\sqrt{\det g}$, obtaining finally

$$\sqrt{\det g}R(g) = \partial_j\mathbb{U}^j + q''\ , \tag{7.1.16}$$

with q'' yet another quadratic expression in ∂g, and

$$\boxed{\mathbb{U}^j := \sqrt{\det g}\ g^{ij}g^{k\ell}\left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right)\ .} \tag{7.1.17}$$

This is the object needed for the definition of mass:

DEFINITION 7.1.2 *Let g be a $W^{1,\infty}_{\mathrm{loc}}$ metric defined on $\mathbb{R}^3 \setminus B(0,R_0)$, we set*

$$\begin{aligned} m &:= \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} \mathbb{U}^j dS_j \\ &= \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} g^{ij} g^{k\ell} \left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right) \sqrt{\det g}\, dS_j \ , \end{aligned} \tag{7.1.18}$$

whenever the limit exists.

We emphasize that we do not assume that the metric is globally defined on $\mathbb{R}^3$, as that would exclude many cases of interest, including the Schwarzschild metric.

REMARK 7.1.3 In dimension two the scalar curvature is always, locally, a total divergence, so that there is no remainder term in (7.1.16), which considerably simplifies the subsequent analysis. We shall say that a two-dimensional manifold M is *finitely connected* if M is diffeomorphic to a compact boundaryless manifold N from which a finite non-zero number of points has been removed. Equivalently, M is diffeomorphic to the union of a compact set with a finite number of exterior regions diffeomorphic to $\mathbb{R}^2 \setminus B(0,R_i)$. Let p be any point in M and let $S_p(t)$ and $B_p(t)$ be the geodesic sphere and ball around p:

$$S_p(t) := \{q \in M : d_g(p,q) = t\} \ , \quad B_p(t) = \{q \in M : d_g(p,q) < t\} \ .$$

We will denote by $L_p(t)$ the length of $S_p(t)$ and by $A_p(t)$ the area of $B_p(t)$. We have the following theorem of Shiohama [99]:•7.1.2

•7.1.2: **ptc:***give proof*

THEOREM 7.1.4 *Let (M,g) be a complete, non-compact, finitely connected two dimensional manifold. If*

$$R(g) \in L^1(M) \ ,$$

then

$$\lim_{t\to\infty} \frac{L_p(t)}{t} = \lim_{t\to\infty} \frac{A_p(t)}{t^2} = 2\pi\chi(M) - \int_M R\, d\mu_g \ . \tag{7.1.19}$$

There are several interesting consequences of this result. First, one notices that the right-hand-side of (7.1.18) does not depend upon p, so that the first two terms are also p–independent. Next, since the left-hand-side (7.1.18) is non-negative, if g is a complete metric on $\mathbb{R}^2$ we obtain the Cohn-Vossen inequality [**?**]

$$\int_M R\, d\mu_g \le 2\pi \ ,$$

with equality if and only if the metric is flat. In order to gain more insight into (7.1.19), consider metrics which are asymptotically Euclidean on a finite number of ends M_i, $i = 1,\ldots,I$, with each of the M_i's diffeomorphic to $[R_i,\infty)\times S^1$, and with the metric asymptotically approaching a flat metric on a cone on M_i:

$$g(\omega_i) = dr^2 + r^2 \left(\frac{\omega_i}{2\pi}\right)^2 d\varphi^2$$

for some positive constant ω_i. Here we parameterize S^1 by an angular variable $\varphi \in [0,2\pi]$, so that the circles $r =$const have $g(\omega_i)$–length equal to ω_i. Under very mild conditions on the convergence of g to $g(\omega_i)$ we will have

$$A(S(t)\cap M_i) = \omega_i t^2 + o(t^2) \ ,$$

for t large. In the simplest case $M = \mathbb{R}^2$ we then obtain

$$\int_M R\, d\mu_g = 2\pi - \omega \ ,$$

with $\omega = \omega_1$. Hence, the integral of R equals the *deficit angle* $2\pi - \omega$. This leads to the *two dimensional positive energy theorem*: for asymptotically flat metrics on $\mathbb{R}^2$ the deficit angle is strictly positive when $R \geq 0$.

More generally, if I is the number of ends, then we have the following relation between the deficit angles and the integral of scalar curvature•7.1.3

•*7.1.3:* **ptc:** *does the first term have a sign?*

$$\int_M R\, d\mu_g = 2\pi(\chi(M) - I) + \sum_{i=1}^{I} (2\pi - \omega_i) \ .$$

It would be of interest to enquire whether the individual deficit angles do have a sign.•7.1.4

•*7.1.4:* **ptc:***an interesting exercise?*

The first question we address is that of convergence of the integral (7.1.18):

Proposition 7.1.5 ([10, 27, 85]) *Let g be a $W^{1,\infty}_{\mathrm{loc}}$ metric defined on $\mathbb{R}^3 \backslash B(0, R_0)$ such that*

$$\forall\ i,j,k,\ell \qquad g_{ij}\ ,\ g^{k\ell} \in L^\infty\ , \qquad \partial_k g_{ij} \in L^2\ . \tag{7.1.20}$$

1. *If*
$$R(g) \in L^1\ ,$$
then m exists, and is finite.

2. [Infinite positive energy theorem] *If $R(g)$ is a non-negative measurable function which is* not *in L^1, then the limit in (7.1.18) exists with*
$$m = \infty\ .$$

Proof: The result follows immediately from the divergence theorem: we write

$$\begin{aligned}
\int_{S(0,R_0)} \mathbb{U}^j dS_j - \int_{S(0,R_)} \mathbb{U}^j dS_j &= \int_{B(0,R)\backslash B(0,R_0)} \partial_j \mathbb{U}^j d^3x \\
&= \int_{B(0,R)\backslash B(0,R_0)} (\sqrt{\det g} R - q'') d^3x\ ,
\end{aligned}$$

with $q'' \in L^1$ since the $\partial_k g_{ij}$'s are in L^2. If $R(g)$ is in L^1, or if $R(g)$ is measurable and positive, the monotone convergence theorem gives

$$\begin{aligned}
\lim_{R\to\infty} \int_{S(0,R)} \mathbb{U}^j dS_j &= \int_{\mathbb{R}^3\backslash B(0,R_0)} \sqrt{\det g} R\ d^3x \\
&\quad - \int_{\mathbb{R}^3\backslash B(0,R_0)} q'' d^3x + \int_{S(0,R_0)} \mathbb{U}^j dS_j\ ,
\end{aligned} \tag{7.1.21}$$

with the last two terms being finite, and the result follows. $\square$

Since the arguments of this section have a purely Riemannian character, the extrinsic curvature tensor K, which would be present if a whole initial data set were considered, is irrelevant for the current purposes. However, it is worthwhile pointing out that similar manipulations can be done with the vector constraint equation, leading to the definition of the *ADM momentum* of an initial data set, as follows: For notational convenience let us set

$$P^{ij} := \mathrm{tr}_g K g^{ij} - K^{ij} , \qquad (7.1.22)$$
$$J^j := T^j{}_\mu n^\mu , \qquad (7.1.23)$$

so that the vector constraint equation (6.3.25) can be rewritten as

$$D_i P^i{}_j = 8\pi J_j . \qquad (7.1.24)$$

The vector field J is usually called the matter momentum vector. Similarly to (7.1.16), we want to obtain a divergence identity involving J. Now, divergence identities involve vector fields, while (7.1.24) involves the divergence of a tensor; this is easily taken care of by choosing some arbitrary vector field X and writing

$$D_i(P^i{}_j X^j) = D_i P^i{}_j X^i + P^i{}_j D_i X^j = 8\pi J_i X^i + P^i{}_j D_i X^j . \qquad (7.1.25)$$

Integrating over large spheres gives

$$\begin{aligned}\int_{S^\infty} P^i{}_j X^j dS_i &= \lim_{R\to\infty} \int_{S(R)} P^i{}_j X^j \\ &= \int_M 8\pi J^i X_i + P^i{}_j D_i X^j , \end{aligned} \qquad (7.1.26)$$

provided that the last integral converges. Let X^i_∞ be any set of constants, the ADM momentum vector p is the set of numbers p_i defined using the boundary integrand above:

$$p_i X^i_\infty := \frac{1}{8\pi} \int_{S^\infty} P^i{}_j X^j_\infty dS_i . \qquad (7.1.27)$$

To analyse convergence, let X be any differentiable vector field which coincides with X_∞ for r large, and which is zero outside of the asymptotic region. It is natural to assume that the total momentum of the fields other than the gravitational one is finite:

$$J \in L^1(M_{\mathrm{ext}}) ,$$

this ensures convergence of the J integral in (7.1.26). The convergence of the second term there is usually taken care of by requiring that

$$P^{ij} , \ \partial_k g_{ij} \in L^2(M_{\mathrm{ext}}) . \qquad (7.1.28)$$

For then we have, for r large,

$$P^{ij} D_i X_j = P^i{}_j D_i X^j = P^i{}_j (\underbrace{\partial_i X^j_\infty}_{=0} + \Gamma^j{}_{ik} X^k_\infty) \le C|P| \sum_{i,j,k} |\partial_i g_{jk}| .$$

Integrating over M and using $2ab \le a^2 + b^2$ gives

$$\left| \int_{M_{\mathrm{ext}}} P^{ij} D_i X_j \right| = \left| \int_{M_{\mathrm{ext}}} P^{ij} D_i X_j \right| \le C \int_{M_{\mathrm{ext}}} \Big(|P|^2 + \sum_{i,j,k} |\partial_i g_{jk}|^2 \Big) ,$$

and convergence follows. We have thus proved

PROPOSITION 7.1.6 *Suppose that*

$$J \in L^1(M_{\rm ext}) , \qquad P^{ij} \ , \ \partial_k g_{ij} \in L^2(M_{\rm ext}) .$$

Then the ADM momentum (7.1.27) is finite.

It seems sensible to test our definition on a few examples. First, if g is the flat Euclidean metric on $\mathbb{R}^3$, and we use the standard Euclidean coordinates, then $m = 0$, which appears quite reasonable. Consider, next, the space-part of the Schwarzschild metric: whether in the form (7.1.10) or (7.1.13) it can be written as

$$g_{ij} = \delta_{ij} + O(r^{-1}) \ , \qquad \text{with} \ \ \partial_k g_{ij} = O(r^{-2}) \tag{7.1.29}$$

(for (7.1.13) this is straightforward; for (7.1.10) one should introduce the obvious pseudo-Euclidean coordinates x^i associated to the spherical coordinates (r, θ, φ). We will use the scalar constraint equation to calculate $R(g)$; this requires calculating the extrinsic curvature tensor K_{ij}. Recall that (see (6.3.9))

$$K(X, Y) := g(P(\nabla_X n), Y) \ ,$$

where P is the orthogonal projection on the space tangent to the hypersurface in consideration; in our case these are the hypersurfaces $t = \text{const}$. From (7.1.8) the field of unit conormals $n_\mu dx^\mu$ to those hypersurfaces takes the form

$$n_\mu dx^\mu = \sqrt{1 - 2m/r} dt \ .$$

Further,

$$P(X^\mu \partial_\mu) = X^i \partial_i \ .$$

Let $X = P(X)$ so that $X = X^i \partial_i$, we calculate

$$\begin{aligned} \nabla_X n_k &= X(\underbrace{n_k}_{=0}) - \Gamma^\nu_{\alpha k} n_\nu X^\alpha \\ &= -n_0 \Gamma^0_{ik} X^i \ . \end{aligned} \tag{7.1.30}$$

Further

$$\Gamma^0_{ik} = \frac{1}{2} {}^4 g^{0\mu} (\partial_i \underbrace{{}^4 g_{0k}}_{=0} + \partial_k \underbrace{{}^4 g_{0i}}_{=0} - \partial_\mu {}^4 g_{ik}) \tag{7.1.31a}$$

$$= -\frac{1}{2} {}^4 g^{00} \partial_0 {}^4 g_{ik} = 0 \ , \tag{7.1.31b}$$

hence

$$K_{ij} = 0 \ .$$

The scalar constraint equation (7.1.15) gives now

$$R(g) = 0 \ .$$

This is obviously in L^1, while r^{-2} is in L^2 on $\mathbb{R}^3 \setminus B(0, 1)$ (since r^{-4} is in $L^1(\mathbb{R}^3 \setminus B(0, 1))$), and convergence of m follows from Proposition 7.1.5. In order to calculate the value of m it is convenient to derive a somewhat simpler form of (7.1.18): generalising somewhat (7.1.29), suppose that

$$g_{ij} = \delta_{ij} + o(r^{-1/2}) , \quad \text{with } \partial_k g_{ij} = O(r^{-3/2}) . \tag{7.1.32}$$

This choice of powers is motivated by the fact that the power $r^{-3/2}$ is the borderline power to be in $L^2(\mathbb{R}^3 \setminus B(0,1))$: the function $r^{-\sigma}$ with $\sigma > 3/2$ will be in L^2, while if $\sigma = 3/2$ it will not. Under (7.1.32) we have

$$\begin{aligned} 16\pi m(R) &:= \int_{S(0,R)} g^{ij} g^{k\ell} \left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right) \sqrt{\det g}\, dS_j \\ &= \int_{S(0,R)} \Big(\delta^{ij} + o(r^{-1/2})\Big) \Big(\delta^{k\ell} + o(r^{-1/2})\Big) \underbrace{\left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right)}_{O(r^{-3/2})} \underbrace{\sqrt{\det g}}_{1+o(r^{-1/2})}\ dS_j \\ &= \int_{S(0,R)} \delta^{ij}\delta^{k\ell} \left(\partial_k g_{\ell i} - \partial_i g_{\ell k}\right)\ dS_j + o(1) , \end{aligned}$$

so that

$$m = m_{ADM} := \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(0,R)} \left(\partial_\ell g_{\ell i} - \partial_i g_{\ell\ell}\right) \sqrt{\det g}\, dS_i . \tag{7.1.33}$$

This formula is known as the *Arnowitt–Deser–Misner* (ADM) expression for the mass of the gravitational field at spatial infinity.

Returning to the Schwarzschild metric consider, first, (7.1.13), or — more generally – metrics which are conformally flat:

$$g_{ij} = (1+2\phi)\delta_{ij} \implies \partial_\ell g_{\ell i} - \partial_i g_{\ell\ell} = 2(\partial_\ell \psi \delta_{\ell i} - \partial_i \phi \underbrace{\delta_{\ell\ell}}_{=3}) = -4\partial_i \phi , \tag{7.1.34}$$

and (7.1.33) reduces to (7.1.4), as desired. The original form given by the left-hand-side of (7.1.10) requires some more work. Again generalising somewhat, we consider general spherically symmetric metrics

$$g = \phi(r) dr^2 + \chi(r) r^2 d\Omega^2 , \tag{7.1.35}$$

with ϕ, χ differentiable, tending to one as r goes to infinity at rates compatible with (7.1.32):

$$\phi - 1 = o(r^{-1/2}) , \quad \chi - 1 = o(r^{-1/2}) , \quad \partial_r \phi = O(r^{-3/2}) , \quad \partial_r \chi = O(r^{-3/2}) . \tag{7.1.36}$$

We need to reexpress the metric in the pseudo-Cartesian coordinate system associate to the spherical coordinate system (r,θ,φ):

$$x = r\sin\theta\cos\varphi ,\ y = r\sin\theta\sin\varphi ,\ z = r\cos\theta . \tag{7.1.37}$$

We have

$$\begin{aligned} g &= \phi\, dr^2 + \chi(dr^2 + r^2 d\Omega^2) - \chi\, dr^2 \\ &= \chi\,\delta + (\phi - \chi) dr^2 \\ &= \chi\,\delta + (\phi-\chi)(\sum_i \frac{x^i}{r} dx^i)^2 , \end{aligned}$$

so that
$$g_{ij} = \chi\delta_{ij} + \frac{(\phi-\chi)x^ix^j}{r^2} \; .$$
The contribution of the first term to the ADM integral (7.1.33) is obtained from the calculation in (7.1.34), while the second one gives
$$\begin{aligned}
&\left[\partial_\ell\left(\frac{(\phi-\chi)x_\ell x_i}{r^2}\right) - \partial_i\left(\frac{(\phi-\chi)x_\ell x_\ell}{r^2}\right)\right]\frac{x^i}{r} \\
&\qquad = (\phi'-\chi') + \Big[(\phi-\chi)\partial_\ell\left(\frac{x_\ell x_i}{r^2}\right) - \underbrace{\partial_i(\phi-\chi)\Big]\frac{x^i}{r}}_{=\phi'-\chi'} \\
&\qquad = (\phi-\chi)\left(\frac{3x^i + x^i - 2x^i}{r^2}\right)\frac{x^i}{r} = 2\frac{\phi-\chi}{r} \; .
\end{aligned}$$
Summing it all up, we obtain the following expression for the ADM mass of a spherically symmetric metric (7.1.35) satisfying (7.1.36):
$$\begin{aligned}
m &= \lim_{R\to\infty}\frac{1}{16\pi}\int_{S(0,R)}\left(-2r^2\chi' + 2r(\phi-\chi)\right)\, d^2S \\
&= \lim_{r\to\infty}\frac{1}{2}\left(-r^2\chi' + r(\phi-\chi)\right) \; . \qquad (7.1.38)
\end{aligned}$$
For the original form of the Schwarzschild metric we have $\chi \equiv 1$ and $\phi = 1/(1-2m/r)$, yielding again the value m for the ADM mass of g.

As another example of calculation of the ADM mass, consider the Kasner metrics on $\{t>0\}\times\mathbb{R}^3$:
$$^4g = -dt^2 + t^{2p_1}dx^2 + t^{2p_2}dy^2 + t^{2p_3}dz^2 \; . \qquad (7.1.39)$$
The metric (7.1.39) is vacuum provided that
$$p_1+p_2+p_3 = p_1^2+p_2^2+p_3^2 = 1 \; . \qquad (7.1.40)$$
All slices $t=$ const are flat, each of them has thus vanishing ADM mass. This seems to be extremely counter-intuitive, because the metric is highly dynamical. In fact, one would be tempted to say that it has infinite kinetic energy: Indeed, let us calculate the extrinsic curvature tensor of the $t=$ const slices: from (7.1.30)–(7.1.31a) we have
$$\begin{aligned}
K &= \nabla_i n_k \, dx^i dx^k \\
&= \frac{1}{2}\partial_t g_{ik}\, dx^i dx^k \\
&= p_1 t^{2p_1-1}dx^2 + p_2 t^{2p_2-1}dy^2 + p_3 t^{2p_3-1}dz^2 \; . \qquad (7.1.41)
\end{aligned}$$
At each value of t we obtain thus a tensor field with entries which are constant in space. The problem here is that while the space slices of the Kasner space-time are asymptotically Euclidean, the space-time metric itself is not asymptotically flat in any sensible way. This example suggests that a physically meaningful notion of total mass can only be obtained for metrics which satisfy asymptotic flatness conditions in a space-time sense; we will return to this question in Section **??** below.

7.2 Coordinate independence

The next example is due to Denissov and Solovyev [39]: let δ be the Euclidean metric on $\mathbb{R}^3$ and introduce a new coordinate system (ρ,θ,ϕ) by changing the radial variable r to

$$r=\rho+c\rho^{1-\alpha}\ , \tag{7.2.1}$$

with some constants $\alpha>0$, $c\in\mathbb{R}$. This gives

$$dr^2+r^2d\Omega^2=(1+(1-\alpha)c\rho^{-\alpha})^2d\rho^2+(1+c\rho^{-\alpha})^2\rho^2d\Omega^2\ .$$

This is of the form (7.1.35) with

$$\phi(\rho)=(1+(1-\alpha)c\rho^{-\alpha})^2\ ,\quad \chi(\rho)=(1+c\rho^{-\alpha})^2\ ,$$

so we can apply (7.1.38):

$$\begin{aligned}
-\rho^2\chi'+\rho(\phi-\chi) &= 2c\alpha\rho^{-\alpha+1}(1+c\rho^{-\alpha})+\rho\left((1+(1-\alpha)c\rho^{-\alpha})^2-(1+c\rho^{-\alpha})^2\right)\\
&= 2c\alpha\rho^{-\alpha+1}(1+c\rho^{-\alpha})+\rho\left((1+c\rho^{-\alpha}-\alpha c\rho^{-\alpha})^2-(1+c\rho^{-\alpha})^2\right)\\
&= 2c\alpha\rho^{-\alpha+1}(1+c\rho^{-\alpha})+\rho\left(-2\alpha c\rho^{-\alpha}(1+c\rho^{-\alpha})+\alpha^2c^2\rho^{-2\alpha}\right)\\
&= \alpha^2c^2\rho^{1-2\alpha}.
\end{aligned}$$

It follows that

$$\begin{aligned}
m_{ADM} &= \lim_{\rho\to\infty}\frac{1}{2}\left(-\rho^2\chi'+\rho(\phi-\chi)\right)\\
&= \begin{cases}\infty\ , & \alpha<1/2\ ,\\ c^2/8\ , & \alpha=1/2\ ,\\ 0\ , & \alpha>1/2\ .\end{cases}
\end{aligned}\tag{7.2.2}$$

Let y^i denote the coordinate system associated to the angular variables (ρ,θ,φ) by replacing r with ρ in (7.1.37). Then the exponent α in (7.2.1) dictates the rate at which the metric components approach δ_{ij}:

$$\delta_{ij}dx^idx^j=g_{ij}dy^idy^j\ ,\ \text{with } g_{ij}-\delta_{ij}=O(\rho^{-\alpha})\ ,\ \partial_kg_{ij}=O(\rho^{-\alpha-1})\ .$$

Note that above we have calculated the ADM mass integral (7.1.33), rather than the original integral (7.1.18). We have already seen that both integrals coincide if $\alpha>1/2$ (compare (7.1.32)), but they do not necessarily do that for $\alpha\le 1/2$. One can similarly calculate the mass m of (7.1.18) obtaining an identical conclusion: the mass m of the flat metric in the coordinate system y^i is infinite if $\alpha<1/2$, can have an arbitrary positive value depending upon c if $\alpha=1/2$, and vanishes for $\alpha>1/2$. The lesson of this is that the mass appears to depend upon the coordinate system chosen, even within the class of coordinate systems in which the metric tends to a constant coefficients matrix as r tends to infinity.

The reader will notice that for $\alpha=1/2$ the metric does not satisfy the conditions of Proposition 7.1.5, as the derivatives of g_{ij} in the new coordinate system will not be in L^2. It follows that the conditions of Proposition 7.1.5 are not necessary for the existence of those limits, though they seem to be very close to be optimal, since — as shown above — allowing α's smaller than 1/2 leads to infinite mass representations for Euclidean space.

In order to clarify the question of dependence of the mass upon coordinates it is useful to include those coordinate systems explicitly in the notation (we follow the approach of [27]). Consider, thus, a pair (g, ϕ), where

1. g is a Riemannian metric on a three dimensional manifold N, N diffeomorphic to $\mathbb{R}^3 \setminus B(R)$, where $B(R)$ is a closed ball. N should be thought of as one of (possible many) "asymptotic ends" of M.•7.2.1

•7.2.1: **ptc:***this should be done on* $\mathbb{R}^n$

2. ϕ is a coordinate system on the complement of a compact set K of N such that, in local coordinates $\phi^i(p) = x^i$ the metric takes the following form:
$$g_{ij} = \delta_{ij} + h_{ij} \ , \tag{7.2.3}$$
with h_{ij} satisfying
$$\forall_{i,j,k} \qquad |h_{ij}| \leq \ c(r+1)^{-\alpha} \ , \quad |\frac{\partial h_{ij}}{\partial x^k}| \leq \ c(r+1)^{-\alpha-1} \ , \tag{7.2.4}$$
for some constant $c \in \mathbb{R}$, where $r(x) = (\sum (x^i)^2 \ )^{1/2}$.

3. Finally, g_{ij} is uniformly equivalent to the flat metric δ: there exists a constant C such that
$$\forall X^i \in \mathbb{R}^3 \qquad C^{-1} \sum (X^i)^2 \leq g_{ij} \ X^i X^j \leq C \sum (X^i)^2 \ . \tag{7.2.5}$$

Such a pair (g, ϕ) will be called α–admissible.

We note that (7.2.5) is equivalent to the requirement that all the g_{ij}'s and g^{ij}'s are uniformly bounded: indeed, at any point we can diagonalise g_{ij} using a rotation; arranging the resulting eigenvalues λ_i in increasing order we have
$$\lambda_1 \sum (X^{\hat{i}})^2 \leq \underbrace{\lambda_1 (X^{\hat{1}})^2 + \lambda_2 (X^{\hat{2}})^2 + \lambda_3 (X^{\hat{3}})^2}_{= g_{ij} \ X^i X^j} \leq \lambda_3 \sum (X^{\hat{i}})^2 \ , \tag{7.2.6}$$
where we have used the symbol $X^{\hat{i}}$ to denote the components of X in the diagonalising frame. Since the X^i's differ from the $X^{\hat{i}}$'s by a rotation we have
$$\sum (X^{\hat{i}})^2 = \sum (X^i)^2 \ ,$$
leading to
$$C = \max(\lambda_1^{-1}, \lambda_3) \ .$$
In order to prove that uniform boundedness of g_{ij}'s leads to the second inequality in (7.2.5) we note that in an arbitrary, not necessarily diagonalising, frame we have
$$\begin{aligned} g_{ij} X^i X^j &\leq \sup_{i,j,x} |g_{ij}(x)| \sum_{i,j} |X^i X^j| \\ &= \sup_{i,j,x} |g_{ij}(x)| ((X^1)^2 + (X^2)^2 + (X^3)^2 + 2|X^1 X^2| + 2|X^2 X^3| + 2|X^3 X^1|) \\ &\leq 2 \sup_{i,j,x} |g_{ij}(x)| ((X^1)^2 + (X^2)^2 + (X^3)^2) \ , \end{aligned}$$

with a similar calculation for g^{ij}, leading to (recall that, after diagonalisation, the largest eigenvalue of g^{ij} is λ_1^{-1})

$$\lambda_3 \leq 2 \sup_{i,j,x} |g_{ij}(x)| \;, \qquad \lambda_1^{-1} \leq 2 \sup_{i,j,x} |g^{ij}(x)| \;. \tag{7.2.7}$$

We thus have the following estimate for the best possible constant C in (7.2.5):

$$C \leq 2 \max(\sup_{i,j,x} |g_{ij}(x)|, \sup_{i,j,x} |g^{ij}(x)|) \;. \tag{7.2.8}$$

To finish the proof of equivalence, we note that (7.2.5) gives directly

$$|g_{ij}| = |g(\partial_i, \partial_j)| \leq 2\lambda_3 \leq 2C \;, \qquad \text{similarly} \;\; |g^{ij}| \leq 2C \;. \tag{7.2.9}$$

We have the following result, we follow the proof in [27]; an independent, completely different proof, under slightly weaker conditions, can be found in [10]:

THEOREM 7.2.1 (Coordinate-independence of $\boldsymbol{m}$ [10, 27]) *Consider two α–admissible coordinate systems ϕ_1 and ϕ_2, with some $\alpha > 1/2$, and suppose that*

$$R(g) \in L^1(N) \;.$$

Let $S(R)$ be any one-parameter family of differentiable spheres, such that $r(S(R)) = \min_{x \in S(R)} r(x)$ tends to infinity, as R does. For $\phi = \phi_1$ and $\phi = \phi_2$ define

$$m(g, \phi) \;\; = \;\; \lim_{R \to \infty} \frac{1}{16\pi} \int_{S(R)} (g_{ik,i} - g_{ii,k}) dS_k \;, \tag{7.2.10}$$

with each of the integrals calculated in the respective local α–admissible coordinates ϕ_a. Then

$$m(g, \phi_1) = m(g, \phi_2) \;.$$

The example of Denissov and Solovyev presented above shows that the condition $\alpha > 1/2$ in Theorem 7.2.1 is sharp.

PROOF: We start with a lemma:

LEMMA 7.2.2 (Asymptotic symmetries of asymptotically Euclidean manifolds) *Let (g, ϕ_1) and (g, ϕ_2) be α_1 and α_2–admissible, respectively, with any $\alpha_a > 0$. Let $\phi_1 \circ \phi_2^{-1} : \mathbb{R}^3 \backslash K_2 \to \mathbb{R}^3 \backslash K_1$ be a twice differentiable diffeomorphism, for some compact sets K_1 and $K_2 \subset \mathbb{R}^3$. Then there exists an $O(3)$ matrix $\omega^i{}_j$ such that, in local coordinates*

$$\phi_1^i(p) = x^i \;, \qquad \phi_2^i(p) = y^i \;,$$

the diffeomorphisms $\phi_1 \circ \phi_2^{-1}$ and $\phi_2 \circ \phi_1^{-1}$ take the form

$$x^i(y) = \omega^i{}_j \; y^i + \eta^i(y) \;, \qquad y^i(x) = (\omega^{-1})^i{}_j \; x^i \;\; + \zeta^i(x) \;,$$

ζ^i and η^i satisfy, for some constant $C \in \mathbb{R}$,

$$|\zeta^i{}_{,j}(x)| \le C(r(x)+1)^{-\alpha} , \quad |\zeta^i(x)| \le \begin{cases} C(\ln r(x)+1) , & \alpha = 1, \\ C(r(x)+1)^{1-\alpha} , & \textit{otherwise}, \end{cases}$$
$$|\eta^i{}_{,j}(y)| \le C(r(y)+1)^{-\alpha} , \quad |\eta^i(y)| \le \begin{cases} C(\ln r(y)+1) , & \alpha = 1, \\ C(r(y)+1)^{1-\alpha} , & \textit{otherwise}, \end{cases}$$
$$r(x) = (\sum (x^i)^2)^{1/2} , \qquad r(y) = (\sum (y^i)^2)^{1/2} ,$$

with $\alpha = \min(\alpha_1, \alpha_2, 1)$.

Proof: Let us first note that both (g, ϕ_1) and (g, ϕ_2) are α–admissible, so that we do not have to worry about two constants α_1 and α_2. Let g^1_{ij} and g^2_{ij} be the representatives of g in local coordinates ϕ_1 and ϕ_2:

$$g = g^1_{ij}(x)dx^i dx^j = g^2_{k\ell}(y)dy^k dy^\ell .$$

In the proof that follows the letters C, C', *etc.*, will denote constants which may vary from line to line, their exact values can be estimated at each step but are irrelevant for further purposes. Let us write down the equations following from the transformation properties of the metric

$$g^2_{ij}(y) = g^1_{k\ell}(x(y)) \frac{\partial x^k}{\partial y^i} \frac{\partial x^\ell}{\partial y^j} , \tag{7.2.11a}$$
$$g^1_{ij}(x) = g^2_{k\ell}(y(x)) \frac{\partial y^k}{\partial x^i} \frac{\partial y^\ell}{\partial x^j} . \tag{7.2.11b}$$

Contracting (7.2.11a) with $g_1^{ij}(x(y))$, where g_1^{ij} denotes the inverse matrix to g^1_{ij}, one obtains

$$g_1^{ij}(x(y)) g^2_{ij}(y) = g_1^{ij}(x(y)) g^1_{k\ell}(x(y)) \frac{\partial x^k}{\partial y^i} \frac{\partial x^\ell}{\partial y^j} . \tag{7.2.12}$$

Now, the function appearing on the right-hand-side above is a strictly positive quadratic form in $\partial x^i / \partial y^j$, and uniform ellipticity of g_1^{ij} gives

$$C^{-1} \sum_{k,i} (\frac{\partial x^k}{\partial y^i})^2 \le g_1^{ij}(x(y)) g^1_{k\ell}(x(y)) \frac{\partial x^k}{\partial y^i} \frac{\partial x^\ell}{\partial y^j} \le C \sum_{k,i} (\frac{\partial x^k}{\partial y^i})^2 .$$

In order to see this, we let $A^i{}_j$ be the tensor field $\partial x^i / \partial x^j$; in a frame diagonalising g^1_{ij}, as in (7.2.6), we have

$$g_1^{ij}(x(y)) g^1_{k\ell}(x(y)) A^k{}_i A^\ell{}_j = \sum_{i,j} \lambda_i^{-1} \lambda_j (A^j{}_i)^2$$

and we conclude with (7.2.7)

Since the function appearing at the left-hand-side of (7.2.12) is uniformly bounded we obtain

$$\sum_{k,i} |\frac{\partial x^k}{\partial y^i}| \le C . \tag{7.2.13}$$

Similar manipulations using (7.2.11b) give

$$\sum_{k,i} |\frac{\partial y^k}{\partial x^i}| \le C \ . \tag{7.2.14}$$

Inequalities (7.2.13)–(7.2.14) show that all the derivatives of $x(y)$ and $y(x)$ are uniformly bounded. Let Γ_x be the ray joining x and K_1, and let $y_0^i(x)$ be the image by $\phi_2 \circ \phi_1^{-1}$ of the intersection point of K_1 with Γ_x (if there is more than one, choose the one which is closest to x). We have, in virtue of (7.2.14),

$$|y^i(x) - y_0^i(x)| = \left| \int_{\Gamma_x} \frac{\partial y^i}{\partial x^k} dx^k \right| \le C \ r(x) \ ,$$

so that

$$r(y(x)) \le C \ r(x) + C_1 \ . \tag{7.2.15}$$

A similar reasoning shows

$$r(x(y)) \le C \ r(y) + C_1 \ . \tag{7.2.16}$$

Equations (7.2.15) and (7.2.16) can be combined into a single inequality

$$r(y(x))/C - C_1 \le r(x) \le C \ r(y(x)) + C_1 \ . \tag{7.2.17}$$

This equation shows that any quantity which is[1] $O(r(x)^{-\beta})$ $(O(r(y)^{-\beta}))$ is also $O(r(y)^{-\beta})$ $(O(r(x)^{-\beta}))$, when composed with $\phi_2 \circ \phi_1^{-1}$ $(\phi_1 \circ \phi_2^{-1})$.

Let us introduce

$$\begin{aligned} A^i{}_j &= \frac{\partial y^i}{\partial x^j} \ , \qquad B^i{}_j = \frac{\partial x^i}{\partial y^j} \ , \\ C_{ijk} &= A^m{}_i g^2_{m\ell} \frac{\partial A^\ell{}_j}{\partial x^k} = g^2_{m\ell} \frac{\partial y^m}{\partial x^i} \frac{\partial^2 y^\ell}{\partial x^j \partial x^k} \ , \\ D_{ijk} &= B^m{}_i g^1_{m\ell} \frac{\partial B^\ell{}_j}{\partial y^k} \ . \end{aligned}$$

Differentiating (7.2.11b) with respect to x, taking into account (7.2.4), (7.2.14) and (7.2.17) leads to

$$C_{ijk} + C_{jik} = O(r^{-\alpha-1}) \ .$$

We perform the usual cyclic permutation calculation:

$$\begin{aligned} C_{ijk} + C_{jik} &= O(r^{-\alpha-1}) \ . \\ -C_{jki} - C_{kji} &= O(r^{-\alpha-1}) \ . \\ C_{kij} + C_{ikj} &= O(r^{-\alpha-1}) \ . \end{aligned}$$

Adding the three equations and using the symmetry of C_{ijk} in the last two indices yields

$$C_{ijk} = O(r^{-\alpha-1}) \ .$$

[1] $f(s) = O(s^\gamma)$ is used here to denote a function satisfying $|f(s)| \le C(s+1)^\gamma$ for some positive constant C.

This equality together with (7.2.13) and the definition of C_{ijk} imply

$$\frac{\partial^2 y^i}{\partial x^j \partial x^k} = O(r^{-\alpha-1}) . \tag{7.2.18}$$

In a similar way one establishes

$$\frac{\partial^2 x^i}{\partial y^j \partial y^k} = O(r^{-\alpha-1}) . \tag{7.2.19}$$

We need a lemma:

LEMMA 7.2.3 *Let $\sigma > 0$ and let $f \in C^1(\mathbb{R}^n \setminus \overline{B(R)})$ satisfy*

$$\partial_i f = O(r^{-\sigma-1}) .$$

Then there exists a constant f_∞ such that

$$f - f_\infty = O(r^{-\sigma}) .$$

PROOF: Integrating along a ray we have

$$f(r_1\vec{n}) - f(r_2\vec{n}) = \int_{r_2}^{r_1} \frac{\partial f}{\partial r}(r\vec{n})dr = \int_{r_2}^{r_1} O(r^{-\sigma-1})dr = O(r_2^{-\sigma}) . \tag{7.2.20}$$

It follows that the sequence $f(i\vec{n})$ is Cauchy, therefore the limit

$$f_\infty(\vec{n}) = \lim_{i\to\infty} f(i\vec{n})$$

exists. Letting $r_1 = i$ in (7.2.20) and passing with i to infinity we obtain

$$f(\vec{x}) - f_\infty(\frac{\vec{x}}{r}) = O(r^{-\sigma}) .$$

Integrating over an arc of circle Γ connecting the vectors $r\vec{n}_1$ and $r\vec{n}_1$ we have

$$|f(r\vec{n}_1) - f(r\vec{n}_1)| = \left|\int_\Gamma df\right| \leq \sup_\Gamma |df||\Gamma| ,$$

where $|\Gamma|$ denotes the Euclidean length of Γ. Since $|\Gamma| \leq 2\pi r$ we obtain

$$|f(r\vec{n}_1) - f(r\vec{n}_1)| \leq 2\pi C r^{-\sigma} .$$

Passing with r to infinity we find

$$f_\infty(\vec{n}_1) = f_\infty(\vec{n}_1) ,$$

so that f_∞ is $\vec{n}$–independent, as desired. □

Lemma 7.2.3 shows that the limits

$$\mathring{A}^i{}_j = \lim_{r\to\infty} A^i{}_j(r\vec{n}) ,$$

$$\mathring{B}^i{}_j = \lim_{r\to\infty} B^i{}_j(r\vec{n}) ,$$

($\vec{n}$ — any vector satisfying $\sum(n^i)^2 = 1$) exist and are n^i independent matrices, with $A = B^{-1}$. Define

$$\zeta^i(x) = y^i(x) - \mathring{A}^i{}_j\ x^j\ , \qquad \eta^i(y) = x^i(y) - \mathring{B}^i{}_j\ y^j\ .$$

Equation (7.2.19) leads to

$$A^i{}_j(r_2\vec{n}) - A^i{}_j(r_1\vec{n}) = \int_{r_1}^{r_2} \frac{\partial^2 x^i}{\partial x^j \partial x^k}(r\vec{n})\ n^k\ dr = O(r_1^{-\alpha})$$

for $r_2 > r_1$. We have $A^i{}_j = \mathring{A}^i{}_j + \zeta^i{}_{,j}$, so that passing with r_2 to infinity one finds

$$\zeta^i{}_{,j}(x) = O(r^{-\alpha})\ .$$

Integrating along rays one obtains

$$\zeta^i(x) = \begin{cases} O(r^{1-\alpha})\ , & 0 < \alpha < 1, \\ O(\ln r)\ , & \alpha = 1, \end{cases}$$

with a similar calculation for η.

Equations (7.2.4) and (7.2.17) allow us to write (7.2.11) in the following form

$$\sum_k \frac{\partial y^k}{\partial x^i}\ \frac{\partial y^k}{\partial x^j} \quad = \quad \delta_{ij} + O(r^{-\alpha})\ , \tag{7.2.21a}$$

$$\sum_k \frac{\partial x^k}{\partial y^i}\ \frac{\partial x^k}{\partial y^j} \quad = \quad \delta_{ij} + O(r^{-\alpha})\ . \tag{7.2.21b}$$

Passing to the limit $r \to \infty$ one obtains that $\mathring{A}^i{}_j$ and $\mathring{B}^i{}_j$ are rotation matrices, which finishes the proof. □

Let us return to the proof of Theorem 7.2.1. We start by noting that the limit in (7.2.10) does not depend upon the family of spheres chosen — this follows immediately from the identity (7.1.21).

Next, let us show that the integrand of the mass has tensorial properties under rotations: if $y^i = \omega^i{}_j x^j$, then

$$g^1_{ij}(x) = g^2_{k\ell}(y(x))\frac{\partial y^k}{\partial x^i}\ \frac{\partial y^\ell}{\partial x^j} = g^2_{k\ell}(\omega x)\omega^k{}_i\ \omega^\ell{}_j\ ,$$

so that

$$\frac{\partial g^1_{ij}(x)}{\partial x^j} - \frac{\partial g^1_{jj}(x)}{\partial x^i} \quad = \quad \frac{\partial g^2_{k\ell}(\omega x)}{\partial y^r}\omega^r{}_j\omega^k{}_i\ \omega^\ell{}_j - \frac{\partial g^2_{k\ell}(\omega x)}{\partial y^r}\omega^r{}_i\omega^k{}_j\ \omega^\ell{}_j\ . \tag{7.2.22}$$

Now, a rotation matrix satisfies

$$\omega^r{}_i\omega^s{}_i = \delta^r_s\ , \tag{7.2.23}$$

so that (7.2.22) can be rewritten as

$$\begin{aligned}\frac{\partial g^1_{ij}(x)}{\partial x^j} - \frac{\partial g^1_{jj}(x)}{\partial x^i} &= \frac{\partial g^2_{k\ell}(\omega x)}{\partial y^\ell}\omega^k{}_i - \frac{\partial g^2_{\ell\ell}(\omega x)}{\partial y^r}\omega^r{}_i \\ &= \left(\frac{\partial g^2_{k\ell}(\omega x)}{\partial y^\ell} - \frac{\partial g^2_{\ell\ell}(\omega x)}{\partial y^k}\right)\omega^k{}_i\,. \qquad (7.2.24)\end{aligned}$$

Finally, the surface forms dS_j also undergo a rotation:

$$\frac{\partial}{\partial x^i}\rfloor dx^1\wedge\ldots\wedge dx^n = \omega^s{}_i\frac{\partial}{\partial y^s}\rfloor\underbrace{\left(\det\frac{\partial x}{\partial y}\right)}_{=1}dy^1\wedge\ldots\wedge dy^n = \omega^s{}_i\frac{\partial}{\partial y^s}\rfloor dy^1\wedge\ldots\wedge dy^n\,.$$

This, together with (7.2.24) and (7.2.23) leads to

$$\begin{aligned}&\left(\frac{\partial g^1_{ij}(x)}{\partial x^j} - \frac{\partial g^1_{jj}(x)}{\partial x^i}\right)\frac{\partial}{\partial x^j}\rfloor dx^1\wedge\ldots\wedge dx^n \\ &\quad= \left(\frac{\partial g^2_{k\ell}(\omega x)}{\partial y^\ell} - \frac{\partial g^2_{\ell\ell}(\omega x)}{\partial y^k}\right)\omega^k{}_i\omega^s{}_i\frac{\partial}{\partial y^s}\rfloor dy^1\wedge\ldots\wedge dy^n \\ &\quad= \left(\frac{\partial g^2_{k\ell}(\omega x)}{\partial y^\ell} - \frac{\partial g^2_{\ell\ell}(\omega x)}{\partial y^k}\right)\frac{\partial}{\partial y^k}\rfloor dy^1\wedge\ldots\wedge dy^n\,.\end{aligned}$$

It follows that the mass will not change if a rigid coordinate rotation is performed.

In particular, replacing the coordinate y^i by $(\omega^{-1})^i{}_j y^j$ will preserve the mass, and to finish the proof it remains to consider coordinate transformations such that the matrix ω in Lemma 7.2.2 is the identity. We then have

$$h^2_{ij} = g^2_{ij} - \delta_{ij} = h^1_{ij}(x(y)) + \eta^k{}_{,i}(y) + \eta^i{}_{,j}(y) + O(r^{-2\alpha}) \qquad (7.2.25)$$

where

$$h^1_{ij} = g^1_{ij} - \delta_{ij}\,.$$

Therefore

$$\begin{aligned}\frac{\partial g^2_{ij}(y)}{\partial y^j} - \frac{\partial g^2_{jj}(y)}{\partial y^i} &= \frac{\partial h^1_{ij}(x(y))}{\partial x^j} - \frac{\partial h^1_{jj}(x(y))}{\partial x^i} \\ &\quad + \frac{\partial}{\partial x^j}\left(\frac{\partial \eta^i}{\partial x^j} - \frac{\partial \eta^j}{\partial x^i}\right) + O(r^{-2\alpha-1})\,. \qquad (7.2.26)\end{aligned}$$

While integrated over the sphere $r(y) = \text{const}$, the last term in (7.2.26) will give no contribution in the limit $r(y)\to\infty$ since $2\alpha+1>2$ by hypothesis. The next to last term in (7.2.26) will give no contribution being the divergence of an antisymmetric quantity: indeed, we have

$$\frac{\partial}{\partial x^j}\left(\frac{\partial \eta^i}{\partial x^j} - \frac{\partial \eta^j}{\partial x^i}\right)\frac{\partial}{\partial x^i}\rfloor dx^1\wedge\ldots\wedge dx^n = d\left(\frac{\partial \eta^i}{\partial x^j}\frac{\partial}{\partial x^j}\rfloor\frac{\partial}{\partial x^j}\rfloor dx^1\wedge\ldots\wedge dx^n\right)\,,$$

and Stokes' theorem shows that the integral of that term over $S(R)$ vanishes. Finally, the first term in (7.2.26) reproduces the ADM mass of the metric g^1_{ij}. □

7.3 The Bartnik-Witten rigidity theorem

•7.3.1: **ptc:***new section, needs proofreading*

•7.3.1 A simple proof of positivity of mass can be given when one assumes that the Ricci tensor of (M, g) is non-negative:

THEOREM 7.3.1 ("Non-existence of gravitational instantons" (Witten [107], Bartnik [10])) *Let (M, g) be a complete Riemannian manifold with an asymptotically flat end, in the sense of (7.2.3)-(7.2.4) with decay rate $\alpha > 1/2$, and suppose that the Ricci tensor of g is non-negative definite:*

$$\forall\ X \qquad \mathrm{Ric}(X, X) \geq 0\ . \tag{7.3.1}$$

Then

$$0 \leq m \leq \infty\ ,$$

with $m = 0$ if and only if $(M, g) = (\mathbb{R}^n, \delta)$.

PROOF: If $R(g) \notin L^1(M)$, the result follows from point 2 of Theorem 7.1.5. From now on we therefore assume that the Ricci scalar of g is integrable over M.

We start by deriving the so-called *Bochner* identity. Suppose that

$$\Delta f = 0\ , \tag{7.3.2}$$

set

$$\varphi := |Df|^2 = D^k f D_l f\ .$$

We have

$$\begin{aligned}
\Delta\varphi &= D^i D_i (D^k f D_k f) \\
&= 2\Big(D^i D^k f D_i D_k f + D^k f \underbrace{D^i D_i D_k f}_{=D^i D_k D_i f = D_k \Delta f + R_i{}^{ji}{}_k D_j f}\Big) \\
&= 2\Big(|\mathrm{Hess}\, f|^2 + \mathrm{Ric}(Df, Df)\Big)\ .
\end{aligned} \tag{7.3.3}$$

This shows that $\Delta\varphi \geq 0$ when (7.3.1) holds.

We shall for simplicity assume that (M, g) has only one asymptotic end, the general case requires some technicalities which are not interesting from the point of view of this work. We will use (7.3.3) with $f = y^i$, where y^i is a solution of the Laplace equation (7.3.2) with the asymptotic condition

$$y^i - x^i = O(r^{1-\alpha})\ . \tag{7.3.4}$$

The existence of such functions is plausible, but a complete proof requires some amount of work, we refer the reader to [10] for details. The results there also show that the functions y^i form an admissible coordinate system, at least for large r, and Theorem 7.2.1 implies that we can use those coordinates to calculate the mass. We denote by φ^i the corresponding φ function, $\varphi^i = |Dy^i|^2$.

In the y–coordinate system we have

$$\varphi^i := g^{kl} \partial_k y^i \partial_l y^i = g^{ii} \ \ \text{(no summation over } i\text{)},$$

so that

$$D^k\varphi^i = g^{kl}\partial_l g^{ii} = -\partial_k g_{ii} + O(r^{-2\alpha-1}) \ \text{ (no summation over } i). \tag{7.3.5}$$

Integrating (7.3.3) with φ replaced by φ^i over (M,g) one has

$$\int_{S_\infty} D^k\varphi^i dS_k = \int_M \Delta\varphi^i = 2\int_M (|\mathrm{Hess}\, y^i| + \mathrm{Ric}(D\varphi^i, D\varphi^i)) \geq 0 \ , \tag{7.3.6}$$

and (7.3.5) gives

$$-\sum_i \int_{S_\infty} \partial_k g_{ii} dS_k \geq 0 \ . \tag{7.3.7}$$

It remains to relate this to the ADM mass. Since the coordinates y^i are harmonic we have

$$0 = \Delta y^i = \frac{1}{\sqrt{\det g}}\partial_k(\sqrt{\det g}g^{kl}\partial_l y^i) = \frac{1}{\sqrt{\det g}}\partial_k(\sqrt{\det g}g^{ki}) \ ,$$

so that

$$0 = \partial_k(\sqrt{\det g}g^{ki}) = -\frac{1}{2}\partial_i g_{jj} - \partial_k g_{ki} + O(r^{-1-2\alpha}) \ ,$$

which leads to•7.3.2

•7.3.2: **ptc:***verify the normalising factor*

$$m = \frac{1}{4\omega_n}\int_{S_\infty}(\partial_i g_{ik} - \partial_k g_{ii})dS_k = -\frac{3}{8\omega_n}\int_{S_\infty}\partial_k g_{ii} dS_k \geq 0$$

by (7.3.7). This establishes non-negativity of m. Now, if the mass vanishes, then (7.3.6) enforces $\mathrm{Hess}\, y^i = 0$ for all i. It follows that the one forms $Y^{(i)} := dy^i$ are covariantly constant,

$$DY^{(i)} = Ddy^i = \mathrm{Hess}\, y^i = 0 \ .$$

This implies

$$0 = D_{[i}D_{j]}Y^{(k)} = \frac{1}{2}R^\ell{}_{kij}\partial_\ell$$

so that the Riemann tensor vanishes. Let $\hat{M}$ be the universal covering space of M with the metric obtained by pull-back from the projection map, the Hadamard-Cartan theorem (see, *e.g.*, [87, Theorem 22, p. 278]) shows that the exponential map of any point $p \in \hat{M}$ is a global diffeomorphism from $\hat{M}$ to $\mathbb{R}^n$. It follows that M is a quotient of Euclidean $\mathbb{R}^n$ by a subgroup G of the Euclidean group. The existence of an asymptotically flat region in M, diffeomorphic to $\mathbb{R}^n \setminus B(R)$, shows that G must be trivial, and the result follows. □

7.4 Some space-time formulae

In this section we review some space-time expressions for the total energy. Let X^μ_∞ be a set of constants, and let X be any vector field on M such that

$X = X^\mu_\infty \partial_\mu$ in a coordinate system x^μ on the asymptotic region such that $\mathscr{S}$ is given by the equation $x^0 = 0$, and such that on $\mathscr{S}$ we have•7.4.1

•7.4.1: **ptc:***the second derivatives restriction is actually only needed for the alternative invariance proof*

$$g_{\mu\nu} = \eta_{\mu\nu} + O(r^{-\alpha}) \ , \quad \partial_\sigma g_{\mu\nu} = O(r^{-\alpha-1}) \ , \quad \partial_\sigma\partial_\rho g_{\mu\nu} = O(r^{-\alpha-2}) \ , \quad (7.4.1)$$

with, as usual, $\alpha > 1/2$. A calculation leads to

$$p_\mu X^\mu_\infty = \lim_{R\to\infty} \frac{3}{16\pi} \int_{S(R)} \delta^{\alpha\beta\gamma}_{\lambda\mu\nu} X^\nu \eta^{\lambda\rho} \eta_{\gamma\sigma} \partial_\rho g^{\sigma\mu} dS_{\alpha\beta} \ . \qquad (7.4.2)$$

Here

$$\delta^{\alpha\beta\gamma}_{\lambda\mu\nu} := \delta^\alpha_{[\lambda}\delta^\beta_\mu\delta^\gamma_{\nu]} \ , \qquad dS_{\alpha\beta} = \frac{1}{2}\epsilon_{\alpha\beta\gamma\delta} dx^\gamma \wedge dx^\delta \ ,$$

with $\epsilon_{0123} = \sqrt{|\det g|}$, *etc.*•7.4.2 Expression (7.4.2) is well suitable for the proof that p_μ is invariant under a certain class of coordinate changes, somewhat similar to that considered in Section **??**: Indeed, suppose that the x^μ's have been replaced by new coordinates y^μ such that

•7.4.2: **ptc:***crosscheck for consistency*

$$y^\mu = x^\mu + \zeta^\mu \ ,$$

with ζ^μ satisfying fall-off conditions analogous to those of Lemma 7.2.2:

$$|\zeta^\mu{}_{,\nu}(x)| \le C(r(x)+1)^{-\alpha} \ , \quad |\zeta^\mu{}_{,\nu\rho}(x)| \le C(r(x)+1)^{-\alpha-1} \ , \quad (7.4.3)$$

$$|\zeta^\mu(x)| \le \begin{cases} C(\ln r(x)+1) \ , & \alpha = 1, \\ C(r(x)+1)^{1-\alpha} \ , & \text{otherwise,} \end{cases} \qquad (7.4.4)$$

This leads to a change of the metric as in (7.2.25),

$$g_{\mu\nu} \longrightarrow g_{\mu\nu} + \zeta_{\mu,\nu} + \zeta_{\nu,\mu} + O(r^{-2\alpha-1}) \ ,$$

with $\zeta_\mu = \eta_{\mu\nu}\zeta^\nu$. Further, up to terms which obviously do not contribute in the limit,

$$\Delta(p_\mu X^\mu_\infty) = \lim_{R\to\infty} \frac{3}{16\pi} \int_{S(R)} (\delta^{\alpha\beta\gamma}_{\lambda\mu\nu} X^\nu \eta^{\lambda\rho} \zeta^\mu{}_{,\rho})_{,\gamma} dS_{\alpha\beta} = 0 \ , \qquad (7.4.5)$$

as the integral of a total divergence integrates out to zero.

With some effort one finds the identity, essentially due to Ashtekar and Hansen [6] (compare also [26])•7.4.3

•7.4.3: **ptc:***this formula does not agree with the one in the printed version of the paper, it might coincide because of some double star identity, but it is not clear*

$$\begin{aligned} p_\mu X^\mu_\infty &= \lim_{R\to\infty} \frac{1}{32\pi} \Big(\int_{S(R)} \epsilon_{\mu\nu\alpha\beta} X^\mu x^\nu R^{\alpha\beta}{}_{\rho\sigma} dx^\rho \wedge dx^\sigma \\ &\qquad + 2 \int_{S(R)} d(\epsilon_{\mu\nu\alpha\beta} X^\mu x^\nu g^{\alpha\gamma} \Gamma^\beta{}_{\gamma\rho} dx^\rho) \Big) \\ &= \lim_{R\to\infty} \frac{1}{32\pi} \int_{S(R)} \epsilon_{\mu\nu\alpha\beta} X^\mu x^\nu R^{\alpha\beta}{}_{\rho\sigma} dx^\rho \wedge dx^\sigma \ , \end{aligned} \qquad (7.4.6)$$

since the integral of the exterior derivative of a one form gives zero by Stokes' theorem.

The expression (7.4.6) looks somewhat more geometric than the more familiar Freud formula (??). However it should be remembered that the coordinate functions x^μ appearing there do not transform as a vector field under coordinate changes so that one still needs to appeal to Lemma 7.2.2 to establish the geometric character of (7.4.6). On the other hand, the proof of coordinate invariance under transformations (7.4.3)-(7.4.4) is immediate,

Formula (7.4.6) can be rewritten in a $3+1$ form, as follows: let, first $X^\mu_\infty = \delta^\mu_0$. Recalling that $\mathscr{S}$ is given by the formula $x^0 = 0$ and that p_0 is the ADM mass m we find

$$m = \lim_{R\to\infty} \frac{1}{32\pi} \int_{S(R)} \epsilon_{ijk} x^i \, {}^{(4)}R^{jk}{}_{\ell m} \, dx^\ell \wedge dx^m \; . \tag{7.4.7}$$

Here we have decorated the space-time Riemann curvature tensor with a subscript four to emphasise its four-dimensional nature. Supposing that the extrinsic curvature tensor K_{ij} falls–off as $r^{-\alpha-1}$, with $\alpha > 1/2$ (which will be the case under (7.4.1)), we then find by the Gauss-Codazzi constraint equation (??)

$${}^{(4)}R^{jk}{}_{\ell m} = {}^{(3)}R^{jk}{}_{\ell m} + O(r^{-2\alpha-2}) \; , \tag{7.4.8}$$

where ${}^{(3)}R^{jk}{}_{\ell m}$ denotes the curvature tensor of the space-metric h induced on $\mathscr{S}$ by g. Since $\alpha > 1/2$ the error terms in (7.4.8) will give no contribution in the limit $r \to \infty$ so that we finally obtain the purely three-dimensional formula

$$m = \lim_{R\to\infty} \frac{1}{32\pi} \int_{S(R)} \epsilon_{ijk} x^i \, {}^{(3)}R^{jk}{}_{\ell m} dx^\ell \wedge dx^m \; . \tag{7.4.9}$$

7.5 Energy in stationary space-times

Yet another way of rewriting (7.4.6) is given by•7.5.1

•7.5.1: **ptc:***justify the double dual identity*

$$p_\mu X^\mu_\infty = \lim_{R\to\infty} \frac{1}{16\pi} \int_{S(R)} X^\mu x^\nu R_{\mu\nu\alpha\beta} dS^{\alpha\beta} \; . \tag{7.5.1}$$

This is particularly convenient when X is a Killing vector as then we have the identity•7.5.2

•7.5.2: **ptc:***cross-reference to the appropriate place*

$$R_{\mu\nu\alpha\beta} X^\mu = \nabla_\nu \nabla_\alpha X_\beta \; .$$

When X is asymptotic to a translation, inserting this into (7.5.1) one obtains

$$\begin{aligned} p_\mu X^\mu_\infty &= \lim_{R\to\infty} \Big(\frac{1}{16\pi} \int_{S(R)} X^{[\beta;\alpha]}{}_{;\gamma} x^\gamma dS_{\alpha\beta} \\ &= \lim_{R\to\infty} \frac{1}{16\pi} \Big(2 \int_{S(R)} X^{[\alpha;\beta]} dS_{\alpha\beta} + 3 \int_{S(R)} (X^{[\alpha;\beta} x^{\gamma]})_{;\gamma} dS_{\alpha\beta} \Big) \\ &= \lim_{R\to\infty} \frac{1}{8\pi} \int_{S(R)} X^{[\alpha;\beta]} dS_{\alpha\beta} \; , \end{aligned} \tag{7.5.2}$$

This last integral is known as the *Komar integral*; the equality of the ADM mass and of the Komar mass for stationary space-times has been first proved by Beig [17].

7.6 Moving hypersurfaces in space-time

By way of example, we consider a family of hypersurfaces $\mathscr{S}_\tau$ in Minkowski space-time given by the equation

$$\mathscr{S}_\tau = \{t = f(\tau, \vec{x})\} ,$$

with

$$f(\tau) = \tau + ar^{1/2}$$

for r large. We then have

$$\begin{aligned} \eta &= -dt^2 + dr^2 + r^2 d\Omega^2 \\ &= (d\tau + \frac{a}{2} r^{-1/2} dr)^2 + dr^2 + r^2 d\Omega^2 \\ &= d\tau^2 + ar^{-1/2} d\tau\, dr + (1 - \frac{a^2}{4r}) dr^2 + r^2 d\Omega^2 . \end{aligned}$$

This shows that, for r large, the metric induced on the $\mathscr{S}_\tau$'s reads

$$(1 - \frac{a^2}{4r}) dr^2 + r^2 d\Omega^2 . \tag{7.6.1}$$

At leading order this is the same as a Schwarzschild metric with mass parameter $-a^2/8$, so that the ADM mass of the slices $\mathscr{S}_\tau$ is negative and equals

$$m_{ADM} = -\frac{a^2}{8} .$$

This example shows that deforming a hypersurface in space-time might lead to a change of mass. The fact that this can happen should already have been clear from the Kasner example (7.1.39), where the space-time itself does not satisfy any asymptotic flatness conditions. But this might seem a little more surprising in Minkowski space-time, which is flat. It should be emphasised that the (strictly negative) mass of the $\mathscr{S}_\tau$'s is not an artifact of a funny coordinate system chosen on $\mathscr{S}_\tau$: indeed, Theorem 7.2.1 shows that m is a geometric invariant of the geometry of $\mathscr{S}_\tau$. Further, one could suspect that negativity of m arises from singularities of the $\mathscr{S}_\tau$'s arising from the singular behaviour of $r^{-1/2}$ at $r = 0$. However, this is not the case, since we are free to modify f at will for r smaller than some constant R to obtain globally smooth spacelike hypersurfaces.

•7.6.1: **ptc:***add details for the Lemaitre example*

•7.6.1 A somewhat similar behavior can be seen when Lemaître coordinates $(\tau, \rho, \theta, \varphi)$ are used in Schwarzschild space-time: in this coordinate system the Schwarzschild metric takes the form [**?**]

$$ds^2 = -d\tau^2 + \left(\frac{\partial Y}{\partial \rho} d\rho\right)^2 + Y^2 d\Omega^2 , \tag{7.6.2}$$

with

$$Y = \left(3\sqrt{\frac{m}{2}}\tau + \rho^{3/2}\right)^{2/3} .$$

On any fixed hypersurface $\tau =$ const we can replace ρ by a new radial coordinate Y; Equation (7.6.2) shows then that the slices $\tau =$ const are flat, hence have zero mass.

All the foliations we have been considering above have mass which does not depend upon the slice $\mathscr{S}_\tau$. This is not true in general, consider a new time τ in Minkowski space-time which, for r large, is given by the formula

$$\tau = \frac{t}{1+ar^{1/2}} \;.$$

We then have

$$\begin{aligned}\eta &= -d\Big(\tau(1+ar^{1/2})\Big)^2 + dr^2 + r^2 d\Omega^2 \\ &= -(1+ar^{1/2})^2 dt^2 - (a+r^{-1/2})a\tau\, d\tau dr + \Big(1-\frac{\tau^2 a^2}{4r}\Big)dr^2 + r^2 d\Omega^2 \;.\end{aligned}$$

It follows that the ADM mass of the slices $\tau =$ const is well defined, equal to

$$m_{ADM} = -\frac{\tau^2 a^2}{8} \;,$$

which clearly changes when going from one slice to another.

7.7 Spherically symmetric positive energy theorem

•7.7.1

•7.7.1: **ptc:** *I don't know where to put this thing*

Theorem 7.7.1 *Consider a complete asymptotically flat spherically symmetric metric g on a*
with

$$R(g) \geq 0 \;.$$

If M contains no compact minimal surfaces, then

$$m \geq 0$$

with equality if and only if $M = \mathbb{R}^3$ with g – the Euclidean metric.

•7.7.2 Proof:

•7.7.2: **ptc:** *one needs to define spherical symmetry, discuss the possible manifolds, discuss the most general metrics, discuss minimal surfaces...*

... We can therefore choose a new radial variable ρ so that

$$e^{\gamma(r)} = \rho$$

Assuming that this has been done, we rewrite the metric in the new coordinate system (ρ, θ, φ), and call r again the new variable, keeping the same symbol β for the new function β appearing in the metric. It is convenient to define a function $m(r)$ by the equation

$$e^{-\beta} = \sqrt{1-\frac{2m(r)}{r}} \quad \Longleftrightarrow \quad m(r) := \frac{r}{2}(1-e^{2\beta}) \;.$$

It is remarkable that this seemingly complicated formula, together with (A.6.19), leads to a very simple formula for R:

$$R = \frac{4m'}{r^2} \;,$$

so that

$$m(r_0) = \frac{1}{4}\int_0^{r_0} r^2 dr \ .$$

It follows that•7.7.3

•7.7.3: **ptc:***needs expanding, and conditions*

$$m = \frac{1}{4}\int_0^{\infty} Rr^2 dr \geq 0 \ ,$$

with equality if and only if $m(r) \equiv 0$ for all r, so that $\beta \equiv 0$, and g is the Euclidean metric, as claimed. □

7.8 Lohkamp's positive energy theorem

•7.8.1: **ptc:***is there a simple proof of that?*

A deep result of Gromov and Lawson asserts that .... •7.8.1

It has been shown by Lohkamp [79] that the positive energy theorem for metrics on an asymptotically flat manifold with compact interior, with one asymptotically flat end[2], and with scalar curvature satisfying

$$R \geq 0 \ ,$$

can be reduced to the Gromov-Lawson theorem.
•7.8.2

•7.8.2: **ptc:***to be finished; this is sections 5, 6 of Lohkamp's paper*

7.9 Li-Tam's proof of positivity of the Brown-York quasi-local mass

7.10 The Riemannian Penrose inequality

7.11 A poor man's positive energy theorem

The positivity results proved so far do not appear to have anything to do with Lorentzian geometry. In this section, based on [31], we prove energy positivity using purely Lorentzian techniques, albeit for a rather restricted class of geometries; it seems that in practice our proof only applies to stationary (with or without black holes) space-times. This is a much weaker statement than the theorems in [96, 108] and their various extensions [14, 20, 54, 61, 63] (some of which are presented here), but the proof below is of interest because the techniques involved are completely different and of a quite elementary nature. Using arguments rather similar in spirit to those of the classical singularity theorems [58], the proof here is a very simple reduction of the problem to the Lorentzian splitting theorem [49] of Section **??**. (In lieu of the Lorentzian splitting theorem, one can impose the *"generic condition"* [58, p. 101], thereby making the proof completely elementary. However, it is not clear how "generic" the generic condition is, when, *e.g.*, vacuum equations are imposed, so it is desirable to have

[2]The argument presented here does not exclude configurations with two ends, with the mass being positive in one end and negative in the second. The full positive energy theorem does exclude such configurations.

results without that condition.) The approach taken here bares some relation to the Penrose-Sorkin-Woolgar [91] argument for positivity of mass, and indeed arose out of an interest in understanding their work.

For $m \in \mathbb{R}$, let g_m denote the $n+1$ dimensional, $n \geq 3$, Schwarzschild metric with mass parameter m; in isotropic coordinates [88],

$$g_m = \left(1 + \frac{m}{2|x|^{n-2}}\right)^{\frac{4}{n-2}} \left(\sum_{1=1}^{n} dx_i^2\right) - \left(\frac{1 - m/2|x|^{n-2}}{1 + m/2|x|^{n-2}}\right)^2 dt^2 \ . \qquad (7.11.1)$$

We shall say that a metric g on $\mathbb{R} \times (\mathbb{R}^n \setminus B(0,R))$, $R^{n-2} > m/2$, is *uniformly Schwarzchildian* if, in the coordinates of (7.11.1),

$$g - g_m = o(|m| r^{-(n-2)}) \ , \quad \partial_\mu (g - g_m) = o(|m| r^{-(n-1)}) \ . \qquad (7.11.2)$$

(Here o is meant at fixed g and m, uniformly in t and in angular variables, with r going to infinity.) It is a flagrant abuse of terminology to allow $m = 0$ in this definition, and we will happily abuse; what is meant in this case is that $g = g_0$, i.e., g is flat [3], for $r > R$.

Some comments about this notion are in order. First, metrics as above have constant Trautman-Bondi mass and therefore do not contain gravitational radiation; one expects such metrics to be stationary if physically reasonable field equations are imposed. Next, every metric in space-time dimension four which is stationary, asymptotically flat and vacuum or electro-vacuum in the asymptotically flat region is uniformly Schwarzschildian there when $m \neq 0$ (cf., e.g., [100]).

The hypotheses of our theorem below are compatible with stationary black hole space-times with non-degenerate Killing horizons.

We say that the matter fields satisfy the *timelike convergence condition* if the Ricci tensor $R_{\mu\nu}$, as expressed in terms of the matter energy-momentum tensor $T_{\mu\nu}$, satisfies the condition

$$R_{\mu\nu} X^\mu X^\nu \geq 0 \text{ for all timelike vectors } X^\mu. \qquad (7.11.3)$$

We define the *domain of outer communications* of $\mathscr{M}$ as the intersection of the past $J^-(\mathscr{M}_{\rm ext})$ of the asymptotic region $\mathscr{M}_{\rm ext} = \mathbb{R} \times (\mathbb{R}^n \setminus B(0,R))$ with its future $J^+(\mathscr{M}_{\rm ext})$.

We need a version of weak asymptotic simplicity [58] for uniformly Schwarzschildian spacetimes. We shall say that such a spacetime $(\mathscr{M}, g)$ is *weakly asymptotically regular* if every null line starting in the domain of outer communications (DOC) either crosses an event horizon (if any), or reaches arbitrarily large values of r in the asymptotically flat region. By definition, a null line in $(\mathscr{M}, g)$ is an inextendible null geodesic that is globally achronal; a timelike line is an inextendible timelike geodesic, each segment of which is maximal. Finally, we shall say that the DOC is *timelike geodesically regular* if every timelike line in $\mathscr{M}$ which is entirely contained in the DOC, and along which r is bounded, is complete.

The main result in this section is the following:

[3] The asymptotic conditions for the case $m = 0$ of our theorem are way too strong for a rigidity statement of real interest, even within a stationary context. So it is fair to say that our result only excludes $m < 0$ for stationary space-times.

THEOREM 7.11.1 *Let* $(\mathscr{M}^{n+1} = \mathscr{M}, g)$ *be an* $(n+1)$*-dimensional space-time with matter fields satisfying the timelike convergence condition (7.11.3), and suppose that* $\mathscr{M}$ *contains a uniformly Schwarzschildian region* $\mathscr{M}_{\rm ext} = \mathbb{R} \times (\mathbb{R}^n \setminus B(0,R))$ *. Assume that* $(\mathscr{M}, g)$ *is weakly asymptotically regular and that the domain of outer communications is timelike geodesically regular. If the domain of outer communications of* $\mathscr{M}$ *has a Cauchy surface* $\mathscr{S}$*, the closure of which is the union of one asymptotic end and of a compact interior region (with a smooth boundary lying at the intersection of the future and past event horizons, if any), then*

$$m > 0$$

unless $(\mathscr{M}, g)$ *isometrically splits as* $\mathbb{R} \times \mathscr{S}$ *with metric* $g = -d\tau^2 + \gamma$*,* $\mathscr{L}_{\partial_\tau}\gamma = 0$*, and* $(\mathscr{S}, \gamma)$ *geodesically complete. Furthermore, the last case does not occur if event horizons are present.*

Before passing to the proof, we note the following Corollary:

COROLLARY 7.11.2 *In addition to the hypotheses of Theorem 7.11.1, assume that*

$$T_{\mu\nu} \in L^1\left(\mathbb{R}^n \setminus B(0,R)\right) , \qquad \partial_\nu \partial_\mu g = O(r^{-\alpha}) , \quad \alpha > 1 + \frac{n}{2} . \qquad (7.11.4)$$

Then $m > 0$ *unless* $\mathscr{M}$ *is the Minkowski space-time.*

PROOF OF THEOREM 7.11.1: The idea is to show that for $m \le 0$ the domain of outer communications contains a timelike line, and the result then follows from Galloway's splitting theorem [49], see Section **??**. A somewhat lengthy but straightforward computation•7.11.1 shows that the Hessian $\mathrm{Hess}\, r = \nabla dr$ of r is given by

•7.11.1: **ptc:***the details of this calculation can be traced back to some calculations in poormanII, to be included*

$$\mathrm{Hess}\, r = -\frac{m}{r^{n-1}}\left((n-2)dt^2 - dr^2 + r^2\, h\right) + r\, h + o(r^{-(n-1)}) , \qquad (7.11.5)$$

where h is the canonical metric on S^n, and the size of the error terms refers to the components of the metric in the coordinates of (7.11.1). Note that when $m < 0$, $\mathrm{Hess}\, r$, when restricted to the hypersurfaces of constant r, is strictly positive definite for $r \ge R_1$, for some sufficiently large R_1. Increasing R_1 if necessary, we can obtain that ∂_t is timelike for $r \ge R_1$. If $m = 0$ we set $R_1 = R$. Let $p_{\pm k}$ denote the points $t = \pm k$, $\vec{x} = (0, 0, R_1)$; by global hyperbolicity there exists a maximal future directed timelike geodesic segment σ_k from p_{-k} to p_{+k}.•7.11.2 We note, first, that the σ_k's are obviously contained in the domain of outer communications and therefore cannot cross the event horizons, if any. If $m = 0$ then σ_k clearly cannot enter $\{r > R_1\}$, since timelike geodesics in that region are straight lines which never leave that region once they have entered it. It turns out that the same occurs for $m < 0$: suppose that σ_k enters $\{r > R_1\}$, then the function $r \circ \sigma_k$ has a maximum. However, if s is an affine parameter along σ_k we have

•7.11.2: **ptc:***should be justified somewhere, and crossreference added*

$$\frac{d^2(r \circ \sigma_k)}{ds^2} = \mathrm{Hess}\, r(\dot\sigma_k, \dot\sigma_k) > 0$$

at the maximum if $m < 0$, since $dr(\dot\sigma_k) = 0$ there, which is impossible. It follows that all the σ_k's (for k sufficiently large) intersect the Cauchy surface $\mathscr{S}$ in the compact set $\overline{\mathscr{S}} \setminus \{r > R_1\}$. A standard argument$^{\bullet 7.11.3}$ shows then that the σ_k's accumulate to a timelike or null line σ through a point $p \in \overline{\mathscr{S}}$. Let $\{p_k\} = \sigma_k \cap \mathscr{S}$; suppose that $p \in \partial\mathscr{S}$, then the portions of σ_k to the past of p_k accumulate at a generator of the past event horizon $\dot J^+(\mathscr{M}_{\rm ext})$, and the portions of σ_k to the future of p_k accumulate at a generator of the future event horizon $\dot J^-(\mathscr{M}_{\rm ext})$. This would result in σ being non-differentiable at p, contradicting the fact that σ is a geodesic. Thus the p_k's stay away from $\partial\mathscr{S}$, and $p \in \mathscr{S}$. By our "weak asymptotic regularity" hypothesis σ cannot be null (as it does not cross the event horizons, nor does it extend arbitrarily far into the asymptotic region). It follows that σ is a timelike line in $\mathscr{M}$ entirely contained in the globally hyperbolic domain of outer communications $\mathscr{D}$, with $r \circ \sigma$ bounded, and hence is complete by the assumed timelike geodesic regularity of $\mathscr{D}$. Thus, one may apply [49] to conclude that $(\mathscr{D}, g|_{\mathscr{D}})$ is a metric product,

•7.11.3: **ptc:**_give crossrefs_

$$g = -d\tau^2 + \gamma\ , \tag{7.11.6}$$

for some τ–independent complete Riemannian metric γ. The completeness of this metric product implies $\mathscr{D} = \mathscr{M}$ (and in particular excludes the existence of event horizons). $\square$

Proof of Corollary 7.11.2: The lapse function N associated with a Killing vector field on a totally geodesic hypersurface $\mathscr{S}$ with induced metric γ and unit normal n satisfies the elliptic equation

$$\Delta_\gamma N - \mathrm{Ric}(n,n)N = 0\ .$$

The vector field ∂_τ is a static Killing vector in $\mathscr{M}_{\rm ext}$, and the usual analysis of groups of isometries of asymptotically flat space-times shows that the metric γ in (7.11.6) is asymptotically flat. Again in (7.11.6) we have $N = 1$ hence $\mathrm{Ric}(n,n) = 0$, and the Komar mass of $\mathscr{S}$ vanishes. By a theorem of Beig [17] (originally proved in dimension four, but the result generalises to any dimensions under (7.11.4)) this implies the vanishing of the ADM mass. Let e_a, $a = 0,\ldots,n$, be an orthonormal frame with $e_0 = \partial_\tau$. The metric product structure implies that $R_{0i} = 0$. Thus, by the energy condition, for any fixed i we have

$$0 \le \mathrm{Ric}(e_0 + e_i, e_0 + e_i) = R_{00} + R_{ii} = R_{ii}\ .$$

But again by the product structure, the components R_{ii} of the space-time Ricci tensor equal those of the Ricci tensor $\mathrm{Ric}_{\mathscr{S}}$ of γ. It follows that $\mathrm{Ric}_{\mathscr{S}} \ge 0$. A generalisation by Bartnik [11] of an argument of Witten [108] shows that $(\mathscr{S}, \gamma)$ is isometric to Euclidean space; we reproduce the proof to make clear its elementary character: Let y^i be global harmonic functions forming an asymptotically rectangular coordinate system near infinity. Let $K^i = \nabla y^i$; then by Bochner's formula,

$$\Delta |K^i|^2 = 2|\nabla K^i|^2 + 2\mathrm{Ric}_{\mathscr{S}}(K^i, K^i)\ .$$

Integrating the sum over $i = 1, \dots, n$ of this gives the ADM mass as boundary term at infinity, and the ∇y^i are all parallel. Since $\mathscr{S}$ is simply connected at infinity, it must be Euclidean space. $\square$

We close this note by showing that the conditions on geodesics in Theorem 7.11.1 are always satisfied in stationary domains of outer communications.

PROPOSITION 7.11.3 *Let the domain of outer communications $\mathscr{D}$ of $(\mathscr{M}, g)$ be globally hyperbolic, with a Cauchy surface $\mathscr{S}$ such that $\overline{\mathscr{S}}$ is the union of a finite number of asymptotically flat regions and of a compact set (with a boundary lying at the intersection of the future and past event horizons, if any). Suppose that there exists on $\mathscr{M}$ a Killing vector field X with complete orbits which is timelike, or stationary-rotating[4] in the asymptotically flat regions. Then the weak asymptotic regularity and the timelike regularity conditions hold.*

REMARK 7.11.4 *We note that there might exist maximally extended null geodesics in $(\mathscr{D}, g)$ which are trapped in space within a compact set (as happens for the Schwarzschild metric), but those geodesics will not be achronal.*

PROOF: By [36, Propositions 4.1 and 4.2] we have $\mathscr{D} = \mathbb{R} \times \mathscr{S}$, with the flow of X consisting of translations along the $\mathbb{R}$ axis:

$$g = \alpha d\tau^2 + 2\beta d\tau + \gamma \;, \quad X = \partial_\tau \;, \tag{7.11.7}$$

where γ is a Riemannian metric on $\mathscr{S}$ and β is a one-form on $\mathscr{S}$. (We emphasise that we do not assume X to be timelike, so that $\alpha = g(X, X)$ can change sign.) Let ϕ_t denote the flow of X and let $\sigma(s) = (\tau(s), p(s)) \in \mathbb{R} \times \mathscr{S}$ be an affinely parameterized causal line in $\mathscr{D}$, then for each $t \in \mathbb{R}$ the curve $\phi_t(\sigma(s)) = (\tau(s) + t, p(s))$ is also an affinely parameterized causal line in $\mathscr{D}$. Suppose that there exists a sequence s_i such that $p(s_i) \to \partial\mathscr{S}$, setting $t_i = -\tau(s_i)$ we have $\tau(\phi_{t_i}(\sigma(s_i)) = 0$, then the points $\{p_{k_i}\} = \phi_{t_i}(\sigma) \cap \mathscr{S}$ accumulate at $\partial\mathscr{S}$, which is not possible as in the proof of Theorem 7.11.1. Therefore there exists an open neighborhood $\mathscr{K}$ of $\partial\mathscr{S}$ such that $\sigma \cap (\mathbb{R} \times \mathscr{K}) = \emptyset$. This implies in turn that σ meets all the level sets of τ. Standard considerations using the fact that $\mathscr{D}$ is a stationary, or stationary-rotating domain of outer communications (cf., e.g., [36]) show that for every $p, q \in \mathscr{S}$ there exists $T > 0$ and a timelike curve from $(0, p)$ to (T, q). The constant T can be chosen independently of p and q within the compact set $\overline{\mathscr{S}} \setminus (\mathscr{K} \cup \{r > R_1\})$, with $R_1 = \sup_\sigma r$. It follows that an inextendible null geodesic which is bounded in space within a compact set cannot be achronal, so that σ has to reach arbitrarily large values of r, and weak asymptotic regularity follows. Similarly, if σ is a timelike line bounded in space within a compact set, then there exists $s_1 > 0$ such that for any point $(\tau(s), p(s))$ with $s = s_1 + u$, $u > 0$ one can find a timelike curve from $(0, p(0))$ to $(\tau(s), p(s))$ by going to the asymptotic region, staying there for a time u, and coming back. The resulting curve will have Lorentzian length larger than $u/2$ if one went sufficiently far into the asymptotic region, and since σ is length-maximising it must be complete. $\square$

[4] See [36] for the definition.

The key point in the proof of Proposition 7.11.3 is non-existence of observer horizons contained in the DOC. Somewhat more generally, we have the following result, which does not assume existence of a Killing vector:

PROPOSITION 7.11.5 *Suppose that causal lines σ, with $r \circ \sigma$ bounded, and which are contained entirely in $\mathscr{D}$, do not have* observer horizons *extending to the asymptotic region $\mathscr{M}_{\rm ext}$ (see (**??**)):*

$$\dot{J}^{\pm}(\sigma;\mathscr{D}) \cap \mathscr{M}_{\rm ext} = \emptyset\ . \tag{7.11.8}$$

Then the weak asymptotic regularity and the timelike regularity conditions hold.

PROOF: It follows from (7.11.8) that for any $u > 0$ and for any s_1 there exists s_2 and a timelike curve Γ_{u,s_1} from $\sigma(s_1)$ to $\sigma(s_2)$ which is obtained by following a timelike curve from $\sigma(s_1)$ to the asymptotic region, then staying there at fixed space coordinate for a coordinate time u, and returning back to σ along a timelike curve. One concludes as in the proof of Proposition 7.11.3. □

7.12 Small data positive energy theorem

•7.12.1

•7.12.2

•7.12.1: **ptc:** *look up ChB, Brill, Fadeev*

•7.12.2: **ptc:** *Bartnik's small data pet*

7.13 The definition of mass: a Hamiltonian approach

So far we have been considering the mass of various hypersurfaces in Lorentzian space-times. Now, from a physical point of view it is desirable to be able to assign a mass to a space-time; for example, one would like to say that Minkowski space-time has zero mass, while the Schwarzschild space-time has mass m. One way of doing this would be to select a preferred family of hypersurfaces in $\mathscr{M}$, calculate the mass for those hypersurfaces, and check that the resulting number is hypersurface–independent within the family of allowed hypersurfaces. The above discussed examples show that some restrictions on the family of hypersurfaces need to be imposed for such a program to succeed. We shall see in Section **??** that such a procedure can indeed be carried through. Before doing that, it is useful to develop a four-dimensional formalism for discussing the notion of mass. This is the object of the next•7.13.1 section.

... •7.13.2

•7.13.1: **ptc:** *next, or this*

•7.13.2: **ptc:** *to be done*

7.14 Some background on spinors

The positive mass theorem asserts that the mass of an asymptotically Euclidean Riemannian manifold with non-negative scalar curvature is non-negative. There exist by now at least four different proofs of this result, the first one due to Schoen and Yau [95], shortly followed by a spinor-based proof by Witten [107]. More recently, a new argument has been given by Lohkamp [79], while positivity of mass can also be obtained from the proof of the Penrose inequality of Huisken and Ilmanen [63]. From those proofs the simplest one by far is that of Witten, and this is the one which we will present here. An advantage thereof is that it is easy to adapt it to obtain further global inequalities; some such inequalities will be presented in Section **??**.

•7.14.1 •*7.14.1:* **ptc:***outline the existing literature*

We start by presenting the background information on spinor fields necessary to carry the argument through. We will adapt a naive calculational approach, renouncing to a proper geometrical treatment which would yield insight into the structures involved, but would complicate considerable the presentation.

•7.14.2: **ptc:** *beginning of trick.tex*

•7.14.2

•7.14.3: **ptc:** *first bug*

•7.14.3

I propose to add, just before the definition of H, the sentence: "Suppose that (M, g) is not the hyperbolic space, otherwise there is nothing to prove." The reason is that the abstract completion of smooth compactly supported spinors with respect to the norm (**??**) (by the way, there is a square missing at the l.h.s. of this equation) does NOT give you a space of spinors in dimension two on hyperbolic space. There are a few ways out:

First possibility: let H be defined as in the paper so far, so H is an abstract Hilbertian completion of smooth compactly supported spinor fields, and ASSUME FURTHER that there are no imaginary spinors on M. The argument in the third possibility below shows that you do get a space of spinors then.

Second possibility: proceed as we suggest in the paper, but ignore whether or not you get a space of spinors. There are then a few things to take care of, like justifying that elliptic estimates give regularity - whatever this means for objects which are not spinors (we are not dealing any more with objects to which the usual theory applies directly), and also justifying that the solution of the second order equation does actually give a solution of the first order one. It is not clear to me that everything can be pushed through, though perhaps it might.

Third possibility: One can show [**?**] that in dimension larger than or equal to three there exists a strictly positive L^∞_{loc} function w on M such that for all H^1_{loc} spinor fields ψ with compact support we have

$$\int_M \|\psi\|_g^2 w d\mu_g \le \int_M \|\widehat{D}\psi\|_g^2 d\mu_g \ . \tag{7.14.1}$$

The function w can be chosen to be constant in the asymptotically hyperbolic end. In dimension two I can only prove (7.14.1) if I assume further that there are no imaginary Killing spinors (without any control of w in the asymptotic region, but such control is not needed in the proof) - but if there exists a Killing spinor then, by Baum, there is nothing to prove, so one might as well suppose that there are no such spinors. Let $\mathscr{H}$ be the space of measurable spinor fields on M such that

$$\|\psi\|^2_{\mathscr{H}} := \int_M \|\psi\|_g^2 (w + \frac{1}{4}(R_g - n(n-1)) d\mu_g + \int_M \|\widehat{D}\psi\|_g^2 d\mu_g < \infty \tag{7.14.2}$$

where the derivative is understood in the distributional sense. Define $\mathring{\mathscr{H}} \subset \mathscr{H}$ as the completion of C^∞_c, in $\mathscr{H}$, with respect to the $\|\cdot\|_{\mathscr{H}}$ norm. It is then easy to verify

PROPOSITION 7.14.1 *The inequality (7.14.1) remains true for all* $\phi \in \mathring{\mathscr{H}}$.

PROOF: Both sides of (7.14.1) are continuous on $(\mathscr{H}, \|\cdot\|_{\mathscr{H}})$. □

PROPOSITION 7.14.2 *If* (M,g) *is complete then* $\mathscr{H} = \mathring{\mathscr{H}}$.

PROOF: If $\phi \in \mathscr{H}$ then the sequence $\chi_i\phi$ converges to ϕ in $(\mathscr{H}, \|\cdot\|_{\mathscr{H}})$, where $\chi_i(p) = \chi(d_{p_0}(p)/i)$, where d_{p_0} is the distance to some chosen point $p_0 \in M$, while $\chi : \mathbb{R} \to [0,1]$ is a smooth function such that $\chi|_{[0,1]} = 1$, $\chi|_{[1,\infty)} = 1$. □

PROPOSITION 7.14.3 *If* (M,g) *is complete then there is a natural continuous bijection between* $(\mathscr{H}, \|\cdot\|_{\mathscr{H}})$ *and* $(H, \|\cdot\|_H)$ *which is the identity on* C^1_c*; in particular, elements of* H *can be identified with spinor fields on* M *which are in* $\mathscr{H}$.

PROOF: By Proposition 7.14.2 both spaces are Hilbert spaces containing C^1_c as a dense subspace, with the norms being equivalent when restricted to C^1_c by Proposition 7.14.1. □

The bottom line is: I know how to justify things if we assume dimension larger than three, while in dimension two we need to assume that we are not in hyperbolic space. •7.14.4 •*7.14.4:* **ptc:***end of first bug*

•7.14.5 Let $F(\psi)$ denote the left-hand-side of Equation (**??**) with $\Phi_u = \chi\varphi_u + \psi$ there, let ψ_i converge to ψ in H, we have •*7.14.5:* **ptc:***second fix: I suggest to add this around Equation (**??**)*

$$\begin{aligned} F(\psi) - F(\psi_i) &= \|\psi\|^2_H - \|\psi_i\|^2_H \\ &\quad +2\int_M \langle \widehat{D}(\chi\varphi_u), \widehat{D}(\psi-\psi_i)\rangle \\ &\quad -2\int_M \langle \widehat{\mathrm{Dirac}}(\chi\varphi_u), \widehat{\mathrm{Dirac}}(\psi-\psi_i)\rangle \\ &\quad +\frac{1}{2}\int_M (R_g - n(n-1))\,\langle \chi\varphi_u, \psi-\psi_i\rangle\ . \end{aligned}$$

It should be clear from the fact that $\widehat{D}(\chi\varphi_u) \in L^2(M)$ that all the terms above converge to zero as i tends to infinity, except perhaps for the last one (recall that we are only assuming that $0 \le (R_g - n(n-1))|V| \in L^1(M_{\mathrm{ext}})$); the convergence of that last term can be justified as follows:

$$\begin{aligned} \left|\int_M (R_g - n(n-1))\,\langle \chi\varphi_u, \psi-\psi_i\rangle\right| &\le \left(\int_M (R_g - n(n-1))\,\|\chi\varphi_u\|^2_g\right)^{1/2} \\ &\qquad \times\left(\int_M (R_g - n(n-1))\,\|\psi-\psi_i\|^2_g\right)^{1/2} \\ &\le \|\chi\varphi_u\|_H \|\psi-\psi_i\|_H\ . \end{aligned}$$

Now, $F(\psi_i) = F(0)$, and we have shown that ...

•7.14.6 •*7.14.6:* **ptc:***end of trick.tex*

7.15 (Local) Schrödinger-Lichnerowicz identities

•7.15.1: **ptc:** *This is for your convenience, in case you wanted to check some of the calculations of the positive mass theorems section, but I suggest NOT to include this section in the paper, which is already too long. I have written this appendix for myself, to avoid calculating those identities over and over again every other year; we may keep it in the paper if you wished to, though I suggest not*

•7.15.1 We use the hypotheses and conventions of Appendix 7.16. Although an explicit representation of the γ-matrices is never used, we shall give one for four-dimensional Minkowski space-time to convince the uninitiated reader that our requirements can indeed be met. For $i = 1, 2, 3$ let the hermitian *Pauli* matrices σ_i be defined as

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} , \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{7.15.1}$$

with $\sigma^i := \sigma_i$. One readily checks

$$\sigma_i\sigma_j = \delta_{ij} + i\epsilon_{ijk}\sigma^k \qquad \Longrightarrow \qquad \{\sigma_i, \sigma_j\} := \sigma_i\sigma_j + \sigma_j\sigma_i = 2\delta_{ij} \ . \tag{7.15.2}$$

Let the 4×4 complex valued matrices be defined as

$$\gamma^0 = \begin{pmatrix} 0 & \mathrm{id}_{\mathbb{C}^2} \\ \mathrm{id}_{\mathbb{C}^2} & 0 \end{pmatrix} = -\gamma_0 \ , \qquad \gamma_i = \begin{pmatrix} 0 & \sigma_i \\ -\sigma_i & 0 \end{pmatrix} = \gamma^i \ . \tag{7.15.3}$$

Clearly γ_0 is hermitian, while the γ_i's are anti-hermitian with respect to the canonical hermitian scalar product $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ on $\mathbb{C}^4$. From Equation (7.15.2) one immediately finds

$$\{\gamma_i, \gamma_j\} = \begin{pmatrix} -\{\sigma_i, \sigma_j\} & 0 \\ 0 & -\{\sigma_i, \sigma_j\} \end{pmatrix} , \quad \{\gamma_i, \gamma_0\} = 0 \ , \quad (\gamma_0)^2 = 1 \ ,$$

and (7.15.3) leads to (7.16.1), as desired. A real representation of the commutation relations (7.16.1) on $\mathbb{R}^8$ can be obtained by viewing $\mathbb{C}^4$ as a vector space over $\mathbb{R}$, so that 1) each 1 above is replaced by $\mathrm{id}_{\mathbb{R}^2}$, and 2) each i is replaced by the antisymmetric 2×2 matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} .$$

More precisely, let us define the 4×4 symmetric matrices $\hat\sigma_i$ by

$$\hat\sigma_1 = \begin{pmatrix} 0 & \mathrm{id}_{\mathbb{R}^2} \\ \mathrm{id}_{\mathbb{R}^2} & 0 \end{pmatrix} , \quad \hat\sigma_3 = \begin{pmatrix} \mathrm{id}_{\mathbb{R}^2} & 0 \\ 0 & -\mathrm{id}_{\mathbb{R}^2} \end{pmatrix} , \tag{7.15.4}$$

$$\hat\sigma_2 = \begin{pmatrix} 0 & -\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} & 0 \end{pmatrix} , \tag{7.15.5}$$

which are clearly symmetric, and the new γ's by

$$\gamma^0 = \begin{pmatrix} 0 & \mathrm{id}_{\mathbb{R}^4} \\ \mathrm{id}_{\mathbb{R}^4} & 0 \end{pmatrix} = -\gamma_0 \ , \qquad \gamma_i = \begin{pmatrix} 0 & \hat\sigma_i \\ -\hat\sigma_i & 0 \end{pmatrix} = \gamma^i \ . \tag{7.15.6}$$

It should be clear that the γ's satisfy (7.16.1), with γ_0 symmetric, and γ_i's - antisymmetric, as desired.

Consider, now, a connection ∇_i of the form

$$\nabla_i = D_i + A_i \;, \tag{7.15.7}$$

where D_i is the standard spin connection for spinor fields which, locally, are represented by fields with values in V. We shall also use the symbol D_i for the usual Levi-Civita derivative associated to the metric g acting on tensors, *etc.* The matrices γ_μ stand for $c_\gamma(e_\mu)$, with c_γ — the canonical injection of TM into the representation under consideration on V of the Clifford algebra associated with the metric (**??**), and are D-covariantly constant,•7.15.2

•*7.15.2:* **ptc:***in an exhaustive treatment this should also be justified; in particular $\mu = 0$ might not be completely clear*

$$D_i\gamma_\mu = 0 \;.$$

Setting

$$\mathrm{MFD} := \gamma^i\nabla_i \;, \qquad \mathrm{MTD} := \gamma^i D_i \;,$$

we calculate

$$\begin{aligned} D_i\langle\phi,\nabla^i\phi\rangle &= |\nabla\phi|^2 + \langle\phi, D_iD^i\phi\rangle \\ &\quad +\langle\phi, D_i(A^i\phi)\rangle - \langle A^i\phi,(D_i+A_i)\phi\rangle \;, \qquad (7.15.8)\\ D_i\langle\phi,\gamma^i\mathrm{MFD}\phi\rangle &= D_i\langle\phi,\gamma^i\gamma^j\nabla_j\phi\rangle \\ &= -|\mathrm{MFD}\phi|^2 + \langle\phi,\mathrm{MTD}^2\phi\rangle \\ &\quad +\langle\phi,\gamma^i\gamma^jD_i(A_j\phi)\rangle + \langle\gamma^iA_i\phi,\gamma^j(D_j+A_j)\phi\rangle (7.15.9) \end{aligned}$$

Adding we obtain

$$\begin{aligned} D_i\mathscr{U}^i &:= D_i\langle\phi,\left(\nabla^i+\gamma^i\mathrm{MFD}\right)\phi\rangle & (7.15.10a)\\ &= |\nabla\phi|^2 - |\mathrm{MFD}\phi|^2 & (7.15.10b)\\ &\quad +\langle\phi,(D_iD^i+\mathrm{MTD}^2)\phi\rangle & (7.15.10c)\\ &\quad +\langle\phi,\left[A^i-(A^i)^t+\gamma^i\gamma^jA_j+(\gamma^jA_j)^t\gamma^i\right]D_i\phi\rangle & (7.15.10d)\\ &\quad +\langle\phi,\left[D_iA^i+\gamma^i\gamma^jD_iA_j\right]\phi\rangle & (7.15.10e)\\ &\quad +\langle\phi,\left[(\gamma^iA_i)^t\gamma^jA_j-(A^i)^tA_i\right]\phi\rangle \;. & (7.15.10f) \end{aligned}$$

The term (7.15.10c) is independent of the A_i's, and is the one that arises in the original Schrödinger-Lichnerowicz identity:

$$\langle\phi,(D_iD^i+\mathrm{MTD}^2)\phi\rangle = \frac{1}{4}R|\phi|^2 \;. \tag{7.15.11}$$

•7.15.3 Let us start by justifying (7.15.11): we have

•*7.15.3:* **ptc:***proof of the original identity added*

$$\begin{aligned} (D_iD^i+\mathrm{MTD}^2)\phi &= (g^{ij}D_iD_j+\gamma^iD_i\gamma^jD_j)\phi \\ &= (g^{ij}+\gamma^i\gamma^j)D_iD_j\phi \\ &= (g^{ij}+\gamma^i\gamma^j)(D_{(i}D_{j)}+D_{[i}D_{j]})\phi \\ &= \gamma^i\gamma^jD_{[i}D_{j]}\phi \;. \end{aligned} \tag{7.15.12}$$

From the definitions

$$
\begin{aligned}
D_k\phi &= e_k(\phi) - \frac{1}{4}\omega_{ijk}\gamma^i\gamma^j\phi\ , \\
\omega_{ijk} &= g(e_i, \nabla_{e_k} e_j)\ , \\
\nabla_{e_k} X^i &= e_k(X^i) + \omega^i{}_{jk}X^k\ , \\
R^i{}_{\ell jk}X^\ell &= \nabla_{e_j}\nabla_{e_k}X^i - \nabla_{e_k}\nabla_{e_j}X^i - \nabla_{[e_j,e_k]}X^i\ ,
\end{aligned}
$$

one readily finds

$$D_{[i}D_{j]}\phi = -\frac{1}{8}R_{ijk\ell}\gamma^k\gamma^\ell\phi \tag{7.15.13}$$

and to calculate (7.15.12) we need to find $R_{ijk\ell}\gamma^i\gamma^j\gamma^k\gamma^\ell$. Now

$$
\begin{aligned}
R_{ijk\ell}\gamma^i\gamma^j\gamma^k\gamma^\ell &= (R_{i(jk)\ell}\gamma^i\gamma^j\gamma^k\gamma^\ell + \underbrace{R_{i[jk]\ell}}_{=\frac{1}{2}(R_{ijk\ell}-R_{ikj\ell})=\frac{1}{2}(R_{ijk\ell}+R_{ik\ell j})=-\frac{1}{2}R_{i\ell jk}}\gamma^i\gamma^j\gamma^k\gamma^\ell) \\
&= R_{i(jk)\ell}\gamma^i\gamma^{(j}\gamma^{k)}\gamma^\ell - \frac{1}{2}R_{i\ell jk}\gamma^i\gamma^j\gamma^k\gamma^\ell \\
&= -R_{i(jk)\ell}\gamma^i g^{jk}\gamma^\ell - \frac{1}{2}R_{i\ell jk}\gamma^i\gamma^j\gamma^k\gamma^\ell \\
&= R_{i\ell}\gamma^i\gamma^\ell - \frac{1}{2}R_{i\ell jk}\gamma^i\gamma^j\gamma^k\gamma^\ell \\
&= -R - \frac{1}{2}R_{i\ell jk}\gamma^i\gamma^j\gamma^k\gamma^\ell\ . \qquad (7.15.14)
\end{aligned}
$$

Next,

$$
\begin{aligned}
R_{i\ell jk}\gamma^i\gamma^j\gamma^k\gamma^\ell &= R_{i\ell jk}\gamma^i\gamma^j(-2g^{k\ell} - \gamma^\ell\gamma^k) \\
&= -2R_{ij}\gamma^i\gamma^j - R_{i\ell jk}\gamma^i\gamma^j\gamma^\ell\gamma^k \\
&= 2R - R_{i\ell jk}\gamma^i(-2g^{j\ell} - \gamma^\ell\gamma^j)\gamma^k \\
&= 2R - 2R_{ik}\gamma^i\gamma^k + R_{i\ell jk}\gamma^i\gamma^\ell\gamma^j\gamma^k \\
&= 4R + R_{ijk\ell}\gamma^i\gamma^j\gamma^k\gamma^\ell\ .
\end{aligned}
$$

Inserting this in (7.15.14) and rearranging terms one obtains

$$R_{ijk\ell}\gamma^i\gamma^j\gamma^k\gamma^\ell = -2R\ ,$$

and (7.15.11) follows from (7.15.12)-(7.15.13).

In order to work out the remaining terms in (7.15.10), an explicit form of the A_i's is needed. Suppose, first, that ∇ is the space-time spin connection as defined in Equation (**??**):

$$A_i = \frac{1}{2}K^j{}_i\gamma_j\gamma_0\ ; \tag{7.15.15}$$

A_i is then symmetric and we have, by symmetry of K_{jk},

$$
\begin{aligned}
A^i - (A^i)^t + \gamma^i\gamma^j A_j + (\gamma^j A_j)^t\gamma^i &= \frac{1}{2}K_{jk}\left(\gamma^i\gamma^j\gamma^k\gamma_0 + \gamma_0\gamma^k\gamma^j\gamma^i\right) \\
&= -\frac{1}{2}\mathrm{tr}_g K\left(\gamma^i\gamma_0 + \gamma_0\gamma^i\right) = 0(7.15.16)
\end{aligned}
$$

so that there is no contribution from (7.15.10d). Next,

$$\begin{aligned} D^iA_i + \gamma^j\gamma^i D_j A_i &= \frac{1}{2}\left(D^iK_{ij}\gamma^j + D_jK_{ik}\gamma^j\gamma^i\gamma^k\right)\gamma_0 \\ &= \frac{1}{2}\left(D^iK_{ij}\gamma^j - D_j\mathrm{tr}_gK\gamma^j\right)\gamma_0 \\ &= \frac{1}{4}\nu_j\gamma^j\gamma_0 \,, \end{aligned} \tag{7.15.17}$$

with ν as in Equation (**??**); this gives the contribution from (7.15.10e). Using symmetry of $K_{ij}K^i{}_k$ we further have

$$\begin{aligned} -(A^i)^tA_i &= \frac{1}{4}K_{ij}K^i{}_k\gamma^j\gamma_0\gamma^k\gamma_0 \\ &= -\frac{1}{4}K_{ij}K^i{}_k\gamma^j\gamma^k \\ &= -\frac{1}{8}K_{ij}K^i{}_k(\gamma^j\gamma^k + \gamma^k\gamma^j) \\ &= \frac{1}{4}|K|_g^2 \,. \end{aligned}$$

Using

$$\gamma^iA_i = -\mathrm{tr}_gK\gamma_0/2 = (\gamma^iA_i)^t \tag{7.15.18}$$

one obtains

$$(\gamma^iA_i)^t\gamma^jA_j - (A^i)^tA_i = \frac{1}{4}\left(|K|_g^2 - (\mathrm{tr}_gK)^2\right) \,.$$

Collecting all this we are led to

$$D_i\langle\phi, \left(\nabla^i + \gamma^i\mathrm{MFD}\right)\phi\rangle = |\nabla\phi|^2 - |\mathrm{MFD}\phi|^2 + \frac{1}{4}\langle\phi, (\mu + \nu_j\gamma^j\gamma_0)\phi\rangle \,, \tag{7.15.19}$$

where μ is given by Equation (**??**).

Consider, now, the vector field $\mathscr{U}^i$ defined by (7.15.10a):

$$\mathscr{U}^i = \langle\phi, \left(\nabla^i + \gamma^i\mathrm{MFD}\right)\phi\rangle = \langle\phi, \left(D^i + \gamma^i\mathrm{MTD}\right)\phi\rangle + \langle\phi, \left(A^i + \gamma^i\gamma^jA_j\right)\phi\rangle \,.$$

We have

$$\begin{aligned} A^i + \gamma^i\gamma^jA_j &= \frac{1}{2}(K_{ij}\gamma^j - \mathrm{tr}_gK\gamma^i)\gamma_0 \\ &= \frac{1}{2}(\mathrm{tr}_gKg^{ij} - K^{ij})\gamma_0\gamma_j \,, \end{aligned}$$

which integrated upon a coordinate sphere S_R in M_{ext} gives

$$\begin{aligned} &\lim_{R\to\infty}\oint_{S(R)}\langle\phi_\infty, \left(A^i + \gamma^i\gamma^jA_j\right)\phi_\infty\rangle dS_i \\ &= \left(\lim_{R\to\infty}\frac{1}{2}\oint_{S(R)}(\mathrm{tr}_gKg^{ij} - K^{ij})dS_i\right)\langle\phi_\infty, \gamma_0\gamma_j\phi_\infty\rangle \\ &= 4\pi p^i\langle\phi_\infty, \gamma_0\gamma_i\phi_\infty\rangle = 4\pi p_i\langle\phi_\infty, \gamma^i\gamma^0\phi_\infty\rangle \,, \end{aligned} \tag{7.15.20}$$

with p^i — the ADM momentum of $M_{\rm ext}$ (recall that $\gamma^0 = -\gamma_0$). Here ϕ_∞ is a covariantly constant spinor field of the euclidean metric associated to the natural coordinates on $M_{\rm ext}$, as described in Section **??**. A classical calculation then gives•7.15.4

•7.15.4: **ptc:**A complete treatment would also require doing this

$$\lim_{R\to\infty}\oint_{S(R)} \mathscr{U}^i dS_i = 4\pi p_\mu \langle \phi_\infty, \gamma^\mu\gamma^0\phi_\infty\rangle \;. \tag{7.15.21}$$

Consider, next the full connection (**??**),

$$A_i = \underbrace{\frac12 K_{ij}\gamma^j\gamma_0}_{A_i(K)} \underbrace{-\frac12 E^k\gamma_k\gamma_i\gamma_0}_{A_i(E)} \underbrace{-\frac14 \epsilon_{jk\ell}B^j\gamma^k\gamma^\ell\gamma_i}_{A_i(B)} \;. \tag{7.15.22}$$

The contribution of K to the linear terms (7.15.10d) and (7.15.10e) has already been worked out, so it remains to evaluate those of E and B. We have:

$$\begin{aligned}
A_i(E)^t &= -\frac12 E^k\gamma_0\gamma_i\gamma_k = -\frac12 E^k\gamma_i\gamma_k\gamma_0 = -\frac12 E^k\left(-\gamma_k\gamma_i - 2g_{ki}\right)\gamma_0 \\
&= -A_i(E) + E_i\gamma_0 \;, &(7.15.23a)\\
\gamma^j A_j(E) &= -\frac12 E^k\gamma^j\gamma_k\gamma_j\gamma_0 = -\frac12 E^k\left(-\gamma_k\gamma^j - 2\delta^j_k\right)\gamma_j\gamma_0 = -\frac12 E^k\left(3\gamma_k - 2\gamma_k\right)\gamma_0 \\
&= -\frac12 E^k\gamma_k\gamma_0 \;, &(7.15.23b)\\
(\gamma^j A_j(E))^t &= \gamma^j A_j(E) \;. &(7.15.23c)
\end{aligned}$$

This gives

$$\begin{aligned}
&A^i(E) - (A^i(E))^t + \gamma^i\gamma^j A_j(E) + (\gamma^j A_j(E))^t\gamma^i \\
&\quad = -E^k(\gamma_k\gamma^i - \frac12\gamma^i\gamma_k + \frac12\gamma_k\gamma^i)\gamma_0 - E^i\gamma_0 = -\frac12 E^k(\gamma^i\gamma_k + \gamma_k\gamma^i + 2\delta^i_k)\gamma_0 \\
&\quad = 0 \;.
\end{aligned} \tag{7.15.24}$$

We note that

$$\begin{aligned}
A_1(B) &= -\frac12\left[B^1\gamma^2\gamma^3 + B^2\gamma^3\gamma^1 + B^3\gamma^1\gamma^2\right]\gamma_1 \\
&= -\frac12\left[B_1\gamma^1\gamma^2\gamma^3 - B^2\gamma^3 + B^3\gamma^2\right] \\
&= -\frac12\left[B_1\gamma^1\gamma^2\gamma^3 - \epsilon_{1jk}B^j\gamma^k\right] \;,
\end{aligned} \tag{7.15.25}$$

so, since the e_1 direction can be chosen at will,

$$A_i(B) \;=\; \frac12\left[\epsilon_{ijk}B^j\gamma^k - B_i\gamma^1\gamma^2\gamma^3\right] \;. \tag{7.15.26}$$

Now

$$(\gamma^1\gamma^2\gamma^3)^t = (\gamma^3)^t(\gamma^2)^t(\gamma^1)^t = -\gamma^3\gamma^2\gamma^1 = -\gamma^2\gamma^1\gamma^3 = \gamma^1\gamma^2\gamma^3 \;, \tag{7.15.27}$$

$$\gamma^i\gamma^1\gamma^2\gamma^3 = -\tfrac12\epsilon^{ijk}\gamma_j\gamma_k \;, \tag{7.15.28}$$

$$\gamma^i\tfrac12\epsilon_{jk\ell}B^j\gamma^k\gamma^\ell = B^i\gamma^1\gamma^2\gamma^3 + \epsilon^{ijk}B_j\gamma_k \;, \tag{7.15.29}$$

where the last two equations have been obtained by a calculation similar to that of Equation (7.15.25). This leads to

$$
\begin{aligned}
A_i(B)^t &= -\frac{1}{2}\left[\epsilon_{ijk}B^j\gamma^k + B_i\gamma^1\gamma^2\gamma^3\right] , && (7.15.30a)\\
\gamma^i A_i(B) &= \frac{1}{2}\left[\epsilon_{ijk}B^j\gamma^i\gamma^k + \frac{1}{2}\epsilon^{ijk}B_i\gamma_j\gamma_k\right] \\
&= -\frac{1}{4}\epsilon_{ijk}B^i\gamma^j\gamma^k , && (7.15.30b)\\
(\gamma^j A_j(B))^t &= -\gamma^j A_j(B) , && (7.15.30c)\\
\gamma^i\gamma^j A_j(B) &= -\frac{1}{2}\left[B^i\gamma^1\gamma^2\gamma^3 + \epsilon^{ijk}B_j\gamma_k\right] \\
&= (A^i(B))^t , && (7.15.30d)\\
(\gamma^j A_j(B))^t\gamma^i &= -A^i(B) . && (7.15.30e)
\end{aligned}
$$

Here (7.15.30d) follows from (7.15.29) and (7.15.30b), while (7.15.30e) is obtained by comparing minus (7.15.30b) multiplied from the right by γ^i, as justified by (7.15.30c), with the definition (7.15.22) of $A_i(B)$. Using (7.15.30d) and (7.15.30e) we conclude that

$$
\begin{aligned}
&A^i(B) - (A^i(B))^t + \gamma^i\gamma^j A_j(B) + (\gamma^j A_j(B))^t\gamma^i \\
&\quad = A^i(B) - (A^i(B))^t + (A^i(B))^t - A^i(B) \\
&\quad = 0 , \qquad (7.15.31)
\end{aligned}
$$

which shows that the contribution (7.15.10d) to (7.15.10) vanishes. We consider next (7.15.10e):

$$
\begin{aligned}
D_iA^i(E) + \gamma^i\gamma^j D_iA_j(E) &= -\frac{1}{2}D_jE^k\{\gamma_k\gamma^j + \gamma^i\gamma^j\gamma_k\gamma_i\}\gamma_0 \\
&= -\frac{1}{2}D_jE^k\{\gamma_k\gamma^j + (-\gamma^j\gamma^i - 2g^{ij})\gamma_k\gamma_i\}\gamma_0 \\
&= -\frac{1}{2}D_jE^k\{-\gamma_k\gamma^j - \gamma^j(-\gamma_k\gamma^i - 2\delta^i_k)\gamma_i\}\gamma_0 \\
&= -\frac{1}{2}D_jE^k\{-\gamma_k\gamma^j - 3\gamma^j\gamma_k + 2\gamma^j\gamma_k\}\gamma_0 \\
&= -D_jE^j\gamma_0 =: -\mathrm{div}(E)\gamma_0 . && (7.15.32)
\end{aligned}
$$

To analyze the contribution of $A_i(B)$ to (7.15.10e) it is convenient to use (7.15.26), which gives

$$
D_iA^i(B) + \gamma^i\gamma^j D_iA_j(B) = \frac{1}{2}D_iB_k[\epsilon^{jk\ell}(\delta^i_j\gamma_\ell + \gamma^i\gamma_j\gamma_\ell) - (g^{ik} + \gamma^i\gamma^k)\gamma^1\gamma^2\gamma^3] . \qquad (7.15.33)
$$

Fortunately it is not necessary to evaluate this expression in detail, because any antisymmetric matrix appearing above will give a zero contribution after insertion into (7.15.10e). Another way of seeing that is that for any linear map F we have

$$
\langle\phi, F\phi\rangle = \langle\phi, F^t\phi\rangle = \langle\phi, \frac{1}{2}(F + F^t)\phi\rangle . \qquad (7.15.34)
$$

Now, by Clifford algebra rules, the right-hand-side of (7.15.33) will be a linear combination of γ^i's and of $\gamma^1\gamma^2\gamma^3$; the γ^i's are antisymmetric and can thus be ignored, and it remains to work out the coefficient in front of the symmetric matrix $\gamma^1\gamma^2\gamma^3$. The first term gives no such contribution, the second one will give one when i equals k, producing then a contribution $-2g^{ik}\gamma^1\gamma^2\gamma^3$, while the only possible contribution from the last two terms could occur when $i = k$, but then they cancel out each other. We are thus led to

$$\begin{aligned} D_iA^i(B) + \gamma^i\gamma^j D_iA_j(B) &= -D_iB^i\gamma^1\gamma^2\gamma^3 + \text{antisymmetric} \\ &=: -\text{div}(B)\gamma^1\gamma^2\gamma^3 + \text{antisymmetric} \end{aligned} \tag{7.15.35}$$

Let us, finally, consider the quadratic term (7.15.10f); from Equations (7.15.18), (7.15.23b) and (7.15.30b) together with (7.15.28) we have

$$\begin{aligned} \gamma_iA^i &= -\frac{1}{2}(\text{tr}_gK + E^i\gamma_i)\gamma_0 - \frac{1}{4}\epsilon_{ijk}B^i\gamma^j\gamma^k \\ &= -\frac{1}{2}\left[(\text{tr}_gK + E^i\gamma_i)\gamma_0 - B^i\gamma_i\gamma^1\gamma^2\gamma^3\right] . \end{aligned} \tag{7.15.36}$$

It follows that

$$(\gamma_iA^i)^t = -\frac{1}{2}\left[(\text{tr}_gK + E^i\gamma_i)\gamma_0 + B^i\gamma_i\gamma^1\gamma^2\gamma^3\right] , \tag{7.15.37}$$

and

$$\begin{aligned} (\gamma_iA^i)^t\gamma_jA^j &= \frac{1}{4}\left[(\text{tr}_gK + E^i\gamma_i)\gamma_0 + B^i\gamma_i\gamma^1\gamma^2\gamma^3\right]\left[(\text{tr}_gK + E^j\gamma_j)\gamma_0 - B^j\gamma_j\gamma^1\gamma^2\gamma^3\right] \\ &= \frac{1}{4}\Big\{(\text{tr}_gK + E^i\gamma_i)\gamma_0(\text{tr}_gK + E^j\gamma_j)\gamma_0 - B^iB^j\gamma_i\gamma^1\gamma^2\gamma^3\gamma_j\gamma^1\gamma^2\gamma^3 \\ &\qquad -E^iB^j\left[\gamma_i\gamma_0\gamma_j\gamma^1\gamma^2\gamma^3 - \gamma_j\gamma^1\gamma^2\gamma^3\gamma_i\gamma_0\right]\Big\} \\ &= \frac{1}{4}\Big\{(\text{tr}_gK + E^i\gamma_i)(\text{tr}_gK - E^j\gamma_j) - B^iB^j\gamma_i\gamma_j(\gamma^1\gamma^2\gamma^3)^2 \\ &\qquad +E^iB^j\left[\gamma_i\gamma_j - \gamma_j\gamma_i\right]\gamma^1\gamma^2\gamma^3\gamma_0\Big\} \\ &= \frac{1}{4}\Big\{(\text{tr}_gK)^2 + |E|_g^2 + |B|_g^2 - 2\epsilon_{ijk}E^iB^j\gamma^k\gamma_0\Big\} . \end{aligned} \tag{7.15.38}$$

Next, from the definition (7.15.22) together with (7.15.23a) and (7.15.30a) we obtain

$$(A_i)^t = \frac{1}{2}\left[(K_{ij}\gamma^j + 2E_i + E^k\gamma_k\gamma_i)\gamma_0 - \epsilon_{ijk}B^j\gamma^k - B_i\gamma^1\gamma^2\gamma^3\right] . \tag{7.15.39}$$

Using the form (7.15.26) of $A_i(B)$ one has

$$\begin{aligned} (A_i)^tA^i &= \frac{1}{4}\left[(K_{ij}\gamma^j + 2E_i + E^k\gamma_k\gamma_i)\gamma_0 - \epsilon_{ijk}B^j\gamma^k - B_i\gamma^1\gamma^2\gamma^3\right] \\ &\qquad \times\left[(K^i{}_j\gamma^j - E^k\gamma_k\gamma^i)\gamma_0 + \epsilon^{ijk}B_j\gamma_k - B^i\gamma^1\gamma^2\gamma^3\right] \end{aligned} \tag{7.15.40}$$

Again, we do not need to calculate all the terms above, only the symmetric part matters. It is straightforward to check that the following symmetry properties hold:

$$\text{symmetric}: \quad \gamma_0\ , \quad \gamma_i\gamma_0\ , \quad \gamma^1\gamma^2\gamma^3\ , \tag{7.15.41a}$$

$$\text{antisymmetric}: \quad \gamma_i\ , \quad \gamma_i\gamma_j \text{ and } \gamma_i\gamma_j\gamma_0 \text{ for } i\neq j\ , \quad \gamma^1\gamma^2\gamma^3\gamma_0 \tag{7.15.41b}$$

For example,

$$(\gamma^1\gamma^2\gamma^3\gamma_0)^t = \gamma_0^t(\gamma^1\gamma^2\gamma^3)^t = \gamma_0\gamma^1\gamma^2\gamma^3 = -\gamma^1\gamma^2\gamma^3\gamma_0\ , \tag{7.15.42}$$

the remaining claims in (7.15.41) being proved similarly, *cf.* also (7.15.27). Using (7.15.41), Equation (7.15.40) can be manipulated[5] as follows

$$\begin{aligned}
(A_i)^tA^i &= \frac14\Big\{(K_{i\ell}\gamma^\ell+2E_i+E^\ell\gamma_\ell\gamma_i)\gamma_0\\
&\qquad\times\Big[(K^i{}_j\gamma^j-E^k\gamma_k\gamma^i)\gamma_0+\epsilon^{ijk}B_j\gamma_k-B^i\gamma^1\gamma^2\gamma^3\Big]\\
&\quad-\epsilon_{i\ell m}B^\ell\gamma^m\Big[(K^i{}_j\gamma^j-E^k\gamma_k\gamma^i)\gamma_0+\epsilon^{ijk}B_j\gamma_k-B^i\gamma^1\gamma^2\gamma^3\Big]\\
&\quad-B_i\gamma^1\gamma^2\gamma^3\Big[(K^i{}_j\gamma^j-E^k\gamma_k\gamma^i)\gamma_0+\epsilon^{ijk}B_j\gamma_k-B^i\gamma^1\gamma^2\gamma^3\Big]\Big\}\\
&= \frac14\Big\{(K_{i\ell}\gamma^\ell+2E_i+E^\ell\gamma_\ell\gamma_i)\\
&\qquad\times\Big[-K^i{}_j\gamma^j-E^k\gamma_k\gamma^i-\epsilon^{ijk}B_j\gamma_k\gamma_0+B^i\gamma^1\gamma^2\gamma^3\gamma_0\Big]\\
&\quad-\epsilon_{i\ell m}B^\ell\gamma^m(K^i{}_j\gamma^j-E^k\gamma_k\gamma^i)\gamma_0+\epsilon_{i\ell m}B^\ell\epsilon^{ijm}B_j\\
&\quad-B_i\gamma^1\gamma^2\gamma^3\Big[-E^k\gamma_k\gamma^i\gamma_0-B^i\gamma^1\gamma^2\gamma^3\Big]\Big\}+\text{antisymmetric}\\
&= \frac14\Big\{(K_{i\ell}\gamma^\ell\Big[-K^i{}_j\gamma^j-E^k\gamma_k\gamma^i-\epsilon^{ijk}B_j\gamma_k\gamma_0\Big]\\
&\quad+2E_i\Big[-E^k\gamma_k\gamma^i-\epsilon^{ijk}B_j\gamma_k\gamma_0\Big]\\
&\quad+E^\ell\gamma_\ell\gamma_i\Big[-K^i{}_j\gamma^j-E^k\gamma_k\gamma^i-\epsilon^{ijk}B_j\gamma_k\gamma_0+B^i\gamma^1\gamma^2\gamma^3\gamma_0\Big]\\
&\quad+\underbrace{\epsilon_{i\ell m}B^\ell K^{im}}_{0}+\underbrace{\underbrace{B^\ell E^k\epsilon_{i\ell m}}_{\text{antisym. in } i,m}\underbrace{(-\gamma^m\delta^i_k-\delta^m_k\gamma^i)}_{\text{sym. in } i,m}}_{0}\gamma_0+2|B|^2_g\\
&\quad+\epsilon_{ik\ell}B^iE^k\gamma^\ell\gamma_0+|B|^2_g\Big\}+\text{antisymmetric}\\
&= \frac14\Big\{|K|^2_g-\underbrace{K_{i\ell}E_k\epsilon^{\ell ki}}_{0}\gamma^1\gamma^2\gamma^3+\underbrace{\epsilon^{ijk}K_{ik}B_j}_{0}\gamma_0\\
&\quad+2|E|^2_g-2\epsilon^{ijk}E_iB_j\gamma_k\gamma_0\\
&\quad-E^\ell E^k\gamma_\ell\underbrace{(-2g_{ik}-\gamma_k\gamma_i)\gamma^i}_{\gamma_k}-\epsilon^{ijk}B_jE^\ell(-g_{\ell i}\gamma_k+g_{k\ell}\gamma_i)\gamma_0-\epsilon_{ijk}B^iE^j\gamma^k\gamma_0\\
&\quad+2|B|^2_g
\end{aligned}$$

[5]The calculation here can be somewhat simplified by noting at the outset that there is no symmetric nonzero matrix which can be built by contraction with K, E, and the γ-matrices with no indices left, similarly for K and B, hence the contributions from $A_i(K)$ and that of $A_i(E)+A_i(B)$ can be computed separately.

$$-\epsilon_{ijk}B^iE^j\gamma^k\gamma_0+|B|_g^2\Big\}+\text{antisymmetric}$$
$$=\frac14\Big\{|K|_g^2+3|E|_g^2+3|B|_g^2-6\epsilon^{ijk}E_iB_j\gamma_k\gamma_0\Big\}\,, \tag{7.15.43}$$

where the last equality is justified by the fact that all the antisymmetric matrices have to cancel out, since the matrix at the left-hand-side of the first line of (7.15.43) is symmetric. Subtracting (7.15.43) from (7.15.38) we thus find the following formula for the term (7.15.10f):

$$\langle\phi,\big\{(\gamma^iA_i)^t\gamma^jA_j-(A^i)^tA_i\big\}\,\phi\rangle=$$
$$-\frac14\langle\phi,\Big\{|K|_g^2-(\mathrm{tr}_gK)^2+2|E|_g^2+2|B|_g^2-4\epsilon^{ijk}E_iB_j\gamma_k\gamma_0\Big\}\phi\rangle. \tag{7.15.44}$$

Summarizing, Equations (7.15.10), (7.15.11), (7.15.16), (7.15.17), (7.15.24), (7.15.31), (7.15.32), (7.15.35) and (7.15.44) lead to

$$D_i\langle\phi,(\nabla^i+\gamma^i\gamma^j\nabla_j)\phi\rangle=|\nabla\phi|^2-|\mathrm{MFD}\phi|^2$$
$$+\frac14\langle\phi,\Big\{\mu+(\nu_i\gamma^i-\mathrm{div}(E))\gamma_0-\mathrm{div}(B)\gamma^1\gamma^2\gamma^3\Big\}\phi\rangle\,, \tag{7.15.45}$$

where

$$\mu:=R-|K|_g^2+(\mathrm{tr}_gK)^2-2|E|_g^2-2|B|_g^2\,, \tag{7.15.46a}$$
$$\nu_j=2D_i(K^{ij}-\mathrm{tr}K g^{ij})+4\epsilon_{jk\ell}E^kB^\ell\,. \tag{7.15.46b}$$

We turn now our attention to the electromagnetic field contribution to the boundary integrand $\langle\phi,(\nabla^i+\gamma^i\gamma^j\nabla_j)\phi\rangle$:

$$A^i(E)+\gamma^i\gamma^jA_j(E)\;=\;E^i\gamma_0+\text{antisymmetric}\,, \tag{7.15.47a}$$
$$A^i(B)+\gamma^i\gamma^jA_j(B)\;=\;-\frac12B^i\gamma^1\gamma^2\gamma^3-\frac14\epsilon_{jk\ell}B^j\gamma^i\gamma^k\gamma^\ell+\text{antisymmetric}$$
$$=\;-B^i\gamma^1\gamma^2\gamma^3+\text{antisymmetric}\,, \tag{7.15.47b}$$

and Equation (7.15.21) gives

$$\lim_{R\to\infty}\oint_{S(R)}\mathscr{U}^idS_i=4\pi\langle\phi_\infty,\left[p_\mu\gamma^\mu\gamma^0+Q\gamma_0-P\gamma^1\gamma^2\gamma^3\right]\phi_\infty\rangle\,. \tag{7.15.48}$$

Let us turn our attention now to the hyperbolic case: let $\alpha\in\mathbb{R}$, we set

$$\mathscr{A}_i=A_i+\underbrace{\frac{\alpha\sqrt{-1}}{2}\gamma_i}_{A_i(\alpha)}\,,\qquad \nabla_i=D_i+\mathscr{A}_i\,, \tag{7.15.49}$$

with A_i as in Equation (7.15.22). (This is the standard way of taking into account a cosmological constant when no Maxwell field is present [54].) Here $\sqrt{-1}:V\to V$ is any map satisfying

$$(\sqrt{-1})^2=-\mathrm{id}_V\,,\quad \sqrt{-1}\gamma_i=\gamma_i\sqrt{-1}\,,\quad (\sqrt{-1})^t=-\sqrt{-1}\,. \tag{7.15.50}$$

(If V is a complex vector space understood as a vector space over $\mathbb{R}$, with the real scalar product $\langle\cdot,\cdot\rangle$ arising from a sesquilinear form $\langle\cdot,\cdot\rangle_{\mathbb{C}}$, then $\sqrt{-1}$ can be taken as multiplication by i. On the other hand, if no such map $\sqrt{-1}$ exists on V, one can always replace V by its complexification $V_{\mathbb{C}} := V\otimes\mathbb{C}$, with new matrices $\gamma_\mu\otimes \mathrm{id}_{\mathbb{C}}$, and use multiplication by $\mathrm{id}_V\otimes(i\ \mathrm{id}_{\mathbb{C}})$ on $V_{\mathbb{C}}$ as the desired map.)•7.15.5 We consider again Equation (7.15.10), with A there replaced by $\mathscr{A}$. Now, the terms (7.15.10d) and (7.15.10e) are linear in $\mathscr{A}$, and they have already been shown to vanish when $\alpha=0$; thus, to show that they vanish for $\alpha\neq 0$ it suffices to show that they do so when $A_i = A_i(\alpha)$. This is obvious for (7.15.10e), while for (7.15.10d) we have

•7.15.5: **ptc:**_this discussion is a little of an overkill, since we are in space-dimension three, so everything can be made explicit_

$$(A_i(\alpha))^t = A_i(\alpha)\ ,$$

$$\gamma^i A_i(\alpha) = -3\frac{\alpha\sqrt{-1}}{2}\ ,\qquad (\gamma^i A_i(\alpha))^t = 3\frac{\alpha\sqrt{-1}}{2}\ ,$$

$$A^i(\alpha)-(A^i(\alpha))^t+\gamma^i\gamma^j A_j(\alpha)+(\gamma^j A_j(\alpha))^t\gamma^i = \frac{\alpha\sqrt{-1}}{2}\left(\gamma^i+(\gamma^i)^t-\frac{3}{2}\gamma^i+\frac{3}{2}\gamma^i\right)=0\ .$$

It follows that the only new terms that can perhaps occur in Equations (7.15.10c)-(7.15.10f) arise from (7.15.10f). We calculate

$$\begin{aligned}\left(\gamma^i A_i+\gamma^i A_i(\alpha)\right)^t\left(\gamma^j A_j+\gamma^j A_j(\alpha)\right) &= \left((\gamma^i A_i)^t+3\frac{\alpha\sqrt{-1}}{2}\right)\left(\gamma^j A_j-3\frac{\alpha\sqrt{-1}}{2}\right)\\ &= (\gamma^i A_i)^t\gamma^j A_j+\frac{9\alpha^2}{4}\\ &\quad +3\frac{\alpha\sqrt{-1}}{2}\left(\gamma^j A_j-(\gamma^i A_i)^t\right)\ , \qquad (7.15.51)\\ \left(A^i+A^i(\alpha)\right)^t\left(A_i+A_i(\alpha)\right) &= \left((A^i)^t+\frac{\alpha\sqrt{-1}}{2}\gamma^i\right)\left(A_i+\frac{\alpha\sqrt{-1}}{2}\gamma_i\right)\\ &= (A^i)^t A_i+\frac{3\alpha^2}{4}\\ &\quad +\frac{\alpha\sqrt{-1}}{2}(\underbrace{(A_i)^t\gamma^i}_{-(A_i)^t(\gamma^i)^t=-(\gamma^i A_i)^t}+\gamma^j A_j)\ , \qquad (7.15.52)\end{aligned}$$

so that

$$\begin{aligned}(\gamma^i\mathscr{A}_i)^t\gamma^j\mathscr{A}_j-(\mathscr{A}^i)^t\mathscr{A}_i &= (\gamma^i A_i)^t\gamma^j A_j-(A^i)^t A_i+\frac{3\alpha^2}{2}\\ &\quad +\alpha\sqrt{-1}\left(\gamma^j A_j-(\gamma^i A_i)^t\right)\ . \qquad (7.15.53)\end{aligned}$$

Equations (7.15.36)-(7.15.37) show that

$$\alpha\sqrt{-1}\left(\gamma^j A_j-(\gamma^i A_i)^t\right)=\alpha\sqrt{-1}B^i\gamma_i\gamma^1\gamma^2\gamma^3\ , \qquad (7.15.54)$$

so that we obtain

$$\begin{aligned}&D_i\langle\phi,(\nabla^i+\gamma^i\gamma^j\nabla_j)\phi\rangle = |\nabla\phi|^2-|\mathrm{MFD}\phi|^2\\ &\quad +\frac{1}{4}\langle\phi,\left\{\mu+6\alpha^2+(\nu_i\gamma^i-\mathrm{div}(E))\gamma_0+\left(4\alpha\sqrt{-1}B^i\gamma_i-\mathrm{div}(B)\right)\gamma^1\gamma^2\gamma^3\right\}\phi\rangle\ ,\end{aligned}$$

(7.15.55)

where μ,ν is as in (7.15.46), div(E) is the divergence of E and div(B) that of B. Somewhat surprisingly, the term $B^i\gamma_i\gamma^1\gamma^2\gamma^3$ occurring above does not seem to combine in any obvious way with the remaining ones to yield a useful identity except when B vanishes. •7.15.6 •7.15.7 In any case the identity (7.15.55) can then be used to prove a mass-charge inequality in an asymptotically hyperboloidal setting, in the spirit of Theorem **??** with $B\equiv 0$ there, both in the boundaryless and in the trapped-boundary cases.•7.15.8 It turns out that a slightly different approach can be used to handle the negative cosmological constant with a general Maxwell field provided that $\mathrm{tr}_g K$ is constant, as follows: let $a\in\mathbb{R}$ and consider the connection (7.15.22) with K_{ij} replaced by

•*7.15.6:* **ptc:***This is very strange - do you see any way of fixing this? I have added new material that makes it work when* $\mathrm{tr}_g K$ *is a constant, but one would expect that no such restriction is needed?*

•*7.15.7:* **ptc:***Warning: this is the file hyperMax, its maintained version is in the levoca directory*

•*7.15.8:* **ptc:***except that I have some difficulties with understanding the associated boundary terms...*

$$\bar K_{ij} := K_{ij} + a g_{ij}\ . \tag{7.15.56}$$

The identity (7.15.45) will then hold with the current ν unchanged, as follows from (7.15.46b), while μ given by (7.15.46a) will be replaced by

$$\begin{aligned}\mu &:= R-|\bar K|^2_g+(\mathrm{tr}_g\bar K)^2-2|E|^2_g-2|B|^2_g\\ &= R-|K|^2_g+(\mathrm{tr}_g K)^2+4a\mathrm{tr}_gK+6a^2-2|E|^2_g-2|B|^2_g\ .\end{aligned}\tag{7.15.57}$$

We will then recover Equation (7.15.55) provided that a is chosen as either one of the following roots

$$a_\pm := -\frac{\mathrm{tr}_gK}{3}\pm\sqrt{\left(\frac{\mathrm{tr}_gK}{3}\right)^2+\alpha^2}\ .$$

Let us turn our attention now to the problem of boundary conditions which one imposes in the asymptotically hyperboloidal case: consider ....•7.15.9 As shown in [32], when

•*7.15.9:* **ptc:***to be finished, several objects to be defined, zomega, ourU*

$$\psi = \psi_u + \chi\ , \tag{7.15.58}$$

with $\chi\in\mathcal{H}$ and with ψ_u satisfying•7.15.10

•*7.15.10:* **ptc:***use systematically complex numbers, or say something about that; fix normalisation*

$$\mathring{D}\psi_u := d\psi_u - \frac14\mathring{\omega}_{ij}\gamma^i\gamma^j\psi_u = -\frac{i}{2}\gamma_j\otimes\mathring{\theta}^j\ \psi_u\ ; \tag{7.15.59}$$

here $\{\mathring{\theta}^i\}$ is a co-frame dual to $\{\mathring{e}_i\}$, while u is any parameter •7.15.11 parameterizing the solutions of (7.15.59). Then we have

•*7.15.11:* **ptc:***describe which*

$$\lim_{r\to\infty}\oint_{S_r}\langle\psi, c_g(n)c_g(e^A)(D_A+i\gamma_A)\psi\rangle \;=\; \frac14\lim_{r\to\infty}\int_{S_r}\mathbb{U}^i\nu_i. \tag{7.15.60}$$

To be able to carry through the positivity argument one needs to be able to solve the equation

$$\mathrm{MFD}\chi = -\mathrm{MFD}(\varphi\psi_u)$$

for a $\chi\in\mathcal{H}$, where φ is a cutoff function supported in M_{ext}, with φ equal to one for large distances. This will be possible when $\mathrm{MFD}\psi_u\in L^2(M_{\mathrm{ext}})$, which in turn will be guaranteed by

$$V\mathrm{tr}_gK\ ,\ VE\ ,\ VB\in L^2(M_{\mathrm{ext}})\ , \tag{7.15.61}$$

where

$$V = V(u) := \langle \psi_u, \psi_u \rangle \ . \tag{7.15.62}$$

It is rather clear that Equation (7.15.61) forces the total electric and magnetic charge to vanish, so that under (7.15.61) the addition of $A_i(E)$ and $A_i(B)$ to the connection ∇ does not lead to any new information, while complicating considerably the argument.•7.15.12 When Equation (7.15.61) holds we might as well take $A_i = A_i(K)$. Now, the boundary contribution from K is obtained from (7.15.20), except that the part of the integrand which arises from ψ_u is not constant over the boundary at infinity in general:•7.15.13

•*7.15.12:* **ptc:** *This is not clear? I need to think this over; what is the correct definition of E*

•*7.15.13:* **ptc:** *maybe this vanishes?*

$$\begin{aligned}
&\lim_{R\to\infty} \oint_{S(R)} \langle \psi_u, \left(A^i + \gamma^i\gamma^j A_j\right) \psi_u \rangle dS_i \\
&\quad = \lim_{R\to\infty} \frac{1}{2} \oint_{S(R)} \left((\mathrm{tr}_g K g^{ij} - K^{ij}) \langle \psi_u, \gamma_0 \gamma_j \psi_u \rangle \right) dS_i \\
&\quad = \lim_{R\to\infty} \frac{1}{2} \oint_{S(R)} \left((\mathrm{tr}_g K \delta^i_j - K^i{}_j) Y^j \right) dS_i \ ,
\end{aligned} \tag{7.15.63}$$

where

$$Y^j := \langle \psi_u, \gamma_0 \gamma^j \psi_u \rangle \ . \tag{7.15.64}$$

•7.15.14 Collecting (7.15.60) and (7.15.63) we obtain

•*7.15.14:* **ptc:** *If* $A_i = A_i(K)$,

$$\lim_{R\to\infty} \oint_{S(R)} \mathscr{U}^i dS_i = 4\pi H(X) \ , \tag{7.15.65}$$

where $H(X)$ is the Hamiltonian energy associated with the space-time vector field X such that its normal component to the initial data hypersurface equals n, and its tangential component equals Y [?];•7.15.15 recall that $\mathscr{U}$ is the vector field defined in Equation (7.15.10a).

•*7.15.15:* **ptc:** *To jest jakies dziwne, bo dla anty de Sitera to i tak jest wszystko styczne? moze to znika?*

Fall-off conditions weaker than those of (7.15.61) can be obtained if we consider comparison spinor fields ψ_u which, instead of (7.15.59), satisfy

$$\mathring{D}\psi_u = -\frac{i}{2}\gamma_j \otimes \mathring{e}^j \ \psi_u - \mathring{A}_j \otimes \mathring{e}^j \ \psi_u \ , \tag{7.15.66}$$

with

$$\mathring{A} := \mathring{A}_i \otimes \mathring{\theta}^i = \Big(\underbrace{\frac{1}{2}\mathring{K}_{ij}\gamma^j\gamma_0}_{A_i(\mathring{K})} \underbrace{-\frac{1}{2}\mathring{E}^k\gamma_k\gamma_i\gamma_0}_{A_i(\mathring{E})} \underbrace{-\frac{1}{4}\epsilon_{jk\ell}\mathring{B}^j\gamma^k\gamma^\ell\gamma_i}_{A_i(\mathring{B})} \Big) \otimes \mathring{\theta}^i \ . \tag{7.15.67}$$

for some background fields $\mathring{K}, \mathring{E}, \mathring{B}$; in (7.15.67) the indices on $\mathring{K}$, $\mathring{E}$ and $\mathring{B}$ are frame indices with respect to the background ON-frame $\{\mathring{e}_i\}$. This leads to the need of reexamining the calculations of [32] for the contribution of the new fields introduced...•7.15.16 In the case of the connection (7.15.22) with the replacement (7.15.56), the asymptotic spinors one needs to consider are not solutions of (7.15.66) anymore, this equation has to be replaced by

•*7.15.16:* **ptc:** *to be done*

$$\mathring{D}_{\mathring{e}_i}\psi_u = -\frac{a}{2}\gamma_i\gamma_0\psi_u - \mathring{A}_i\psi_u \ . \tag{7.15.68}$$

In the simplest case $\mathring{A}=0$ there is a one-to-one correspondence between solutions of (7.15.66) (which coincides then with (7.15.59)) and those of (7.15.68), obtained as follows: Let ϕ_u satisfy (7.15.59), and set

$$\psi_u := \frac{1}{\sqrt{2}}(1 \pm i\gamma_0)\phi_u \quad \Longleftrightarrow \quad \phi_u := \frac{1}{\sqrt{2}}(1 \mp i\gamma_0)\psi_u \; . \tag{7.15.69}$$

We then have

$$\begin{aligned}
\mathring{D}_{\mathring{e}_i}\psi_u &= -\frac{i}{2\sqrt{2}}(1 \pm i\gamma_0)\gamma_i\phi_u \\
&= \frac{i}{2\sqrt{2}}\gamma_i(1 \mp i\gamma_0)\phi_u \\
&= \frac{i}{2\sqrt{2}}\gamma_i(\gamma_0)^2(1 \mp i\gamma_0)\phi_u \\
&= \frac{i}{2\sqrt{2}}\gamma_i\gamma_0 \underbrace{(\gamma_0 \mp i)}_{\mp i(1 \pm i\gamma_0)} \phi_u \\
&= \pm\frac{1}{2}\gamma_i\gamma_0\psi_u \; .
\end{aligned}$$

In the hyperboloidal case, there are still issues about the boundary integrals which I do not understand at all when K does not vanish; so, for the while, I am leaving things as they are. We can finish off the paper without having to expand the cosmological constant case. I will come back to this when I have more time

*•7.15.17: **ptc:** This paragraph does not lead anywhere, and is kept here only for recycling purposes*

•7.15.17 In fact, Equation (7.15.55) will always hold when the offending term (**??**) vanishes, there is an important case when this happens. Namely, suppose that V is the real form of a complex vector space, still denoted by V, with

$$\sqrt{-1} = i\ \mathrm{id}_V \; , \quad \langle \cdot,\cdot \rangle = \Re(\langle \cdot,\cdot \rangle_{\mathbb{C}}) \; ,$$

where $\Re$ denotes the real part, and $\langle \cdot,\cdot \rangle_{\mathbb{C}}$ is a sesquilinear scalar product on V. Suppose further that the γ^i's are anti-Hermitian, and γ_0 is Hermitian. Then all the calculations performed so far remain correct with trivial modifications, such as the replacement of

7.16 Eigenvalues of certain matrices

In the proofs of the energy-momentum inequalities of Section **??** the positivity properties of several matrices acting on the space of spinors have to be analyzed. It is sufficient to make a pointwise analysis, so we consider a real vector space V equipped with a scalar product $\langle\cdot,\cdot\rangle$ together with matrices γ_μ, $\mu = 0,1,2,3$ satisfying

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = -2\eta_{\mu\nu} , \tag{7.16.1}$$

where $\eta = \mathrm{diag}(-1,1,1,1)$. We further suppose that the matrices γ_μ^t, transposed with respect to $\langle\cdot,\cdot\rangle$, satisfy

$$\gamma_0^t = \gamma_0 , \quad \gamma_i^t = -\gamma_i ,$$

where the index i runs from one to three. Let us start with

$$a^\mu\gamma_0\gamma_\mu = a^0 + a^i\gamma_0\gamma_i , \quad (a^\mu) = (a^0,\vec{a}) = (a^0,(a^i)) .$$

The matrix $a^i\gamma_0\gamma_i$ is symmetric and satisfies

$$(a^i\gamma_0\gamma_i)^2 = a^ia^j\gamma_0\gamma_i\gamma_0\gamma_j = -a^ia^j\gamma_0\gamma_0\gamma_i\gamma_j = |\vec{a}|^2_\delta ,$$

so that its eigenvalues belong to the set $\{\pm|\vec{a}|_\delta\}$. Since γ_0 anticommutes with $a^i\gamma_0\gamma_i$, it interchanges the eigenspaces with positive and negative eigenvalues. Let ψ_i, $i = 1,\cdot,N$ be an ON basis of the $|\vec{a}|_\delta$ eigenspace of $a^i\gamma_0\gamma_i$, set

$$\phi_{2i-1} = \psi_i , \quad \phi_{2i} = \gamma_0\psi_i .$$

It follows that $\{\phi_i\}_{i=1}^{2N}$ forms an ON basis of V (in particular $\mathrm{dim}V = 2N$), and in that basis $a^\mu\gamma_0\gamma_\mu$ is diagonal with entries $a^0 \pm |\vec{a}|_\delta$. We have thus proved

PROPOSITION 7.16.1 *The quadratic form $\langle\psi, a^\mu\gamma_0\gamma_\mu\psi\rangle$ is non-negative if and only if $a^0 \geq |\vec{a}|_\delta$.*

Let us consider, next, the symmetric matrix

$$A := a^\mu\gamma_0\gamma_\mu + b\gamma_0 + c\gamma_1\gamma_2\gamma_3 . \tag{7.16.2}$$

Let ψ_1 be an eigenvecteur of $a^i\gamma_0\gamma_i$ with eigenvalue $|\vec{a}|_\delta$, set

$$\phi_1 = \psi_1 , \quad \phi_2 = \gamma_0\psi_1 , \quad \phi_3 = \gamma_1\gamma_2\gamma_3\psi_1 , \quad \phi_4 = \gamma_1\gamma_2\gamma_3\gamma_0\psi_1 .$$

From the commutation relations (7.16.1) one easily finds

$$a^i\gamma_0\gamma_i\phi_1 = |\vec{a}|_\delta\phi_1 , \quad a^i\gamma_0\gamma_i\phi_2 = -|\vec{a}|_\delta\phi_2 , \quad a^i\gamma_0\gamma_i\phi_3 = -|\vec{a}|_\delta\phi_3 , \quad a^i\gamma_0\gamma_i\phi_4 = |\vec{a}|_\delta\phi_4 ,$$

$$\gamma_0\phi_1 = \phi_2 , \quad \gamma_0\phi_2 = \phi_1 , \quad \gamma_0\phi_3 = -\phi_4 , \quad \gamma_0\phi_4 = -\phi_3 ,$$

$$\gamma_1\gamma_2\gamma_3\phi_1 = \phi_3 , \quad \gamma_1\gamma_2\gamma_3\phi_2 = \phi_4 , \quad \gamma_1\gamma_2\gamma_3\phi_3 = \phi_1 , \quad \gamma_1\gamma_2\gamma_3\phi_4 = \phi_2 .$$

It is simple to check that the ϕ_i's so defined are ON; proceeding by induction one constructs an ON-basis $\{\phi_i\}_{i=1}^{2N}$ of V (in particular dim$V$ is a multiple of 4) in which A is block-diagonal, built-out of blocks of the form

$$\begin{pmatrix} a^0+|\vec{a}|_\delta & b & c & 0 \\ b & a^0-|\vec{a}|_\delta & 0 & c \\ c & 0 & a^0-|\vec{a}|_\delta & -b \\ 0 & c & -b & a^0+|\vec{a}|_\delta \end{pmatrix} .$$

The eigenvalues of this matrix are easily found to be $a^0 \pm \sqrt{|\vec{a}|_\delta^2+b^2+c^2}$. We thus have:

PROPOSITION 7.16.2 *We have the sharp inequality*

$$\langle\psi, (a^\mu\gamma_0\gamma_\mu + b\gamma_0 - c\gamma_1\gamma_2\gamma_3)\,\psi\rangle \geq \left(a^0 - \sqrt{|\vec{a}|_\delta^2+b^2+c^2}\right)|\psi|^2 ,$$

in particular the quadratic form $\langle\psi, A\psi\rangle$, with A defined in (7.16.2), is non-negative if and only if

$$a^0 \geq \sqrt{|\vec{a}|_\delta^2+b^2+c^2} .$$

Chapter 8

Black holes

•8.0.1

•*8.0.1:* **ptc:***A figure by Anil Zeninoglu from Helmut's Cargese volume*

Figure 8.1: A Schwarzschild-Kruskal space-time in a conformal Gauss gauge. This is not a schematic picture but quantitatively correct. In Schwarzschild coordinates t and r the lower horizontal line, the initial hypersurface $\mathcal{S} \sim S^3$, corresponds for $\pi/2 < \chi < \pi$ to the hypersurface $\{t = 0\}$ of a Schwarzschild space-time. On $\mathcal{S}$ the coordinate χ satisfies $r = \tan(\chi/2)$, takes the value $\pi/2$ at the throat and the value π at one of the asymptotically flat ends. The parameter τ on the conformal geodesics vanishes on $\mathcal{S}$. With $\Omega = \sin^2\chi/2\,(1 + \sin\chi)$ the physical metric induced on $\tilde{\mathcal{S}}$ is $\tilde{h} = \Omega^{-2}\,d\omega^2$ with $d\omega^2$ the standard line element on S^3. The initial conditions of section **??** are satisfied with $\kappa = \sin\chi$ so that $\Theta = \kappa^{-1}\,\Omega\,\{1 - \tau^2[\cos\chi(2 + \sin\chi)/2\,(1 + \sin\chi)]^2\}$. The rescaled space-time and the conformal Gauss gauge extend smoothly through null infinity, where $\Theta = 0$. The expression for Θ stops being meaningful when the conformal geodesics hit the singularity. The behaviour of the hypersurfaces of constant retarded and advanced time w and v shows that along curves which approach the cylinder $\mathcal{I}$ the null cones collapse. Along curves on $\mathscr{I}^+$ which approach the critical set $\mathcal{I}^+$ this behaviour does not occur. This indicates a degeneracy at $\mathcal{I}^+$ of the set of characteristics.

•8.0.2
•8.0.3
•8.0.4

•*8.0.2:* **ptc:***check Nicholas's spelling; this is from his Dissertationes Mathematica paper*

•*8.0.3:* **ptc:***Figures stolen from Giulini [55]*

Figure 8.9: bipolar coordinates

•*8.0.4:* **ptc:***the encyclopedia file should be used here*

Figure 8.2: The Kruskal-Szekeres extension of the Schwarzschild solution.[1]

Figure 8.3: The Carter-Penrose diagram of the Schwarzschild-Kruskal-Szekeres space-time.[??]

Figure 8.4: One black hole

Figure 8.5: Two black holes well separated

Figure 8.6: Two black holes after merging

Figure 8.7: Multi-Schwarzschild

Figure 8.8: Einstein-Rosen manifold

Figure 8.10: The Misner Wormhole representing two black holes

Appendix A

Some elementary facts in differential geometry

A.1 Vector fields

Let M be an n-dimensional manifold. Physicists often think of vector fields in terms of coordinate systems: a vector field X is an object which in a coordinate system $\{x^i\}$ is represented by a collection of functions X^i. In a new coordinate system $\{y^j\}$ the field X is represented by a new set of functions:

$$X^i(x) \to X^j(y) := X^j(x(y))\frac{\partial y^i}{\partial x^j} \ . \tag{A.1.1}$$

(The summation convention is used throughout, so that the index j has to be summed over.) In modern differential geometry a different approach is taken: one identifies vector fields with homogeneous first order differential operators acting on real valued functions $f : M \to \mathbb{R}$. In local coordinates $\{x^i\}$ a vector field X will be written as $X^i\partial_i$, where the X^i's are the "physicists's functions" just mentioned. This means that the action of X on functions is given by the formula

$$\boxed{X(f) := X^i\partial_i f} \tag{A.1.2}$$

(recall that ∂_i is the partial derivative with respect to the coordinate x^i). Conversely, given some abstract derivative operator X, the (perhaps locally defined) functions X^i in (A.1.2) can be found by acting on the coordinate functions:

$$X(x^i) = X^i \ . \tag{A.1.3}$$

One justification for the differential operator approach is the fact that the tangent $\dot\gamma$ to a curve γ can be calculated — in a way independent of the coordinate system $\{x^i\}$ chosen to represent γ — using the equation

$$\dot\gamma(f) := \frac{d(f\circ\gamma)}{dt} \ .$$

Indeed, if γ is represented as $\gamma(t) = \{x^i = \gamma^i(t)\}$ within a coordinate patch, then we have

$$\frac{d(f\circ\gamma)(t)}{dt} = \frac{d(f(\gamma(t)))}{dt} = \frac{d\gamma^i(t)}{dt}(\partial_i f)(\gamma(t)) \ ,$$

recovering the usual coordinate formula $\dot\gamma = (d\gamma^i/dt)$. An alternative justification is that this approach does encode the transformation law in a natural way: indeed, from (A.1.3) and (A.1.2) we have

$$X(y^i) = X^j \frac{\partial y^i}{\partial x^j} ,$$

reproducing (A.1.1).

Covector fields are fields dual to vector fields. It is convenient to define

$$\boxed{dx^i(X) := X^i} ,$$

where X^i is as in (A.1.2). With this definition the (locally defined) bases $\{\partial_i\}_{i=1,\ldots,\dim M}$ of TM and $\{dx^j\}_{i=1,\ldots,\dim M}$ of T^*M are dual to each other:

$$\langle dx^i, \partial_j\rangle := dx^i(\partial_j) = \delta^i_j ,$$

where δ^i_j is the Kronecker delta, equal to one when $i = j$ and zero otherwise.

Vector fields can be added and multiplied by functions in the obvious way. Another useful operation is the *Lie bracket*, or *commutator*, defined as

$$\boxed{[X,Y](f) := X(Y(f)) - Y(X(f))} . \tag{A.1.4}$$

One needs to check that this does indeed define a new vector field: the simplest way is to use local coordinates,

$$\begin{aligned}
[X,Y](f) &= X^j\partial_j(Y^i\partial_i f) - Y^j\partial_j(X^i\partial_i f)\\
&= X^j(\partial_j(Y^i)\partial_i f + Y^i\partial_j\partial_i f) - Y^j(\partial_j(X^i)\partial_i f + X^i\partial_j\partial_i f)\\
&= (X^j\partial_j Y^i - Y^j\partial_j X^i)\partial_i f + \underbrace{X^jY^i\partial_j\partial_i f - Y^jX^i\partial_j\partial_i f}_{=X^jY^i\underbrace{(\partial_j\partial_i f - \partial_i\partial_j f)}_{0}}\\
&= (X^j\partial_j Y^i - Y^j\partial_j X^i)\partial_i f , \qquad (A.1.5)
\end{aligned}$$

which is indeed a homogeneous first order differential operator. Here we have used the symmetry of the matrix of second derivatives of twice differentiable functions. We note that the last line of (A.1.5) also gives an explicit coordinate expression for the commutator of two differentiable vector fields.

A.2 Raising and lowering of indices

Let g be a symmetric two-covariant tensor field on M, by definition such an object is the assignment to each point $p \in M$ of a bilinear map $g(p)$ from $T_pM \times T_pM$ to $\mathbb{R}$, with the additional property

$$g(X,Y) = g(Y,X) .$$

In this work the symbol g will be reserved to *non-degenerate* symmetric two-covariant tensor fields. It is usual to simply write g for $g(p)$, the point p being implicitly understood. We will sometimes write g_p for $g(p)$ when referencing p will be useful.

The usual Sylvester's inertia theorem tells us that at each p the map g will have a well defined signature; clearly this signature will be point-independent on a connected manifold when g is non-degenerate. A pair (M,g) is said to be a *Riemannian manifold* when the signature of g is $(\dim M, 0)$; equivalently, when g is a positive definite bilinear form on every product $T_pM \times T_pM$. A pair (M,g) is said to be a *Lorentzian manifold* when the signature of g is $(\dim M - 1, 1)$. One talks about *pseudo-Riemannian* manifolds whatever the signature of g, as long as g is non-degenerate, but we will only encounter Riemannian and Lorentzian metrics in this work.

Since g is non-degenerate it induces an isomorphism

$$\flat : T_pM \to T_p^*M$$

by the formula

$$\boxed{X_\flat(Y) = g(X,Y)}\ .$$

In local coordinates this gives

$$X_\flat = g_{ij}X^i dx^j =: X_j dx^j \ . \tag{A.2.1}$$

This last equality defines X_j — "the vector X^j with the index j lowered":

$$\boxed{X_i := g_{ij}X^j}\ . \tag{A.2.2}$$

The operation (A.2.2) is called the *lowering of indices* in the physics literature and, again in the physics literature, one does not make a distinction between the one-form $X_\flat$ and the vector X.

The inverse map will be denoted by $\sharp$ and is called the *raising of indices*; from (A.2.1) we obviously have

$$\alpha^\sharp = g^{ij}\alpha_i\partial_j =: \alpha^i\partial_i \quad \Longleftrightarrow \quad dx^i(\alpha^\sharp) = \boxed{\alpha^i = g^{ij}\alpha_j}\ ,$$

where g^{ij} is the matrix inverse to g_{ij}. For example,

$$(dx^i)^\sharp = g^{ik}\partial_k \ .$$

Clearly g^{ij}, understood as the matrix of a bilinear form on T_p^*M, has the same signature as g, and can be used to define a scalar product $g^\sharp$ on $T_p^*(M)$:

$$g^\sharp(\alpha,\beta) := g(\alpha^\sharp,\beta^\sharp) \quad \Longleftrightarrow \quad g^\sharp(dx^i,dx^j) = g^{ij} \ .$$

This last equality is justified as follows:

$$g^\sharp(dx^i,dx^j) = g((dx^i)^\sharp,(dx^j)^\sharp) = g(g^{ik}\partial_k, g^{j\ell}\partial_\ell) = \underbrace{g^{ik}g_{k\ell}}_{=\delta^i_\ell}g^{j\ell} = g^{ji} = g^{ij} \ .$$

It is convenient to use the same letter g for $g^\sharp$ — physicists do it all the time — or for scalar products induced by g on all the remaining tensor bundles, and we will sometimes do so.

A.3 Lie derivatives (to be done)

A.4 Covariant derivatives

When dealing with $\mathbb{R}^n$, or subsets thereof, there exists an obvious prescription how to differentiate tensor fields: we have then at our disposal the canonical trivialization $\{\partial_i\}_{i=1,\dots,n}$ of $T\mathbb{R}^n$, together with its dual trivialization $\{dx^j\}_{i=1,\dots,n}$ of $T^*\mathbb{R}^n$. We can expand a tensor field T of valence (k,ℓ) in terms of those bases,

$$\begin{aligned} T &= T^{i_1\dots i_k}{}_{j_1\dots j_\ell}\partial_{i_1}\otimes\dots\otimes\partial_{i_k}\otimes dx^{j_1}\otimes\dots\otimes dx^{j_\ell} \\ \iff \quad & T^{i_1\dots i_k}{}_{j_1\dots j_\ell} = T(dx^{i_1},\dots,dx^{i_k},\partial_{j_1},\dots,\partial_{j_\ell})\ , \end{aligned} \tag{A.4.1}$$

and differentiate each component $T^{i_1\dots i_k}{}_{j_1\dots j_\ell}$ of T separately:

$$X(T) := X^i\partial_i(T^{i_1\dots i_k}{}_{j_1\dots j_\ell})\partial_{i_1}\otimes\dots\otimes\partial_{i_k}\otimes dx^{j_1}\otimes\dots\otimes dx^{j_\ell}\ . \tag{A.4.2}$$

The resulting object does, however, *not* behave as a tensor under coordinate transformations: as an example, consider the one-form $T = dx$ on $\mathbb{R}^n$, which has vanishing derivative as defined by (A.4.2). When expressed in spherical coordinates we have

$$T = d(\rho\cos\varphi) = \rho\sin\varphi d\varphi + \cos\varphi d\rho\ ,$$

the partial derivatives of which are non-zero, both with respect to the original cartesian coordinates (x,y) and to the new spherical ones (ρ,φ). The notion of a *covariant derivative*, sometimes also referred to as a *connection*, is introduced precisely to obtain a notion of derivative which has tensorial properties. By definition, a covariant derivative is a map which to a vector field X and a tensor field T assigns a tensor field of the same type as T, denoted by $\nabla_X T$, with the following properties:

1. $\nabla_X T$ is linear with respect to addition both with respect to X and T:

$$\nabla_{X+Y}T = \nabla_X T + \nabla_Y T\ , \qquad \nabla_X(T+Y) = \nabla_X T + \nabla_X S\ , \tag{A.4.3}$$

2. $\nabla_X T$ is linear with respect to multiplication of X by functions f,

$$\nabla_{fX}T = f\nabla_X T\ , \tag{A.4.4}$$

3. and, finally, $\nabla_X T$ satisfies the *Leibniz rule* under multiplication of T by a differentiable function f:

$$\nabla_X(fT) = f\nabla_X T + X(f)T\ . \tag{A.4.5}$$

It is natural to ask whether covariant derivatives do exist at all in general and, if so, how many of them can there be. First, it immediately follows from the axioms above that if D and ∇ are two covariant derivatives, then

$$\Delta(X,T) := D_X T - \nabla_X T$$

is multi-linear both with respect to addition and multiplication by functions — the non-homogeneous terms $X(f)T$ in (A.4.5) cancel out — and is thus a tensor field. Reciprocally, if D is a covariant derivative and $\Delta(X,T)$ is bilinear with respect to addition and multiplication by functions, then

$$\nabla_X T := D_X T + \Delta(X,T) \tag{A.4.6}$$

is a new covariant derivative.

We note that the sum of two covariant derivatives is *not* a covariant derivative. However, *convex* combinations of covariant derivatives, with coefficients which may vary from point to point, are again covariant derivatives. This remark allows one to construct covariant derivatives using partitions of unity: Let, indeed, $\{\mathscr{O}_i\}_{i\in\mathbb{N}}$ be an open covering of M by coordinate patches and let φ_i be an associated partition of unity. In each of those coordinate patches we can decompose a tensor field T as in (A.4.1), and define

$$D_X T := \sum_i \varphi_i X^j \partial_j (T^{i_1\ldots i_k}{}_{j_1\ldots j_\ell})\partial_{i_1}\otimes\ldots\otimes\partial_{i_k}\otimes dx^{j_1}\otimes\ldots\otimes dx^{j_\ell}\ . \tag{A.4.7}$$

This procedure, which depends upon the choice of the coordinate patches and the choice of the partition of unity, defines *one* covariant derivative; all other covariant derivatives are then obtained from D using (A.4.6). Note that (A.4.2) is a special case of (A.4.7) when there exists a global coordinate system on M. Thus (A.4.2) *does* define a covariant derivative. However, the associated operation on tensor fields will *not* take the simple form (A.4.2) when we go to a different coordinate system $\{y^i\}$ in general.

As an illustration, let us describe all possible covariant derivatives on functions: first, it is straightforward to check that the assignment

$$(X,f) \longrightarrow X(f) \tag{A.4.8}$$

is a covariant derivative. It then follows that prescribing a covariant derivative on functions is equivalent to prescribing a field γ of one-forms with

$$\nabla_X f = X(f) + \gamma(X)f\ . \tag{A.4.9}$$

Clearly, any one-form

$$\gamma(X) = \nabla_X 1$$

determines a unique covariant derivative on functions by (A.4.9). We are free to choose γ arbitrarily, and each covariant derivative on functions is uniquely determined by some γ. For functions the generalization obtained by adding a γ piece is not very useful, and throughout this work only the covariant derivative (A.4.8) will be used for functions. The addition of a lower order term in ∇ becomes, however, a necessity when one wishes to construct tensors by differentiation of tensors other than functions.

The simplest next possibility is that of a covariant derivative of vector fields. We will first assume that we are working on a set $\Omega\subset M$ over which we have a *global trivialization* of the tangent bundle TM; by definition, this means that there exist vector fields e_a, $a = 1, \ldots, \dim M$, such that at every point $p \in \Omega$ the fields $e_a(p) \in T_pM$ form a basis of T_pM.[1] Let θ^a denote the dual trivialization of T^*M — by definition the θ^a's satisfy

$$\boxed{\theta^a(e_b) = \delta^a_b} \,.$$

Given a covariant derivative ∇ on vector fields we set

$$\begin{aligned} {\Gamma^a}_b(X) := \theta^a(\nabla_X e_b) \quad &\Longleftrightarrow \quad \nabla_X e_b = {\Gamma^a}_b(X) e_a \,, & (A.4.10a)\\ \boxed{{\Gamma^a}_{bc} := {\Gamma^a}_b(e_c) = \theta^a(\nabla_{e_c} e_b)} \quad &\Longleftrightarrow \quad \nabla_X e_b = {\Gamma^a}_{bc} X^c e_a \,. & (A.4.10b) \end{aligned}$$

The (locally) defined functions ${\Gamma^a}_{bc}$ are called *connection coefficients.* If $\{e_a\}$ is the coordinate basis $\{\partial_\mu\}$ we shall write

$$ {\Gamma^\mu}_{\alpha\beta} := dx^\mu(\nabla_{\partial_\beta}\partial_\alpha) \quad \Big(\Longleftrightarrow \quad \nabla_{\partial_\mu}\partial_\nu = {\Gamma^\sigma}_{\nu\mu}\partial_\sigma\Big) \,, \qquad \text{(A.4.11)}$$

etc. In this particular case the connection coefficients are usually called *Christoffel symbols.* We will sometimes write $\Gamma^\sigma_{\nu\mu}$ instead of ${\Gamma^\sigma}_{\nu\mu}$. By using the Leibniz rule (A.4.5) we find

$$\begin{aligned} \nabla_X Y &= \nabla_X (Y^a e_a) \\ &= X(Y^a) e_a + Y^a \nabla_X e_a \\ &= X(Y^a) e_a + Y^a {\Gamma^b}_a(X) e_b \\ &= (X(Y^a) + {\Gamma^a}_b(X) Y^b) e_a \\ &= (X(Y^a) + {\Gamma^a}_{bc} Y^b X^c) e_a \,, \qquad \text{(A.4.12)} \end{aligned}$$

which gives various equivalent ways of writing $\nabla_X Y$. The (perhaps only locally defined) ${\Gamma^a}_b$'s are linear in X, and the collection $({\Gamma^a}_b)_{a,b=1,\ldots,\dim M}$ is sometimes referred to as the *connection one-form.* The one-covariant, one-contravariant tensor field ∇Y is defined as

$$\nabla Y := \nabla_a Y^b \theta^a \otimes e_b \Longleftrightarrow \nabla_a Y^b := \theta^b(\nabla_{e_a} Y) \Longleftrightarrow \boxed{\nabla_a Y^b = e_a(Y^b) + {\Gamma^b}_{ca} Y^c} \,. \qquad \text{(A.4.13)}$$

We will sometimes write ∇_a for ∇_{e_a}. Further, $\nabla_a Y^b$ will sometimes be written as ${Y^b}_{;a}$. It should be stressed that the notation $\nabla_a Y^b$ does *not* mean the action of a derivative operator ∇_a on a component Y^b of a vector field (as would have been the case if the Y^a's were treated as functions, as in (A.4.9)), but represents the tensor field ∇Y as in (A.4.13).

Suppose that we are given a covariant derivative on vector fields, there is a natural way of inducing a covariant derivative on one-forms by imposing the condition that *the duality operation be compatible with the Leibniz rule*: given two vector fields X and Y together with a field of one-forms α one sets

$$\boxed{(\nabla_X \alpha)(Y) := X(\alpha(Y)) - \alpha(\nabla_X Y)} \,. \qquad \text{(A.4.14)}$$

[1]This is the case when Ω is a coordinate patch with coordinates (x^i), then the $\{e_a\}_{a=1,\ldots,\dim M}$ can be chosen to be equal to $\{\partial_i\}_{a=1,\ldots,\dim M}$. Recall that a manifold is said to be parallelizable if a basis of TM can be chosen globally over M — in such a case Ω can be taken equal to M. We emphasize that we are *not* assuming that M is parallelizable, so that equations such as (A.4.10) have only a local character in general.

Let us, first, check that (A.4.14) defines indeed a field of one-forms. The linearity, in the Y variable, with respect to addition is obvious. Next, for any function f we have

$$\begin{aligned}(\nabla_X\alpha)(fY) &= X(\alpha(fY)) - \alpha(\nabla_X(fY)) \\ &= X(f)\alpha(Y) + fX(\alpha(Y)) - \alpha(X(f)Y + f\nabla_X Y) \\ &= f(\nabla_X\alpha)(Y)\;,\end{aligned}$$

as should be the case for one-forms. Next, we need to check that ∇ defined by (A.4.14) does satisfy the remaining axioms imposed on covariant derivatives. Again multi-linearity with respect to additions is obvious, as well as linearity with respect to multiplication of X by a function. Finally,

$$\begin{aligned}\nabla_X(f\alpha)(Y) &= X(f\alpha(Y)) - f\alpha(\nabla_X Y) \\ &= X(f)\alpha(Y) + f(\nabla_X\alpha)(Y)\;,\end{aligned}$$

as desired.

The duality pairing

$$T_p^*M \times T_pM \ni (\alpha, X) \to \alpha(X) \in \mathbb{R}$$

is sometimes called *contraction*. As already pointed out, the operation ∇ on one forms has been defined in (A.4.14) so as to satisfy the *Leibniz rule under duality pairing*:

$$X(\alpha(Y)) = (\nabla_X\alpha)(Y) + \alpha(\nabla_X Y)\;; \tag{A.4.15}$$

this follows directly from (A.4.14). This should not be confused with the Leibniz rule under multiplication by functions, which is part of the definition of a covariant derivative, and therefore always holds. It should be kept in mind that (A.4.15) does not necessarily hold for all covariant derivatives: if ${}^v\nabla$ is some covariant derivative on vectors, and ${}^f\nabla$ is some covariant derivative on one-forms, in general one will have

$$X(\alpha(Y)) \neq ({}^f\nabla_X)\alpha(Y) + \alpha({}^v\nabla_X Y)\;.$$

Using the basis-expression (A.4.12) of $\nabla_X Y$ and the definition (A.4.14) we have

$$\nabla_X\alpha = X^a\nabla_a\alpha_b\,\theta^b\;,$$

with

$$\begin{aligned}\boxed{\nabla_a\alpha_b} &:= (\nabla_{e_a}\alpha)(e_b) \\ &= e_a(\alpha(e_b)) - \alpha(\nabla_{e_a}e_b) \\ &= \boxed{e_a(\alpha_b) - \Gamma^c{}_{ba}\alpha_c}\;.\end{aligned}$$

It should now be clear how to extend ∇ to tensors of arbitrary valence: if T is r covariant and s contravariant one sets

$$\begin{aligned}(\nabla_X T)(X_1,\ldots,X_r,\alpha_1,\ldots\alpha_s) &:= X\Big(T(X_1,\ldots,X_r,\alpha_1,\ldots\alpha_s)\Big) \\ &-T(\nabla_X X_1,\ldots,X_r,\alpha_1,\ldots\alpha_s) - \ldots - T(X_1,\ldots,\nabla_X X_r,\alpha_1,\ldots\alpha_s) \\ &-T(X_1,\ldots,X_r,\nabla_X\alpha_1,\ldots\alpha_s) - \ldots - T(X_1,\ldots,X_r,\alpha_1,\ldots\nabla_X\alpha_s)\;.\end{aligned} \tag{A.4.16}$$

The verification that this defines a covariant derivative proceeds in a way identical to that for one-forms. In a basis we have

$$\nabla_X T = X^a \nabla_a T_{a_1\ldots a_r}{}^{b_1\ldots b_s} \theta^{a_1} \otimes \ldots \otimes \theta^{a_r} \otimes e_{b_1} \otimes \ldots \otimes e_{b_s} \ ,$$

and (A.4.16) gives

$$\begin{aligned} \nabla_a T_{a_1\ldots a_r}{}^{b_1\ldots b_s} &:= (\nabla_{e_a} T)(e_{a_1}, \ldots, e_{a_r}, \theta^{b_1}, \ldots, \theta^{b_s}) \\ &= e_a(T_{a_1\ldots a_r}{}^{b_1\ldots b_s}) - \Gamma^c{}_{a_1 a} T_{c\ldots a_r}{}^{b_1\ldots b_s} - \ldots - \Gamma^c{}_{a_r a} T_{a_1\ldots c}{}^{b_1\ldots b_s} \\ &\quad + \Gamma^{b_1}{}_{ca} T_{a_1\ldots a_r}{}^{c\ldots b_s} + \ldots + \Gamma^{b_s}{}_{ca} T_{a_1\ldots a_r}{}^{b_1\ldots c} \ . \end{aligned} \tag{A.4.17}$$

Carrying over the last two lines of (A.4.16) to the left-hand-side of that equation one obtains the Leibniz rule for ∇ under pairings of tensors with vectors or forms. It should be clear from (A.4.16) that ∇ defined by that equation is the *only covariant derivative which agrees with the original one on vectors, and which satisfies the Leibniz rule under the pairing operation.* We will only consider such covariant derivatives in this work.

A.4.1 Torsion

Let ∇ be a covariant derivative defined for vector fields, the *torsion tensor* T is defined by the formula

$$\boxed{T(X,Y) := \nabla_X Y - \nabla_Y X - [X,Y]} \ , \tag{A.4.18}$$

where $[X,Y]$ is the Lie bracket defined in (A.1.4). We obviously have

$$T(X,Y) = -T(Y,X) \ . \tag{A.4.19}$$

Let us check that T is actually a tensor field: multi-linearity with respect to addition is obvious. To check what happens under multiplication by functions, in view of (A.4.19) it is sufficient to do the calculation for the first slot of T. We then have

$$\begin{aligned} T(fX,Y) &= \nabla_{fX} Y - \nabla_Y (fX) - [fX,Y] \\ &= f\Big(\nabla_X Y - \nabla_Y X\Big) - Y(f)X - [fX,Y] \ . \end{aligned} \tag{A.4.20}$$

To work out the last commutator term we compute, for any function g,

$$[fX,Y](g) = fX(Y(g)) - \underbrace{Y(fX(g))}_{=Y(f)X(g)+fY(X(g))} = f[X,Y](g) - Y(f)X(g) \ ,$$

hence

$$[fX,Y] = f[X,Y] - Y(f)X \ , \tag{A.4.21}$$

and the last term here cancels the undesirable before-last term in (A.4.20), as required.

In a coordinate basis ∂_μ we have $[\partial_\mu, \partial_\nu] = 0$ and one finds from (A.4.11)

$$\boxed{T_{\mu\nu} := T(\partial_\mu, \partial_\nu) = (\Gamma^\sigma{}_{\nu\mu} - \Gamma^\sigma{}_{\mu\nu})\partial_\sigma} \ , \tag{A.4.22}$$

which shows that — in coordinate frames — T is determined by twice the antisymmetrization of the $\Gamma^\sigma{}_{\mu\nu}$'s over the lower indices. In particular that last antisymmetrization produces a tensor field.

A.5 Curvature

Let ∇ be a covariant derivative defined for vector fields, the curvature tensor is defined by the formula

$$\boxed{R(X,Y)Z := \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z}\;, \tag{A.5.1}$$

where, as elsewhere, $[X,Y]$ is the Lie bracket defined in (A.1.4). We note the anti-symmetry

$$R(X,Y)Z = -R(Y,X)Z\;. \tag{A.5.2}$$

Multi-linearity with respect to addition is obvious, as before; multiplications by functions require more work. First, we have (see (A.4.21))

$$\begin{aligned}
R(fX,Y)Z &= \nabla_{fX}\nabla_Y Z - \nabla_Y\nabla_{fX}Z - \nabla_{[fX,Y]}Z\\
&= f\nabla_X\nabla_Y Z - \nabla_Y(f\nabla_X Z) - \underbrace{\nabla_{f[X,Y]-Y(f)X}Z}_{=f\nabla_{[X,Y]}Z-Y(f)\nabla_X Z}\\
&= fR(X,Y)Z\;.
\end{aligned}$$

Next,

$$\begin{aligned}
R(X,Y)(fZ) &= \nabla_X\nabla_Y(fZ) - \nabla_Y\nabla_X(fZ) - \nabla_{[X,Y]}(fZ)\\
&= \Big\{\nabla_X\Big(Y(f)Z + f\nabla_Y Z\Big)\Big\} - \Big\{\cdots\Big\}_{X\leftrightarrow Y}\\
&\quad -[X,Y](f)Z - f\nabla_{[X,Y]}Z\\
&= \Big\{\underbrace{X(Y(f))Z}_{a} + \underbrace{Y(f)\nabla_X Z + X(f)\nabla_Y Z}_{b} + f\nabla_X\nabla_Y Z\Big\} - \Big\{\cdots\Big\}_{X\leftrightarrow Y}\\
&\quad - \underbrace{[X,Y](f)Z}_{c} - f\nabla_{[X,Y]}Z\;.
\end{aligned}$$

Now, a together with its counterpart with X and Y interchanged cancel out with c, while b is symmetric with respect to X and Y and therefore cancels out with its counterpart with X and Y interchanged, leading to the desired equality

$$R(X,Y)(fZ) = fR(X,Y)Z\;.$$

In a coordinate basis $\{e_a\} = \{\partial_\mu\}$ we find[2] (recall that $[\partial_\mu,\partial_\nu]=0$)

$$\begin{aligned}
R^\alpha{}_{\beta\gamma\delta} &:= \langle dx^\alpha, R(\partial_\gamma,\partial_\delta)\partial_\beta\rangle\\
&= \langle dx^\alpha, \nabla_\gamma\nabla_\delta\partial_\beta\rangle - \langle\cdots\rangle_{\delta\leftrightarrow\gamma}\\
&= \langle dx^\alpha, \nabla_\gamma(\Gamma^\sigma{}_{\beta\delta}\partial_\sigma)\rangle - \langle\cdots\rangle_{\delta\leftrightarrow\gamma}\\
&= \langle dx^\alpha, \partial_\gamma(\Gamma^\sigma{}_{\beta\delta})\partial_\sigma + \Gamma^\rho{}_{\sigma\gamma}\Gamma^\sigma{}_{\beta\delta}\partial_\rho\rangle - \langle\cdots\rangle_{\delta\leftrightarrow\gamma}\\
&= \{\partial_\gamma\Gamma^\alpha{}_{\beta\delta} + \Gamma^\alpha{}_{\sigma\gamma}\Gamma^\sigma{}_{\beta\delta}\} - \{\cdots\}_{\delta\leftrightarrow\gamma}\;,
\end{aligned}$$

[2] The reader is warned that certain authors use a different sign convention either for $R(X,Y)Z$, or for $R^\alpha{}_{\beta\gamma\delta}$, or both. A useful table that lists the sign conventions for a series of standard GR references can be found on the backside of the front cover of [83].

leading finally to

$$\boxed{\boxed{R^\alpha{}_{\beta\gamma\delta} = \partial_\gamma \Gamma^\alpha{}_{\beta\delta} - \partial_\delta \Gamma^\alpha{}_{\beta\gamma} + \Gamma^\alpha{}_{\sigma\gamma}\Gamma^\sigma{}_{\beta\delta} - \Gamma^\alpha{}_{\sigma\delta}\Gamma^\sigma{}_{\beta\gamma}}} \ . \tag{A.5.3}$$

In a general frame some supplementary commutator terms will appear in the formula for $R^a{}_{bcd}$.

Equation (A.5.1) is most frequently used "upside-down", not as a definition of the Riemann tensor, but as a tool for calculating what happens when one changes the order of covariant derivatives. Recall that for partial derivatives we have

$$\partial_\mu \partial_\nu Z^\sigma = \partial_\nu \partial_\mu Z^\sigma \ ,$$

but this is not true in general if partial derivatives are replaced by covariant ones:

$$\nabla_\mu \nabla_\nu Z^\sigma \neq \nabla_\nu \nabla_\mu Z^\sigma \ .$$

To find the correct formula let us consider the tensor field S defined as

$$Y \longrightarrow S(Y) := \nabla_Y Z \ .$$

In local coordinates, S takes the form

$$S = \nabla_\mu Z^\nu \, dx^\mu \otimes \partial_\nu \ .$$

It follows from the Leibniz rule — or, equivalently, from the definitions in Section A.4 — that we have

$$\begin{aligned}
(\nabla_X S)(Y) &= \nabla_X (S(Y)) - S(\nabla_X Y) \\
&= \nabla_X \nabla_Y Z - \nabla_{\nabla_X Y} Z \ .
\end{aligned}$$

The commutator of the derivatives can then be calculated as

$$\begin{aligned}
(\nabla_X S)(Y) - (\nabla_Y S)(X) &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{\nabla_X Y} Z + \nabla_{\nabla_Y X} Z \\
&= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\
&\quad + \nabla_{[X,Y]} Z - \nabla_{\nabla_X Y} Z + \nabla_{\nabla_Y X} Z \\
&= R(X,Y)Z - \nabla_{T(X,Y)} Z \ .
\end{aligned} \tag{A.5.4}$$

Writing ∇S in the usual form

$$\nabla S = \nabla_\sigma S_\mu{}^\nu \, dx^\sigma \otimes dx^\mu \otimes \partial_\nu = \nabla_\sigma \nabla_\mu Z^\nu \, dx^\sigma \otimes dx^\mu \otimes \partial_\nu \ ,$$

we are thus led to

$$\nabla_\mu \nabla_\nu Z^\alpha - \nabla_\nu \nabla_\mu Z^\alpha = R^\alpha{}_{\sigma\mu\nu} Z^\sigma - T^\sigma{}_{\mu\nu} \nabla_\sigma Z^\alpha \ . \tag{A.5.5}$$

In the important case of vanishing torsion, the coordinate-component equivalent of (A.5.1) is thus

$$\boxed{\nabla_\mu \nabla_\nu X^\alpha - \nabla_\nu \nabla_\mu X^\alpha = R^\alpha{}_{\sigma\mu\nu} X^\sigma} \ . \tag{A.5.6}$$

An identical calculation gives, still for torsionless connections,

$$\nabla_\mu \nabla_\nu a_\alpha - \nabla_\nu \nabla_\mu a_\alpha = -R^\sigma{}_{\alpha\mu\nu} a_\sigma \; . \tag{A.5.7}$$

For a general tensor t and torsion-free connection each tensor index comes with a corresponding Riemann tensor term:

$$\begin{aligned}
&\nabla_\mu \nabla_\nu t_{\alpha_1\ldots\alpha_r}{}^{\beta_1\ldots\beta_s} - \nabla_\nu \nabla_\mu t_{\alpha_1\ldots\alpha_r}{}^{\beta_1\ldots\beta_s} = \\
&\quad -R^\sigma{}_{\alpha_1\mu\nu} t_{\sigma\ldots\alpha_r}{}^{\beta_1\ldots\beta_s} - \ldots - R^\sigma{}_{\alpha_r\mu\nu} t_{\alpha_1\ldots\sigma}{}^{\beta_1\ldots\beta_s} \\
&\quad +R^{\beta_1}{}_{\sigma\mu\nu} t_{\alpha_1\ldots\alpha_r}{}^{\sigma\ldots\beta_s} + \ldots + R^{\beta_s}{}_{\sigma\mu\nu} t_{\alpha_1\ldots\alpha_r}{}^{\beta_1\ldots\sigma} \; .
\end{aligned} \tag{A.5.8}$$

A.5.1 Bianchi identities

We have already seen the anti-symmetry property of the Riemann tensor, which in the index notation corresponds to the equation

$$R^\alpha{}_{\beta\gamma\delta} = -R^\alpha{}_{\beta\delta\gamma} . \tag{A.5.9}$$

There are a few other identities satisfied by the Riemann tensor, we start with the *first Bianchi identity*. Let $A(X, Y, Z)$ be any expression depending upon three vector fields X, Y, Z which is antisymmetric in X and Y, we set

$$\sum_{[XYZ]} A(X, Y, Z) := A(X, Y, Z) + A(Y, Z, X) + A(Z, X, Y) , \tag{A.5.10}$$

thus $\sum_{[XYZ]}$ is a sum over cyclic permutations of the vectors X, Y, Z. Clearly,

$$\sum_{[XYZ]} A(X, Y, Z) = \sum_{[XYZ]} A(Y, Z, X) = \sum_{[XYZ]} A(Z, X, Y) . \tag{A.5.11}$$

Suppose, first, that X, Y and Z commute. Using (A.5.11) together with the definition (**??**) of the torsion tensor T we calculate

$$\begin{aligned}
\sum_{[XYZ]} R(X,Y)Z &= \sum_{[XYZ]} \Big(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z\Big) \\
&= \sum_{[XYZ]} \Big(\nabla_X\nabla_Y Z - \nabla_Y \underbrace{(\nabla_Z X + T(X,Z))}_{\text{we have used } [X,Z]=0, \text{ see } (\mathbf{??})}\Big) \\
&= \underbrace{\sum_{[XYZ]} \nabla_X\nabla_Y Z - \sum_{[XYZ]} \nabla_Y\nabla_Z X}_{=0 \text{ (see (A.5.11))}} - \sum_{[XYZ]} \nabla_Y(\underbrace{T(X,Z)}_{=-T(Z,X)}) \\
&= \sum_{[XYZ]} \nabla_X(T(Y,Z)) ,
\end{aligned}$$

and in the last step we have again used (A.5.11). This can be somewhat rearranged by using the definition of the covariant derivative of a higher order tensor (compare (A.4.16)) — equivalently, using the Leibniz rule rewritten upside-down:

$$(\nabla_X T)(Y, Z) = \nabla_X(T(Y, Z)) - T(\nabla_X Y, Z) - T(Y, \nabla_X Z) .$$

•*A.5.1:* **ptc:***there is a sign wrong in the whole calculation*

This leads to•A.5.1

$$\begin{aligned}
\sum_{[XYZ]} \nabla_X(T(Y,Z)) &= \sum_{[XYZ]} \Big((\nabla_X T)(Y,Z) - T(\nabla_X Y, Z) - T(Y, \underbrace{\nabla_X Z}_{=T(X,Z)+\nabla_Z X})\Big) \\
&= \sum_{[XYZ]} \Big((\nabla_X T)(Y,Z) + T(\underbrace{T(X,Z)}_{=-T(Z,X)}, Y)\Big) \\
&\quad \underbrace{- \sum_{[XYZ]} T(\nabla_X Y, Z) - \sum_{[XYZ]} \underbrace{T(Y, \nabla_Z X)}_{=-T(\nabla_Z X, Y)}}_{=0 \text{ (see (A.5.11))}}
\end{aligned}$$

$$= \sum_{[XYZ]} \Big((\nabla_X T)(Y,Z) - T(T(X,Y),Z) \Big) .$$

Summarizing, we have obtained the first Bianchi identity:

$$\sum_{[XYZ]} R(X,Y)Z = \sum_{[XYZ]} \Big((\nabla_X T)(Y,Z) - T(T(X,Y),Z) \Big) , \qquad (A.5.12)$$

under the hypothesis that X, Y and Z commute. However, both sides of this equation are tensorial with respect to X, Y and Z, so that they remain correct without the commutation hypothesis.

We are mostly interested in connections with vanishing torsion, in which case (A.5.12) can be rewritten as

$$\boxed{R^\alpha{}_{\beta\gamma\delta} + R^\alpha{}_{\gamma\delta\beta} + R^\alpha{}_{\beta\gamma\delta} = 0} . \qquad (A.5.13)$$

Our next goal is the *second Bianchi identity*. We consider four vector fields X, Y, Z and W and we assume again that everybody commutes with everybody else. We calculate

$$\begin{aligned}
\sum_{[XYZ]} \nabla_X (R(Y,Z)W) &= \sum_{[XYZ]} \Big(\underbrace{\nabla_X \nabla_Y \nabla_Z W}_{=R(X,Y)\nabla_Z W + \nabla_Y \nabla_X \nabla_Z W} - \nabla_X \nabla_Z \nabla_Y W \Big) \\
&= \sum_{[XYZ]} R(X,Y)\nabla_Z W \\
&\quad + \underbrace{\sum_{[XYZ]} \nabla_Y \nabla_X \nabla_Z W - \sum_{[XYZ]} \nabla_X \nabla_Z \nabla_Y W}_{=0} \\
&\quad .
\end{aligned} \qquad (A.5.14)$$

Next,

$$\begin{aligned}
\sum_{[XYZ]} (\nabla_X R)(Y,Z)W &= \sum_{[XYZ]} \Big(\nabla_X (R(Y,Z)W) - R(\nabla_X Y, Z)W \\
&\quad - R(Y, \underbrace{\nabla_X Z}_{=\nabla_Z X + T(X,Z)})W - R(Y,Z)\nabla_X W \Big) \\
&= \sum_{[XYZ]} \nabla_X (R(Y,Z)W) \\
&\quad \underbrace{- \sum_{[XYZ]} R(\nabla_X Y, Z)W - \sum_{[XYZ]} \underbrace{R(Y, \nabla_Z X)W}_{=-R(\nabla_Z X, Y)W}}_{=0} \\
&\quad - \sum_{[XYZ]} \Big(R(Y, T(X,Z))W + R(Y,Z)\nabla_X W \Big) \\
&= \sum_{[XYZ]} \Big(\nabla_X (R(Y,Z)W) - R(T(X,Y),Z)W - R(Y,Z)\nabla_X W \Big) .
\end{aligned}$$

It follows now from (A.5.14) that the first term cancels out the third one, leading to

$$\sum_{[XYZ]} (\nabla_X R)(Y,Z)W = -\sum_{[XYZ]} R(T(X,Y),Z)W \;, \tag{A.5.15}$$

which is the desired second Bianchi identity for commuting vector fields. As before, because both sides are multi-linear with respect to addition and multiplication by functions, the result remains valid for arbitrary vector fields.

For torsionless connections the components equivalent of (A.5.15) reads

$$\boxed{R^\alpha{}_{\mu\beta\gamma;\delta} + R^\alpha{}_{\mu\gamma\delta;\beta} + R^\alpha{}_{\mu\beta\gamma;\delta} = 0} \;. \tag{A.5.16}$$

A.5.2 The Levi-Civita connection

One of the fundamental results in pseudo-Riemannian geometry is that of existence of a torsion-free connection which preserves the metric:

THEOREM A.5.1 *Let g be a two-covariant symmetric non-degenerate tensor field on a manifold M. Then there exists a unique connection ∇ such that*

1. *$\nabla g = 0$,*
2. *the torsion tensor T of ∇ vanishes.*

PROOF: Let us start with uniqueness. Suppose, thus, that a connection satisfying the above is given, by the Leibniz rule we then have for any vector fields X, Y and Z,

$$0 = (\nabla_X g)(Y,Z) = X(g(Y,Z)) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z) \;. \tag{A.5.17}$$

One then rewrites the same equation applying cyclic permutations to X, Y, and Z, with a minus sign for the last equation:

$$\begin{aligned} +g(\nabla_X Y, Z) + g(Y, \nabla_X Z) &= X(g(Y,Z)) \;, \\ +g(\nabla_Y Z, X) + g(Z, \nabla_Y X) &= Y(g(Z,X)) \;, \\ -g(\nabla_Z X, Y) - g(X, \nabla_Z Y) &= -Z(g(X,Y)) \;. \end{aligned} \tag{A.5.18}$$

As the torsion tensor vanishes, the sum of the left-hand-sides of these equations can be manipulated as follows:

$$\begin{aligned} & g(\nabla_X Y, Z) + g(Y, \nabla_X Z) + g(\nabla_Y Z, X) + g(Z, \nabla_Y X) - g(\nabla_Z X, Y) - g(X, \nabla_Z Y) \\ &= g(\nabla_X Y + \nabla_Y X, Z) + g(Y, \nabla_X Z - \nabla_Z X) + g(X, \nabla_Y Z - \nabla_Z Y) \\ &= g(2\nabla_X Y - [X,Y], Z) + g(Y, [X,Z]) + g(X, [Y,Z]) \\ &= 2g(\nabla_X Y, Z) - g([X,Y], Z) + g(Y, [X,Z]) + g(X, [Y,Z]) \;. \end{aligned}$$

This shows that the sum of the three equations (A.5.18) can be rewritten as

$$\begin{aligned} 2g(\nabla_X Y, Z) &= g([X,Y], Z) - g(Y, [X,Z]) - g(X, [Y,Z]) \\ &\quad + X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) \;. \end{aligned} \tag{A.5.19}$$

Since Z is arbitrary and g is non-degenerate, the left-hand-side of this equation determines the vector field $\nabla_X Y$ uniquely, and uniqueness of ∇ follows.

To prove existence, let $S(X,Y)(Z)$ be defined as one half of the right-hand-side of (A.5.19),

$$\begin{aligned} S(X,Y)(Z) &= \frac{1}{2}\Big(X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) \\ &\quad + g(Z,[X,Y]) - g(Y,[X,Z]) - g(X,[Y,Z])\Big) . \end{aligned} \tag{A.5.20}$$

Clearly S is linear with respect to addition in all fields involved. It is straightforward to check that it is linear with respect to multiplication of Z by a function, and since g is non-degenerate there exists a unique vector field $W(X,Y)$ such that

$$S(X,Y)(Z) = g(W(X,Y),Z) .$$

One readily checks that the assignment

$$(X,Y) \to W(X,Y)$$

satisfies all the requirements imposed on a covariant derivative $\nabla_X Y$. □

Consider (A.5.19) with $X = \partial_\gamma$, $Y = \partial_\beta$ and $Z = \partial_\sigma$,

$$\begin{aligned} 2g(\nabla_\gamma \partial_\beta, \partial_\sigma) &= 2g(\Gamma^\rho{}_{\beta\gamma}\partial_\rho, \partial_\sigma) \\ &= 2g_{\rho\sigma}\Gamma^\rho{}_{\beta\gamma} \\ &= \partial_\gamma g_{\beta\sigma} + \partial_\beta g_{\gamma\sigma} - \partial_\sigma g_{\beta\gamma} \end{aligned}$$

Multiplying this equation by $g^{\alpha\sigma}/2$ we then obtain

$$\boxed{\Gamma^\alpha{}_{\beta\gamma} = \tfrac{1}{2} g^{\alpha\sigma}\{\partial_\beta g_{\sigma\gamma} + \partial_\gamma g_{\sigma\beta} - \partial_\sigma g_{\beta\gamma}\}} . \tag{A.5.21}$$

There is one more identity satisfied by the curvature tensor which is specific to the curvature tensor associated with the Levi-Civita connection, namely

$$g(X, R(Y,Z)W) = g(Y, R(X,W)Z) . \tag{A.5.22}$$

If one sets

$$\boxed{R_{abcd} := g_{ae}R^e{}_{bcd}} , \tag{A.5.23}$$

then (A.5.22) is equivalent to

$$\boxed{R_{abcd} = R_{cdab}} . \tag{A.5.24}$$

In order to prove (A.5.22) it is convenient to first establish some preliminary results, which are of interest on their own:

Proposition A.5.2 *1. Let g be a continuous Lorentzian metric, for every $p \in M$ there exists a neighborhood thereof with a coordinate system such that $g_{\mu\nu} = \eta_{\mu\nu} = \mathrm{diag}(-1,+1,\cdots,+1)$ at p.*

2. If g is differentiable, then the coordinates can be further chosen so that

$$\partial_\sigma g_{\alpha\beta} = 0 \tag{A.5.25}$$

at p, while preserving the degree of differentiability of g.

REMARK A.5.3 The properties spelled-out above do of course hold in the normal coordinates discussed in Section 4.2. However, the introduction of normal coordinates *does lead* to a loss of differentiability of the metric.

PROOF: 1. Let y^μ be any coordinate system around p, and let $e_a = e_a{}^\mu \partial/\partial y^\mu$ be any frame at p such that $g(e_a, e_b) = \eta_{ab}$ — such frames can be found by, *e.g.*, a Gram-Schmidt orthogonalisation. Calculating the determinant of both sides of the equation

$$g_{\mu\nu} e_a{}^\mu e_b{}^\nu = \eta_{ab}$$

we obtain

$$\det(g_{\mu\nu}) \det(e_a{}^\mu)^2 = -1 \; ,$$

which shows that $\det(e_a{}^\mu)$ is non-vanishing. It follows that the formula

$$y^\mu = e^\mu{}_a x^a$$

defines a (linear) diffeomorphism. In the new coordinates we have at p

$$g\Big(\frac{\partial}{\partial x^a}, \frac{\partial}{\partial x^b}\Big) = e^\mu{}_a e^\nu{}_b g\Big(\frac{\partial}{\partial y^\mu}, \frac{\partial}{\partial y^\nu}\Big) = \eta_{ab} \; . \tag{A.5.26}$$

2. Let x^μ be the coordinates described in point 1., shifting by a constant if necessary one can without loss of generality assume that p lies at the origin of those coordinates. The new coordinates z^α will be implicitly defined by the equations

$$x^\mu = z^\mu + \frac{1}{2} A^\mu{}_{\alpha\beta} z^\alpha z^\beta \; ,$$

where $A^\mu{}_{\alpha\beta}$ is a set of constants, symmetric with respect to the interchange of α and β . Set

$$g'_{\alpha\beta} := g\Big(\frac{\partial}{\partial z^\alpha}, \frac{\partial}{\partial z^\beta}\Big) \; , \qquad g_{\alpha\beta} := g\Big(\frac{\partial}{\partial x^\alpha}, \frac{\partial}{\partial x^\beta}\Big) \; .$$

Recall the transformation law

$$g'_{\mu\nu}(z^\sigma) = g_{\alpha\beta}(x^\rho(z^\sigma)) \frac{\partial x^\alpha}{\partial z^\mu} \frac{\partial x^\beta}{\partial z^\nu} \; .$$

By differentiation one obtains at $x^\mu = z^\mu = 0$,

$$\begin{aligned} \frac{\partial g'_{\mu\nu}}{\partial z^\rho}(0) &= \frac{\partial g_{\mu\nu}}{\partial x^\rho}(0) + g_{\alpha\beta}(0)\Big(A^\alpha{}_{\mu\rho}\delta^\beta_\nu + \delta^\alpha_\mu A^\beta{}_{\nu\rho}\Big) \\ &= \frac{\partial g_{\mu\nu}}{\partial x^\rho}(0) + A_{\nu\mu\rho} + A_{\mu\nu\rho} \; , \end{aligned} \tag{A.5.27}$$

where

$$A_{\alpha\beta\gamma} = g_{\alpha\sigma}(0) A^\sigma{}_{\beta\gamma} \; .$$

It remains to show that we can choose $A^\sigma{}_{\beta\gamma}$ so that the left-hand-side can be made to vanish at p. An explicit formula for $A_{\sigma\beta\gamma}$ can be obtained from

(A.5.27) by a cyclic permutation calculation similar to that in (A.5.18). After raising the first index, the final result is

$$A^\alpha{}_{\beta\gamma} = \frac{1}{2} g^{\alpha\rho} \left\{ \frac{\partial g_{\beta\gamma}}{\partial x^\rho} - \frac{\partial g_{\beta\rho}}{\partial x^\gamma} - \frac{\partial g_{\rho\gamma}}{\partial x^\beta} \right\} (0) \; ;$$

the reader may wish to check directly that this does indeed lead to a vanishing right-hand-side of (A.5.27). □

We are ready now to pass to the proof of (A.5.24). We suppose that the metric is twice-differentiable, by point 2. of Proposition A.5.2 in a neighborhood of any point $p \in M$ there exists a coordinate system in which the connection coefficients $\Gamma^\alpha{}_{\beta\gamma}$ vanish at p. Equation (A.5.3) evaluated at p therefore reads

$$\begin{aligned} R^\alpha{}_{\beta\gamma\delta} &= \partial_\gamma \Gamma^\alpha{}_{\beta\delta} - \partial_\delta \Gamma^\alpha{}_{\beta\gamma} \\ &= \frac{1}{2} \Big\{ g^{\alpha\sigma} \partial_\gamma (\partial_\delta g_{\sigma\beta} + \partial_\beta g_{\sigma\delta} - \partial_\sigma g_{\beta\delta}) \\ &\quad - g^{\alpha\sigma} \partial_\delta (\partial_\gamma g_{\sigma\beta} + \partial_\beta g_{\sigma\gamma} - \partial_\sigma g_{\beta\gamma}) \Big\} \\ &= \frac{1}{2} g^{\alpha\sigma} \left\{ \partial_\gamma \partial_\beta g_{\sigma\delta} - \partial_\gamma \partial_\sigma g_{\beta\delta} - \partial_\delta \partial_\beta g_{\sigma\gamma} + \partial_\delta \partial_\sigma g_{\beta\gamma} \right\} . \end{aligned}$$

Equivalently,

$$R_{\sigma\beta\gamma\delta}(0) = \frac{1}{2} \Big\{ \partial_\gamma \partial_\beta g_{\sigma\delta} - \partial_\gamma \partial_\sigma g_{\beta\delta} - \partial_\delta \partial_\beta g_{\sigma\gamma} + \partial_\delta \partial_\sigma g_{\beta\gamma} \Big\} (0) \quad \text{(A.5.28)}$$

This last expression is obviously symmetric under the exchange of $\sigma\beta$ with $\gamma\delta$, and (A.5.24) follows.

A.5.3 Further algebraic identities?

•A.5.2 A natural question is whether there are any further identities satisfied by the curvature tensor, or by its derivatives, in addition to those that we have established so far. We start with the following[3]: •*A.5.2:* **ptc:***new material, to be paper reread*

Proposition A.5.4 *Let $p \in M$ and let $\rho_{\mu\nu\rho\sigma}$ be a tensor at p with all the symmetries of the Riemann tensor, then there exists a pseudo-Riemannian metric g for which $\rho_{\mu\nu\rho\sigma}$ equals the curvature tensor of g at p.*

Proof: Let g be given by the formula

$$g_{\mu\nu} = \eta_{\mu\nu} + A_{\mu\nu\rho\sigma} x^\rho x^\sigma \; . \quad \text{(A.5.29)}$$

Here $\eta_{\mu\nu}$ is a diagonal matrix with entries plus or minus on the diagonal, as appropriate for the signature at hand, while A is a set of constants. We clearly want A to be symmetric in its first two and in its last two indices. Formula (A.5.28) applies and gives

$$R_{\sigma\beta\gamma\delta}(0) = A_{\sigma\delta\gamma\beta} - A_{\beta\delta\gamma\sigma} - A_{\sigma\gamma\delta\beta} + A_{\beta\gamma\delta\sigma} \; . \quad \text{(A.5.30)}$$

[3] Our treatment of the problems that arise here is based on results that can be found in [104] and [84].

It turns out to be convenient to impose a further condition as follows: define

$$\mathscr{A}_0 := \{A_{\alpha\beta\gamma\delta} \mid A_{[\alpha\beta]\gamma\delta} = A_{\alpha\beta[\gamma\delta]} = A_{\alpha(\beta\gamma\delta)} = 0\} \ . \tag{A.5.31}$$

At this stage the reader might think of the last condition in the definition of $\mathscr{A}_0$ as of an ansatz; its origin will become clear in Section A.5.4 below, *cf.* the comments following Proposition A.5.13. In order to establish Proposition A.5.4 we need to prove that any tensor with the symmetries of the Riemann tensor can be obtained from some $A \in \mathscr{A}_1$. So define

$$\mathscr{R}_0 := \{\rho_{\alpha\beta\gamma\delta} \mid \rho_{(\alpha\beta)\gamma\delta} = \rho_{\alpha\beta(\gamma\delta)} = \rho_{\alpha[\beta\gamma\delta]} = \rho_{\alpha\beta\gamma\delta} - \rho_{\gamma\delta\alpha\beta} = 0\} \ . \tag{A.5.32}$$

Let Φ_0 denote the map defined by the right-hand-side of (A.5.30), we have the following:

Lemma A.5.5 *The map*

$$\mathscr{A}_0 \ni A \to \Phi_0(A)_{\alpha\beta\gamma\delta} := A_{\alpha\delta\beta\gamma} - A_{\alpha\gamma\beta\delta} + A_{\beta\gamma\alpha\delta} - A_{\beta\delta\alpha\gamma} \tag{A.5.33}$$

is a bijection from $\mathscr{A}_0$ to $\mathscr{R}_0$. Its inverse is given by

$$\Psi_0(\rho)_{\alpha\beta\gamma\delta} = -\frac{1}{6}(\rho_{\alpha\gamma\beta\delta} + \rho_{\alpha\delta\beta\gamma}) \ . \tag{A.5.34}$$

Remark A.5.6 For further purposes it is useful to derive a somewhat simpler form of (A.5.33). We start by noting that the identity $A_{\alpha(\beta\gamma\delta)} = 0$ is, in view of the symmetry of A in its last two indices, equivalent to

$$A_{\alpha\beta\gamma\delta} + A_{\alpha\gamma\delta\beta} + A_{\alpha\delta\beta\gamma} = 0 \ , \tag{A.5.35}$$

which can also be written as

$$A_{\alpha\beta\gamma\delta} = -2A_{\alpha(\gamma\delta)\beta} \ . \tag{A.5.36}$$

It follows that the first two terms in (A.5.33) can be rewritten as

$$\underbrace{A_{\alpha\delta\beta\gamma}}_{=-A_{\alpha\gamma\delta\beta}-A_{\alpha\beta\gamma\delta}} - A_{\alpha\gamma\beta\delta} = -2A_{\alpha\gamma\delta\beta} - A_{\alpha\beta\gamma\delta} \ .$$

An identical calculation for the last two terms in (A.5.33) gives

$$A_{\beta\gamma\alpha\delta} - A_{\beta\delta\alpha\gamma} = -2A_{\beta\delta\alpha\gamma} - A_{\beta\alpha\delta\gamma} \ .$$

Adding one obtains an alternative formula, where only additions are performed:

$$\Phi_0(A)_{\alpha\beta\gamma\delta} = -2(A_{\beta\delta\alpha\gamma} + A_{\beta\alpha\delta\gamma} + A_{\alpha\gamma\delta\beta}) \ . \tag{A.5.37}$$

Equation (A.5.35) shows that the first two terms add up to minus $A_{\beta\gamma\delta\alpha}$, leading to an equation which only involves two terms

$$\Phi_0(A)_{\alpha\beta\gamma\delta} = 2(A_{\beta\gamma\delta\alpha} - A_{\alpha\gamma\delta\beta}) \ . \tag{A.5.38}$$

PROOF OF LEMMA A.5.5: We will work directly with (A.5.33), we leave it as an exercise to the reader to obtain a simpler proof based on (A.5.38).•A.5.3 The property that Φ_0 maps $\mathscr{A}_0$ into $\mathscr{R}_0$ follows immediately from the facts that a) (A.5.28) has been obtained by calculating the curvature tensor of some metric, and b) that $\mathscr{R}_0$ has been defined precisely by using the symmetries of the Riemann tensor that we have derived so far. (It is in any case a simple exercice to verify directly that $\Phi(\mathscr{A}_0) \subset \mathscr{R}_0$.) To prove bijectivity we need to verify that the map Ψ_0 of (A.5.91) equals Φ_0^{-1}. Indeed:

•*A.5.3:* **ptc:***I am not sure that this remark makes sense*

1. $\Psi_0(\rho)_{\alpha\beta\gamma\delta}$ is clearly symmetric in γ and δ;

2. It is also symmetric in α and β:

$$\Psi_0(\rho)_{\alpha\beta\gamma\delta} = -\frac{1}{6}(\rho_{\alpha\gamma\beta\delta} + \underbrace{\rho_{\alpha\delta\beta\gamma}}_{=\rho_{\beta\gamma\alpha\delta}}) = \Psi_0(\rho)_{\beta\alpha\gamma\delta} \, .$$

3. The symmetrisation constraint holds because ρ is anti-symmetric in its last two indices:

$$\Psi_0(\rho)_{\alpha(\beta\gamma\delta)} = -\frac{1}{6}(\underbrace{\rho_{\alpha(\gamma\beta\delta)}}_{=0} + \underbrace{\rho_{\alpha(\delta\beta\gamma)}}_{=0}) = 0 \, .$$

As for $\Phi_0 \circ \Psi_0$, we have•A.5.4

•*A.5.4:* **ptc:***the proof would be simpler with the new form?*

$$\begin{aligned}
(\Phi_0 \circ \Psi_0)(\rho)_{\alpha\beta\gamma\delta} &= -\frac{1}{6}\Big\{(\rho_{\alpha\beta\delta\gamma} + \rho_{\alpha\gamma\delta\beta} + \underbrace{\rho_{\beta\alpha\gamma\delta}}_{=\rho_{\alpha\beta\delta\gamma}} + \underbrace{\rho_{\beta\delta\gamma\alpha}}_{=\rho_{\alpha\gamma\delta\beta}}) - \gamma \leftrightarrow \delta\Big\} \\
&= -\frac{1}{3}\Big\{(\rho_{\alpha\beta\delta\gamma} + \rho_{\alpha\gamma\delta\beta}) - \gamma \leftrightarrow \delta\Big\} \\
&= -\frac{1}{3}(2\rho_{\alpha\beta\delta\gamma} + \rho_{\alpha\gamma\delta\beta} - \underbrace{\rho_{\alpha\delta\gamma\beta}}_{=-\rho_{\alpha\delta\beta\gamma}}) \, . \qquad \text{(A.5.39)}
\end{aligned}$$

Using the cyclic identity the last two terms add up to $\rho_{\alpha\beta\delta\gamma}$, which shows that Ψ_0 is a right inverse for Φ_0. We note that this suffices for the proof of Proposition A.5.4. However, to finish the proof of the current lemma it remains to show that the composition in the reverse order also produces the identity. This could be done by counting dimensions; instead, we calculate directly. We will need an identity which is derived as follows: applying (A.5.35) twice we have

$$\begin{aligned}
A_{\beta\delta\alpha\gamma} &= -\underbrace{A_{\beta\alpha\gamma\delta}}_{=A_{\alpha\beta\gamma\delta}} - A_{\beta\gamma\delta\alpha} \\
&= A_{\alpha\gamma\delta\beta} + A_{\alpha\delta\beta\gamma} - A_{\beta\gamma\delta\alpha} \qquad \text{(A.5.40)}
\end{aligned}$$

It follows that

$$(\Psi_0 \circ \Phi_0)(A)_{\alpha\gamma\beta\delta} = -\frac{1}{6}\Big\{(A_{\alpha\delta\beta\gamma} - A_{\alpha\gamma\beta\delta} + A_{\beta\gamma\alpha\delta} - \underbrace{A_{\beta\delta\alpha\gamma}}_{\text{replace using (A.5.40)}}) + \beta \leftrightarrow \delta\Big\}$$

$$\begin{aligned} &= -\frac{1}{6}\Big\{(-2A_{\alpha\gamma\beta\delta} + 2A_{\beta\gamma\alpha\delta}) + \beta \leftrightarrow \delta\Big\} \\ &= -\frac{1}{6}(-4A_{\alpha\gamma\beta\delta} + \underbrace{4A_{\gamma(\beta\delta)\alpha}}_{=-2A_{\gamma\alpha\beta\delta}}) = A_{\alpha\gamma\beta\delta} \ . \end{aligned}$$

□

Proposition A.5.4 follows now by taking g to be of the form (A.5.29) with $A = \Psi_0(\rho)$. □

The bijectivity part of Lemma A.5.5 gives the following result:

•*A.5.5:* **ptc:***is this up to isometry or something?*

PROPOSITION A.5.7 *For any $\rho \in \mathscr{R}_0$ there exists a unique*•A.5.5 *metric of the form (A.5.29) with $A \in \mathscr{A}_0$, where $\mathscr{A}_0$ is defined by (A.5.31), such that* $\mathrm{Riem}(0) = \rho$.

We shall say that a tensor X, calculated out of the metric and its derivatives, is *freely prescribable at a point* if there exists a metric g such that for any value of X there exists a metric g such that $X = g(0)$. Proposition A.5.4 can be restated as

PROPOSITION A.5.8 (Thomas [104]) *Any $\rho \in \mathscr{R}_0$ is freely prescribable at a point.*

In Thomas's terminology [104], Proposition A.5.8 is phrased as the property that the identities appearing in the definition of $\mathscr{R}_0$ form a *complete set of identities.*

To get further insight into the structure of the Riemann tensor, let us start by observing that

PROPOSITION A.5.9 *The collection of components*

$$\{R_{\alpha\beta\alpha\beta} \ \textit{(no summation)}\}_{\alpha<\beta} \tag{A.5.41}$$

is freely prescribable at a point.

PROOF: Clearly no non-trivial permutation of α and β can take one element from this collection into another, similarly for symmetry under the exchange of the first pair with the second pair. Finally the cyclic identity for such elements reads

$$R_{\alpha\beta\alpha\beta} + \underbrace{R_{\alpha\alpha\beta\beta}}_{=0} + R_{\alpha\beta\beta\alpha} = 0 \ ,$$

which simply reflects the anti-symmetry of the Riemann tensor in the last two indices, and therefore does not impose any constraint on the collection (A.5.41). □

In dimension one the Ricci scalar vanishes, as well as the whole curvature tensor. In higher dimensions we have:

PROPOSITION A.5.10 *In dimensions $n > 1$ the Ricci scalar is freely prescribable at a point.*

PROOF: We have

$$R = \sum_{\alpha\beta} R^{\alpha\beta}{}_{\alpha\beta} = 2\sum_{\alpha<\beta} R^{\alpha\beta}{}_{\alpha\beta} \;, \tag{A.5.42}$$

thus R is a linear combination of independent objects freely prescribable at a point, and hence also freely prescribable at a point. □

Consider, next the Ricci tensor. In dimension two it is *not* freely prescribable: indeed, in this dimension the only non-trivial component of the Riemann tensor is

$$R_{1212} = \frac{\epsilon}{2} R \tag{A.5.43}$$

(which has actually already been shown to be freely prescribable at a point), with $\epsilon = \pm 1$ depending upon the signature of g; to obtain (A.5.43) we are assuming that the metric g is diagonal with entries ± 1 on the diagonal. In other words, the symmetry properties of elements of $\mathscr{R}_0$ in dimension two imply that every other component of the Riemann tensor is proportional to R_{1212}, or vanishes. Define, then, a tensor field D by the formula

$$D_{ijkl} = R_{ijkl} - \frac{R}{2}(g_{ik}g_{jl} - g_{il}g_{jk}) \;.$$

It has all the symmetries of the Riemann tensor, with $D_{1212} = 0$. It follows that D vanishes, which shows that in dimension two we necessarily have

$$R_{ijkl} = \frac{R}{2}(g_{ik}g_{jl} - g_{il}g_{jk}) \;. \tag{A.5.44}$$

This implies

$$R_{ij} = \frac{R}{2} g_{ij} \;, \tag{A.5.45}$$

which is of rather special form, clearly not freely specifiable.

In order to analyse the Ricci tensor in dimensions $n > 2$ it is convenient to define the following tensor:

$$P_{\alpha\beta} = \frac{1}{n-2}\Big(R_{\alpha\beta} - \frac{R}{2(n-1)} g_{\alpha\beta}\Big) \;. \tag{A.5.46}$$

One then defines the *Weyl tensor* by the formula

$$C_{\alpha\beta\gamma\delta} = R_{\alpha\beta\gamma\delta} - g_{\alpha\gamma}P_{\beta\delta} + g_{\alpha\delta}P_{\beta\gamma} - g_{\beta\delta}P_{\alpha\gamma} + g_{\beta\gamma}P_{\alpha\delta} \;. \tag{A.5.47}$$

The definition of P has been tailored so that

$$C^{\alpha}{}_{\beta\alpha\gamma} = 0 \;, \tag{A.5.48}$$

which can be checked by a straightforward calculation. (For the purpose for this calculation it is useful to work out first the trace of (A.5.46), obtaining

$$\mathrm{tr}_g P := g^{\alpha\beta}P_{\alpha\beta} = \frac{R}{2(n-1)} \;.\Big) \tag{A.5.49}$$

The Weyl tensor has useful covariance properties under conformal transformations, discussed in Section **??**, which are of no concern to us here.

It follows directly from (A.5.47) that the Weyl tensor has all the symmetries of the Riemann one: the anti-symmetry in α and β, and the symmetry under the exchange of $\alpha\beta$ with $\gamma\delta$ are clear. Finally each of the last four terms is symmetric under the permutation of *some pair* of the indices in the set $\beta\gamma\delta$, so an anti-symmetrisation upon those will necessarily give zero, as desired.

Equation (A.5.46) can be inverted using (A.5.49),

$$R_{\alpha\beta} = (n-2)P_{\alpha\beta} + \mathrm{tr}_g P g_{\alpha\beta} \ , \tag{A.5.50}$$

which shows that the correspondence $P \leftrightarrow \mathrm{Ric}$ is a bijection. This implies that P is freely prescribably at a point if and only if the Ricci tensor is.

We are ready now to prove

PROPOSITION A.5.11 *In dimensions $n \geq 3$ the Ricci tensor is freely prescribable at a point. The same holds for the Einstein tensor or for the P tensor (A.5.46).*

PROOF: For any symmetric P consider the tensor

$$\rho_{\alpha\beta\gamma\delta} = g_{\alpha\gamma}P_{\beta\delta} - g_{\alpha\delta}P_{\beta\gamma} + g_{\beta\delta}P_{\alpha\gamma} - g_{\beta\gamma}P_{\alpha\delta} \ .$$

As already pointed out it has all the symmetries of the Riemann tensor, $\rho \in \mathscr{R}_0$, and the result follows from Proposition A.5.8. (We note that this argument does not work in dimension two because of the $(n-2)$ factor in (A.5.46).) □

The argument just given naturally leads to the question whether the Weyl tensor is freely prescribable at a point. Let, then, $\mathscr{C}_0$ be the collection of "Weyl-tensor-candidates":

$$\mathscr{C}_0 := \{\kappa_{\alpha\beta\gamma\delta} \mid \kappa_{(\alpha\beta)\gamma\delta} = \kappa_{\alpha\beta(\gamma\delta)} = \kappa_{\alpha[\beta\gamma\delta]} = \kappa_{\alpha\beta\gamma\delta} - \kappa_{\gamma\delta\alpha\beta} = \kappa^{\alpha}{}_{\beta\alpha\gamma} = 0\} \ . \tag{A.5.51}$$

We have

PROPOSITION A.5.12 *1. In dimension $n = 3$ the Weyl tensor vanishes.*

2. In dimensions $n \geq 4$ both the Weyl tensor and P are freely prescribable at a point.

PROOF: 1. Multiplying g by -1 if necessary, we may without loss of generality assume that there exists an ON frame such that $g_{11} = \epsilon = \pm 1$, $g_{22} = g_{33} = 1$. We then have

$$\begin{aligned} 0 &= C^{\alpha}{}_{1\alpha 1} = C_{2121} + C_{3131} \ , \\ 0 &= C^{\alpha}{}_{2\alpha 2} = \epsilon C_{2121} + C_{3232} \ , \\ 0 &= C^{\alpha}{}_{3\alpha 3} = \epsilon C_{3131} + C_{3232} \ . \end{aligned}$$

This can be written as a matrix equation,

$$\begin{bmatrix} 1 & 0 & 1 \\ \epsilon & 1 & 0 \\ 0 & 1 & \epsilon \end{bmatrix} \begin{bmatrix} C_{1212} \\ C_{2323} \\ C_{1313} \end{bmatrix} = 0 \ .$$

The determinant of the matrix above is 2ϵ, hence all components of the Weyl tensor which involve only two different indices vanish. Next

$$0 = C^{\alpha}{}_{1\alpha 2} = C_{3132} \ ,$$

and one similarly shows that all other components of C that involve three different indices vanish. Since in dimension three there cannot be four different indices, the result follows.

2. Let $\kappa \in \mathscr{C}_0$ and let P by an arbitrary symmetric tensor, repeat the proof of Proposition (A.5.11) with

$$\rho_{\alpha\beta\gamma\delta} = \kappa_{\alpha\beta\gamma\delta} + g_{\alpha\gamma}P_{\beta\delta} - g_{\alpha\delta}P_{\beta\gamma} + g_{\beta\delta}P_{\alpha\gamma} - g_{\beta\gamma}P_{\alpha\delta} \ .$$

□

This gives a rather exhaustive treatment of the properties of the possible forms of the Riemann tensor at a point.

As promised we pass now to the origin of the last restriction in the definition of $\mathscr{A}_0$, (A.5.31). This will be discussed in the next section.•A.5.6

•*A.5.6:* **ptc:***this remark might have to be discarded after rearrangements*

A.5.4 How to recognise that coordinates are normal

•A.5.7

•*A.5.7:* **ptc:***Move this subsection to the normal coordinates one*

The purpose of this section is to give a few simple necessary and sufficient condition for a coordinate system to be normal:

Proposition A.5.13 (Thomas [104]) *Let $\{x^\mu\}$ be a local coordinate system defined on a star shaped domain containing the origin. The following conditions are equivalent:*

1. *For every $a^\mu \in \mathbb{R}^n$ the rays $s \to sa^\mu$ are geodesics;*
2. $\Gamma^{\mu}{}_{\alpha\beta}(x)x^\alpha x^\beta = 0$*;*
3. $\dfrac{\partial g_{\gamma\alpha}}{\partial x^\beta}(x)x^\alpha x^\beta = 0$*;*
4. $g_{\alpha\beta}(x)x^\beta = g_{\alpha\beta}(0)x^\beta$.

•A.5.8 Before we pass to the proof of Proposition A.5.13, let us apply it to the metric (A.5.29): we will have $g_{\mu\nu}x^\nu = \eta_{\mu\nu}x^\nu$ if and only if $A_{\mu\nu\rho\sigma}x^\nu x^\rho x^\sigma = 0$. Differentiating three times this is equivalent to $A_{\mu(\nu\rho\sigma)}$. Thus, point 4 of Proposition A.5.13 shows that the last condition in (A.5.31) is equivalent to the statement that the coordinates x in (A.5.29) are normal. We emphasise that while the formula (A.5.30) does not require the normality condition, bijectivity of the map Φ_0 will be lost if normality is not assumed.

•*A.5.8:* **ptc:***this should be removed in the oberwolfach version*

Proof of Proposition A.5.13: 1. ⇔ 2.: The rays $\gamma^\mu(s) = sa^\mu$ are geodesics if and only if

$$0 = \underbrace{\frac{d^2\gamma^\mu}{ds^2}}_{=0} + \Gamma^{\mu}{}_{\alpha\beta}(sa^\sigma)\frac{d\gamma^\alpha}{ds}\frac{d\gamma^\beta}{ds} = \Gamma^{\mu}{}_{\alpha\beta}(sa^\sigma)a^\alpha a^\beta \ ,$$

multiplying by s^2 and setting $x^\mu = sa^\mu$ the result follows.

3. $\Leftrightarrow$ 4.:

$$\begin{aligned} g_{\mu\alpha}(x^\sigma)x^\alpha = g_{\mu\alpha}(0)x^\alpha \quad &\Longleftrightarrow \quad g_{\mu\alpha}(sa^\sigma)a^\alpha = g_{\mu\alpha}(0)a^\alpha \\ &\Longleftrightarrow \quad \frac{d}{ds}\left(g_{\mu\alpha}(sa^\sigma)a^\alpha\right) = 0 \\ &\Longleftrightarrow \quad \frac{\partial g_{\mu\alpha}(x^\sigma)}{\partial x^\beta}x^\alpha x^\beta = 0\ . \end{aligned}$$

2. $\Rightarrow$ 4.: From the formula for the Christoffel symbols in terms of the metric we have

$$\Gamma^\mu{}_{\alpha\beta}(x)x^\alpha x^\beta = 0 \quad \Longleftrightarrow \quad \left(2\frac{\partial g_{\mu\alpha}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\mu}\right)x^\alpha x^\beta = 0\ . \tag{A.5.52}$$

Multiplying by x^μ we obtain

$$\begin{aligned} \frac{\partial g_{\mu\alpha}(x^\sigma)}{\partial x^\beta}x^\alpha x^\beta x^\mu = 0 \quad &\Longleftrightarrow \quad \frac{\partial g_{\mu\alpha}(sa^\sigma)}{\partial x^\beta}a^\alpha a^\beta a^\mu = 0 \\ &\Longleftrightarrow \quad \frac{d}{ds}\left(g_{\mu\alpha}(sa^\sigma)a^\alpha a^\mu\right) = 0 \\ &\Longleftrightarrow \quad g_{\mu\alpha}(sa^\sigma)a^\alpha a^\mu = g_{\mu\alpha}(0)a^\alpha a^\mu \\ &\Longleftrightarrow \quad g_{\mu\alpha}(x^\sigma)x^\alpha x^\mu = g_{\mu\alpha}(0)x^\alpha x^\mu\ . \end{aligned}$$

Differentiating it follows that

$$\frac{\partial g_{\mu\alpha}(x^\sigma)}{\partial x^\gamma}x^\alpha x^\mu + 2g_{\gamma\alpha}(x^\sigma)x^\alpha = 2g_{\gamma\alpha}(0)x^\alpha\ .$$

Substituting this into the last term in (A.5.52) one obtains

$$\frac{\partial g_{\mu\alpha}}{\partial x^\beta}(x^\sigma)x^\alpha x^\beta + g_{\mu\alpha}(x^\sigma)x^\alpha - g_{\mu\alpha}(0)x^\alpha = 0\ . \tag{A.5.53}$$

This implies that

$$\frac{d}{ds}\left(g_{\mu\alpha}(sa^\mu)sa^\alpha - g_{\mu\alpha}(0)sa^\alpha\right) = 0\ ,$$

and the result follows by integration.

3.&4. $\Rightarrow$ 2.: Point 4. implies

$$g_{\alpha\beta}(x^\gamma)x^\alpha x^\beta = g_{\alpha\beta}(0)x^\alpha x^\beta\ .$$

Differentiating one obtains

$$\frac{\partial g_{\alpha\beta}(x^\gamma)}{\partial x^\mu}x^\alpha x^\beta + 2g_{\alpha\mu}(x^\gamma)x^\alpha = 2g_{\alpha\mu}(0)x^\alpha\ .$$

The last two terms are equal by point 4. so that

$$\frac{\partial g_{\alpha\beta}(x^\gamma)}{\partial x^\mu}x^\alpha x^\beta = 0\ .$$

This shows that the last term in (A.5.52) vanishes, so does the next-to-last by point 3., and the proof is complete. $\square$

A.5.5 Taylor expansions of the metric in normal coordinates

•A.5.9: **ptc:**_this should go as a subsection of the normal coordinates section; there should be somewhere a section on Jacobi fields_

•A.5.10: **ptc:**_do Vince equations help here?_

•A.5.9

•A.5.10

Consider any coordinate system y^i defined around a point $p \in M$. Let x denote the normal coordinates around p, and let $e_i = e^k{}_i \partial_{y^k}$ be an ON-frame at p. On an appropriate subset of the intersection of the domains of definition of both coordinate systems, the exponential map $y^i(x^k)$ is obtained by solving the system of ODE's

$$\frac{d^2 z^i(s,x^j)}{ds^2} = -\Gamma^i{}_{mn}(z^\ell(s,x^j))\frac{dz^m(s,x^j)}{ds}\frac{dz^n(s,x^j)}{ds} \ , \tag{A.5.54}$$

with initial values

$$z^i(0,x^i) = 0 \ , \qquad \frac{dz^m(0,x^j)}{ds} = e^m{}_k x^k \ , \tag{A.5.55}$$

and then setting

$$y^i(x^k) = z^i(1,x^k) \ .$$

It is a standard fact in the theory of ODE's that the derivatives $\partial y^i/\partial x^k$ are obtained by solving the system of equations

$$\begin{aligned} \frac{d^2 \frac{\partial z^i(s,x^j)}{\partial x^k}}{ds^2} &= -\frac{\partial}{\partial x^k}\Big\{\Gamma^i{}_{mn}(z^\ell(s,x^j))\frac{dz^m(s,x^j)}{ds}\frac{dz^n(s,x^j)}{ds}\Big\} \\ &= -\frac{\partial \Gamma^i{}_{mn}}{\partial y^r}(z^\ell(s,x^j))\frac{\partial z^r(s,x)}{\partial x^k}\frac{dz^m(s,x^j)}{ds}\frac{dz^n(s,x^j)}{ds} \\ &\quad -2\Gamma^i{}_{mn}(z^\ell(s,x^j))\frac{d\frac{\partial z^m(s,x^j)}{\partial x^k}}{ds}\frac{dz^n(s,x^j)}{ds} \ , \end{aligned} \tag{A.5.56}$$

with initial values

$$\frac{\partial z^i(0,x^i)}{\partial x^k} = 0 \ , \qquad \frac{d\frac{\partial z^m(0,x^j)}{\partial x^k}}{ds} = e^m{}_k \ , \tag{A.5.57}$$

and then setting

$$\frac{\partial y^i(x^r)}{\partial x^k} = \frac{\partial z^i(1,x^r)}{\partial x^k} \ .$$

Equation (A.5.56) has a natural geometric interpretation which we now derive. Let $\gamma(s)$ be any curve and let $Z(s)$ be any vector field defined along that curve: by definition, this means that we have $Z(s) \in T_{\gamma(s)}M$. (The collection

$$\cup_s T_{\gamma(s)}M$$

is sometimes denoted by $\gamma^* TM$, or by $\gamma! TM$, and is called the pull-back by γ of the bundle TM.)•A.5.11 For such vector fields one sets

•A.5.11: **ptc:**_discard and/or reword those comments if pull back bundles are discussed somewhere_

$$\frac{DZ^i}{ds}(s) := \frac{dZ^i}{ds} + \Gamma^i{}_{jk}(\gamma(s))\dot\gamma^j(s)Z^k(s) \in T_{\gamma(s)}M \ , \tag{A.5.58}$$

giving again a vector field defined along γ. Equation (A.5.58) looks formally like $\nabla_{\dot\gamma} Z$, but recall that the operation $\nabla_X Y$ requires Y to be a vector field defined over M, while Z is only defined over γ; this is the reason why we did not write it in this way. (The operation D/ds is actually the pull-back to γ^*TM of the connection ∇, and is a connection acting on sections of the pulled-back bundle.)

Suppose, now, that we have a one parameter family of geodesics $\gamma(s,\lambda)$, where s is the parameter along the geodesic, and λ is a parameter which distinguishes the geodesics. (In (A.5.54)-(A.5.55) we actually have an n-parameter family of geodesics, which reduces to the case here by taking as λ any one of the coordinates x^i.) Set

$$Z = \gamma_* \partial_\lambda \quad (\text{equivalently,} \quad Z(s,\lambda) = \frac{\partial \gamma^i(s,\lambda)}{\partial\lambda}\partial_i)\;,$$

for each λ this defines a vector field Z along $\gamma(\lambda)$. The vector fields $\dot\gamma$ and Z defined over $\cup_{s,\lambda}\gamma(s,\lambda)$ commute:

$$[Z,\dot\gamma] = \left[\frac{\partial\gamma^k}{\partial\lambda}\partial_k, \frac{\partial\gamma^i}{\partial s}\partial_i\right] = [\gamma_*\partial_\lambda, \gamma_*\partial_s] = \gamma_* \underbrace{[\partial_\lambda,\partial_s]}_{=0} = 0\;.$$

Since ∇ has no torsion this implies

$$\nabla_{\dot\gamma} Z = \nabla_Z \dot\gamma\;,$$

leading to

$$\frac{D^2 Z}{ds^2}(s) = \nabla_{\dot\gamma}\nabla_{\dot\gamma} Z = \nabla_{\dot\gamma}\nabla_Z\dot\gamma = \nabla_{\dot\gamma}\nabla_Z\dot\gamma - \nabla_Z \underbrace{\nabla_{\dot\gamma}\dot\gamma}_{=0} = R(\dot\gamma, Z)\dot\gamma\;.$$

We have obtained an equation known as the *Jacobi equation*, or as the *geodesic deviation equation*:

$$\boxed{\frac{D^2 Z}{ds^2}(s) = R(\dot\gamma, Z)\dot\gamma}\;. \tag{A.5.59}$$

Solutions of (A.5.59) are called *Jacobi fields* along γ.

Now, (A.5.56) coincides with (A.5.59) in a coordinate system in which the Christoffel symbols vanish along γ.•A.5.12 Since (A.5.59) is coordinate independent, it follows that the fields

•*A.5.12:* **ptc:***one should justify existence of such systems*

$$Z_k(s,x^k) := \frac{\partial z^i(s,x^j)}{\partial x^k}\partial_{z^i} \tag{A.5.60}$$

appearing in (A.5.56), are Jacobi field defined along the geodesics $s \to sx^i$.

While the above geometric considerations settle the question, they leave one wondering by what magic a single derivative of the Christoffel symbols appearing in

(A.5.56) has turned into a Riemann tensor. The answer is that the missing derivative of Γ is hidden in the D^2/ds^2 symbol, which may be seen as follows: Iterating (A.5.58) and assuming that γ is a geodesic one finds

$$\begin{aligned}
\frac{D^2 Z^i}{ds^2}(s) &= \frac{D}{ds}\Big(\frac{dZ^i}{ds} + \Gamma^i{}_{jk}(\gamma(s))\dot\gamma^j(s)Z^k(s)\Big) \\
&= \frac{d}{ds}\Big(\frac{dZ^i}{ds} + \Gamma^i{}_{jk}(\gamma(s))\dot\gamma^j(s)Z^k(s)\Big) \\
&\quad +\Gamma^i{}_{mn}(\gamma(s))\dot\gamma^m(s)\Big(\frac{dZ^n}{ds} + \Gamma^n{}_{jk}(\gamma(s))\dot\gamma^j(s)Z^k(s)\Big) \\
&= \frac{d^2 Z^i}{ds^2} + \frac{\partial \Gamma^i{}_{jk}}{\partial y^\ell}(\gamma(s))\dot\gamma^\ell(s)\dot\gamma^j(s)Z^k(s) + \Gamma^i{}_{jk}(\gamma(s))\underbrace{\ddot\gamma^j(s)}_{=-\Gamma^j{}_{uv}(\gamma(s))\dot\gamma^u(s)\dot\gamma^v(s)}Z^k(s) \\
&\quad +\Gamma^i{}_{mn}(\gamma(s))\dot\gamma^m(s)\Big(2\frac{dZ^n}{ds} + \Gamma^n{}_{jk}(\gamma(s))\dot\gamma^j(s)Z^k(s)\Big)\ . \qquad \text{(A.5.61)}
\end{aligned}$$

When Z is one of the fields Z_k of (A.5.60) one can use (A.5.56) to replace d^2Z^i/ds^2 in the last line above, and it then suffices to check that all the $\partial\Gamma$ and $\Gamma\Gamma$ terms combine to a Riemann tensor as in (A.5.59).

We note that for any $X = X^i e_i \in T_pM$ the vector

$$\frac{\partial y^i(x^r)}{\partial x^k}X^k \in T_{y(x)=exp_p(x)}M$$

is, by definition of the push-forward operation, the coordinate-equivalent of

$$\Big(\exp_p(x)\Big)_* X\ .$$

We have thus proved:

Proposition A.5.14 *The fields* $Y_i := \Big(\exp_p(x)\Big)_* e_i$ *are given by*

$$Y_i = Z_i(1)\ , \qquad \text{(A.5.62)}$$

where the $Z_i(s)$*'s are Jacobi fields* $Z^i(s)$ *along the geodesic* $s \to sx^i$ *with initial values* $Z_i(0) = 0$, $\dot Z_i(0) = e_i$.

As a straightforward corollary we obtain:

Proposition A.5.15 *The coordinate components* g_{ij} *of the metric tensor in a normal coordinate systems are given by the formula*

$$g_{ij}(x) = g(Z_i(1), Z_j(1))\ ,$$

where $Z_i(s)$ *is a Jacobi field along the geodesic* $s \to sx^i$ *with initial values* $Z_i(0) = 0$, $\dot Z_i(0) = e_i$.

Proof: By definition,

$$\begin{aligned}
g_{ij}(x) &= \Big(\Big(\exp_p(x)\Big)^* g\Big)(\partial_{x^i}, \partial_{x^j}) \\
&= \Big(\Big(\exp_p(x)\Big)^* g\Big)(e_i, e_j) \\
&= g\Big(\Big(\exp_p(x)\Big)_* e_i, \Big(\exp_p(x)\Big)_* e_i\Big) \\
&= g(Y_i, Y_j)\ ,
\end{aligned}$$

and the result follows from Proposition A.5.14.

In coordinate notation the above calculation runs as follows: let $g_{x^ix^j}$ denote the coordinate components of the metric tensor in the normal coordinate system x^i, and let $g_{y^iy^j}$ denote those in the coordinate system y^i. In normal coordinates we have $e_i(0)=\partial_{x^i}$, and Proposition A.5.14 gives

$$Y_i(x) = Y^k{}_i(x)\partial_{y^k} = \frac{\partial z^k}{\partial x^i}(1,x)\partial_{y^k} = \frac{\partial y^k}{\partial x^i}(x)\partial_{y^k} \ , \tag{A.5.63}$$

so that

$$g_{x^ix^j}(x) \quad = \quad g_{y^ky^\ell}(y(x))\frac{\partial y^k}{\partial x^i}\frac{\partial y^\ell}{\partial x^j} = g_{y^ky^\ell}(y(x))Y^k{}_iY^\ell{}_j = g(Y_i,Y_j) \ ,$$

as desired. By the way, (A.5.63) shows that the matrix formed out of the vectors $(Y_1,\cdots,Y_k)$ is simply the Jacobi matrix of the exponential map in the coordinate system $\{y^i\}$.

□

*•A.5.13: **ptc:**the problem with this derivation is that one obtains an expansion of the Jacobi fields in a wrong basis, not very helpful, is it? while the MSV guys seem to have the right one directly?*

•A.5.13 We continue by deriving the Taylor coefficients of the Jacobi fields Z_k in a frame f_i which is parallel propagated along the radial geodesics $s\to\gamma_{x^k}(s)=\{sx^k\}$. More precisely, we take

$$f_i(0)=e_i \ ,$$

and we propagate the f_i's along each of the γ_{x^k}'s using the equation

$$\nabla_{\dot\gamma_{x^k}}f_i = 0 \ . \tag{A.5.64}$$

Let $\dot\gamma^i_{x^k}$ denote the components of $\dot\gamma_{x^k}$ in the basis f_i, we have

$$0 = \nabla_{\dot\gamma_{x^k}}\dot\gamma_{x^i} = \nabla_{\dot\gamma_{x^k}}(\dot\gamma^i_{x^k}f_i) = \frac{d\dot\gamma^i_{x^k}}{ds}f_i + \gamma^i_{x^k}\underbrace{\nabla_{\dot\gamma_{x^k}}f_i}_{=0} \ ,$$

so that

$$\dot\gamma^i_{x^k}(s) = \dot\gamma^i_{x^k}(0) = x^i \quad \Longrightarrow \dot\gamma_{x^k} = x^if_i \ . \tag{A.5.65}$$

For any vector field along γ_{x^k} we have

$$\frac{DZ}{ds} = \frac{D(Z^if_i)}{ds} = \frac{dZ^i}{ds}f_i + Z^i\underbrace{\frac{Df_i}{ds}}_{=\nabla_{\dot\gamma_{x^k}}f_i=0} = \frac{dZ^i}{ds}f_i \ .$$

This shows that if we decompose the Z_r's in the $\{f_i\}$ basis,

$$Z_r(s,x^k) = Z^i{}_r(s,x^k)f_i \ ,$$

then the Jacobi equation (**??**) reads

$$\frac{d^2Z^i{}_r}{ds^2}(s,x^k) = \Big(R(\dot\gamma_{x^k},Z_r)\dot\gamma_{x^k}\Big)^i = R^i{}_{j\ell m}(sx^k)x^jx^mZ^\ell{}_r(s,x^k) \ . \tag{A.5.66}$$

We wish to derive the coefficients of a Taylor expansion of the Z_r's. Recall that (see Proposition A.5.14)

$$Z_r(0,x^k)=0\ .$$

Equation (A.5.66) gives thus

$$\frac{d^2Z^i{}_r}{ds^2}(0,x^k) \quad = \quad 0\ . \qquad \text{(A.5.67)}$$

Next,

$$e_r=\dot Z_r(0,x^k)=\dot Z^i{}_r(0,x^k)f_k(0)=\dot Z^i{}_r(0,x^k)e_k\ ,$$

which yields

$$\dot Z^i_r(0,x^k)=\delta^i_k\ . \qquad \text{(A.5.68)}$$

In order to continue, recall that for any function φ we have

$$\frac{d}{ds}\varphi=x^if_i(\varphi)\ .$$

Differentiating (A.5.66) one thus obtains

$$\begin{aligned}\frac{d^3Z^i{}_r}{ds^3}(s,x^k) &= \frac{d}{ds}\Big(R^i{}_{j\ell m}(sx^k)x^jx^mZ^\ell{}_r(s,x^k)\Big)\\ &= f_{m_1}(R^i{}_{j\ell m})(sx^k)x^jx^mx^{m_1}Z^\ell{}_r(s,x^k)+R^i{}_{j\ell m}(sx^k)x^jx^m\frac{dZ^\ell{}_r(s,x^k)}{ds}\ ,\end{aligned} \qquad \text{(A.5.69)}$$

leading to

$$\frac{d^3Z^i{}_r}{ds^3}(0,x^k) \quad = \quad R^i{}_{jrm}(0)x^jx^m=R^i{}_{(j|r|m)}(0)x^jx^m\ . \qquad \text{(A.5.70)}$$

One more derivative of (A.5.69) gives

$$\frac{d^4Z^i{}_r}{ds^4}(0,x^k) \quad = \quad 2f_{m_1}(R^i{}_{jrm})(0)x^jx^mx^{m_1}\ .$$

Let us show that the partial derivative above can be replaced by a covariant one. In order to do that, let φ^i be a basis dual to the f_i's, and let $\omega^i{}_j=\omega^i{}_{jk}\varphi^k$ be the connection one-forms associated with the frame f_i, as discussed•A.5.14 in Appendix A.6, it follows from (A.6.2) that

•*A.5.14:* **ptc:***this should be after Appendix A.6? or somewhere in the main text? or as an appendix to the relevant chapter?*

$$\omega^i{}_{jk}x^k=0\ . \qquad \text{(A.5.71)}$$

This implies that for any tensor field $T^{i_1\ldots i_k}{}_{j_1\ldots j_\ell}$ we have

$$\begin{aligned}x^m\nabla_{f_m}T^{i_1\ldots i_k}{}_{j_1\ldots j_\ell} &= x^mf_m(T^{i_1\ldots i_k}{}_{j_1\ldots j_\ell})+\underbrace{x^m\omega^{i_1}{}_{im}}_{=0}T^{i\ldots i_k}{}_{j_1\ldots j_\ell}+\ldots\\ &\quad -\underbrace{x^m\omega^j{}_{j_1m}}_{=0}T^{i_1\ldots i_k}{}_{j\ldots j_\ell}-\ldots\\ &= x^mf_m(T^{i_1\ldots i_k}{}_{j_1\ldots j_\ell})\ .\end{aligned}$$

It follows that

$$\begin{aligned}\frac{d^4 Z^i{}_r}{ds^4}(0,x^k) &= 2R^i{}_{jrm;m_1}(0)x^j x^m x^{m_1} \\ &= 2R^i{}_{(j|r|m;m_1)}(0)x^j x^m x^{m_1}\,. \qquad \text{(A.5.72)}\end{aligned}$$

More generally, the formula

$$\frac{d^n(fg)}{ds^n} = \frac{d^n f}{ds^n}g + n\frac{d^{n-1}f}{ds^{n-1}}\frac{dg}{ds} + \cdots$$

leads to•A.5.15

•*A.5.15:* **ptc:***fix the first lines; the x dependence of S needs justification*

$$\begin{aligned}\frac{d^{3+n} Z^i{}_r}{ds^{3+n}}(0,x^k) &= nf_{m_1}(R^i{}_{jrm})(0)x^j x^m x^{m_1} + S^i j \\ &= nR^i{}_{jrm;m_1}(0)x^j x^m x^{m_1} \\ &= \left(nR^i{}_{(m_1|r|m_2;m_3\cdots m_{n+2})}(0) + S^i{}_{rm_1\cdots m_{n+2}}\right)x^{m_1}x^{m_2}\cdots x^{m_{n+2}}\,, \\ & \qquad \text{(A.5.73)}\end{aligned}$$

where the coefficients $S^i{}_{rm_1\cdots m_{n+2}}$ are polynomials in the Riemann tensor and its derivatives of order less than or equal to $n-2$.
•A.5.16

•*A.5.16:* **ptc:***this needs finishing; the point is to derive an identity between Riemann and A*

As a corollary of the results derived so far, we obtain the following:•A.5.17

•*A.5.17:* **ptc:***add on, make sure that the relevant information is there*

PROPOSITION A.5.16 *Let Q be a polynomial in the inverse metric, the metric, and a finite number of their partial derivatives, with the property that the numerical value of Q does not depend upon the coordinate system chosen. Then Q is obtained by contractions of g, $g^\sharp$, the volume tensor $\sqrt{\det g}\,\epsilon_{a_1\ldots a_n}$ (where $\epsilon_{a_1\ldots a_n}$ is totally anti-symmetric and equals the sign of $(1\ldots n)\to(a_1\ldots a_n)$ for permutations), the Riemann tensor and its covariant derivatives.*

PROOF: In normal coordinates centred at p all partial derivatives of the metric at p are polynomials in the Riemann tensor and a finite number of its derivatives.
•A.5.18

•*A.5.18:* **ptc:***this is not quite complete, what about the volume form? why do they have to be contractions; there is an analysis of that by G.B. Gurevich, Foundations of the Theory of Algebraic invariants, Noordhoff, Groningen 1964 : from an email by Malcom:*

```
 { No I don't have a quick fix for that - it's
exactly why the 'I think' appeared. Even contracting with the
(raised) metric is not so clear because it involves, in the metric
itself, something of order n-1 (in n dimensions) divided by
something of order n in the components, so the result may be
rational rather than polynomial in the components. In fact
classical invariant theory tends to be couched in terms of
rational rather than polynomial invariants, possibly for this
reason. The invariants people take in GR are usually polynomial in
the Riemann tensor and its derivatives but not in the metric
components (for the same reason?). But every polynomial is
rational of course...
\\
A quick further look however turned up what is needed to complete
```

```
the proof - in the sense of where to look.  From a secondary
source (Siklos' thesis) Gurevich gives as the Fundamental Theorem
a theorem that every rational invariant of a system of arbitrary
tensors is a linear combination of terms each of which is obtained
from the tensors with the aid of outer multiplication, total
alternation (i.e. using the volume form) and contraction.  I would
guess the reason for the limitation on contractions in our case
can be understood as follows: the objects giving the coefficients
have to be tensors to get the invariance, and have to depend
solely on the metric and its derivatives (to be
polynomial/rational). The derivatives could be replaced by
derivatives with indices down multiplied by metric terms so we can
forget them as coefficients. So one needs only to ask what tensors
can one form solely from polynomials in the metric? But I don't
know if the full proof runs on that sort of line.

Regards, Malcolm }
```

□

A.5.6 Further first-order differential identities?

At the end of Section A.5.1 we have derived a differential identity satisfied by the Riemann tensor – the second Bianchi identity (A.5.16) – and again one would like to know whether any more first order differential identities exist. This turns out not to be the case. Similarly to (A.5.29), the idea now is to prescribe two terms in a Taylor expansion of the metric in normal coordinates. In the next section we will consider expansions with an arbitrary number of terms, so it is convenient to set-up an appropriate general formalism now.
•A.5.19

•*A.5.19:* **ptc:***this should be moved to a separate section on linearised equations in the body*

Consider a metric of the form

$$g_{\alpha\beta} = \mathring{g}_{\alpha\beta} + h_{\alpha\beta} \;, \tag{A.5.74}$$

here one should think of h as a perturbation. Let ϵ_0 be a parameter which measures the size of h, and for $i > 0$ let $\epsilon_i \geq 0$ be a parameter which measures the size of the $i-th$ •A.5.20 derivatives of h. By increasing the parameters if necessary one can without loss of generality assume that

•*A.5.20:* **ptc:***should his be partial or covariant? I think here is partial, but covariant would be more elegant? or this does not matter? should be said in the calculation*

$$0 \leq \epsilon_i \leq \epsilon_{i+1} \;, \tag{A.5.75}$$

and this assumption will be made throughout. For example, when h is the ℓ-th term from a Taylor expansion in normal coordinates together with the accompanying remainder, then for a smooth metric g one has

$$\epsilon_0 = C_0|x|^\ell \;, \quad \epsilon_1 = C_1|x|^{\ell-1} \;, \quad \ldots \;, \quad \epsilon_{\ell-1} = C_{\ell-1}|x| \;, \quad \epsilon_i = C_i \text{ for } i \geq \ell \;. \tag{A.5.76}$$

Assuming the metric to be C^1, to calculate the Christoffel symbols of g one can proceed as follows: choose a coordinate system[4] so that

$$\mathring{g}(0) = \eta \text{ and } \partial_\sigma \mathring{g}_{\mu\nu}(0) = 0, \text{ which gives } \mathring{\Gamma}^\alpha{}_{\beta\gamma} = o(1) \,, \tag{A.5.77}$$

hence

$$\begin{aligned} \Gamma^\alpha{}_{\beta\gamma} &= \frac{1}{2} g^{\alpha\sigma}\{\partial_\beta g_{\sigma\gamma} + \partial_\gamma g_{\sigma\beta} - \partial_\sigma g_{\beta\gamma}\} \\ &= \frac{1}{2}(\mathring{g}^{\alpha\sigma} + O(\epsilon_0))\{\partial_\beta h_{\sigma\gamma} + \partial_\gamma h_{\sigma\beta} - \partial_\sigma h_{\beta\gamma} + o(1)\} \\ &= \frac{1}{2}(\mathring{g}^{\alpha\sigma} + O(\epsilon_0))\{\mathring{\nabla}_\beta h_{\sigma\gamma} + \mathring{\nabla}_\gamma h_{\sigma\beta} - \mathring{\nabla}_\sigma h_{\beta\gamma} + o(1+\epsilon_0)\} \,. \end{aligned}$$

At $x = 0$ we thus obtain

$$\begin{aligned} \Gamma^\alpha{}_{\beta\gamma} &= \Gamma^\alpha{}_{\beta\gamma} - \mathring{\Gamma}^\alpha{}_{\beta\gamma} \\ &= \frac{1}{2}\mathring{g}^{\alpha\sigma}\{\mathring{\nabla}_\beta h_{\sigma\gamma} + \mathring{\nabla}_\gamma h_{\sigma\beta} - \mathring{\nabla}_\sigma h_{\beta\gamma}\} + O(\epsilon_0\epsilon_1) \,. \end{aligned}$$

The reader should note that it follows immediately from the axioms for a connection in Section **??** that the difference of two connections is a tensor, so that regardless of the coordinate system we have

$$C^\alpha{}_{\beta\gamma} := \Gamma^\alpha{}_{\beta\gamma} - \mathring{\Gamma}^\alpha{}_{\beta\gamma} = \frac{1}{2}\mathring{g}^{\alpha\sigma}\{\mathring{\nabla}_\beta h_{\sigma\gamma} + \mathring{\nabla}_\gamma h_{\sigma\beta} - \mathring{\nabla}_\sigma h_{\beta\gamma}\} + O(\epsilon_0\epsilon_1) \,. \tag{A.5.78}$$

To calculate the curvature we continue in coordinates satisfying (A.5.77). Calculating directly at $x = 0$ we have

$$\begin{aligned} R^\alpha{}_{\beta\gamma\delta} &= \partial_\gamma \Gamma^\alpha{}_{\beta\delta} - \partial_\delta \Gamma^\alpha{}_{\beta\gamma} + \Gamma^\alpha{}_{\sigma\gamma}\Gamma^\sigma{}_{\beta\delta} - \Gamma^\alpha{}_{\sigma\delta}\Gamma^\sigma{}_{\beta\gamma} \\ &= \partial_\gamma(\mathring{\Gamma}^\alpha{}_{\beta\delta} + C^\alpha{}_{\beta\delta}) - \partial_\delta(\mathring{\Gamma}^\alpha{}_{\beta\gamma} + C^\alpha{}_{\beta\gamma}) + C^\alpha{}_{\sigma\gamma}C^\sigma{}_{\beta\delta} - C^\alpha{}_{\sigma\delta}C^\sigma{}_{\beta\gamma} \\ &= \mathring{R}^\alpha{}_{\beta\gamma\delta} + \mathring{\nabla}_\gamma C^\alpha{}_{\beta\delta} - \mathring{\nabla}_\delta C^\alpha{}_{\beta\gamma} + C^\alpha{}_{\sigma\gamma}C^\sigma{}_{\beta\delta} - C^\alpha{}_{\sigma\delta}C^\sigma{}_{\beta\gamma} \,. \end{aligned}$$

The object that appears at the left of this string of equalities is a tensor, and so is the one on the right, so that we obtain in all coordinate systems

$$R^\alpha{}_{\beta\gamma\delta} = \mathring{R}^\alpha{}_{\beta\gamma\delta} + \mathring{\nabla}_\gamma C^\alpha{}_{\beta\delta} - \mathring{\nabla}_\delta C^\alpha{}_{\beta\gamma} + C^\alpha{}_{\sigma\gamma}C^\sigma{}_{\beta\delta} - C^\alpha{}_{\sigma\delta}C^\sigma{}_{\beta\gamma} \,. \tag{A.5.79}$$

We emphasise that this is an exact formula when C is defined by the first equality in (A.5.78), no error terms have been neglected. In the calculations that follow we will assume that terms which are $O(\epsilon_0\epsilon_1)$ are also $O(\epsilon_1)$; recall that we are thinking about a situation in which the ϵ's are small, in which case this property holds. The second equality in (A.5.78) gives then•A.5.21

•*A.5.21:* **ptc:***I have no idea where the first term comes from*

$$C^\alpha{}_{\beta\gamma} = O(\epsilon_0 + \epsilon_1 + \epsilon_0\epsilon_1) = O(\epsilon_1) \,, \tag{A.5.80}$$

[4]For smooth metrics normal coordinates for $\mathring{g}$ could be used. Since those require more differentiability than C^1, and further lead to a loss of differentiability of the metric, it is more convenient to use an approximate normal coordinate system adapted to $\mathring{g}$ as in Proposition A.5.2.

and

$$
\begin{aligned}
R_{\mu\beta\gamma\delta} &= g_{\mu\alpha}R^{\alpha}{}_{\beta\gamma\delta} \\
&= (\mathring{g}_{\mu\alpha}+h_{\mu\alpha})\left(\mathring{R}^{\alpha}{}_{\beta\gamma\delta}+\mathring{\nabla}_{\gamma}C^{\alpha}{}_{\beta\delta}-\mathring{\nabla}_{\delta}C^{\alpha}{}_{\beta\gamma}+O(\epsilon_1^2)\right) \\
&= \mathring{R}_{\mu\beta\gamma\delta}+h_{\mu\alpha}\mathring{R}^{\alpha}{}_{\beta\gamma\delta}+\mathring{\nabla}_{\gamma}(\mathring{g}_{\mu\alpha}C^{\alpha}{}_{\beta\delta})-\mathring{\nabla}_{\delta}(\mathring{g}_{\mu\alpha}C^{\alpha}{}_{\beta\gamma})+O(\epsilon_1\epsilon_2) \\
&= \mathring{R}_{\mu\beta\gamma\delta}+h_{\mu\alpha}\mathring{R}^{\alpha}{}_{\beta\gamma\delta}+\frac{1}{2}\mathring{\nabla}_{\gamma}\{\mathring{\nabla}_{\beta}h_{\mu\delta}+\mathring{\nabla}_{\delta}h_{\mu\beta}-\mathring{\nabla}_{\mu}h_{\beta\delta}\} \\
&\quad -\frac{1}{2}\mathring{\nabla}_{\delta}\{\mathring{\nabla}_{\beta}h_{\mu\gamma}+\mathring{\nabla}_{\gamma}h_{\mu\beta}-\mathring{\nabla}_{\mu}h_{\beta\gamma}\}+O(\epsilon_1\epsilon_2)\ .
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
R^{\mu}{}_{\beta\gamma\delta} &= \mathring{R}^{\mu}{}_{\beta\gamma\delta}+\frac{1}{2}\mathring{\nabla}_{\gamma}\{\mathring{\nabla}_{\beta}h^{\mu}{}_{\delta}+\mathring{\nabla}_{\delta}h^{\mu}{}_{\beta}-\mathring{\nabla}^{\mu}h_{\beta\delta}\} \\
&\quad -\frac{1}{2}\mathring{\nabla}_{\delta}\{\mathring{\nabla}_{\beta}h^{\mu}{}_{\gamma}+\mathring{\nabla}_{\gamma}h^{\mu}{}_{\beta}-\mathring{\nabla}^{\mu}h_{\beta\gamma}\}+O(\epsilon_1\epsilon_2)\ , \qquad \text{(A.5.81)}
\end{aligned}
$$

where $h^{\alpha}{}_{\beta}:=\mathring{g}^{\alpha\sigma}h_{\sigma\beta}$.

Rewriting a commutator of derivatives in terms of the Riemann tensor one finally obtains

$$
\begin{aligned}
R_{\alpha\beta\gamma\delta} &= \mathring{R}_{\alpha\beta\gamma\delta}+\frac{1}{2}\Big\{\mathring{\nabla}_{\gamma}\mathring{\nabla}_{\beta}h_{\alpha\delta}-\mathring{\nabla}_{\gamma}\mathring{\nabla}_{\alpha}h_{\beta\delta}-\mathring{\nabla}_{\delta}\mathring{\nabla}_{\beta}h_{\alpha\gamma}+\mathring{\nabla}_{\delta}\mathring{\nabla}_{\alpha}h_{\beta\gamma}\Big\} \\
&\quad +\mathring{R}^{\sigma}{}_{\beta\gamma\delta}h_{\alpha\sigma}-\frac{1}{2}\mathring{R}_{\alpha}{}^{\sigma}{}_{\gamma\delta}h_{\sigma\beta}-\frac{1}{2}\mathring{R}_{\beta}{}^{\sigma}{}_{\gamma\delta}h_{\alpha\sigma}+O(\epsilon_1\epsilon_2)\ . \qquad \text{(A.5.82)}
\end{aligned}
$$

For further reference we note

$$
\begin{aligned}
R_{\alpha\gamma} &= \mathring{R}_{\alpha\gamma}+\frac{1}{2}\Big\{\mathring{\nabla}_{\gamma}\mathring{\nabla}^{\beta}h_{\alpha\beta}-\mathring{\nabla}_{\gamma}\mathring{\nabla}_{\alpha}(\mathring{g}^{\beta\delta}h_{\beta\delta})-\mathring{\nabla}^{\beta}\mathring{\nabla}_{\beta}h_{\alpha\gamma}+\mathring{\nabla}^{\beta}\mathring{\nabla}_{\alpha}h_{\beta\gamma}\Big\} \\
&\quad -\mathring{R}_{\alpha}{}^{\beta}{}_{\gamma}{}^{\delta}h_{\beta\delta}+\mathring{R}^{\sigma}{}_{\gamma}h_{\alpha\sigma}-\frac{1}{2}\mathring{R}_{\alpha}{}^{\sigma}{}_{\gamma}{}^{\beta}h_{\sigma\beta}+\frac{1}{2}\mathring{R}^{\sigma}{}_{\gamma}h_{\alpha\sigma}+O(\epsilon_1\epsilon_2) \\
&= \mathring{R}_{\alpha\gamma}+\frac{1}{2}\Big\{\mathring{\nabla}_{\gamma}\mathring{\nabla}^{\beta}h_{\alpha\beta}-\mathring{\nabla}_{\gamma}\mathring{\nabla}_{\alpha}(\mathring{g}^{\beta\delta}h_{\beta\delta})-\mathring{\nabla}^{\beta}\mathring{\nabla}_{\beta}h_{\alpha\gamma}+\mathring{\nabla}^{\beta}\mathring{\nabla}_{\alpha}h_{\beta\gamma}\Big\} \\
&\quad -\frac{3}{2}\mathring{R}_{\alpha}{}^{\sigma}{}_{\gamma}{}^{\beta}h_{\sigma\beta}+\frac{3}{2}\mathring{R}^{\sigma}{}_{\gamma}h_{\alpha\sigma}+O(\epsilon_1\epsilon_2)\ , \qquad \text{(A.5.83)}
\end{aligned}
$$

where we have used

$$
g^{\alpha\beta}=\mathring{g}^{\alpha\beta}-\mathring{g}^{\alpha\delta}\mathring{g}^{\beta\gamma}h_{\delta\gamma}+O(\epsilon_0)\ .
$$

Moreover,

$$
R = \mathring{R}-\mathring{\nabla}^{\alpha}\mathring{\nabla}_{\alpha}(\mathring{g}^{\beta\delta}h_{\beta\delta})+\mathring{\nabla}^{\alpha}\mathring{\nabla}^{\beta}h_{\alpha\beta}-\mathring{R}^{\sigma\beta}h_{\sigma\beta}+O(\epsilon_1\epsilon_2)\ . \qquad \text{(A.5.84)}
$$

An alternative, equivalent, expression for the Ricci tensor can be obtained working directly with (A.5.81)

$$
R_{\beta\delta} = \mathring{R}_{\beta\delta}+\frac{1}{2}\mathring{\nabla}_{\gamma}\{\mathring{\nabla}_{\beta}h^{\gamma}{}_{\delta}+\mathring{\nabla}_{\delta}h^{\gamma}{}_{\beta}-\mathring{\nabla}^{\gamma}h_{\beta\delta}\}-\frac{1}{2}\mathring{\nabla}_{\delta}\mathring{\nabla}_{\beta}h^{\gamma}{}_{\gamma}+O(\epsilon_1\epsilon_2)\ . \qquad \text{(A.5.85)}
$$

If we denote by Ric' the derivative of the Ricci tensor with respect to the metric, (A.5.85) implies the formula

$$
\begin{aligned}
(\mathrm{Ric}'\cdot h)_{\beta\delta} &= \frac{1}{2}\left(-\mathring{\nabla}_{\gamma}\mathring{\nabla}^{\gamma}h_{\beta\delta}-\mathring{\nabla}_{\delta}\mathring{\nabla}_{\beta}\mathrm{tr}_{\mathring{g}}h\right)+\mathring{\nabla}_{\gamma}\mathring{\nabla}_{(\beta}h^{\gamma}{}_{\delta)} \\
&= -\frac{1}{2}\Delta h_{\beta\delta}+\mathring{\nabla}_{(\beta}\mathring{\nabla}_{\gamma}\left(h^{\gamma}{}_{\delta)}-\frac{1}{2}\mathrm{tr}_{\mathring{g}}h\ \delta^{\gamma}_{\delta)}\right) \\
&\quad +R^{\sigma}{}_{(\beta}h_{\gamma)\sigma}-R_{\beta}{}^{\sigma}{}_{\delta}{}^{\gamma}h_{\sigma\gamma}\ . \qquad \text{(A.5.86)}
\end{aligned}
$$

•A.5.22: **ptc:***some useful formulae in formules.tex by erwann*

•A.5.22

To continue, we introduce the set $\mathscr{R}_1$ of candidates for a first derivative of the Riemann tensor:

$$\mathscr{R}_1 := \{\rho_{\alpha\beta\gamma\delta\mu} \mid \rho_{(\alpha\beta)\gamma\delta\mu} = \rho_{\alpha\beta(\gamma\delta)\mu} = \rho_{\alpha[\beta\gamma\delta]\mu} = \rho_{\alpha\beta\gamma\delta\mu} - \rho_{\gamma\delta\alpha\beta\mu} = \rho_{\alpha\beta[\gamma\delta\mu]} = 0\} . \tag{A.5.87}$$

We write the following ansatz for the new metric

$$g_{\alpha\beta} = \mathring{g}_{\alpha\beta} + \frac{2}{3!} A_{\alpha\beta\gamma\delta\mu} x^\gamma x^\delta x^\mu , \tag{A.5.88}$$

so that the tensor h of (A.5.74) is a third-order polynomial in x. We assume that A lives in a space $\mathscr{A}_1$ of tensors satisfying

$$\mathscr{A}_1 := \{A_{\alpha\beta\gamma\delta\mu} \mid A_{[\alpha\beta]\gamma\delta\mu} = A_{\alpha\beta\gamma\delta\mu} - A_{\alpha\beta(\gamma\delta\mu)} = A_{\alpha(\beta\gamma\delta\mu)} = 0\} . \tag{A.5.89}$$

The coordinates x might, but do not have to, be chosen to be normal for $\mathring{g}$; if they are, then point 4. of Proposition A.5.13 shows that they remain normal for g, though this plays no role in the considerations that follow. The associated sequence of ϵ_i's satisfies (A.5.76) with $\ell = 3$. Noting that $\epsilon_1\epsilon_2 = C|x|^3$, Equation (A.5.82) gives

$$\begin{aligned} R_{\alpha\beta\gamma\delta}(0) &= \mathring{R}_{\alpha\beta\gamma\delta}(0) \\ R_{\alpha\beta\gamma\delta;\mu}(0) &= \mathring{R}_{\alpha\beta\gamma\delta;\mu}(0) + A_{\alpha\delta\gamma\beta\mu} - A_{\beta\delta\gamma\alpha\mu} - A_{\alpha\gamma\delta\beta\mu} + A_{\beta\gamma\delta\alpha\mu} . \end{aligned} \tag{A.5.90}$$

As before, the right-hand-side of (A.5.90) defines a map Φ_1 from $\mathscr{A}_1$ to $\mathscr{R}_1$:

Lemma A.5.17 *The map*

$$\Phi_1(A)_{\alpha\beta\gamma\delta\mu} = A_{\alpha\delta\gamma\beta\mu} - A_{\beta\delta\gamma\alpha\mu} - A_{\alpha\gamma\delta\beta\mu} + A_{\beta\gamma\delta\alpha\mu} .$$

is a bijection from $\mathscr{A}_1$ to $\mathscr{R}_1$.

Proof: As in the proof of Lemma A.5.5, we check that the map

$$\Psi_1(\rho)_{\alpha\beta\gamma\delta\mu} = -\rho_{\alpha(\gamma|\beta|\delta\mu)} , \tag{A.5.91}$$

which clearly has the right symmetries to take values in $\mathscr{A}_1$ (compare the beginning of proof of Lemma A.5.5), equals Φ_1^{-1}. In order to do this we write $\Psi_1(\rho)$ in detail,

$$\Psi_1(\rho)_{\alpha\beta\gamma\delta\mu} = -\frac{1}{3!}\Big\{\rho_{\alpha\gamma\beta\delta\mu} + \rho_{\alpha\delta\beta\gamma\mu} + \underbrace{\rho_{\alpha\gamma\beta\mu\delta} + \rho_{\alpha\mu\beta\gamma\delta}}_{B} + \underbrace{\rho_{\alpha\mu\beta\delta\gamma} + \rho_{\alpha\delta\beta\mu\gamma}}_{C}\Big\} .$$

The first two terms have the same form as in the definition of Ψ_0 in (A.5.91) except for a supplementary index μ added at the end, while Φ_1 leaves the last index alone, and those terms can therefore be handled as in (A.5.39), reproducing $\rho_{\alpha\beta\gamma\delta\mu}$. It is convenient to rewrite Φ_1 as

$$\Phi_1(A)_{\alpha\beta\gamma\delta\mu} = \Big\{A_{\alpha\delta\beta\gamma\mu} + A_{\beta\gamma\alpha\delta\mu}\Big\} - \gamma \leftrightarrow \delta ,$$

which says: translate the fourth index to the second slot by jumping over the previous ones, in the result exchange the first pair of indices with the second, and then anti-symmetrise the result over the original third and fourth indices. Applying this rule to calculate the contribution of B to $\Phi_1(\Psi_1(\rho))$ gives

$$\begin{aligned} B &\rightarrow \Big\{(\rho_{\alpha\mu\gamma\beta\delta}+\underbrace{\rho_{\gamma\beta\alpha\mu\delta}}_{=\rho_{\alpha\mu\gamma\beta\delta}})-\mu\leftrightarrow\beta\Big\}+\Big\{(\rho_{\alpha\gamma\mu\beta\delta}+\underbrace{\rho_{\mu\beta\alpha\gamma\delta}}_{=\rho_{\alpha\gamma\mu\beta\delta}})-\gamma\leftrightarrow\beta\Big\} \\ &= 2(\rho_{\alpha\mu\gamma\beta\delta}-\rho_{\alpha\beta\gamma\mu\delta}+\rho_{\alpha\gamma\mu\beta\delta}-\rho_{\alpha\beta\mu\gamma\delta}) \\ &= 2(\rho_{\alpha\mu\gamma\beta\delta}-2\rho_{\alpha\beta\gamma\mu\delta}+\rho_{\alpha\gamma\mu\beta\delta})\ . \end{aligned} \tag{A.5.92}$$

C coincides with B except for a symmetrisation in γ and δ, so that

$$B+C \quad\rightarrow\quad 2(\rho_{\alpha\mu\gamma\beta\delta}-2\rho_{\alpha\beta\gamma\mu\delta}+\rho_{\alpha\gamma\mu\beta\delta}+\rho_{\alpha\mu\delta\beta\gamma}-2\rho_{\alpha\beta\delta\mu\gamma}+\rho_{\alpha\delta\mu\beta\gamma})\ .$$

•A.5.23

•A.5.23: **ptc:***help, I couldn't finish this*

PROPOSITION A.5.18 *Any $\rho\in\mathscr{R}_1$ is freely prescribable at a point.*

•A.5.24

•A.5.24: **ptc:***do I want to say anything about the Einstein tensor in dimension three and / or higher dimensions?*

As shown in Proposition A.5.12, in dimension three the Riemann tensor is determined uniquely by the Ricci one. It follows that the second Bianchi identity can be rewritten as an identity for the derivatives of the Ricci tensor. Now, in the identity $R_{\alpha\beta[\delta\gamma;\mu]}=0$ in dimension three the only values of $\delta\gamma\mu$ that possibly lead to a non-trivial equation are — up to permutation — $\delta\gamma\mu=123$. Consider, for example, this identity with $\alpha\beta=12$:

$$R_{1212;3}+R_{1223;1}+R_{1231;2}=0\ . \tag{A.5.93}$$

In an ON frame with $g_{11}=\epsilon=\pm1$, $g_{22}=g_{33}=1$ Equation (A.5.47) gives

$$R_{1212}=\epsilon P_{22}+P_{11}\ ,\quad R_{1223}=-P_{13}\ ,\quad R_{1231}=-\epsilon P_{23}\ ,$$

which inserted into (A.5.93) gives

$$0=\epsilon\Big((\underbrace{P_{22}+\epsilon P_{11}}_{=\mathrm{tr}_g P-P_3{}^3})_{;3}-\underbrace{\epsilon P_{13;1}}_{=P_3{}^1{}_{;1}}-\underbrace{P_{23;2}}_{=P_3{}^2{}_{;2}}\Big)=\epsilon\Big(\mathrm{tr}_g P_{;3}-P_3{}^\alpha{}_{;\alpha}\Big)\ .$$

It should be clear, or can be checked by similar calculations, that the two remaining non-trivial identities combine together to the identity

$$\mathrm{tr}_g P_{;\beta}-P_\beta{}^\alpha{}_{;\alpha}=0\ , \tag{A.5.94}$$

equivalent to the second Bianchi identity in dimension three. Using (A.5.46) and (A.5.49) this is further equivalent to the usual divergence identity

$$\Big(R_\alpha{}^\beta-\frac{R}{2}\delta^\beta_\alpha\Big)_{;\beta}=0\ . \tag{A.5.95}$$

This leads to

PROPOSITION A.5.19 *In dimension three, the first derivatives of the Ricci tensor satisfying (A.5.95) are freely prescribable.* •A.5.25

•A.5.25: **ptc:** *what about higher dimensions?*

PROOF: The proof of Corollary A.5.12 shows that in dimension three a tensor field $\rho_{\alpha\beta\gamma\delta\mu}$ will satisfy all but the last equation in (A.5.87) if and only if $\rho_{\alpha\beta\gamma\delta\mu}$ is of the form

$$\rho_{\alpha\beta\gamma\delta\mu} = g_{\alpha\gamma} P_{\beta\delta\mu} - g_{\alpha\delta} P_{\beta\gamma\mu} + g_{\beta\delta} P_{\alpha\gamma\mu} - g_{\beta\gamma} P_{\alpha\delta\mu} \; ,$$

for some P which is symmetric in the first two indices. The calculation leading to (A.5.92) shows that the last condition in (A.5.87) will hold if and only if

$$P_{\alpha}{}^{\alpha}{}_{\beta} = P_{\beta}{}^{\alpha}{}_{\alpha} \; ,$$

which is equivalent to (A.5.95) when ρ is interpreted as the derivative of the Riemann tensor. □

A.5.7 Higher differential identities?

For $a, b \in T_p\mathscr{M}$ set

$$\begin{aligned} A_{\alpha\beta\gamma\delta}(a,b) &= a_\alpha a_\beta b_\gamma b_\delta + b_\alpha b_\beta a_\gamma a_\delta - a_\alpha b_\beta a_{(\gamma} b_{\delta)} - b_\alpha a_\beta a_{(\gamma} b_{\delta)} \\ &= a_\alpha a_{[\beta} b_{\gamma]} b_\delta + a_\alpha a_{[\beta} b_{\delta]} b_\gamma + b_\alpha b_{[\beta} a_{\gamma]} a_\delta + b_\alpha b_{[\beta} a_{\delta]} a_\gamma \; . \end{aligned} \tag{A.5.96}$$

The first line shows that A has the right symmetries both in the first pair of indices and in the last one, while the second immediately implies that the cyclic identity holds — remember that a symmetrisation of an anti-symmetrisation gives zero.

This generalises to more indices as follows:

$$\begin{aligned} A_{\alpha\beta\mu_1\mu_2\ldots\mu_n}(a,b) &= a_\alpha a_\beta b_{\mu_1} b_{\mu_2} \ldots b_{\mu_n} + b_\alpha b_\beta a_{(\mu_1} a_{\mu_2} b_{\mu_3} \ldots b_{\mu_n)} \\ &\quad -(a_\alpha b_\beta + b_\alpha a_\beta) a_{(\mu_1} b_{\mu_2} \ldots b_{\mu_n)} \; . \end{aligned} \tag{A.5.97}$$

In order to verify the only not-entirely obvious symmetry it suffices to rewrite (A.5.97) as

$$\begin{aligned} A_{\alpha\beta\mu_1\mu_2\ldots\mu_n}(a,b) &= a_\alpha a_\beta b_{\mu_1} b_{\mu_2} \ldots b_{\mu_n} - a_\alpha b_\beta a_{(\mu_1} b_{\mu_2} \ldots b_{\mu_n)} \\ &\quad + b_\alpha b_\beta a_{(\mu_1} a_{\mu_2} b_{\mu_3} \ldots b_{\mu_n)} - b_\alpha a_\beta a_{(\mu_1} b_{\mu_2} \ldots b_{\mu_n)} \; , \end{aligned}$$

and symmetrisation over all but the first index clearly gives zero on each of the lines above.

The next question we wish to address is whether there exist any differential identities involving more than one derivative which are satisfied by the Riemann tensor. Clearly, there are the obvious identities which result from differentiating the Bianchi identities; those can hardly be considered as new conditions. Next, there are identities which arise when one exchanges the order of two covariant derivatives, leading to contractions with the Riemann tensor. Those are of course trivial, and to get rid of those it is natural in our context to focus attention to completely symmetrised derivatives of the Riemann tensor. We shall use the symbol S for such derivatives:

$$S_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} := R_{\alpha\beta\gamma\delta;(\mu_1\dots\mu_\ell)}\ . \tag{A.5.98}$$

We have:

PROPOSITION A.5.20 *The symmetrised derivatives of the Riemann tensor satisfying the Bianchi identities are freely prescribable at a point.*•A.5.26

•A.5.26: **ptc:***I have no idea whether this is true*

Here the space $\mathscr{S}_\ell$ of candidate S tensors is the space•A.5.27

•A.5.27: **ptc:***the last condition seems wrong, and it seems to scary to write it down*

$$\begin{aligned}\mathscr{S}_\ell \ := \ & \{s_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} \mid s_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} - s_{\alpha\beta\gamma\delta(\mu_1\dots\mu_\ell)} = s_{(\alpha\beta)\gamma\delta\mu_1\dots\mu_\ell} = \\ & = s_{\alpha\beta(\gamma\delta)\mu_1\dots\mu_\ell} = s_{\alpha[\beta\gamma\delta]\mu_1\dots\mu_\ell} = s_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} - s_{\gamma\delta\alpha\beta\mu_1\dots\mu_\ell} = \\ & = s_{\alpha\beta[\gamma\delta\mu_1]\mu_2\dots\mu_\ell} = 0\}\ .\end{aligned} \tag{A.5.99}$$

The proofs are rather obvious generalisations of those in the preceding sections. We proceed by induction on ℓ, writing the metric in the form

$$g_{\alpha\beta} = \mathring{g}_{\alpha\beta} + \frac{2}{3!} A_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} x^\gamma x^\delta x^{\mu_1}\cdots x^{\mu_\ell}\ , \tag{A.5.100}$$

so that the tensor h of (A.5.74) is an ℓ'th order polynomial in x. We assume that A lives in a space $\mathscr{A}_1$ of tensors satisfying

$$\mathscr{A}_\ell := \{A_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} \mid A_{[\alpha\beta]\gamma\delta\mu_1\dots\mu_\ell} = A_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} - A_{\alpha\beta(\gamma\delta\mu_1\dots\mu_\ell)} = A_{\alpha(\beta\gamma\delta\mu_1\dots\mu_\ell)} = 0\}\ . \tag{A.5.101}$$

As before the coordinates x might be chosen to be normal for $\mathring{g}$; and they will be normal for g if and only if they are normal for $\mathring{g}$ by point 4. of Proposition A.5.13. The associated sequence of ϵ_i's satisfies (A.5.76). As $\epsilon_1\epsilon_2 = C|x|^{2\ell-3}$, Equation (A.5.82) gives

$$\begin{aligned} R_{\alpha\beta\gamma\delta}(0) &= \mathring{R}_{\alpha\beta\gamma\delta}(0) \\ &\dots \\ R_{\alpha\beta\gamma\delta;\mu_1\dots\mu_{\ell-1}}(0) &= \mathring{R}_{\alpha\beta\gamma\delta;\mu_1\dots\mu_{\ell-1}}(0) \end{aligned} \tag{A.5.102}$$

$$\begin{aligned} R_{\alpha\beta\gamma\delta;\mu_1\dots\mu_\ell}(0) &= \mathring{R}_{\alpha\beta\gamma\delta;\mu_1\dots\mu_\ell}(0) + A_{\alpha\delta\gamma\beta\mu_1\dots\mu_\ell} - A_{\beta\delta\gamma\alpha\mu_1\dots\mu_\ell} \\ & \quad - A_{\alpha\gamma\delta\beta\mu_1\dots\mu_\ell} + A_{\beta\gamma\delta\alpha\mu_1\dots\mu_\ell} \end{aligned} \tag{A.5.103}$$

As before, the right-hand-side of (A.5.103) defines a map Φ_ℓ from $\mathscr{A}_\ell$ to $\mathscr{S}_\ell$:

LEMMA A.5.21 *The map*

$$\Phi_\ell(A)_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} = A_{\alpha\delta\gamma\beta\mu_1\dots\mu_\ell} - A_{\beta\delta\gamma\alpha\mu_1\dots\mu_\ell} - A_{\alpha\gamma\delta\beta\mu_1\dots\mu_\ell} + A_{\beta\gamma\delta\alpha\mu_1\dots\mu_\ell}\ .$$

is a bijection from $\mathscr{A}_\ell$ to $\mathscr{S}_\ell$.

PROOF: The formula for Φ_ℓ^{-1} is

$$A_{\alpha\beta\gamma\delta\mu_1\dots\mu_\ell} = ?S_{\alpha|\gamma|\beta\delta\mu_1\dots\mu_\ell)}$$

(this should follow by abstract arguments from [84] if they are correct). □

A.6 Moving frames

A formalism which is very convenient for practical calculations is that of *moving frames*; it also plays a key role when considering spinors, compare Section **??**. By definition, a moving frame is a (locally defined) field of bases $\{e_a\}$ of TM such that the scalar products

$$g_{ab} := g(e_a, e_b) \tag{A.6.1}$$

are point independent. In most standard applications one assumes that the e_a's form an orthonormal basis, so that g_{ab} is a diagonal matrix with plus and minus ones on the diagonal. However, it is sometimes convenient to allow other such frames, *e.g.* with isotropic vectors being members of the frame.

It is customary to denote by ${\omega^a}_{bc}$ the associated connection coefficients:

$${\omega^a}_{bc} := \theta^a(\nabla_{e_c} e_b) \quad \Longleftrightarrow \quad \nabla_X e_b = {\omega^a}_{bc} X^c e_a \ , \tag{A.6.2}$$

where, as elsewhere, $\{\theta^a(p)\}$ is a basis of T^*_pM dual to $\{e_a(p)\} \subset T_pM$. The *connection one forms* ${\omega^a}_b$ are defined as

$${\omega^a}_b(X) := \theta^a(\nabla_X e_b) \quad \Longleftrightarrow \quad \nabla_X e_b = {\omega^a}_b(X) e_a \ . \tag{A.6.3}$$

As always we use the metric to raise and lower indices, so that

$$\omega_{abc} := g_{ad}{\omega^e}_{bc} \ , \quad \omega_{ab} := g_{ad}{\omega^e}_b \ . \tag{A.6.4}$$

When ∇ is metric compatible, the ω_{ab}'s are anti-antisymmetric: indeed, as the g_{ab}'s are point independent, for any vector field X we have

$$\begin{aligned}
0 = X(g_{ab}) = X(g(e_a, e_b)) &= g(\nabla_X e_a, e_b) + g(e_a, \nabla_X e_b) \\
&= g({\omega^c}_a(X) e_c, e_b) + g(e_a, {\omega^d}_b(X) e_d) \\
&= g_{cb}{\omega^c}_a(X) + g_{ad}{\omega^d}_b(X) \\
&= \omega_{ba}(X) + \omega_{ab}(X) \ .
\end{aligned}$$

Hence

$$\boxed{\omega_{ab} = -\omega_{ba} \quad \Longleftrightarrow \quad \omega_{abc} = -\omega_{bac} \ .} \tag{A.6.5}$$

If the connection is the Levi-Civita connection of g, this equation will allow us to algebraically express the ω_{ab}'s in terms of the Lie brackets of the vector fields e_a. In order to see this, we note that

$$g(e_a, \nabla_{e_c} e_b) = g(e_a, {\omega^d}_{bc} e_d) = g_{ad}{\omega^d}_{bc} = \omega_{abc} \ .$$

Rewritten the other way round this gives an alternative equation for the ω's with all indices down:

$$\omega_{abc} = g(e_a, \nabla_{e_c} e_b) \quad \Longleftrightarrow \quad \omega_{ab}(X) = g(e_a, \nabla_X e_b) \ . \tag{A.6.6}$$

If ∇ has no torsion we find

$$\omega_{abc} - \omega_{acb} = g(e_a, \nabla_{e_c} e_b - \nabla_{e_b} e_c) = g(e_a, [e_c, e_b]) \ .$$

We can now do the usual cyclic permutation calculation to obtain

$$\begin{aligned}\omega_{abc} - \omega_{acb} &= g(e_a, [e_c, e_b]) \ , \\ -(\omega_{bca} - \omega_{bac}) &= -g(e_b, [e_a, e_c]) \ , \\ -(\omega_{cab} - \omega_{cba}) &= -g(e_c, [e_b, e_a]) \ .\end{aligned}$$

Summing the three equations and using (A.6.5) we obtain

$$\boxed{\omega_{abc} = \frac{1}{2}\Big(g(e_a, [e_c, e_b]) - g(e_b, [e_a, e_c]) - g(e_c, [e_b, e_a])\Big) \ .} \tag{A.6.7}$$

Equation (A.6.7) provides an explicit expression for the ω's. While it is useful to know that there is one, and while this expression is useful to estimate things, it is rarely used for practical calculations; see Example A.6.1 for more comments about that last issue.

It turns out that one can obtain a simple expression for the torsion of ω using exterior differentiation. Recall that if α is a one-form, then its exterior derivative $d\alpha$ can be defined using the formula

$$d\alpha(X, Y) = X(\alpha(Y)) - Y(\alpha(X)) - \alpha([X, Y]) \ . \tag{A.6.8}$$

We set

$$T^a(X, Y) := \theta^a(T(X, Y)) \ ,$$

and using (A.6.8) together with the definition (A.4.18) of the torsion tensor T we calculate as follows:

$$\begin{aligned}T^a(X, Y) &= \theta^a(\nabla_X Y - \nabla_Y X - [X, Y]) \\ &= X(Y^a) + \omega^a{}_b(X)Y^b - Y(X^a) - \omega^a{}_b(Y)X^b - \theta^a([X, Y]) \\ &= X(\theta^a(Y)) - Y(\theta^a(X)) - \theta^a([X, Y]) + \omega^a{}_b(X)\theta^b(Y) - \omega^a{}_b(Y)\theta^b(X) \\ &= d\theta^a(X, Y) + (\omega^a{}_b \wedge \theta^b)(X, Y) \ .\end{aligned}$$

It follows that

$$T^a = d\theta^a + \omega^a{}_b \wedge \theta^b \ . \tag{A.6.9}$$

In particular when the torsion vanishes we obtain the so-called *Cartan's first structure equation*

$$\boxed{d\theta^a + \omega^a{}_b \wedge \theta^b = 0} \ . \tag{A.6.10}$$

EXAMPLE A.6.1 As an example of the moving frame technique we consider (the most general) three-dimensional spherically symmetric metric

$$g = e^{2\beta(r)}dr^2 + e^{2\gamma(r)}d\theta^2 + e^{2\gamma(r)}\sin^2\theta d\varphi^2 \ . \tag{A.6.11}$$

There is an obvious choice of ON co-frame•A.6.1 for g given by

•*A.6.1:* **ptc:***should this be frame or coframe; watch out for consistency throughout the section, and elsewhere*

$$\theta^1 = e^{\beta(r)}dr \ , \ \theta^2 = e^{\gamma(r)}d\theta \ , \ \theta^3 = e^{\gamma(r)}\sin\theta d\varphi \ , \tag{A.6.12}$$

leading to

$$g = \theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2 + \theta^3 \otimes \theta^3 \ ,$$

so that the frame e_a dual to the θ^a's will be ON, as desired:

$$g_{ab} = g(e_a, e_b) = \mathrm{diag}(1,1,1) \ .$$

The idea of the calculation which we are about to do is the following: there is only one connection which is compatible with the metric, and which is torsion free. If we find a set of one forms ω_{ab} which exhibit the properties just mentioned, then they have to be the connection forms of the Levi-Civita connection. As shown in the calculation leading to (A.6.5), the compatibility with the metric will be ensured if we require

$$\omega_{11} = \omega_{22} = \omega_{33} = 0 \ ,$$

$$\omega_{12} = -\omega_{21} \ , \quad \omega_{13} = -\omega_{31} \ , \quad \omega_{23} = -\omega_{32} \ .$$

Next, we have the equations for the vanishing of torsion:

$$\begin{aligned}
0 = d\theta^1 &= -\underbrace{\omega^1{}_1}_{=0}\theta^1 - \omega^1{}_2\theta^2 - \omega^1{}_3\theta^3 \\
&= -\omega^1{}_2\theta^2 - \omega^1{}_3\theta^3 \ , \\
d\theta^2 &= \gamma' e^{\gamma} dr\wedge d\theta = \gamma' e^{-\beta}\theta^1\wedge\theta^2 \\
&= -\underbrace{\omega^2{}_1}_{=-\omega^1{}_2}\theta^1 - \underbrace{\omega^2{}_2}_{=0}\theta^2 - \omega^2{}_3\theta^3 \\
&= \omega^1{}_2\theta^1 - \omega^2{}_3\theta^3 \ , \\
d\theta^3 &= \gamma' e^{\gamma}\sin\theta\, dr\wedge d\varphi + e^{\gamma}\cos\theta\, d\theta\wedge d\varphi = \gamma' e^{-\beta}\theta^1\wedge\theta^3 + e^{-\gamma}\cot\theta\,\theta^2\wedge\theta^3 \\
&= -\underbrace{\omega^3{}_1}_{=-\omega^1{}_3}\theta^1 - \underbrace{\omega^3{}_2}_{=-\omega^2{}_3}\theta^2 - \underbrace{\omega^3{}_3}_{=0}\theta^3 \\
&= \omega^1{}_3\theta^1 + \omega^2{}_3\theta^2 \ .
\end{aligned}$$

Summarising,

$$\begin{aligned}
-\omega^1{}_2\theta^2 - \omega^1{}_3\theta^3 \qquad\qquad &= 0 \ , \\
\omega^1{}_2\theta^1 \qquad\qquad - \omega^2{}_3\theta^3 &= \gamma' e^{-\beta}\theta^1\wedge\theta^2 \ , \\
\omega^1{}_3\theta^1 + \omega^2{}_3\theta^2 &= \gamma' e^{-\beta}\theta^1\wedge\theta^3 + e^{-\gamma}\cot\theta\,\theta^2\wedge\theta^3 \ .
\end{aligned}$$

It should be clear from the first and second line that an $\omega^1{}_2$ proportional to θ^2 should do the job; similarly from the first and third line one sees that an $\omega^2{}_3$ proportional to θ^3 should work. It is then easy to find the relevant coefficient, as well as to find $\omega^2{}_3$:

$$\omega^1{}_2 = -\gamma' e^{-\beta}\theta^2 = -\gamma' e^{-\beta+\gamma} d\theta \ , \tag{A.6.13a}$$
$$\omega^1{}_3 = -\gamma' e^{-\beta}\theta^3 = -\gamma' e^{-\beta+\gamma}\sin\theta\, d\varphi \ , \tag{A.6.13b}$$
$$\omega^2{}_3 = -e^{-\gamma}\cot\theta\,\theta^3 = -\cos\theta\, d\varphi \ . \tag{A.6.13c}$$

It is convenient to define *curvature two-forms*:

$$\Omega^a{}_b = \frac{1}{2} R^a{}_{bcd}\theta^c\wedge\theta^d \ . \tag{A.6.14}$$

The *second Cartan structure equation* then reads

$$\boxed{\Omega^a{}_b = d\omega^a{}_b + \omega^a{}_c\wedge\omega^c{}_b \ .} \tag{A.6.15}$$

This identity is easily verified using (A.6.8):

$$\begin{aligned}
\Omega^a{}_b(X,Y) &= \frac{1}{2}R^a{}_{bcd}\underbrace{\theta^c\wedge\theta^d(X,Y)}_{=X^cY^d-X^dY^c}\\
&= R^a{}_{bcd}X^cY^d\\
&= \theta^a(\nabla_X\nabla_Y e_b-\nabla_Y\nabla_X e_b-\nabla_{[X,Y]}e_b)\\
&= \theta^a(\nabla_X(\omega^c{}_b(Y)e_c)-\nabla_Y(\omega^c{}_b(X)e_c)-\omega^c{}_b([X,Y])e_c)\\
&= \theta^a\Big(X(\omega^c{}_b(Y))e_c+\omega^c{}_b(Y)\nabla_X e_c\\
&\quad -Y(\omega^c{}_b(X))e_c-\omega^c{}_b(X)\nabla_Y e_c-\omega^c{}_b([X,Y])e_c\Big)\\
&= X(\omega^a{}_b(Y))+\omega^c{}_b(Y)\omega^a{}_c(X)\\
&\quad -Y(\omega^a{}_b(X))-\omega^c{}_b(X)\omega^a{}_c(Y)-\omega^a{}_b([X,Y])\\
&= \underbrace{X(\omega^a{}_b(Y))-Y(\omega^a{}_b(X))-\omega^a{}_b([X,Y])}_{=d\omega^a{}_b(X,Y)}\\
&\quad +\omega^a{}_c(X)\omega^c{}_b(Y)-\omega^a{}_c(Y)\omega^c{}_b(X)\\
&= (d\omega^a{}_b+\omega^a{}_c\wedge\omega^c{}_b)(X,Y)\ .
\end{aligned}$$

Equation (A.6.15) provides an efficient way of calculating the curvature tensor of any metric.

EXAMPLE A.6.1 CONTINUED: From (A.6.13) we find:

$$\begin{aligned}
\Omega^1{}_2 &= d\omega^1{}_2+\underbrace{\omega^1{}_1}_{=0}\wedge\omega^1{}_2+\omega^1{}_2\wedge\underbrace{\omega^2{}_2}_{=0}+\underbrace{\omega^1{}_3\wedge\omega^3{}_2}_{\sim\theta^3\wedge\theta^3=0}\\
&= -d(\gamma' e^{-\beta+\gamma}d\theta)\\
&= -(\gamma' e^{-\beta+\gamma})'dr\wedge d\theta\\
&= -(\gamma' e^{-\beta+\gamma})'e^{-\beta-\gamma}\theta^1\wedge\theta^2\\
&= \sum_{a<b}R^1{}_{2ab}\theta^a\wedge\theta^b\ ,
\end{aligned}$$

which shows that the only non-trivial coefficient (up to permutations) with the pair 12 in the first two slots is

$$R^1{}_{212}=-(\gamma' e^{-\beta+\gamma})'e^{-\beta-\gamma}\ . \tag{A.6.16}$$

A similar calculation, or arguing by symmetry, leads to

$$R^1{}_{313}=-(\gamma' e^{-\beta+\gamma})'e^{-\beta-\gamma}\ . \tag{A.6.17}$$

Finally,

$$\begin{aligned}
\Omega^2{}_3 &= d\omega^2{}_3+\omega^2{}_1\wedge\omega^1{}_3+\underbrace{\omega^2{}_2}_{=0}\wedge\omega^2{}_3+\omega^2{}_3\wedge\underbrace{\omega^3{}_3}_{=0}\\
&= -d(\cos\theta\, d\varphi)+(\gamma' e^{-\beta}\theta^2)\wedge(-\gamma' e^{-\beta}\theta^3)\\
&= (e^{-2\gamma}-(\gamma')^2e^{-2\beta})\theta^2\wedge\theta^3\ ,
\end{aligned}$$

yielding

$$R^2{}_{323}=e^{-2\gamma}-(\gamma')^2e^{-2\beta}\ . \tag{A.6.18}$$

The curvature scalar can easily be calculated now to be

$$
\begin{aligned}
R = R^{ij}{}_{ij} &= 2(R^{12}{}_{12} + R^{13}{}_{13} + R^{23}{}_{23}) \\
&= -4(\gamma' e^{-\beta+\gamma})' e^{-\beta-\gamma} + 2(e^{-2\gamma} - (\gamma')^2 e^{-2\beta}) \, . \qquad \text{(A.6.19)}
\end{aligned}
$$

Appendix B

Some interesting space-times

B.1 Minkowski space-time

The simplest example possible of a Lorentzian metric is that of Minkowski space-time: by definition, this is $\mathbb{R}^{n+1}$ with the metric

$$\eta \equiv \eta_{\mu\nu} dx^\mu dx^\nu := -(dx^0)^2 + (dx^1)^2 + \cdots (dx^n)^2 \ . \qquad \text{(B.1.1)}$$

The symbol $\eta_{\mu\nu}$ will be used throughout this work as above, and the coordinates x^μ in (B.1.1) will be referred to as *Minkowskian*; we will sometimes write t for x^0. The inverse metric $\eta^{\mu\nu}$ coincides with η:

$$\eta^{\mu\nu} = \text{diag}(-1, +1, \cdots, +1) \ . \qquad \text{(B.1.2)}$$

In the Minkowskian coordinates the connection coefficients, as given by (A.5.21), vanish, which implies the vanishing of the Riemann tensor.

B.2 further examples (Schwarzschild, Kerr)...

B.3 Taub-NUT space-times

•B.3.1: **ptc:***I am not sure that this is proof read, in final form, in any case one could add here some results about non-equivalent extensions, non-equivalent scris, and so on*

•B.3.1 The Taub–NUT metrics are solutions of vacuum Einstein equations on space-time manifolds $\mathscr{M}_I$ of the form

$$\mathscr{M}_I := I \times S^3 ,$$

where I an interval. They take the form [82]

$$-U^{-1}dt^2 + (2L)^2 U\sigma_1^2 + (t^2+L^2)(\sigma_2^2+\sigma_3^2) , \tag{B.3.1}$$

$$U(t) = -1 + \tfrac{2(mt+L^2)}{t^2+L^2} . \tag{B.3.2}$$

Here L and m are real numbers with $L > 0$. Further, the one-forms σ_1, σ_2 and σ_3 are left invariant one–forms on $SU(2) \approx S^3$: If

$$i_{S^3} : S^3 \to \mathbb{R}^4$$

is the standard embedding of S^3 into $\mathbb{R}^4$, then

$$\begin{aligned} \sigma_1 &= 2i^*_{S^3}(x\,dw - w\,dx + y\,dz - z\,dy) , \\ \sigma_2 &= 2i^*_{S^3}(z\,dx - x\,dz + y\,dw - w\,dy) , \\ \sigma_3 &= 2i^*_{S^3}(x\,dy - y\,dx + z\,dw - w\,dz) . \end{aligned}$$

The function U always has two zeros,

$$U(t) = \frac{(t_+ - t)(t - t_-)}{t^2+L^2} ,$$

where

$$t_\pm := m \pm \sqrt{m^2+L^2} .$$

It follows that I has to be chosen so that $t_\pm \notin I$.

It is convenient to parameterize S^3 with Euler angles

$$(\mu, \theta, \varphi) \in [0, 2\pi] \times [0, 2\pi] \times [0, \pi] ,$$

so that

$$x + iy = \sin\left(\frac{\theta}{2}\right) e^{i(\mu-\phi)/2} , \quad z + iw = \cos\left(\frac{\theta}{2}\right) e^{i(\mu+\phi)/2} . \tag{B.3.3}$$

This leads to the following form of the metric

$$g = -U^{-1}dt^2 + (2L)^2 U(d\mu + \cos\theta d\varphi)^2 + (t^2+L^2)(d\theta^2 + \sin^2\theta d\varphi^2) \tag{B.3.4}$$

There is a natural action of S^1 on circles obtained by varying μ at fixed t, θ and φ:

$$\mu \to \mu + 2\alpha .$$

It follows from (B.3.3) that this corresponds to the following action of S^1 on $\mathbb{R}^4$

$$(x + iy, z + iw) \to e^{i\alpha}(x + iy, z + iw) .$$

This shows that the circles with fixed t, θ and φ form the *Hopf fibration* of S^3 by S^1's.•B.3.2

•B.3.2: **ptc:***I'm note completely sure; have to crosscheck*

Clearly

$$g^\#(dt, dt) = g^{tt} = -U , \tag{B.3.5}$$

which shows that the level sets of t are

- spacelike for $t \in (t_-, t_+)$, and
- timelike for $t < t_-$ or $t > t_+$.

Equation (B.3.5) further shows that

$$\nabla t = g^{\mu t}\partial_\mu = -U\partial_t \ ,$$

so that

$$g(\nabla t, \nabla t) = -U(t) < 0 \quad \text{for} \quad t \in (t_-, t_+) \ .$$

Equivalently, t is a time function in that range of t's. Theorem **??** implies that

$$(\mathscr{M}_{(t_-,t_+)}, g) \quad \text{is globally hyperbolic} \ .$$

From (B.3.4) we further find

$$g(\partial_\mu, \partial_\mu) = 2L^2 U \ ,$$

so that the Hopf circles — to which ∂_μ is tangent — are

- spacelike for $t \in (t_-, t_+)$, and
- timelike for $t < t_-$ or $t > t_+$.

In particular

$$(\mathscr{M}_{(-\infty,t_-)}, g) \quad \text{and} \quad (\mathscr{M}_{(t_+,\infty)}, g) \quad \text{contain closed timelike curves} \ .$$

Let γ be the metric induced by g on the level sets of t,

$$\gamma = (2L)^2 U\sigma_1^2 + (t^2 + L^2)(\sigma_2^2 + \sigma_3^2) \ .$$

Again in the t-range (t_-, t_+), the volume $|\mathscr{S}_\tau|$ of the level sets $\mathscr{S}_\tau$ of t equals

$$|\mathscr{S}_\tau| = \int_{\mathscr{S}_\tau} d\mu_\gamma = \sqrt{\frac{U(t)(t^2+L^2)}{U(0)L^2}}|\mathscr{S}_0| \ .$$

Here $d\mu_\gamma$ is the volume element of the metric γ — in local coordinates

$$d\mu_\gamma = \sqrt{\det \gamma_{ij}}\, d^3x \ .$$

This is a typical "big-bang — big-crunch" behaviour, where the volume of the space-slices of the universe "starts" at zero, expands to a maximum, and collapses again to a zero value.

The standard way of performing extensions across the Cauchy horizons $t_\pm$ is to introduce new coordinates

$$(t, \mu, \theta, \varphi) \to (t, \mu \pm \int_{t_0}^{t} [2LU(s)]^{-1} ds, \theta, \varphi) \ , \tag{B.3.6}$$

which gives

$$\begin{aligned} g_\pm \quad = \quad & \pm 4L(d\mu + \cos\theta d\varphi)dt \\ & +(2L)^2 U(d\mu + \cos\theta d\varphi)^2 + (t^2 + L^2)(d\theta^2 + \sin^2\theta d\varphi^2) \ . \end{aligned} \tag{B.3.7}$$

Somewhat surprisingly, the metrics $g_\pm$ are non-singular for all $t \in \mathbb{R}$. In order to see that, let

$$\mathscr{M} := \mathbb{R} \times S^3 ,$$

and on $\mathscr{M}$ consider the one-forms θ^a, $a = 0, \ldots, 3$, defined as

$$\begin{aligned} \theta^0 &= dt , \\ \theta^1 &= 2\epsilon L(d\mu + \cos\theta d\phi) = 2L\epsilon\sigma_3 , \\ \theta^2 &= \sqrt{t^2 + L^2}\, d\theta , \\ \theta^3 &= \sqrt{t^2 + L^2}\, \sin\theta d\varphi . \end{aligned}$$

θ^0 and θ^1 are smooth everywhere on $\mathscr{M}$, while θ^2 and θ^3 are smooth except for the usual spherical coordinates singularity at the south and north poles of S^3. In this frame the metrics g_ϵ take the form

$$\begin{aligned} g_\epsilon &= \theta^0 \otimes \theta^1 + \theta^1 \otimes \theta^0 + U\theta^1 \otimes \theta^1 + \theta^2 \otimes \theta^2 + \theta^3 \otimes \theta^3 \\ &= g_{ab}\theta^a \otimes \theta^b . \end{aligned} \tag{B.3.8}$$

Each of the metrics $g_\pm$ can be smoothly conformally extended to the boundary at infinity "$t = \infty$" by introducing

$$x = 1/t ,$$

so that (B.3.7) becomes

$$\begin{aligned} g_\pm = x^{-2}\Big(&\mp 4L(d\mu + \cos\theta d\varphi)dx \\ &+(2L)^2 x^2 U (d\mu + \cos\theta d\varphi)^2 + (1 + L^2x^2)(d\theta^2 + \sin^2\theta d\varphi^2)\Big) \end{aligned} \tag{B.3.9}$$

In each case this leads to a Scri diffeomorphic to S^3. There is a simple isometry between g_+ and g_- given by

$$(x, \mu, \theta, \varphi) \to (x, -\mu, \theta, -\varphi)$$

(this does correspond to a smooth map of the region $t \in (t_+, \infty)$ into itself, cf. [33]), so that the two Scri's so obtained are isometric. However, in addition to the two ways of attaching Scri to the region $t \in (t_+, \infty)$ there are the two corresponding ways of extending this region across the Cauchy horizon $t = t_+$, leading to four possible manifolds with boundary. It can then be seen, using e.g. the arguments of [33], that the four possible manifolds split into two pairs, each of the manifolds from one pair *not* being isometric to one from the other. Taking into account the corresponding completion at "$t = -\infty$", and the two extensions across the Cauchy horizon $t = t_-$, one is led to four inequivalent conformal completions of each of the two inequivalent [33] time-oriented, maximally extended, standard Taub-NUT space-times.

Appendix C

Conformal rescalings of the metric

Consider a metric $\widetilde{g}$ related to g by a conformal rescaling:

$$\widetilde{g}_{ij} = \varphi^{\ell} g_{ij} \quad \Longleftrightarrow \quad \widetilde{g}^{ij} = \varphi^{-\ell} g^{ij} \ . \tag{C.0.1}$$

This gives the following transformation law for the Christoffel symbols:

$$\begin{aligned}
\widetilde{\Gamma}^{i}{}_{jk} &= \frac{1}{2}\widetilde{g}^{im}(\partial_j \widetilde{g}_{km} + \partial_k \widetilde{g}_{jm} - \partial_m \widetilde{g}_{jk}) \\
&= \frac{1}{2}\varphi^{-\ell} g^{im}(\partial_j(\varphi^{\ell}\widetilde{g}_{km}) + \partial_k(\varphi^{\ell}\widetilde{g}_{jm}) - \partial_m(\varphi^{\ell} g_{jk})) \\
&= \Gamma^{i}{}_{jk} + \frac{\ell}{2\varphi}(\delta^i_k \partial_j \varphi + \delta^i_j \partial_k \varphi - g_{jk} D^i \varphi) \ ,
\end{aligned} \tag{C.0.2}$$

where D denotes the covariant derivative of g. Equation (C.0.2) can be rewritten as

$$\tilde{D}_X Y = D_X Y + C(X,Y) \ , \tag{C.0.3}$$

with

$$C(X,Y) = \frac{\ell}{2\varphi}\Big(Y(\varphi)X + X(\varphi)Y - g(X,Y)D\varphi\Big) \tag{C.0.4a}$$

$$= \frac{\ell}{2\varphi}\Big(Y(\varphi)X + X(\varphi)Y - \tilde{g}(X,Y)\tilde{D}\varphi\Big) \ . \tag{C.0.4b}$$

C.1 The curvature

Further,•C.1.1

•C.1.1: **ptc:***these are formulae from Mielke, checked by Walter, but perhaps a derivation would make sense, especially since we don't have riemann here; one should also introduce the schouten tensor, the conformal geodesics, etc*

$$\begin{aligned}
\tilde{R}_{ij} &:= \mathrm{Ric}(\tilde{g})_{ij} \\
&= R_{ij} - \frac{(n-2)\ell}{2\varphi} D_i D_j \varphi + \frac{(n-2)\ell(\ell+2)}{4\varphi^2} D_i\varphi D_j\varphi - \frac{\ell}{2\varphi}\Delta_g \varphi\, g_{ij} \\
&\quad - \frac{(n-2)\ell^2 - 2\ell}{4\varphi^2} D^k\varphi D_k \varphi\, g_{ij} \ .
\end{aligned} \tag{C.1.1}$$

By contraction one obtains

$$\begin{aligned}\tilde{R} &:= \tilde{g}^{ij}\tilde{R}_{ij} = \varphi^{\ell} g^{ij}\tilde{R}_{ij} \\ &= \varphi^{\ell}\left(R - \frac{(n-1)\ell}{\varphi}\Delta_g\varphi - \frac{(n-1)\ell\,\{(n-2)\ell-4\}}{4\varphi^2}D^i\varphi D_i\varphi\right) .\end{aligned} \tag{C.1.2}$$

Clearly a very convenient choice is

$$(n-2)\ell = 4 ,$$

leading to

$$\tilde{g}_{ij} = \varphi^{\frac{4}{n-2}} g_{ij} , \qquad \tilde{R} = \varphi^{\ell}\left(R - \frac{4(n-1)}{(n-2)\varphi}\Delta_g\varphi\right) . \tag{C.1.3}$$

C.2 Non-characteristic hypersurfaces

Let $\mathscr{S}$ be a non-characteristic hypersurface in $\mathscr{M}$, under (C.0.1) the unit normal to $\mathscr{S}$ transforms as

$$\tilde{n}^i = \varphi^{-\ell/2} n^i \quad \Longleftrightarrow \quad \tilde{n}_i = \varphi^{\ell/2} n_i . \tag{C.2.1}$$

The projection tensor P defined in (6.3.4) is invariant under (C.0.1),

$$\tilde{P} = P .$$

From the definition (6.3.6) of the Weingarten map we obtain, for $X \in T\mathscr{S}$,

$$\begin{aligned}\tilde{B}(X) &= \tilde{P}(\tilde{D}_X\tilde{n}) = P\Big(D_X(\varphi^{-\ell/2}n) + C(X,\tilde{n})\Big) \\ &= P\Big(X(\varphi^{-\ell/2})n + \varphi^{-\ell/2}D_X n + \frac{\ell}{2\varphi}\Big(\tilde{n}(\varphi)X + X(\varphi)\tilde{n} - \underbrace{\tilde{g}(X,\tilde{n})}_{0}\tilde{D}\varphi\Big) \\ &= P\Big(\varphi^{-\ell/2}D_X n + \frac{\ell}{2\varphi}\tilde{n}(\varphi)X\Big) \\ &= \varphi^{-\ell/2}B(X) + \frac{\ell}{2\varphi}\tilde{n}(\varphi)X .\end{aligned}$$

The definition (6.3.9) of the extrinsic curvature tensor (second fundamental form) K leads to, for $X, Y \in T\mathscr{S}$,

$$\tilde{K}(X,Y) = \tilde{g}\Big(\tilde{B}(X), Y\Big) = \tilde{g}\Big(\varphi^{-\ell/2}B(X) + \frac{\ell}{2\varphi}\tilde{n}(\varphi)X, Y\Big) ,$$

which can be rewritten in the following three equivalent forms

$$\tilde{K}(X,Y) = \varphi^{\ell/2}K(X,Y) + \frac{\ell\tilde{n}(\varphi)}{2\varphi}\tilde{g}(X,Y) \tag{C.2.2a}$$

$$= \varphi^{\ell/2}K(X,Y) + \frac{\ell\varphi^{(-\ell-2)/2}n(\varphi)}{2}\tilde{g}(X,Y) \tag{C.2.2b}$$

$$= \varphi^{\ell/2} K(X,Y) + \frac{\ell \varphi^{(\ell-2)/2} n(\varphi)}{2} g(X,Y)\ . \tag{C.2.2c}$$

(C.2.3)

(C.2.4)

Appendix D

Sobolev spaces on manifolds

D.1 Weighted Sobolev spaces on manifolds

•D.1.1: **ptc:***this is an auxiliary section, which will NOT be in the final version of the paper, just to be sure that things are right*

•D.1.1 In this section we shall recall the main elements of the construction of Sobolev spaces on manifolds. The material here is well known to most researchers; however, in other works assumptions on completeness or on radius of injectivity of the manifold are made — those ensure some further properties of the spaces constructed. Our point here is to make it clear that no supplementary properties are needed for the claims below. We will actually be slightly more general and construct weighted Sobolev spaces, the standard ones are obtained by setting to one the functions ϕ and ψ in Equation (D.1.2). The reader is referred to [7, 8, 59] for a discussion of further properties of the usual Sobolev spaces on Riemannian manifolds.

Let, then, M be a differentiable manifold[1]; by definition, M is paracompact and Hausdorff. The metric g on M will be assumed to be Riemannian and continuous, higher differentiability of g will be indicated whenever needed. Paracompactness of M implies that there exists a (countable) covering $\mathscr{O}_i$, $i \in \mathbb{N}$, of M with open sets on which local coordinate systems $\Phi_i : \mathscr{O}_i \to \mathbb{R}^n$ are defined. Let ϕ_i be a partition of unity subordinate to that covering. Let $\mathscr{B}(M)$ be the σ-algebra of Borel sets of M, for $\Omega \in \mathscr{B}(M)$ we set

$$\mu_i(\Omega) = \int_{\Phi_i(\mathscr{O}_i)} (\chi_\Omega \phi_i) \circ \Phi_i^{-1} \sqrt{\det g_{ij}}\, dx^1 \ldots dx^n \ ,$$

where χ_Ω denotes the characteristic function of the set Ω. By standard theory[2] on $\mathbb{R}^n$ the μ_i's are measures on $\mathscr{B}(M)$, and thus so is

$$\mathring{\mu} = \sum_{i \in \mathbb{N}} \mu_i \ .$$

Let $\Lambda(M, g)$ be the σ-algebra obtained by Caratheodory completion [93] of $(\mathscr{B}(M), \mathring{\mu})$, and let μ_g denote the resulting extension of $\mathring{\mu}$ to $\Lambda(M, g)$; $\Lambda(M, g)$ does actually not depend upon g, but we shall keep g in its argument to emphasize that the metric has been used in its construction•D.1.2 . For any $1 \le p < \infty$ and any non-negative $\Lambda(M, g)$-measurable function α on M the spaces $L^p(\alpha\mu_g)$ are then obtained by the usual construction of Lebesgue, starting from the measured space $(M, \Lambda(M, g), \alpha\mu_g)$. Here

•D.1.2: **ptc:***perhaps not?*

$$\alpha\mu_g(\Omega) := \int_M \chi_\Omega \alpha d\mu_g \ .$$

The Fischer-Riesz [94, Thm. 3.11] theorem shows that the $L^p(\alpha\mu_g)$'s are complete Banach spaces when equipped with the norm

$$\|f\|_{L^p} := \left(\int_M |f|^p \alpha d\mu_g \right)^{1/p} \ . \tag{D.1.1}$$

[1] We use the analysts' convention that a manifold M is always open; thus a manifold M with non-empty boundary ∂M does not contain its boundary; instead, $\overline{M} := M \cup \partial M$ is a manifold with boundary in the differential geometric sense.

[2] The elementary facts from measure theory which we use here can be found *e.g.* in [94] and [93].

(Throughout this section we shall freely mix the notions of functions and of equivalence classes of functions, depending upon the context, where two functions are equivalent if they are equal $\alpha\mu_g$ almost everywhere.) A function f is in $L^p(\alpha\mu_g)$ if and only if f is $\Lambda(M,g)$-measurable with the right-hand-side of (D.1.1) being finite. All continuous functions on M are $\Lambda(M,g)$-measurable, and all compactly supported continuous functions are in $L^p(\alpha\mu_g)$.

Let u be a section of the tensor bundle

$$T^r_s M := \underbrace{TM \otimes \cdots \otimes TM}_{r \text{ times}} \otimes \underbrace{T^*M \otimes \cdots \otimes T^*M}_{s \text{ times}} ;$$

u will be said to be measurable if for all collections X_1, X_2, …, X_s of continuous vector fields and all collections α_1, α_2, …, α_r of continuous one-forms the function $u(\alpha_1, \ldots, \alpha_r, X_1, \ldots, X_r)$ is measurable. u will be said to be in $L^p(\alpha\mu_g)$ if u is measurable and if the g-Riemannian norm $|u|_g$ of u is in $L^p(\alpha\mu_g)$. A tensor field u will be said to be in $L^p_{\rm loc}(\alpha\mu_g)$ if for every compact K the tensor field $\chi_K u$ is in $L^p(\alpha\mu_g)$. All continuous tensor fields are measurable; pointwise limits of measurable tensor fields are measurable tensor fields.

For i in $\mathbb{N}$ and for i-times differentiable tensor fields u we shall denote by $\nabla^{(i)}u$ the section of the tensor bundle $T^r_{s+i}M$ obtained by covariantly differentiating i times the tensor field u using the Levi-Civita connection of g, $\nabla^{(i)} := \underbrace{\nabla \ldots \nabla}_{i \text{ times}} u$, where ∇ is the Levi-Civita covariant derivative of g; this requires g to be i times differentiable when $r+s>0$; $i-1$ differentiability of g suffices if u is a function. In local coordinates, if

$$u = u^{i_1\cdots i_r}{}_{j_1\cdots j_s}\, \partial_{i_1} \otimes \cdots \otimes \partial_{i_r} \otimes dx^{j_1} \cdots \otimes dx^{j_s} ,$$

then

$$\nabla^{(i)}u := \nabla_{k_1} \cdots \nabla_{k_i} u^{i_1\cdots i_r}{}_{j_1\cdots j_s}\, \partial_{i_1} \otimes \cdots \otimes \partial_{i_r} \otimes dx^{k_1} \cdots \otimes dx^{k_i} \otimes dx^{j_1} \cdots \otimes dx^{j_s} .$$

Let us denote by $\langle \cdot, \cdot \rangle$ the pairing between $T^r_s M$ and $T^s_r M$ tensors:

$$\langle v, u \rangle := v^{j_1\cdots j_s}{}_{i_1\cdots i_r} u^{i_1\cdots i_r}{}_{j_1\cdots j_s} ;$$

recall that the summation convention is used throughout, which means that repeated indices, one up and one down, have to be summed over. Let $(\nabla^{(i)})^t$ denote the formal adjoint of $\nabla^{(i)}$ obtained by integrating i-times by parts:

$$\int \langle (\nabla^{(i)})^t v, u \rangle d\mu_g = \int \langle v, \nabla^{(i)} u \rangle d\mu_g ;$$

here we have assumed that both u and v are i-times differentiable and that at least one of them has compact support. For example, if u is a compactly supported differentiable field of one forms, then for differentiable v's we have

$$\int v^{ij} \nabla_i u_j = - \int u_j \nabla_i v^{ij} ,$$

so that

$$(\nabla^{(1)})^t v = -\nabla_i v^{ij} \partial_j .$$

Similarly one checks that whatever the rank of u, in our conventions above the operator $(\nabla^{(1)})^t$ is minus the divergence on the first slot of v. The higher order operators $(\nabla^{(i)})^t$ are obtained by iteration of the formula for $(\nabla^{(1)})^t$.

Let u and K be two tensor fields in $L^1_{\mathrm{loc}}(d\mu_g)$ and suppose that for any smooth compactly supported v we have

$$\int \langle (\nabla^{(i)})^t v, u\rangle d\mu_g = \int \langle v, K\rangle d\mu_g \ .$$

(This equation is well posed for metrics which are in $C^{i-1,1} = W^{i,\infty}_{\mathrm{loc}}$.) We then say that K is the i-th covariant derivative of u in the sense of distributions, and we set

$$\nabla^{(i)} u := K \ .$$

Let ϕ and ψ be two strictly positive $\Lambda(M,g)$-measurable functions on M. Let u be a tensor field in L^1_{loc} such that its distributional derivatives $\nabla^{(i)}u$, $i = 1,\ldots,k$ are also in L^1_{loc}, we shall say that $u \in W^{k,p}_{\phi,\psi}$ if the norm

$$\|u\|_{W^{k,p}_{\phi,\psi}} := (\int_M (\sum_{i=0}^k \left(\phi^i |\nabla^{(i)} u|_g \psi\right)^p d\mu_g)^{\frac{1}{p}} \tag{D.1.2}$$

is finite. We will write $W^{k,p}_{\phi,\psi}(g)$ for $W^{k,p}_{\phi,\psi}$ if ambiguities are likely to occur.

We claim that $W^{k,p}_{\phi,\psi}$ with the above norm is a Banach space. Indeed, let u_n be a Cauchy sequence in $W^{k,p}_{\phi,\psi}$, thus for $0 \le i \le k$ the sequences $\nabla^{(i)}u_n$ are Cauchy in $L^p(\phi^{ip}\psi^p d\mu_g)$. The Fischer-Riesz theorem implies that there exist tensor fields $K^{(i)}$ such that the $\nabla^{(i)}u_n$'s converge to $K^{(i)}$ in $L^p(\phi^{ip}\psi^p d\mu_g)$. It then follows by elementary considerations (using L^1-continuity of the integral together with Hölder's inequality if $p > 1$) that for any smooth compactly supported v's we have

$$\begin{aligned}
\int \langle (\nabla^{(i)})^t v, K^{(0)}\rangle d\mu_g &= \lim_{n\to\infty} \int \langle (\nabla^{(i)})^t v, u_n\rangle d\mu_g \\
&= \lim_{n\to\infty} \int \langle v, \nabla^{(i)} u_n\rangle d\mu_g \\
&= \int \langle v, K^{(i)}\rangle d\mu_g \ .
\end{aligned}$$

This shows that $K^{(i)}$ is the i-th distributional derivative of $K^{(0)}$, so that $K^{(0)}$ is in $W^{k,p}_{\phi,\psi}$, as desired, with u_n converging to $K^{(0)}$ in $W^{k,p}_{\phi,\psi}$.

When $p = 2$ the $W^{k,p}_{\phi,\psi}$'s are Hilbert spaces with the obvious scalar product associated to the norm (D.1.2).

We denote by $\mathring{W}^{k,p}_{\phi,\psi}$ the closure in $W^{k,p}_{\phi,\psi}$ of the space of $W^{k,p}_{\phi,\psi}$ functions or tensors which are compactly (up to a negligible set) supported in M, with the norm induced from $W^{k,p}_{\phi,\psi}$; it is easily seen that smooth compactly supported tensor fields are dense in $\mathring{W}^{k,p}_{\phi,\psi}$.

•*D.1.3:* **ptc:***added*

•D.1.3 For $k = 0$ the spaces $W^{k,p}_{\phi,\psi}$ and $\mathring{W}^{k,p}_{\phi,\psi}$ coincide: indeed, let $X_n \subset M$ be any family of open relatively compact sets exhausting M, it is easily seen from the Lebesgue dominated convergence theorem that for $u \in W^{0,p}_{\phi,\psi}$ the family of tensor fields $\chi_{X_n} u$ converges to u in $W^{0,p}_{\phi,\psi}$. We shall often use the notation L^p_ψ for this space.

The interesting question whether or not $\mathring{W}^{k,p}_{\phi,\psi}$ coincides with $W^{k,p}_{\phi,\psi}$ for $k \ge 1$ is irrelevant for our purposes, but we mention that if there exists a compact set $\mathscr{K} \subset M$ and a constant C such that we have

$$\phi \le C\sigma$$

on $M \setminus \mathscr{K}$, where σ denotes the distance from, say, some point $x_0 \in M$, then•D.1.4 the set C^∞_c of compactly supported smooth functions or tensor fields is dense in $W^{1,p}_{\phi,\psi}$. The reader is referred to [7, 8, 59] for further results concerning the density question when $\phi \equiv \psi \equiv 1$.

•*D.1.4:* **ptc:***justify, actually this might be wrong; probably want $x_0 \in \mathscr{K}$; what about completeness, injectivity radius?*

Appendix E

$W^{k,p}$ manifolds

Appendix F

A pot-pourri of formulae

•F.0.5 Here are some formulae for the commutation of derivatives:

•F.0.5: **ptc:**_these are formulae by Erwann, I haven't checked them_

$$\nabla_m\nabla_l t_{ik} - \nabla_l\nabla_m t_{ik} = R^p{}_{klm} t_{ip} + R^p{}_{ilm} t_{kp} ,$$

$$\nabla_i\nabla_j V^l - \nabla_j\nabla_i V^l = R^l{}_{kij} V^k ,$$

$$\nabla^k\nabla_k |df|^2 = 2(\nabla^l f \nabla_l \nabla^k \nabla_k f + Ric(\nabla f, \nabla f) + |\nabla\nabla f|^2) ,$$

$$\begin{aligned} \nabla^k\nabla_k\nabla_i\nabla_j f - \nabla_i\nabla_j\nabla^k\nabla_k f - R_{kj}\nabla^k\nabla_i f - R_{ki}\nabla^k\nabla_j f + 2R_{qjli}\nabla^q\nabla^l f \\ = (\nabla_i R_{kj} + \nabla_j R_{ki} - \nabla_k R_{ij})\nabla^k f . \end{aligned}$$

The Bianchi identities for a Levi-Civita connection:

$$R^i{}_{jkl} + R^i{}_{ljk} + R^i{}_{klj} = 0 ,$$

$$\nabla_l R^t{}_{ijk} + \nabla_k R^t{}_{ilj} + \nabla_j R^t{}_{ikl} = 0 ,$$

$$\nabla_t R^t{}_{ijk} + \nabla_k R_{ij} - \nabla_j R_{ik} = 0 ,$$

$$\nabla^k R_{ik} - \frac{1}{2}\nabla_k R = 0 .$$

Linearisations for various objects of interest:

$$D_g\Gamma^k_{ij}(g)h = \frac{1}{2}(\nabla_i h^k_j + \nabla_j h^k_i - \nabla^k h_{ij}) ,$$

$$2[D_g\mathrm{Riem}(g)h]_{sklm} = \nabla_l\nabla_k h_{sm} - \nabla_l\nabla_s h_{km} + \nabla_m\nabla_s h_{kl} - \nabla_m\nabla_k h_{sl} + R^p{}_{klm}h_{ps} + R^p{}_{sml}h_{pk} ,$$

$$2[D_g\mathrm{Riem}(g)h]^i_{klm} = \nabla_l\nabla_k h^i_m - \nabla_l\nabla^i h_{km} + \nabla_m\nabla^i h_{kl} - \nabla_m\nabla_k h^i_l + g^{is}R^p{}_{sml}h_{pk} - R^p{}_{klm}h^i_p ,$$

$$D_g Ric(g)h = \frac{1}{2}\Delta_L h - \mathrm{div}^*\mathrm{div}(Gh) ,$$

$$\Delta_L h = -\nabla^k\nabla_k h_{ij} + R_{ik}h^k_j + R_{jk}h^k_j - 2R_{ikjl}h^{kl} ,$$

$$Gh = h - \frac{1}{2}tr\ hg, \quad (\mathrm{div}h)_i = -\nabla^k h_{ik}, \quad \mathrm{div}^* w = \frac{1}{2}(\nabla_i w_j + \nabla_j w_i) ,$$

$$D_g R(g)h = -\nabla^k\nabla_k(tr\ h) + \nabla^k\nabla^l h_{kl} - R^{kl}h_{kl} ,$$

$$[D_g R(g)]^* f = -\nabla^k\nabla_k f g + \nabla\nabla f - f\ Ric(g) .$$

Warped products: Let (M,g), $\nabla := \nabla_g$, $f : M \to \mathbb{R}$ and

$$(\mathcal{M} = M \times_f I, \widetilde{g} = -f^2 dt^2 + g) ,$$

then for X, Y tangent to M and V, W tangent to I, we have

$$Ric(\widetilde{g})(X,Y) = Ric(g)(X,Y) - f^{-1}\nabla\nabla f(X,Y) ,$$

$$Ric(\widetilde{g})(X,V) = 0 = \widetilde{g}(X,V) ,$$
$$Ric(\widetilde{g})(V,W) = -f^{-1}\nabla^k\nabla_k f\widetilde{g}(V,W) .$$

Let (M,g), $\nabla := \nabla_g$, $f : M \to \mathbb{R}$ and let $(\mathcal{M} = M \times_f I, \widetilde{g} = \epsilon f^2 dt^2 + g)$, $\epsilon = \pm 1$. $x^a = (x^0 = t, x^i = (x^1, \ldots, x^n))$.

$$\widetilde{\Gamma}^0_{00} = \widetilde{\Gamma}^0_{ij} = \widetilde{\Gamma}^k_{i0} = 0,\ \ \widetilde{\Gamma}^0_{i0} = f^{-1}\partial_i f,\ \ \widetilde{\Gamma}^k_{00} = -\epsilon f \nabla^k f,\ \ \widetilde{\Gamma}^k_{ij} = \Gamma^k_{ij} ,$$

$$\widetilde{R}^l{}_{ijk} = R^l_{ijk},\ \ \widetilde{R}^l{}_{0j0} = -\epsilon f \nabla_j \nabla^l f,\ \ \widetilde{R}^0{}_{ij0} = f^{-1}\nabla_j\nabla_i f ,$$

$$\widetilde{R}^0{}_{ijk} = \widetilde{R}l_{ij0} = \widetilde{R}l_{0jk} = \widetilde{R}^0{}_{0jk} = \widetilde{R}^0{}_{0j0} = 0 ,$$

$$\widetilde{R}_{mijk} = R_{mijk},\ \ \widetilde{R}_{0ijk} = 0,\ \ \widetilde{R}_{0ij0} = \epsilon f \nabla_j \nabla_i f ,$$

$$\widetilde{R}_{ik} = R_{ik} - f^{-1}\nabla_k\nabla_i f,\ \ \widetilde{R}_{0k} = 0,\ \ \widetilde{R}_{00} = -\epsilon f \nabla^i\nabla_i f ,$$

$$\widetilde{R} = R - 2f^{-1}\nabla^i\nabla_i f .$$

Hypersurfaces: Let M be a non-isotropic hypersurface in $\widetilde{M}$, with ν normal, and u, v tangent to M at m, we have

$$II(u,v) = (\widetilde{\nabla}_U V - \nabla_U V)_m = (\widetilde{\nabla}_U V)^\perp_m = II(v,u) = -l(u,v)\nu_m .$$

Setting $S(u) = \widetilde{\nabla}_u \nu \in T_m M$, one has

$$< S(u), v >=< \widetilde{\nabla}_u \nu, v >=< -\nu, \widetilde{\nabla}_u V >= l(u,v) .$$

•F.0.6: **ptc:** *there are bound to be sign problems here*

If x, y, u, v are tangent to M, then•F.0.6

$$R(x,y,u,v) = \widetilde{R}(x,y,u,v) + l(x,u)l(y,v) - l(x,v)l(y,u) .$$

The *Gauss-Codazzi* equations read

$$\widetilde{R}(x,y,u,\nu) = \nabla_y l(x,u) - \nabla_x l(y,u) .$$

The Ricci tensor can be decomposed as:

$$\widetilde{R}(y,v) = R(y,v) + II \circ II(y,v) - trII\ II(y,v) + \widetilde{R}(\nu,y,\nu,v) ,$$

$$\widetilde{R}(y,\nu) = -\nabla_y trII + y^j\nabla^i II_{ij} ,$$
$$\widetilde{R} = R + |II|^2 - (trII)^2 + 2\widetilde{R}(\nu,\nu) .$$

The Weyl tensor:

$$W_{ijkl} = R_{ijkl} - \frac{1}{n-2}(R_{ik}g_{jl} - R_{il}g_{jk} + R_{jl}g_{ik} - R_{jk}g_{il}) + \frac{R}{(n-1)(n-2)}(g_{jl}g_{ik} - g_{jk}g_{il}) .$$

We have

$$W_i{}^j{}_{kl}(e^f g) = W_i{}^j{}_{kl}(g) \; .$$

The Schouten tensor

$$S_{ij} = \frac{1}{n-2}[2R_{ij} - \frac{R}{n-1} g_{ij}] \; .$$

Under a conformal transformation $g' = e^f g$, we have

$$\Gamma'^k_{ij} - \Gamma^k_{ij} = \frac{1}{2}(\delta^k_j \partial_i f + \delta^k_i \partial_j f - g_{ij}\nabla^k f) \; .$$

$$R'_{ij} = R_{ij} - \frac{n-2}{2}\nabla_i\nabla_j f + \frac{n-2}{4}\nabla_i f \nabla_j f - \frac{1}{2}(\nabla^k\nabla_k f + \frac{n-2}{2}|df|^2) g_{ij}$$

$$R' = e^{-f}[R - (n-1)\nabla^i\nabla_i f - \frac{(n-1)(n-2)}{4}\nabla^i f \nabla_i f] \; .$$

Specialising to $g' = e^{\frac{2}{n-2}u} g$,

$$R'_{ij} = R_{ij} - \nabla_i\nabla_j u + \frac{1}{n-2}\nabla_i u \nabla_j u - \frac{1}{n-2}(\nabla^k\nabla_k u + |du|^2) g_{ij} \; .$$

In the notation $g' = v^{\frac{2}{n-2}} g$,

$$R'_{ij} = R_{ij} - v^{-1}\nabla_i\nabla_j v + \frac{n-1}{n-2} v^{-2}\nabla_i v \nabla_j v - \frac{1}{n-2} v^{-1}(\nabla^k\nabla_k v) g_{ij} \; .$$

If we write instead $g' = \phi^{4/(n-2)} g$, then

$$R'_{ij} = R_{ij} - 2\phi^{-1}\nabla_i\nabla_j\phi + \frac{2n}{n-2}\phi^{-2}\nabla_i\phi\nabla_j\phi - \frac{2}{n-2}\phi^{-1}(\nabla^k\nabla_k\phi + \phi^{-1}|d\phi|^2) g_{ij} \; ,$$

$$R'\phi^{(n+2)/(n-2)} = -\frac{4(n-1)}{n-2}\nabla^k\nabla_k\phi + R\phi \; .$$

When we have two metrics g and g' at our disposal, then

$$T^k_{ij} := \Gamma'^k_{ij} - \Gamma^k_{ij} = \frac{1}{2} g'^{kl}(\nabla_i g'_{lj} + \nabla_j g'_{li} - \nabla_l g'_{ij}) \; .$$

$$\mathrm{Riem}'^i{}_{klm} - \mathrm{Riem}^i{}_{klm} = \nabla_l T^i_{km} - \nabla_m T^i_{kl} + T^i_{jl} T^j_{km} - T^i_{jm} T^j_{kl} \; .$$

Under $g' = e^f g$, the Laplacian acting on functions becomes•F.0.7

•*F.0.7:* **ptc:***quelle signe pour le laplacien*

$$\nabla'^k\nabla'_k v = e^{-f}(\nabla^k\nabla_k v + \frac{n-2}{2}\nabla^k f \nabla_k v) \; .$$

For symmetric tensors we have instead

$$\begin{aligned}
\nabla'^k\nabla'_k u_{ij} \quad = \quad & e^{-f}\Big[\nabla^k\nabla_k u_{ij} + \frac{n-6}{2}\nabla^k f \nabla_k u_{ij} - (\nabla_i f \nabla^k u_{kj} + \nabla_j f \nabla^k u_{ki}) \\
& + (\nabla^k f \nabla_i u_{kj} + \nabla^k f \nabla_j u_{ki}) + (\frac{3-n}{2}\nabla^k f \nabla_k f - \nabla^k\nabla_k f) u_{ij} \\
& - \frac{n}{4}(\nabla_i f \nabla^k f u_{kj} + \nabla_j f \nabla^k f u_{ki}) + \frac{1}{2}\nabla_i f \nabla_j f u^k_k + \frac{1}{2} g_{ij} u_{kl} \nabla^k f \nabla^l f\Big] \; .
\end{aligned}$$

Laplacians : For symmetric u's and arbitrary T's let

$$(Du)_{kij} := \frac{1}{\sqrt{2}}(\nabla_k u_{ij} - \nabla_j u_{ik}),$$

then

$$(D^*T)_{ij} = \frac{1}{2\sqrt{2}}(-\nabla^k T_{kij} - \nabla^k T_{kji} + \nabla^k T_{ijk} + \nabla^k T_{jik})\ .$$

Further

$$D^*Du_{ij} = -\nabla^k\nabla_k u_{ij} + \frac{1}{2}(\nabla^k\nabla_i u_{jk} + \nabla^k\nabla_j u_{ik})\ ,$$

and

$$\mathrm{div}^*\mathrm{div} u = -\frac{1}{2}(\nabla_i\nabla^k u_{jk} + \nabla_j\nabla^k u_{ik})\ ,$$

thus

$$(D^*D + \mathrm{div}^*\mathrm{div})u_{ij} = -\nabla^k\nabla_k u_{ij} + \frac{1}{2}(R_{kj}u_i^k + R_{ki}u_j^k - 2R_{qjli}u^{ql})\ .$$

Stationary metrics : (M,γ) Riemannian, in dimension three, let $\lambda : M \to \mathbb{R}$, $\xi : M \to T^*M$, $(N = I \times M, g)$ with

$$g(t,x) = \begin{pmatrix} \lambda & {}^t\xi \\ \xi & \lambda^{-1}(\xi\ {}^t\xi - \gamma) \end{pmatrix} = \lambda(dt + \lambda^{-1}\xi_i dx^i)^2 - \lambda^{-1}\gamma_{ij}dx^i dx^j\ .$$

Let $w = -\lambda^2 *_\gamma d(\lambda^{-1}\xi)$. $\nabla = \nabla_g$, $E^i = \gamma^{is}E_s$. Then

$$\mathrm{Ric}(\gamma)_{ij} = \frac{1}{2}\lambda^{-1}(\nabla_i\lambda\nabla_j\lambda + w_i w_j) + \lambda^{-2}(Ric(g)_{ij} - Ric(g)_{cd}\xi^c\xi^d\gamma_{ij})\ ,$$

$$\nabla^i\nabla_i\lambda = \lambda^{-1}(|d\lambda|^2 - |w|^2) - 2\lambda^{-1}Ric(g)_{ab}\xi^a\xi^b\ ,$$

$$\nabla^i(\lambda^{-2}w_i) = 0\ ,$$

$$\lambda(*_\gamma dw)^i = -2\lambda^{-1}T(g)^i_c\xi^c, \quad Ric(g) = G(T(g))\ .$$

Bibliography

[1] R.A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.

[2] L. Andersson and P.T. Chruściel, *On "hyperboloidal" Cauchy data for vacuum Einstein equations and obstructions to smoothness of null infinity*, Phys. Rev. Lett. **70** (1993), 2829–2832.

[3] L. Andersson and P.T. Chruściel, *On asymptotic behavior of solutions of the constraint equations in general relativity with "hyperboloidal boundary conditions"*, Dissert. Math. **355** (1996), 1–100.

[4] L. Andersson, P.T. Chruściel, and H. Friedrich, *On the regularity of solutions to the Yamabe equation and the existence of smooth hyperboloidal initial data for Einsteins field equations*, Comm. Math. Phys. **149** (1992), 587–612.

[5] L. Andersson, G.J. Galloway, and R. Howard, *A strong maximum principle for weak solutions of quasi-linear elliptic equations with applications to Lorentzian and Riemannian geometry*, Comm. Pure Appl. Math. **51** (1998), 581–624.

[6] A. Ashtekar and R.O. Hansen, *A unified treatment of null and spatial infinity in general relativity. i. universal structure, asymptotic symmetries and conserved quantities at spatial infinity*, J. Math. Phys. **19** (1978), 1542–1566.

[7] T. Aubin, *Espaces de Sobolev sur les variétés Riemanniennes*, Bull. Sci. Math., II. Ser. **100** (1976), 149–173.

[8] ———, *Nonlinear analysis on manifolds. Monge–Ampère equations*, Springer, New York, Heidelberg, Berlin, 1982.

[9] H. Bahouri and J.-Y. Chemin, *Équations d'ondes quasilineáires et estimations de Strichartz*, Am. Jour. Math. **121** (1999), 1337–1377.

[10] R. Bartnik, *The mass of an asymptotically flat manifold*, Comm. Pure Appl. Math. **39** (1986), 661–693.

[11] ———, *The mass of an asymptotically flat manifold*, Commun. Pure Appl. Math. **39** (1986), 661–693.

[12] ———, *Remarks on cosmological spacetimes and constant mean curvature surfaces*, Comm. Math. Phys. **117** (1988), 615–624.

[13] ———, *Quasi-spherical metrics and prescribed scalar curvature*, Jour. Diff. Geom. **37** (1993), 31–71.

[14] R. Bartnik and P.T. Chruściel, *Boundary value problems for Dirac-type equations*, (2003), math.DG/0307278.

[15] R. Bartnik and G. Fodor, *On the restricted validity of the thin sandwich conjecture*, Phys. Rev. D **48** (1993), 3596–3599.

[16] J. K. Beem, P. E. Ehrlich, and K. L. Easley, *Global Lorentzian geometry*, 2 ed., Pure and Applied Mathematics, vol. 202, Marcel Dekker, New York, 1996.

[17] R. Beig, *Arnowitt–Deser–Misner energy and g_{00}*, Phys. Lett. **69A** (1978), 153–155.

[18] R. Beig and W. Simon, *Proof of a multipole conjecture due to Geroch*, Commun. Math. Phys. **78** (1980), 75–82.

[19] J.-P. Bourguignon, D.G. Ebin, and J.E. Marsden, *Sur le noyau des opérateurs pseudo-différentiels à symbole surjectif et non injectif*, C. R. Acad. Sci. Paris Sér. A **282** (1976), 867–870.

[20] H.L. Bray, *Proof of the Riemannian Penrose conjecture using the positive mass theorem*, Jour. Diff. Geom. **59** (2001), 177–267, math.DG/9911173.

[21] J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of non-negative curvature*, Jour. Diff. Geom. **6** (1971), 119–128.

[22] Y. Choquet-Bruhat, J. Isenberg, and J.W. York, Jr, *Einstein constraints on asymptotically Euclidean manifolds*, Phys. Rev. D **61** (2000), 084034 (20 pp.), gr-qc/9906095.

[23] Y. Choquet-Bruhat and J. York, *The Cauchy problem*, General Relativity (A. Held, ed.), Plenum Press, New York, 1980.

[24] D. Christodoulou and S. Klainermann, *On the global nonlinear stability of Minkowski space*, Princeton University Press, Princeton, 1995.

[25] D. Christodoulou and N. Ó Murchadha, *The boost problem in general relativity*, Comm. Math. Phys. **80** (1980), 271–300.

[26] P.T. Chruściel, *A remark on the positive energy theorem*, Class. Quantum Grav. **33** (1986), L115–L121.

[27] ———, *Boundary conditions at spatial infinity from a Hamiltonian point of view*, Topological Properties and Global Structure of Space–Time (P. Bergmann and V. de Sabbata, eds.), Plenum Press, New York, 1986, pp. 49–59, URL `http://www.phys.univ-tours.fr/~piotr/scans`.

[28] P.T. Chruściel, *On uniqueness in the large of solutions of Einstein equations ("Strong Cosmic Censorship")*, Australian National University Press, Canberra, 1991.

[29] P.T. Chruściel and E. Delay, *On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications*, Mém. Soc. Math. de France. **94** (2003), 1–103, gr-qc/0301073v2.

[30] P.T. Chruściel, E. Delay, G. Galloway, and R. Howard, *Regularity of horizons and the area theorem*, Annales Henri Poincaré **2** (2001), 109–178, gr-qc/0001003.

[31] P.T. Chruściel and G.J. Galloway, *A poor man's positive energy theorem*, Class. Quantum Grav. **21** (2004), L59–L63, gr-qc/0402106.

[32] P.T. Chruściel and M. Herzlich, *The mass of asymptotically hyperbolic Riemannian manifolds*, Pacific Jour. Math. **212** (2003), 231–264, dg-ga/0110035.

[33] P.T. Chruściel and J. Isenberg, *Non–isometric vacuum extensions of vacuum maximal globally hyperbolic space–times*, Phys. Rev. **D48** (1993), 1616–1628.

[34] P.T. Chruściel, J. Jezierski, and J. Kijowski, *Hamiltonian field theory in the radiating regime*, Lect. Notes in Physics, vol. m70, Springer, Berlin, Heidelberg, New York, 2001, URL `http://www.phys.univ-tours.fr/~piotr/papers/hamiltonian_structure`.

[35] P.T. Chruściel and O. Lengard, *Solutions of Einstein equations polyhomogeneous at Scri*, in preparation.

[36] P.T. Chruściel and R.M. Wald, *Maximal hypersurfaces in stationary asymptotically flat space–times*, Commun. Math. Phys. **163** (1994), 561–604, gr–qc/9304009.

[37] J. Corvino, *Scalar curvature deformation and a gluing construction for the Einstein constraint equations*, Commun. Math. Phys. **214** (2000), 137–189.

[38] J. Corvino and R. Schoen, *Vacuum spacetimes which are identically Schwarzschild near spatial infinity*, talk given at the Santa Barbara Conference on Strong Gravitational Fields, June 22-26, 1999, `http://doug-pc.itp.ucsb.edu/online/gravity_c99/schoen/`.

[39] V.I. Denisov and V.O. Solov′ev, *Energy defined in general relativity on the basis of the traditional Hamiltonian approach has no physical meaning*, Teoret. Mat. Fiz. **56** (1983), 301–314.

[40] A. Einstein, *Die Grundlagen der allgemeinen Relativitätstheorie*, Ann. Phys. **49** (1916), 769–822.

[41] J.-H. Eschenburg, *Comparison theorems and hypersurfaces*, Manuscripta Math. **59** (1987), no. 3, 295–323.

[42] J.-H. Eschenburg, *The splitting theorem for space–times with strong energy condition*, Jour. Diff. Geom. **27** (1988), 477–491.

[43] J.-H. Eschenburg and G.J. Galloway, *Lines in space-times*, Commun. Math. Phys. **148** (1992), 209–216.

[44] L.C. Evans and R.F. Gariepy, *Measure theory and fine properties of functions*, CRC Press, Boca Raton, FL, 1992.

[45] Y. Fourès-Bruhat, *Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires*, Acta Math. **88** (1952), 141–225.

[46] H. Friedrich, *On the existence of n–geodesically complete or future complete solutions of Einstein's field equations with smooth asymptotic structure*, Commun. Math. Phys. **107** (1986), 587–609.

[47] ______, *Einstein's equation and geometric asymptotics*, Gravitation and Relativity: At the turn of the Millennium, N. Dahdich and J. Narlikar (eds.), IUCAA, Pune, 1998, Proceedings of GR15, pp. 153–176.

[48] ______, *Gravitational fields near space-like and null infinity*, Jour. Geom. Phys. **24** (1998), 83–163.

[49] G.J. Galloway, *The Lorentzian splitting theorem without the completeness assumption*, Jour. Diff. Geom. **29** (1989), 373–387.

[50] ______, *Maximum principles for null hypersurfaces and null splitting theorems*, Stockholm preprint, 1999.

[51] ______, *Maximum principles for null hypersurfaces and null splitting theorems*, Ann. Henri Poincaré **1** (2000), 543–567.

[52] Gregory J. Galloway, *Some results on Cauchy surface criteria in Lorentzian geometry*, Illinois Jour. Math. **29** (1985), 1–10.

[53] R. Geroch, *Domain of dependence*, Jour. Math. Phys. **11** (1970), 437–449.

[54] G.W. Gibbons, S.W. Hawking, G.T. Horowitz, and M.J. Perry, *Positive mass theorem for black holes*, Commun. Math. Phys. **88** (1983), 295–308.

[55] Domenico Giulini, *Close encounters of black holes*, (2003), gr-qc/0302111.

[56] R. Glassey and M. Tsutsumi, *On uniqueness of weak solutions to semilinear wave equations*, Commun. Partial Differential Equations **7** (1982), no. 2, 153–195.

[57] Steven G. Harris, *What is the shape of space in a spacetime?*, Greene, Robert (ed.) et al., Differential geometry. Part 2: Geometry in mathematical physics and related topics. Proceedings of a summer research institute, held at the University of California, Los Angeles, CA, USA, July 8-28, 1990. Providence, RI: American Mathematical Society. Proc. Symp. Pure Math. 54, Part 2, 287-296 , 1993.

[58] S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press, Cambridge, 1973.

[59] E. Hebey, *Sobolev spaces on Riemannian manifolds*, Lecture Notes in Mathematics, vol. 1635. Berlin: Springer, 1996.

[60] U. Heilig, *On the existence of rotating stars in general relativity*, Commun. Math. Phys. **166** (1995), 457–493.

[61] M. Herzlich, *A Penrose-like inequality for the mass on Riemannian asymptotically flat manifolds*, Commun. Math. Phys. **188** (1997), 121–133.

[62] H. Hopf and W. Rinow, *Ueber den Begriff der vollstäendingen differentialgeometrischen Fläche*, Comment. Math. Helv. **3** (1931), 209–225.

[63] G. Huisken and T. Ilmanen, *The inverse mean curvature flow and the Riemannian Penrose inequality*, Jour. Diff. Geom. **59** (2001), 353–437, URL `http://www.math.nwu.edu/~ilmanen`.

[64] J. Isenberg, *Constant mean curvature solutions of the Einstein constraint equations on closed manifolds*, Class. Quantum Grav. **12** (1995), 2249–2274.

[65] J. Isenberg, *The constraint problem for Einstein's equations*, Living Reviews in Relativity (1998), URL `http://www.livingreviews.org`.

[66] J. Isenberg, R. Mazzeo, and D. Pollack, *Gluing and wormholes for the Einstein constraint equations*, Commun. Math. Phys. **231** (2002), 529–568, gr-qc/0109045.

[67] ______, *On the topology of vacuum spacetimes*, Annales Henri Poincaré **4** (2003), 369–383, gr-qc/0206034.

[68] J. Isenberg and V. Moncrief, *Some results on non–constant mean curvature solutions of the Einstein constraint equations*, Physics on Manifolds (M. Flato, R. Kerner, and A. Lichnerowicz, eds.), Kluwer Academic Publishers, Dordrecht, 1994, Y. Choquet–Bruhat Festschrift, pp. 295–302.

[69] J. Isenberg and J. Park, *Asymptotically hyperbolic non-constant mean curvature solutions of the Einstein constraint equations*, Classical Quantum Grav. **14** (1997), A189–A201.

[70] D. Joyce, *Constant scalar curvature metrics on connected sums*, Int. Jour. Math. Math. Sci. (2003), 405–450, math.DG/0108022.

[71] S. Klainerman and F. Nicolò, *On local and global aspects of the Cauchy problem in general relativity*, Class. Quantum Grav. **16** (1999), R73–R157.

[72] S. Klainerman and I. Rodnianski, *The causal structure of microlocalized Einstein metrics*, (2001), math.ap/0109174.

[73] ______, *Rough solution for the Einstein vacuum equations*, (2001), math.ap/0109173.

[74] V.I. Kondrashov, *On certain estimates of families of functions satisfying integral inequalities*, Dokl. Akad. Nauk SSSR **48** (1945), 565–566 (Russian).

[75] M. Kriele, *Spacetime*, Lecture Notes in Physics. New Series m: Monographs, vol. 59, Springer-Verlag, Berlin, 1999, Foundations of general relativity and differential geometry.

[76] D.N. Kupeli, *On null submanifolds in spacetimes*, Geom. Dedicata **23** (1987), 33–51.

[77] J. Lelong-Ferrand, *Geometrical interpretations of scalar curvature and regularity of conformal homeomorphisms* , Differ. Geom. Relativ., Vol. Honour A. Lichnerowicz 60th Birthday, 1976, pp. 91–105.

[78] A. Lichnerowicz, *L'intégration des équations de la gravitation relativiste et le problème des n corps*, Jour. Math. Pures Appli. **23** (1944), 37–63.

[79] J. Lohkamp, *Scalar curvature and hammocks*, Math. Ann. **313** (1999), 385–407.

[80] L. Markus, *Line element fields and lorentz structures on differentiable manifolds*, Annals of Math. **62** (1955), 411–417.

[81] J. Milnor, *Morse theory*, Annals of Mathematics Studies, vol. 51, Princeton Univ. Press, 1963.

[82] C.W. Misner, *Taub–NUT space as a counterexample to almost anything*, Relativity Theory and Astrophysics, AMS, Providence, Rhode Island, 1967, Lectures in Appl. Math., vol. 8, pp. 160–169.

[83] C.W. Misner, K. Thorne, and J.A. Wheeler, *Gravitation*, Freeman, San Fransisco, 1973.

[84] U. Mueller, C. Schubert, and A.E.M. van de Ven, *A closed formula for the Riemann normal coordinate expansion*, Gen. Rel. Gravitation **31** (1999), 1759–1781.

[85] N. Ó Murchadha, *Total energy momentum in GR*, Jour. Math. Phys. **27** (1986), 2111–2128.

[86] R.P.A.C. Newman, *A proof of the splitting conjecture of S.-T. Yau*, Jour. Diff. Geom. **31** (1990), 163–184.

[87] B. O'Neill, *Semi–Riemannian geometry*, Academic Press, New York, 1983.

[88] L.K. Patel, R. Tikekar, and N. Dadhich, *Higher-dimensional analogue of McVittie solution*, Grav. Cosmol. **6** (2000), 335–336, gr-qc/9909069.

[89] R. Penrose, *Techniques of differential topology in relativity*, SIAM, Philadelphia, 1972, (Regional Conf. Series in Appl. Math., vol. 7).

[90] R. Penrose and W. Rindler, *Spinors and space-time, vol. i,ii*, Cambridge UP, 1984,1986.

[91] R. Penrose, R.D. Sorkin, and E. Woolgar, *A positive mass theorem based on the focusing and retardation of null geodesics*, (1993), gr-qc/9301015.

[92] V. Perlick, *On Fermat's principle in general relativity: I. The general case*, Class. Quantum Grav. **7** (1990), 1319–1331.

[93] H.L. Royden, *Real Analysis*, McMillan, New York, 1988.

[94] W. Rudin, *Analyse réelle et complexe*, Dunod, Paris, 1998.

[95] R. Schoen and S.-T. Yau, *On the proof of the positive mass conjecture in general relativity*, Commun. Math. Phys. **65** (1979), 45–76.

[96] ———, *Proof of the positive mass theorem II*, Commun. Math. Phys. **79** (1981), 231–260.

[97] H.J. Seifert, *Smoothing and extending cosmic time functions*, Gen. Rel. Grav. **8** (77), 815–831.

[98] Y. Shi and L.-F. Tam, *Positive mass theorem and the boundary behaviors of compact manifolds with nonnegative scalar curvature*, Jour. Diff. Geom. **62** (2002), 79–125, math.DG/0301047.

[99] K. Shiohama, *Total curvature and minimal area of complete open surfaces*, Proc. Am. Math. Soc. **94** (1985), 310–316.

[100] W. Simon, *The multipole expansion of stationary Einstein-Maxwell fields*, Jour. Math. Phys. **25** (1984), 1035–1038.

[101] B. Smith and G. Weinstein, *On the connectedness of the space of initial data for the Einstein equations*, Electron. Res. Announc. Am. Math. Soc. **6** (2000), 52–63.

[102] H. Smith and D. Tataru, *Sharp local well-posedness results for the nonlinear wave equation*, (2001), `http://www.math.berkeley.edu/~tataru/nlw.html`.

[103] M.E. Taylor, *Partial differential equations*, Springer, New York, Berlin, Heidelberg, 1996.

[104] T.Y. Thomas, *The differential invariants of generalized spaces*, Cambridge University Press, 1934.

[105] R.M. Wald, *General relativity*, University of Chicago Press, Chicago, Ill., 1984.

[106] J. Weidmann, *Linear operators in Hilbert spaces*, Graduate Texts in Mathematics, vol. 68, Springer Verlag, New York, Heidelberg, Berlin, 1980.

[107] E. Witten, *A simple proof of the positive energy theorem*, Commun. Math. Phys. **80** (1981), 381–402.

[108] E. Witten, *A simple proof of the positive energy theorem*, Commun. Math. Phys. **80** (1981), 381–402.

[109] J.W. York, *Energy and momentum of the gravitational field*, Essays in General Relativity (A Festschrift for Abraham Taub) (F.J. Tipler, ed.), Academic Press, New York, 1980, pp. 39–58.