A brief introduction to Semi-Riemannian geometry and general relativity

Hans Ringström

April 18, 2017

Contents

Chapter 1

Scalar product spaces

A semi-Riemannian manifold (M, g) is a manifold M with a metric g. A smooth covariant 2-tensor field g is a metric if it induces a scalar product on T_pM for each $p \in M$. Before proceeding to the subject of semi-Riemannian geometry, it is therefore necessary to define the notion of a scalar product on a vector space and to establish some of the basic properties of scalar products.

1.1 Scalar products

Definition 1. Let V be a finite dimensional real vector space and let g be a bilinear form on V (i.e., an element of $L(V, V; \mathbb{R})$). Then g is called a *scalar product* if the following conditions hold:

- g is *symmetric*; i.e., $g(v, w) = g(w, v)$ for all $v, w \in V$.
- g is *non-degenerate*; i.e., $g(v, w) = 0$ for all w implies that $v = 0$.

A vector space V with a scalar product g is called a *scalar product space.*

Remark 2. Since a scalar product space is a vector space V with a scalar product g, it is natural to write it (V, g). However, we sometimes, in the interest of brevity, simply write V.

The two basic examples are the Euclidean scalar product and the Minkowski scalar product.

Example 3. The *Euclidean scalar product* on $\mathbb{R}^n$, $1 \le n \in \mathbb{Z}$, here denoted g_{Eucl}, is defined as follows. If $v = (v^1, \ldots, v^n)$ and $w = (w^1, \ldots, w^n)$ are two elements of $\mathbb{R}^n$, then

$$g_{\text{Eucl}}(v, w) = \sum_{i=1}^{n} v^i w^i.$$

The vector space $\mathbb{R}^n$ equipped with the Euclidean scalar product is called the (n-dimensional) *Euclidean scalar product space.* The *Minkowski scalar product* on $\mathbb{R}^{n+1}$, $1 \le n \in \mathbb{Z}$, here denoted g_{Min}, is defined as follows. If $v = (v^0, v^1, \ldots, v^n)$ and $w = (w^0, w^1, \ldots, w^n)$ are two elements of $\mathbb{R}^{n+1}$, then

$$g_{\text{Min}}(v, w) = -v^0 w^0 + \sum_{i=1}^{n} v^i w^i.$$

The vector space $\mathbb{R}^{n+1}$ equipped with the Minkowski scalar product is called the (n+1-dimensional) *Minkowski scalar product space.*

In order to distinguish between different scalar products, it is convenient to introduce the notion of an index.

Definition 4. Let (V, g) be a scalar product space. Then the *index*, say ι, of g is the largest integer that is the dimension of a subspace $W \subseteq V$ on which g is negative definite.

As in the case of Euclidean geometry, it is in many contexts convenient to use particular bases, such as an *orthonormal basis*; in other words, a basis $\{e_i\}$ such that $g(e_i, e_j) = 0$ for $i \neq j$ and $g(e_i, e_i) = \pm 1$ (no summation on i).

Lemma 5. *Let (V, g) be a scalar product space. Then there is an integer $d \leq n := \dim V$ and a basis $\{e_i\}$, $i = 1, \dots, n$, of V such that*

- $g(e_i, e_j) = 0$ *if* $i \neq j$.
- $g(e_i, e_i) = -1$ *if* $i \leq d$.
- $g(e_i, e_i) = 1$ *if* $i > d$.

Moreover, if $\{e_i\}$ is a basis satisfying these three properties for some $d \leq n$, then d equals the index of g.

Proof. Let $\{v_i\}$ be a basis for V and let $g_{ij} = g(v_i, v_j)$. If G is the matrix with components g_{ij}, then G is a symmetric matrix. There is thus an orthogonal matrix T so that TGT^t is diagonal. If T_{ij} are the components of T, then the ij'th component of TGT^t is given by

$$\sum_{k,l} T_{ik} G_{kl} T_{jl} = \sum_{k,l} T_{ik} g(v_k, v_l) T_{jl} = g\left(\sum_k T_{ik} v_k, \sum_l T_{jl} v_l\right).$$

Introducing the basis $\{w_i\}$ according to

$$w_i = \sum_k T_{ik} v_k,$$

it thus follows that $g(w_i, w_j) = 0$ if $i \neq j$. Due to the non-degeneracy of the scalar product, $g(w_i, w_i) \neq 0$. We can thus define a basis $\{E_i\}$ according to

$$E_i = \frac{1}{|g(w_i, w_i)|^{1/2}} w_i.$$

Then $g(E_i, E_i) = \pm 1$. By renumbering the E_i, one obtains a basis with the properties stated in the lemma.

If g is definite, the last statement of the lemma is trivial. Let us therefore assume that $0 < d < n$. Clearly, the index ι of g satisfies $\iota \geq d$. In order to prove the opposite inequality, let W be a subspace of V such that g is negative definite on W and such that $\dim W = \iota$. Let N be the subspace of V spanned by $\{e_i\}$, $i = 1, \dots, d$, and $\varphi : W \to N$ be the map defined by

$$\varphi(w) = -\sum_{i=1}^{d} g(w, e_i) e_i.$$

If φ is injective, the desired conclusion follows. Moreover,

$$w = -\sum_{i=1}^{d} g(w, e_i) e_i + \sum_{i=d+1}^{n} g(w, e_i) e_i; \tag{1.1}$$

this equality is a consequence of the fact that if we take the scalar product of e_i with the left hand side minus the right hand side, then the result is zero for all i (so that non-degeneracy implies that (1.1) holds). If $\varphi(w) = 0$, we thus have

$$w = \sum_{i=d+1}^{n} g(w, e_i) e_i.$$

Compute

$$g(w,w) = \sum_{i,j=d+1}^{n} g(w,e_i)g(w,e_j)g(e_i,e_j) = \sum_{i=d+1}^{n} g(w,e_i)^2 \geq 0.$$

Since g is negative definite on W, this implies that $w = 0$. Thus φ is injective, and the lemma follows. □

Let g and h be a scalar products on V and W respectively. A linear map $T : V \to W$ is said to *preserve scalar products* if $h(Tv_1, Tv_2) = g(v_1, v_2)$. If T preserves scalar products, then it is injective (exercise). A linear isomorphism $T : V \to W$ that preserves scalar products is called a *linear isometry*.

Lemma 6. *Scalar product spaces V and W have the same dimension and index if and only if there exists a linear isometry from V to W.*

Exercise 7. Prove Lemma 6.

1.2 Orthonormal bases adapted to subspaces

Two important special cases of the notion of a scalar product space are the following.

Definition 8. A scalar product with index 0 is called a *Riemannian scalar product* and a vector space with a Riemannian scalar product is called *Riemannian scalar product space*. A scalar product with index 1 is called a *Lorentz scalar product* and a vector space with a Lorentz scalar product is called *Lorentz scalar product space*.

If V is an n-dimensional Riemannian scalar product space, then there is a linear isometry from V to the n-dimensional Euclidean scalar product space. If V is an $n+1$-dimensional Lorentz scalar product space, then there is a linear isometry from V to the $n+1$-dimensional Minkowski scalar product space. Due to this fact, and the fact that the reader is assumed to be familiar with Euclidean geometry, we here focus on the Lorentz setting.

In order to understand Lorentz scalar product spaces better, it is convenient to make a few more observations of a linear algebra nature. To begin with, if (V, g) is a scalar product space and W is a subspace of V, then

$$W^\perp = \{v \in V : g(v,w) = 0 \ \ \forall w \in W\}.$$

In contrast with the Riemannian setting, $W + W^\perp$ does not equal V in general.

Exercise 9. Give an example of a Lorentz scalar product space (V, g) and a subspace W of V such that $W + W^\perp \neq V$.

On the other hand, we have the following result.

Lemma 10. *Let W be a subspace of a scalar product space V. Then*

1. $\dim W + \dim W^\perp = \dim V$.
2. $(W^\perp)^\perp = W$.

Exercise 11. Prove Lemma 10.

Another useful observation is the following.

Exercise 12. Let W be a subspace of a scalar product space V. Then

$$\dim(W + W^\perp) + \dim(W \cap W^\perp) = \dim W + \dim W^\perp. \tag{1.2}$$

Even though $W + W^{\perp}$ does not in general equal V, it is of interest to find conditions on W such that the relation holds. One such condition is the following.

Definition 13. Let W be a subspace of a scalar product space V. Then W is said to be *non-degenerate* if $g|_W$ is non-degenerate.

We then have the following observation.

Lemma 14. *Let W be a subspace of a scalar product space V. Then W is non-degenerate if and only if $V = W + W^{\perp}$.*

Proof. Due to Lemma 10 and (1.2), it is clear that $W + W^{\perp} = V$ if and only if $W \cap W^{\perp} = \{0\}$. However, $W \cap W^{\perp} = \{0\}$ is equivalent to W being non-degenerate. □

One important consequence of this observation is the following.

Corollary 15. *Let W_1 be a subspace of a scalar product space (V, g). If W_1 is non-degenerate, then $W_2 = W_1^{\perp}$ is also non-degenerate. Thus W_i, $i = 1, 2$, are scalar product spaces with indices ι_i; the scalar product on W_i is given by $g_i = g|_{W_i}$. If ι is the index of V, then $\iota = \iota_1 + \iota_2$. Moreover, there is an orthonormal basis $\{e_i\}$, $i = 1, \dots, n$, of V which is adapted to W_1 and W_2 in the sense that $\{e_i\}$, $i = 1, \dots, d$, is a basis for W_1 and $\{e_i\}$, $i = d + 1, \dots, n$, is a basis for W_2.*

Proof. Since W_1 is non-degenerate and $W_2^{\perp} = W_1$ (according to Lemma 10), Lemma 14 implies that

$$V = W_1 + W_2 = W_2^{\perp} + W_2.$$

Applying Lemma 14 again implies that W_2 is non-degenerate. Defining g_i as in the statement of the corollary, it is clear that (W_i, g_i), $i = 1, 2$, are scalar product spaces. Due to Lemma 5, we know that each of these scalar product spaces have an orthonormal basis. Let $\{e_i\}$, $i = 1, \dots, d$, be an orthonormal basis for W_1 and $\{e_i\}$, $i = d + 1, \dots, n$, be an orthonormal basis for W_2. Then $\{e_i\}$, $i = 1, \dots, n$, is an orthonormal basis of V. Since ι_1 equals the number of elements of $\{e_i\}$, $i = 1, \dots, d$, with squared norm equal to -1, and similarly for ι_2 and ι, it is clear that $\iota = \iota_1 + \iota_2$. □

1.3 Causality for Lorentz scalar product spaces

One important notion in Lorentz scalar product spaces is that of causality, or causal character of a vector.

Definition 16. Let (V, g) be a Lorentz scalar product space. Then a vector $v \in V$ is said to be

1. *timelike* if $g(v, v) < 0$,
2. *spacelike* if $g(v, v) > 0$ or $v = 0$,
3. *lightlike* or *null* if $g(v, v) = 0$ and $v \neq 0$.

The classification of a vector $v \in V$ according to the above is called the *causal character* of the vector v.

The importance of this terminology stems from its connection to the notion of causality in physics. According to special relativity, no information can travel faster than light. Assuming γ to be a curve in the Minkowski scalar product space (γ should be thought of as the trajectory of a physical object; a particle, a spacecraft, light etc.), the speed of the corresponding object relative to that of light is characterized by the causal character of $\dot{\gamma}$ with respect to the Minkowski scalar product.

If $\dot{\gamma}$ is timelike, the speed is strictly less than that of light, if $\dot{\gamma}$ is lightlike, the speed equals that of light.

In Minkowski space, if $v = (v^0, \bar{v}) \in \mathbb{R}^{n+1}$, where $\bar{v} \in \mathbb{R}^n$, then

$$g(v,v) = -(v^0)^2 + |\bar{v}|^2,$$

where $|\bar{v}|$ denotes the usual norm of an element $\bar{v} \in \mathbb{R}^n$. Thus v is timelike if $|v^0| > |\bar{v}|$, lightlike if $|v^0| = |\bar{v}| \neq 0$ and spacelike if $|v^0| < |\bar{v}|$ or $v = 0$. The set of timelike vectors consists of two components; the vectors with $v^0 > |\bar{v}|$ and the vectors with $-v^0 > |\bar{v}|$. Choosing one of these components corresponds to a choice of so-called time orientation (a choice of what is the future and what is the past). Below we justify these statements and make the notion of a time orientation more precise. However, to begin with, it is convenient to introduce some additional terminology.

Definition 17. Let (V, g) be a scalar product space and $W \subseteq V$ be a subspace. Then W is said to be *spacelike* if $g|_W$ is positive definite; i.e., if $g|_W$ is nondegenerate of index 0. Moreover, W is said to be *lightlike* if $g|_W$ is degenerate. Finally, W is said to be *timelike* if $g|_W$ is nondegenerate of index 1.

It is of interest to note the following consequence of Corollary 15.

Lemma 18. *Let (V, g) be a Lorentz scalar product space and $W \subseteq V$ be a subspace. Then W is timelike if and only if $W^{\perp}$ is spacelike.*

Remark 19. The words timelike and spacelike can be interchanged in the statement.

Let (V, g) be a Lorentz scalar product space. If $u \in V$ is a timelike vector, the *timecone* of V containing u, denoted $C(u)$, is defined by

$$C(u) = \{v \in V : g(v,v) < 0,\ g(v,u) < 0\}.$$

The *opposite timecone* is defined to be $C(-u)$. Note that $C(-u) = -C(u)$. If $v \in V$ is timelike, then v has to belong to $C(u)$ or $C(-u)$. The reason for this is that $(\mathbb{R}u)^{\perp}$ is spacelike; cf. Lemma 18. The following observation will be of importance in the discussion of the existence of Lorentz metrics.

Lemma 20. *Let (V, g) be a Lorentz scalar product space and $v, w \in V$ be timelike vectors. Then v and w are in the same timecone if and only if $g(v, w) < 0$.*

Proof. Consider a timecone $C(u)$ (where we, without loss of generality, can assume that u is a unit timelike vector). Due to Corollary 15, there is an orthonormal basis $\{e_\alpha\}$, $\alpha = 0, \dots, n$, of V such that $e_0 = u$. Then $v \in C(u)$ if and only if $v^0 > 0$, where $v = v^\alpha e_\alpha$. Note also that if $x = x^\alpha e_\alpha$, then x is timelike if and only if $|x^0| > |\bar{x}|$, where $\bar{x} = (x^1, \dots, x^n)$ and $|\bar{x}|$ denotes the ordinary Euclidean norm of $\bar{x} \in \mathbb{R}^n$.

Let v and w be timelike and define v^α, w^α, $\bar{v}$ and $\bar{w}$ in analogy with the above. Compute

$$g(v,w) = -v^0 w^0 + \bar{v} \cdot \bar{w}, \tag{1.3}$$

where $\cdot$ denotes the ordinary dot product on $\mathbb{R}^n$. Since v and w are timelike, $|v^0| > |\bar{v}|$ and $|w^0| > |\bar{w}|$, so that

$$|\bar{v} \cdot \bar{w}| \leq |\bar{v}||\bar{w}| < |v^0 w^0|.$$

Thus the first term on the right hand side of (1.3) is bigger in absolute value than the second term. In particular, $g(v, w) < 0$ if and only if v^0 and w^0 have the same sign.

Assume that v and w are in the same timecone; say $C(u)$. Then $v^0, w^0 > 0$, so that $g(v, w) < 0$ by the above. Assume that $g(v, w) < 0$ and fix a timelike unit vector u. Then v^0 and w^0 have the same sign by the above. If both are positive, $v, w \in C(u)$. If both are negative, $v, w \in C(-u)$. In particular, v, w are in the same timecone. The lemma follows. □

As a consequence of Lemma 20, timecones are convex; in fact, if $0 \leq a, b \in \mathbb{R}$ are not both zero and $v, w \in V$ are in the same timecone, then $av + bw$ is timelike and in the same timecone as v and w. In particular, it is clear that the timelike vectors can be divided into two components. A choice of *time orientation* of a Lorentz scalar product space is a choice of timecone, say $C(u)$. A Lorentz scalar product space with a time orientation is called a *time oriented Lorentz scalar product space*. Given a choice of time orientation, the timelike vectors belonging to the corresponding timecone are said to be *future oriented*. Let v be a null vector and $C(u)$ be a timecone. Then $g(v, u) \neq 0$. If $g(v, u) < 0$, then v is said to be future oriented, and if $g(v, u) > 0$, then v is said to be past oriented.

Chapter 2

Semi-Riemannian manifolds

The main purpose of the present chapter is to define the notion of a semi-Riemannian manifold and to describe some of the basic properties of such manifolds.

2.1 Semi-Riemannian metrics

To begin with, we need to define the notion of a metric.

Definition 21. Let M be a smooth manifold and g be a smooth covariant 2-tensor field on M. Then g is called a *metric* on M if the following holds:

- g induces a scalar product on T_pM for each $p \in M$.
- the index ι of the scalar product induced on T_pM by g is independent of p.

The constant index ι is called *the index of the metric g*.

Definition 22. A *semi-Riemannian manifold* is a smooth manifold M together with a metric g on M.

Two important special cases are Riemannian and Lorentz manifolds.

Definition 23. Let (M, g) be a semi-Riemannian manifold. If the index of g is 0, the metric is called *Riemannian*, and (M, g) is called a *Riemannian manifold*. If the index equals 1, the metric is called a *Lorentz metric*, and (M, g) is called a *Lorentz manifold*.

Let (M, g) be a semi-Riemannian manifold. If (x^i) are local coordinates, the corresponding components of g are given by

$$g_{ij} = g(\partial_{x^i}, \partial_{x^j}),$$

and g can be written

$$g = g_{ij} dx^i \otimes dx^j.$$

Since g_{ij} are the components of a non-degenerate matrix, there is a matrix with components g^{ij} such that

$$g^{ij} g_{jk} = \delta^i_k.$$

Note that the functions g^{ij} are smooth, whenever they are defined. Moreover, $g^{ij} = g^{ji}$. In fact, g^{ij} are the components of a smooth, symmetric contravariant 2-tensor field. As will become clear, this construction is of central importance in many contexts.

Again, the basic examples of metrics are the Euclidean metric and the Minkowski metric.

Definition 24. Let (x^i), $i = 1, \dots, n$, be the standard coordinates on $\mathbb{R}^n$. Then the *Euclidean metric* on $\mathbb{R}^n$, denoted g_{E}, is defined as follows. Let $(v^1, \dots, v^n), (w^1, \dots, w^n) \in \mathbb{R}^n$ and

$$v = v^i \left.\frac{\partial}{\partial x^i}\right|_p \in T_p\mathbb{R}^n, \quad w = w^i \left.\frac{\partial}{\partial x^i}\right|_p \in T_p\mathbb{R}^n.$$

Then

$$g_{\mathrm{E}}(v, w) = \sum_{i=1}^{n} v^i w^i.$$

Let (x^α), $\alpha = 0, \dots, n$, be the standard coordinates on $\mathbb{R}^{n+1}$. Then the *Minkowski metric* on $\mathbb{R}^{n+1}$, denoted g_{M}, is defined as follows. Let $(v^0, \dots, v^n), (w^0, \dots, w^n) \in \mathbb{R}^{n+1}$ and

$$v = v^\alpha \left.\frac{\partial}{\partial x^\alpha}\right|_p \in T_p\mathbb{R}^{n+1}, \quad w = w^\alpha \left.\frac{\partial}{\partial x^\alpha}\right|_p \in T_p\mathbb{R}^{n+1}.$$

Then

$$g_{\mathrm{M}}(v, w) = -v^0 w^0 + \sum_{i=1}^{n} v^i w^i.$$

In Section 2.7 we discuss the relevance of these metrics.

2.2 Pullback, isometries and musical isomorphisms

Let M and N be smooth manifolds and h be a semi-Riemannian metric on N. If $F : M \to N$ is a smooth map, F^*h is smooth symmetric covariant 2-tensor field. However, it is not always a semi-Riemannian metric. If h is a Riemannian metric, then F^*h is a Riemannian metric if and only if F is a smooth immersion; cf. [3, Proposition 13.9, p 331]. However, if h is a Lorentz metric, F^*h need not be a Lorentz metric even if F is a smooth immersion (on the other hand, it is necessary for F to be a smooth immersion in order for F^*h to be a Lorentz metric).

Exercise 25. Give an example of a smooth manifold M, a Lorentz manifold (N, h) and a smooth immersion $F : M \to N$ such that F^*h is not a semi-Riemannian metric on M.

Due to this complication, the definition of a semi-Riemannian submanifold is slightly different from that of a Riemannian submanifold; cf. [3, p. 333].

Definition 26. Let S be a submanifold of a semi-Riemannian manifold (M, g) with inclusion $\iota : S \to M$. If $\iota^* g$ is a metric on S, then S, equipped with this metric, is called a *semi-Riemannian submanifold* of (M, g). Moreover, the metric $\iota^* g$ is called the *induced metric* on S.

Two fundamental examples are the following.

Example 27. Let $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ denote the n-sphere and $\iota_{\mathbb{S}^n} : \mathbb{S}^n \to \mathbb{R}^{n+1}$ the corresponding inclusion. Then the *round metric on* $\mathbb{S}^n$, $g_{\mathbb{S}^n}$, is defined by $g_{\mathbb{S}^n} = \iota^*_{\mathbb{S}^n} g_{\mathrm{E}}$; cf. Definition 24. Let H^n denote the set of $x \in \mathbb{R}^{n+1}$ such that $g_{\mathrm{Min}}(x, x) = -1$ and let $\iota_{H^n} : H^n \to \mathbb{R}^{n+1}$ denote the corresponding inclusion. Then the *hyperbolic metric on* H^n, g_{H^n}, is defined by $g_{H^n} = \iota^*_{H^n} g_{\mathrm{M}}$; cf. Definition 24.

Remark 28. Both $(\mathbb{S}^n, g_{\mathbb{S}^n})$ and (H^n, g_{H^n}) are Riemannian manifolds (we shall not demonstrate this fact in these notes; the interested reader is referred to, e.g., [2, Chapter 4] for a more detailed discussion). Note that there is a certain symmetry in the definitions: $\mathbb{S}^n$ is the set of $x \in \mathbb{R}^{n+1}$ such that $g_{\mathrm{Eucl}}(x, x) = 1$ and H^n is the set of $x \in \mathbb{R}^{n+1}$ such that $g_{\mathrm{Min}}(x, x) = -1$.

Another important example is obtained by considering a submanifold, say S, of $\mathbb{R}^n$. If $\iota : S \to \mathbb{R}^n$ is the corresponding inclusion, then $\iota^* g_{\mathrm{E}}$ is a Riemannian metric on S (the Riemannian metric induced by the Euclidean metric). If S is oriented, there is also a way to define a Euclidean notion of volume of S (in specific cases, it may of course be more natural to speak of length or area). In order to justify this observation, note, first of all, that on an oriented Riemannian manifold (M, g), there is a (uniquely defined) Riemannian volume form, say ω_g; cf. [3, Proposition 15.29]. The Riemannian volume of (M, g) is then given by

$$\mathrm{Vol}(M, g) = \int_M \omega_g,$$

assuming that this integral makes sense. If S is an oriented submanifold of $\mathbb{R}^n$, the volume of S is then defined to be the volume of the oriented Riemannian manifold $(S, \iota^* g_{\mathrm{E}})$.

A fundamental notion in semi-Riemannian geometry is that of an isometry.

Definition 29. Let (M, g) and (N, h) be semi-Riemannian manifolds and $F : M \to N$ be a smooth map. Then F is called an *isometry* if F is a diffeomorphism such that $F^* h = g$.

Remark 30. If F is an isometry, than so is F^{-1}. Moreover, the composition of two isometries is an isometry. Finally, the identity map on M is an isometry. As a consequence of these observations, the set of isometries of a semi-Riemannian manifold is a group, referred to as the *group of isometries.*

It will be useful to keep in mind that a semi-Riemannian metric induces an isomorphism between the sections of the tangent bundle and the sections of the cotangent bundle.

Lemma 31. *Let (M, g) be a semi-Riemannian manifold. If $X \in \mathfrak{X}(M)$, then $X^\flat$ is defined to be the one-form given by*

$$X^\flat(Y) = g(X, Y)$$

for all $Y \in \mathfrak{X}(M)$. The map taking X to $X^\flat$ is an isomorphism between $\mathfrak{X}(M)$ and $\mathfrak{X}^*(M)$. Moreover, this isomorphism is linear over the functions. In particular, given a one-form η, there is thus a unique $X \in \mathfrak{X}(M)$ such that $X^\flat = \eta$. The vectorfield X is denoted $\eta^\sharp$.

Remark 32. Here $\mathfrak{X}^*(M)$ denotes the smooth sections of the cotangent bundle; i.e., the one-forms.

Proof. Note that, given $X \in \mathfrak{X}(M)$, it is clear that $X^\flat$ is linear over $C^\infty(M)$. Due to the tensor characterization lemma, [3, Lemma 12.24, p. 318], is is thus clear that $X^\flat \in \mathfrak{X}^*(M)$. In addition, it is clear that the map taking X to $X^\flat$ is linear over $C^\infty(M)$.

In order to prove injectivity of the map, assume that $X^\flat = 0$. Then $g(X, Y) = 0$ for all $Y \in \mathfrak{X}(M)$. In particular, given $p \in M$, $g(X_p, v) = 0$ for all $v \in T_p M$. Due to the non-degeneracy of the metric, this implies that $X_p = 0$ for all $p \in M$. Thus $X = 0$ and the map is injective.

In order to prove surjectivity, let $\eta \in \mathfrak{X}^*(M)$. To begin with, let us try to find a vectorfield X on a coordinate neighbourhood U such that $X^\flat = \eta$ on U. If η_i are the components of η with respect to local coordinates, then we can define a vectorfield on U by

$$X = g^{ij} \eta_j \frac{\partial}{\partial x^i}.$$

In this expression, g^{ij} are the components of the inverse of the matrix with components g_{ij}. Then X is a smooth vectorfield on U. Moreover,

$$X^\flat(Y) = g(X, Y) = g_{ik} X^i Y^k = g_{ik} g^{ij} \eta_j Y^k = \delta^j_k \eta_j Y^k = \eta_j Y^j = \eta(Y).$$

Thus $X^\flat = \eta$. Due to the uniqueness, the local vectorfields can be combined to give an $X \in \mathfrak{X}(M)$ such that $X^\flat = \eta$. This proves surjectivity. □

The maps

$$\flat : \mathfrak{X}(M) \to \mathfrak{X}^*(M), \quad \sharp : \mathfrak{X}^*(M) \to \mathfrak{X}(M)$$

are sometimes referred to as *musical isomorphisms*. In the physics literature, where authors prefer to write everything in coordinates, the maps $\sharp$ and $\flat$ are referred to as raising and lowering indices using the metric; if η_i are the components of a one-form, then $g^{ij}\eta_j$ are the components of the corresponding vectorfield; if X^i are the components of a vectorfield, then $g_{ij}X^j$ are the components of the corresponding one-form. However, the musical isomorphisms are just a special case of a general construction. If A is a tensor field of mixed (k,l)-type, then we can, for example, lower one of the indices of A according to

$$g_{ii_1}A^{i_1\cdots i_k}_{j_1\cdots j_l}.$$

The result defines a tensor field of mixed $(k-1,l+1)$-type. Again, this is just a special case of a construction called *contraction*. The idea is the following. If $A^{i_1\cdots i_k}_{j_1\cdots j_l}$ are the components of a tensor field of mixed (k,l)-type with respect to local coordinates, then setting $i_r = j_s = i$ and summing over i yields a tensor field of mixed $(k-1,l-1)$-type. For example $g^{i_1i_2}\eta_{j_1}$ are the components of a tensor field of mixed $(2,1)$-type. Applying the contraction construction to i_2 and j_1 yields the components of $\eta^\sharp$.

Exercise 33. Let A be a tensor field of mixed $(2,2)$-type. The components of A with respect to local coordinates are $A^{i_1i_2}_{j_1j_2}$. Prove that $A^{i_1i}_{j_1i}$ (where Einstein's summation convention is enforced) are the components of a tensor field of mixed $(1,1)$-type.

2.3 Causal notions in Lorentz geometry

In Definition 16 we assigned a causal character to vectors in a Lorentz scalar product space. The same can be done for vectors, curves etc. in a Lorentz manifold. However, before we do so, let us discuss the notion of a time orientation in the context of Lorentz manifolds.

Definition 34. Let (M,g) be a Lorentz manifold. A *time orientation* of (M,g) is a choice of time orientation of each scalar product space (T_pM, g_p), $p \in M$, such that the following holds. For each $p \in M$, there is an open neighbourhood U of p and a smooth vectorfield X on U such that X_q is future oriented for all $q \in U$. A Lorentz metric g on a manifold M is said to be *time orientable* if (M,g) has a time orientation. A Lorentz manifold (M,g) is said to be *time orientable* if (M,g) has a time orientation. A Lorentz manifold with a time orientation is called a *time oriented Lorentz manifold*.

Remark 35. Here g_p denotes the scalar product induced on T_pM by g. The requirement that there be a local vectorfield with the properties stated in the definition is there to ensure the "continuity" of the choice of time orientation.

A choice of time orientation for a Lorentz manifold corresponds to a choice of which time direction corresponds to the future and which time direction corresponds to the past. In physics, time oriented Lorentz manifolds are of greater interest than non-time oriented ones. For this reason, the following terminology is sometimes introduced.

Definition 36. A time oriented Lorentz manifold is called a *spacetime*.

Let us now introduce some of the notions of causality that we shall use.

Definition 37. Let (M,g) be a Lorentz manifold. A vector $v \in T_pM$ is said to be *timelike*, *spacelike* or *lightlike* if it is timelike, spacelike or lightlike, respectively, with respect to the scalar product g_p induced on T_pM by g. A vector field X on M is said to be *timelike*, *spacelike* or *lightlike* if X_p is timelike, spacelike or lightlike, respectively, for all $p \in M$. A smooth curve $\gamma : I \to M$ (where I is an open interval) is said to be *timelike*, *spacelike* or *lightlike* if $\dot\gamma(t)$ is timelike, spacelike or lightlike, respectively, for all $t \in I$. A submanifold S of M is said to be *spacelike* if S is a semi-Riemannian submanifold of M such that the induced metric is Riemannian. A tangent vector which is either timelike of lightlike is said to be *causal*. The terminology concerning vectorfields and curves is analogous.

In case (M, g) is a spacetime, it is also possible to speak of future directed timelike vectors etc. Note, however, that a causal curve is said to be future directed if and only if $\dot{\gamma}(t)$ is future oriented for all t in the domain of definition of γ. Our requirements concerning vector fields is similar.

2.4 Warped product metrics

One construction which is very important in the context of general relativity is that of a so-called warped product metric.

Definition 38. Let (M_i, g_i), $i = 1, 2$, be semi-Riemannian manifolds, $\pi_i : M_1 \times M_2 \to M_i$ be the projection taking (p_1, p_2) to p_i, and $f \in C^\infty(M_1)$ be strictly positive. Then the *warped product*, denoted $M = M_1 \times_f M_2$, is the manifold $M = M_1 \times M_2$ with the metric

$$g = \pi_1^* g_1 + (f \circ \pi_1)^2 \pi_2^* g_2.$$

Exercise 39. Prove that the warped product is a semi-Riemannian manifold.

One special case of this construction is obtained by demanding that $f = 1$. In that case, the resulting warped product is referred to as a *semi-Riemannian product manifold*. One basic example of a warped product is the following.

Example 40. Let $M_1 = I$ (where I is an open interval), $M_2 = \mathbb{R}^3$, $g_1 = -dt \otimes dt$, $g_2 = g_E$ (the Euclidean metric on $\mathbb{R}^3$) and f be a smooth strictly positive function on I. Then the resulting warped product is the manifold $M = I \times \mathbb{R}^3$ with the metric

$$g = -dt \otimes dt + f^2(t) \sum_{i=1}^{3} dx^i \otimes dx^i,$$

where t is the coordinate on the first factor in $I \times \mathbb{R}^3$ and x^i, $i = 1, 2, 3$, are the coordinates on the last three factors. The geometry of most models of the universe used by physicists today are of the type (M, g). What varies from model to model is the function f.

2.5 Existence of metrics

In semi-Riemannian geometry, a fundamental question to ask is: given a manifold, is there a semi-Riemannian metric on it? In the Riemannian setting, this question has a simple answer.

Proposition 41. *Every smooth manifold with or without boundary admits a Riemannian metric.*

Proof. The proof can be found in [3, p. 329]. □

In the Lorentzian setting, the situation is more complicated.

Proposition 42. *A manifold M, $n := \dim M \geq 2$, admits a time orientable Lorentz metric if and only if there is an $X \in \mathfrak{X}(M)$ such that $X_p \neq 0$ for all $p \in M$.*

Proof. Assume that there is a nowhere vanishing smooth vectorfield X on M. Let h be a Riemannian metric on M (such a metric exists due to Proposition 41). By normalizing X if necessary, we can assume $h(X,X) = 1$. Define g according to

$$g = -2X^\flat \otimes X^\flat + h,$$

where $X^\flat$ is defined in Lemma 31. Then

$$g(X,X) = -2[X^\flat(X)]^2 + 1 = -1.$$

Given $p \in M$, let $e_2|_p, \dots, e_n|_p \in T_pM$ be such that $e_1|_p, \dots, e_n|_p$ is an orthonormal basis for (T_pM, h_p), where $e_1|_p = X_p$. Then $\{e_i|_p\}$ is an orthonormal basis for (T_pM, g_p). Moreover, it is clear that the index of g_p is 1. Thus (M,g) is a Lorentz manifold. Since we can define a time orientation by requiring X_p to be future oriented for all $p \in M$, it is clear that M admits a time orientable Lorentz metric.

Assume now that M admits a time orientable Lorentz metric g. Fix a time orientation. Let $\{U_\alpha\}$ be an open covering of M such that on each U_α there is a timelike vector field X_α which is future pointing (that such a covering exists is a consequence of the definition of a time orientation; cf. Definition 34). Let $\{\phi_\alpha\}$ be a partition of unity subordinate to the covering $\{U_\alpha\}$. Define X by

$$X = \sum_\alpha \phi_\alpha X_\alpha.$$

Fix a $p \in M$. At this point, the sum consists of finitely many terms, so that

$$X_p = \sum_{i=1}^k a_i X_{i,p},$$

where $0 < a_i \in \mathbb{R}$ and $X_{i,p} \in T_pM$ are future oriented timelike vectors. Due to Lemma 20 it is then clear that X_p is a future oriented timelike vector. In particular, X is thus a future oriented timelike vectorfield. Since such a vectorfield is nowhere vanishing, it is clear that M admits a non-zero vector field. □

It is of course natural to ask what happens if we drop the condition that (M,g) be time orientable. However, in that case there is a double cover which is time orientable (for those unfamiliar with covering spaces, we shall not make any use of this fact). It is important to note that the existence of a Lorentz metric *is a topological restriction*; not all manifolds admit Lorentz metrics. As an orientation in the subject of Lorentz geometry, it is also of interest to make the following remark (we shall not make any use of the statements made in the remark in what follows).

Remark 43. If (M,g) is a spacetime such that M is a closed manifold (in other words, M is compact and without boundary), then there is a closed timelike curve in M. In other words, there is a future oriented timelike curve γ in M such that $\gamma(t_1) = \gamma(t_2)$ for some $t_1 < t_2$ in the domain of definition of γ. This means that it is possible to travel into the past. Since this is not very natural in physics, spacetimes (M,g) such that M is closed are not very natural (in contrast with the Riemannian setting). For a proof of this statement, see [2, Lemma 10, p. 407]. In general relativity, one often requires spacetimes to satisfy an additional requirement called global hyperbolicity (which we shall not define here) which involves additional conditions concerning causality. Moreover, globally hyperbolic spacetimes (M,g), where $n+1 = \dim M$, are topologically products $M = \mathbb{R} \times \Sigma$ where Σ is an n-dimensional manifold.

2.6 Riemannian distance function

Let (M,g) be a Riemannian manifold. Then it is possible to associate a distance function

$$d : M \times M \to [0, \infty)$$

with (M, g). Since the basic properties of the Riemannian distance function are described in [3, pp. 337–341], we shall not do so here.

2.7 Relevance of the Euclidean and the Minkowski metrics

It is of interest to make some comments concerning the relevance of the Euclidean and the Minkowski metrics. The Euclidean metric gives rise to Euclidean geometry, and the relevance of this geometry is apparent in much of mathematics. For that reason, we here focus on the Minkowski metric.

Turning to Minkowski space, it is of interest to recall the origin of the special theory of relativity (for those uninterested in physics, the remainder of this section can be skipped). In special relativity, there are frames of reference (in practice, coordinate systems) which are preferred, the so-called *inertial frames*. These frames should be thought of as the "non-accelerated" frames, and two inertial frames travel at "constant velocity" relative to each other. It is of interest to relate measurements made with respect to different inertial frames. Let us consider the classical and the special relativistic perspective separately.

The classical perspective. In the classical perspective, the transformation laws are obtained by demanding that time is absolute. The relation between two inertial frames is then specified by fixing the relative velocity, an initial translation, and a rotation. More specifically, given two inertial frames F and F', there are $t_0 \in \mathbb{R}$, $v, x_0 \in \mathbb{R}^3$ and $A \in \mathrm{SO}(3)$ such that if $(t, x) \in \mathbb{R}^4$ are the time and space coordinates of an event with respect to the inertial frame F and $(t', x') \in \mathbb{R}^4$ are the coordinates of the same event with respect to the inertial frame F', then

$$t' = t + t_0, \tag{2.1}$$

$$x' = Ax + vt + x_0. \tag{2.2}$$

The corresponding transformations are referred to as the *Galilean transformations.*

The special relativistic perspective. The classical laws of physics transform well under changes of coordinates of the form (2.1)–(2.2). However, it turns out that Maxwell's equations do not. This led Einstein to use a different starting point, namely that the speed of light is the same in all inertial frames. One consequence of this assumption is that time is no longer absolute. Moreover, if one wishes to compute the associated changes of coordinates when going from one inertial frame to another, they are different from (2.1)–(2.2). The group of transformations (taking the coordinates of one inertial frame to the coordinates of another frame) that arise when taking this perspective is called the group of Lorentz transformations. The main point of introducing Minkowski space is that the group of isometries of Minkowski space are exactly the group of Lorentz transformations.

Chapter 3

The Levi-Civita connection, parallel translation and geodesics

Einstein's equations of general relativity relate the curvature of a spacetime with the matter content of the spacetime. In order to understand this equation, it is therefore important to understand the notion of curvature. This subject has a rich history, and here we only give a quite formal and brief introduction to it. One way to define curvature is to examine how a vector is changed when parallel translating it along a closed curve in the manifold. In order for this to make sense, it is of course necessary to assign a meaning to the notion of "parallel translation". In the case of Euclidean space, the notion is perhaps intuitively clear; we simply fix the components of the vector with respect to the standard coordinate frame and then change the base point. Transporting a vector in Euclidean space (along a closed curve) in this way yields the identity map; one returns to the vector one started with. This is one way to express that the curvature of Euclidean space vanishes. Using a rather intuitive notion of parallel translation on the 2-sphere, one can convince oneself that the same is not true of the 2-sphere.

In order to proceed to a formal development of the subject, it is necessary to clarify what is meant by parallel translation. One natural way to proceed is to define an "infinitesimal" version of this notion. This leads to the definition of a so-called connection.

3.1 The Levi-Civita connection

In the end we wish to define the notion of a Levi-Civita connection, but we begin by defining what a connection is in general.

Definition 44. Let M be a smooth manifold. A map $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \to \mathfrak{X}(M)$ is called a *connection* if

- $\nabla_X Y$ is linear over $C^\infty(M)$ in X,
- $\nabla_X Y$ is linear over $\mathbb{R}$ in Y,
- $\nabla_X(fY) = X(f)Y + f\nabla_X Y$ for all $X, Y \in \mathfrak{X}(M)$ and all $f \in C^\infty(M)$.

The expression $\nabla_X Y$ is referred to as the *covariant derivative* of Y with respect to X for the connection ∇.

Note that the first condition of Definition 44 means that

$$\nabla_{f_1 X_1 + f_2 X_2} Y = f_1 \nabla_{X_1} Y + f_2 \nabla_{X_2} Y$$

for all $f_i \in C^\infty(M)$, $X_i, Y \in \mathfrak{X}(M)$, $i = 1, 2$. The second condition of Definition 44 means that

$$\nabla_X(a_1 Y_1 + a_2 Y_2) = a_1 \nabla_X Y_1 + a_2 \nabla_X Y_2$$

for all $a_i \in \mathbb{R}$, $X, Y_i \in \mathfrak{X}(M)$, $i = 1, 2$. On $\mathbb{R}^n$ there is a natural connection.

Definition 45. Let (x^i), $i = 1, \ldots, n$, be the standard coordinates on $\mathbb{R}^n$. Let $X, Y \in \mathfrak{X}(\mathbb{R}^n)$ and define

$$\nabla_X Y = X(Y^i)\frac{\partial}{\partial x^i},$$

where

$$Y = Y^i \frac{\partial}{\partial x^i}.$$

Then ∇ is referred to as the *standard connection on* $\mathbb{R}^n$.

Exercise 46. Prove that the standard connection on $\mathbb{R}^n$ is a connection in the sense of Definition 44.

Let (M, g) be a semi-Riemannian manifold. Our next goal is to argue that there is preferred connection, given the metric g. However, in order to single out a preferred connection, we have to impose additional conditions. One such condition would be to require that

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z) \tag{3.1}$$

for all $X, Y, Z \in \mathfrak{X}(M)$. In what follows, it is going to be a bit cumbersome to use the notation $g(X, Y)$. We therefore define $\langle \cdot, \cdot \rangle$ by

$$\langle X, Y \rangle = g(X, Y);$$

we shall use this notation both for vectorfields and for individual vectors. With this notation, (3.1) can be written

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle. \tag{3.2}$$

A connection satisfying this requirement is said to be *metric*. However, it turns out that the condition (3.2) does not determine a unique connection. In fact, we are free to add further conditions. One such condition would be to impose that $\nabla_X Y - \nabla_Y X$ can be expressed in terms of only X and Y, without any reference to the connection. Since $\nabla_X Y - \nabla_Y X$ is antisymmetric, one such condition would be to require that

$$\nabla_X Y - \nabla_Y X = [X, Y] \tag{3.3}$$

for all $X, Y \in \mathfrak{X}(M)$. A connection satisfying this criterion is said to be *torsion free*. Remarkably, it turns out that conditions (3.2) and (3.3) uniquely determine a connection, referred to as the Levi-Civita connection.

Theorem 47. *Let (M, g) be a semi-Riemannian manifold. Then there is a unique connection ∇ satisfying (3.2) and (3.3) for all $X, Y, Z \in \mathfrak{X}(M)$. It is called the Levi-Civita connection of (M, g). Moreover, it is characterized by the Koszul formula:*

$$2\langle \nabla_X Y, Z \rangle = X\langle Y, Z \rangle + Y\langle Z, X \rangle - Z\langle X, Y \rangle - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + \langle Z, [X, Y] \rangle. \tag{3.4}$$

Proof. Assume that ∇ is a connection satisfying (3.2) and (3.3) for all $X, Y, Z \in \mathfrak{X}(M)$. Compute

$$\begin{aligned}
\langle \nabla_X Y, Z \rangle =& X\langle Y, Z \rangle - \langle Y, \nabla_X Z \rangle = X\langle Y, Z \rangle - \langle Y, [X, Z] \rangle - \langle Y, \nabla_Z X \rangle \\
=& X\langle Y, Z \rangle - \langle Y, [X, Z] \rangle - Z\langle Y, X \rangle + \langle \nabla_Z Y, X \rangle \\
=& X\langle Y, Z \rangle - \langle Y, [X, Z] \rangle - Z\langle Y, X \rangle + \langle [Z, Y], X \rangle + \langle \nabla_Y Z, X \rangle \\
=& X\langle Y, Z \rangle - Z\langle Y, X \rangle - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + Y\langle Z, X \rangle - \langle Z, \nabla_Y X \rangle \\
=& X\langle Y, Z \rangle + Y\langle Z, X \rangle - Z\langle Y, X \rangle - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle - \langle Z, [Y, X] \rangle - \langle Z, \nabla_X Y \rangle,
\end{aligned}$$

where we have applied (3.2) and (3.3). In the fourth and fifth steps, we also rearranged the terms and used the antisymmetry of the Lie bracket. Note that this equation implies that (3.4) holds. This leads to the uniqueness of the Levi-Civita connection. The reason for this is the following. Assume that ∇ and $\hat{\nabla}$ both satisfy (3.2) and (3.3). Then, due to the Koszul formula,

$$\langle \nabla_X Y - \hat{\nabla}_X Y, Z\rangle = 0$$

for all $X, Y, Z \in \mathfrak{X}(M)$. Due to Lemma 31, this implies that

$$\nabla_X Y = \hat{\nabla}_X Y$$

for all $X, Y \in \mathfrak{X}(M)$. In other words, there is at most one connection satisfying the conditions (3.2) and (3.3). Given $X, Y \in \mathfrak{X}(M)$, let $\theta_{X,Y}$ be defined by the condition that $2\theta_{X,Y}(Z)$ is given by the right hand side of (3.4). It can then be demonstrated that $\theta_{X,Y}$ is linear over $C^\infty(M)$; in other words,

$$\theta_{X,Y}(f_1 Z_1 + f_2 Z_2) = f_1 \theta_{X,Y}(Z_1) + f_2 \theta_{X,Y}(Z_2)$$

for all $X, Y, Z_i \in \mathfrak{X}(M)$, $f_i \in C^\infty(M)$, $i = 1, 2$ (we leave it as an exercise to verify that this is true). Due to the tensor characterization lemma, [3, Lemma 12.24, p. 318], it thus follows that $\theta_{X,Y}$ is a one-form. By appealing to Lemma 31, we conclude that there is a smooth vectorfield $\theta^\sharp_{X,Y}$ such that

$$\langle \theta^\sharp_{X,Y}, Z\rangle = \theta_{X,Y}(Z),$$

where the right hand side is given by the right hand side of (3.4). We define $\nabla_X Y$ by

$$\nabla_X Y = \theta^\sharp_{X,Y}.$$

Then ∇ is a function from $\mathfrak{X}(M) \times \mathfrak{X}(M)$ to $\mathfrak{X}(M)$. However, it is not obvious that it satisfies the conditions of Definition 44. Moreover, it is not obvious that it satisfies (3.2) and (3.3). In other words, there are five conditions we need to verify. Let us verify the first condition in the definition of a connection. Note, to this end, that $2\langle \nabla_{fX} Y, Z\rangle$ is given by the right hand side of (3.4), with X replaced by fX. However, a straightforward calculation shows that if you replace X by fX in (3.4), then you obtain f times the right hand side of (3.4). In other words,

$$\langle \nabla_{fX} Y, Z\rangle = \langle f\nabla_X Y, Z\rangle.$$

Due to Lemma 31, it follows that $\nabla_{fX} Y = f\nabla_X Y$. We leave it as an exercise to prove that $\nabla_X Y$ is linear in X and Y over $\mathbb{R}$, and conclude that the first two conditions of Definition 44 are satisfied. To prove that the third condition holds, compute

$$\begin{aligned}
2\langle \nabla_X(fY), Z\rangle =& X\langle fY, Z\rangle + fY\langle Z, X\rangle - Z\langle X, fY\rangle - \langle X, [fY, Z]\rangle + \langle fY, [Z, X]\rangle + \langle Z, [X, fY]\rangle \\
=& X(f)\langle Y, Z\rangle + fX\langle Y, Z\rangle + fY\langle Z, X\rangle - Z(f)\langle X, Y\rangle - fZ\langle X, Y\rangle \\
&+ Z(f)\langle X, Y\rangle - f\langle X, [Y, Z]\rangle + f\langle Y, [Z, X]\rangle + X(f)\langle Z, Y\rangle + f\langle Z, [X, Y]\rangle \\
=& 2f\langle \nabla_X Y, Z\rangle + 2\langle X(f)Y, Z\rangle = 2\langle f\nabla_X Y + X(f)Y, Z\rangle.
\end{aligned}$$

Appealing to Lemma 31 yields

$$\nabla_X(fY) = f\nabla_X Y + X(f)Y.$$

Thus the third condition of Definition 44 is satisfied. We leave it to the reader to verify that (3.2) and (3.3) are satisfied. □

Exercise 48. Prove that the connection ∇ constructed in the proof of Theorem 47 satisfies the conditions (3.2) and (3.3) for all $X, Y, Z \in \mathfrak{X}(M)$.

Let (M, g) be a semi-Riemannian manifold and ∇ be the associated Levi-Civita connection. It is of interest to express ∇ with respect to local coordinates (x^i). Introduce, to this end, the notation Γ^k_{ij} by

$$\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k,$$

where we use the short hand notation

$$\partial_i = \frac{\partial}{\partial x^i}.$$

The smooth functions Γ^k_{ij}, defined on the domain of the coordinates, are called the *Christoffel symbols*. Using the notation $g_{ij} = g(\partial_i, \partial_j)$, let us compute

$$g_{lk}\Gamma^k_{ij} = \langle \partial_l, \Gamma^k_{ij}\partial_k\rangle = \langle \nabla_{\partial_i}\partial_j, \partial_l\rangle = \frac{1}{2}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}),$$

where we have used the Koszul formula, (3.4), in the last step. Multiplying this equality with g^{ml} and summing over l yields

$$\Gamma^m_{ij} = \frac{1}{2}g^{ml}(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}). \tag{3.5}$$

Note that $\Gamma^m_{ij} = \Gamma^m_{ji}$. If $X = X^i\partial_i$ and $Y = Y^i\partial_i$ are vectorfields, we obtain

$$\begin{aligned}\nabla_X Y =& \nabla_X(Y^j\partial_j) = X(Y^j)\partial_j + Y^j\nabla_X\partial_j = X(Y^j)\partial_j + Y^jX^i\nabla_{\partial_i}\partial_j \\ =& X(Y^j)\partial_j + Y^jX^i\Gamma^k_{ij}\partial_k.\end{aligned}$$

Thus

$$\nabla_X Y = \left[X(Y^k) + \Gamma^k_{ij}X^iY^j\right]\partial_k.$$

When defining parallel transport, we shall use the following consequence of this formula.

Lemma 49. *Let (M, g) be a semi-Riemannian manifold and ∇ be the associated Levi-Civita connection. Let $v \in T_pM$ for some $p \in M$, and $Y \in \mathfrak{X}(M)$. Let $X_i \in \mathfrak{X}(M)$, $i = 1, 2$, be such that $X_{i,p} = v$. Then*

$$(\nabla_{X_1}Y)_p = (\nabla_{X_2}Y)_p.$$

This lemma justifies defining $\nabla_v Y$ in the following way.

Definition 50. Let (M, g) be a semi-Riemannian manifold and ∇ be the associated Levi-Civita connection. Let $v \in T_pM$ for some $p \in M$, and $Y \in \mathfrak{X}(M)$. Given any vectorfield $X \in \mathfrak{X}(M)$ such that $X_p = v$, define $\nabla_v Y$ by

$$\nabla_v Y = (\nabla_X Y)_p.$$

3.2 Parallel translation

At the beginning of the present chapter, we justified the introduction of the notion of a connection by the (vague) statement that it would constitute an "infinitesimal" version of a notion of parallel translation. In the present section, we wish to justify this statement by using the Levi-Civita connection to define parallel translation. To begin with, let us introduce some terminology.

Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. Then a smooth map from I to the tangent bundle of M, say X, is said to be an element of $\mathfrak{X}(\gamma)$ if the base point of $X(t)$ is $\gamma(t)$. If $X \in \mathfrak{X}(M)$, we let X_γ denote the element of $\mathfrak{X}(\gamma)$ which assigns the vector $X_{\gamma(t)}$ to the number $t \in I$.

Proposition 51. *Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. Then there is a unique function taking $X \in \mathfrak{X}(\gamma)$ to*

$$X' = \frac{\nabla X}{dt} \in \mathfrak{X}(\gamma),$$

satisfying the following properties:

$$(a_1X_1 + a_2X_2)' = a_1X_1' + a_2X_2', \tag{3.6}$$
$$(fX)' = f'X + fX', \tag{3.7}$$
$$(Y_\gamma)'(t) = \nabla_{\gamma(t)}Y, \tag{3.8}$$

for all $X, X_i \in \mathfrak{X}(\gamma)$, $a_i \in \mathbb{R}$, $f \in C^\infty(I)$ *and* $Y \in \mathfrak{X}(M)$, $i = 1, 2$. *Moreover, this map has the property that*

$$\frac{d}{dt}\langle X_1, X_2\rangle = \langle X_1', X_2\rangle + \langle X_1, X_2'\rangle. \tag{3.9}$$

Remark 52. How to interpret the expression $\nabla_{\gamma(t)}Y$ appearing in (3.8) is explained in Lemma 49 and Definition 50.

Proof. Let us begin by proving uniqueness. Let $X \in \mathfrak{X}(\gamma)$. Then we can write X as

$$X(t) = X^i(t)\partial_i|_{\gamma(t)}, \tag{3.10}$$

where the X^i are smooth functions on $X^{-1}[\pi^{-1}(U)]$ (where U is the set on which the coordinates (x^i) are defined and $\pi : TM \to M$ is the projection taking a tangent vector to its base point). Assume now that we have derivative operator satisfying (3.6)–(3.8). Applying (3.6)–(3.8) to (3.10) yields

$$X'(t) = \frac{dX^i}{dt}(t)\partial_i|_{\gamma(t)} + X^i(t)(\partial_i|_\gamma)'(t) = \frac{dX^i}{dt}(t)\partial_i|_{\gamma(t)} + X^i(t)\nabla_{\gamma'(t)}\partial_i. \tag{3.11}$$

Since the right hand side only depends on the Levi-Civita connection, we conclude that uniqueness holds.

In order to prove existence, we can define $X'(t)$ by (3.11) for $t \in X^{-1}[\pi^{-1}(U)]$. It can then be verified that the corresponding derivative operator satisfies the conditions (3.6)–(3.9); we leave this as an exercise. Due to uniqueness, these coordinate representations can be patched together to produce an element $X' \in \mathfrak{X}(\gamma)$. □

Exercise 53. Prove that the derivative operator defined by the formula (3.11) has the properties (3.6)–(3.9).

It is of interest to write down a formula for X' in local coordinates. Let (x^i) be local coordinates, $\gamma^i = x^i \circ \gamma$ and X^i be defined by (3.10). Then, since

$$\gamma'(t) = \frac{d\gamma^i}{dt}(t)\partial_i|_{\gamma(t)},$$

(3.11) implies

$$\begin{aligned} X'(t) =& \frac{dX^i}{dt}(t)\partial_i|_{\gamma(t)} + X^i(t)\frac{d\gamma^j}{dt}(t)\Gamma^k_{ji}[\gamma(t)]\partial_k|_{\gamma(t)} \\ =& \left(\frac{dX^k}{dt}(t) + X^i(t)\frac{d\gamma^j}{dt}(t)\Gamma^k_{ji}[\gamma(t)]\right)\partial_k|_{\gamma(t)}. \end{aligned} \tag{3.12}$$

Given the derivative operator of Proposition 51, we are now in a position to assign a meaning to the expression parallel translation used in the introduction to the present chapter.

Definition 54. Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. Then $X \in \mathfrak{X}(\gamma)$ is said to be *parallel along* γ if and only if $X' = 0$.

Note that, in local coordinates, the equation $X' = 0$ is a *linear equation* for the components of X; cf. (3.12). For this reason, we have the following proposition (cf. also the arguments used to prove the existence of integral curves of vectorfields).

Proposition 55. *Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. If $t_0 \in I$ and $\xi \in T_{\gamma(t_0)}M$, then there is a unique $X \in \mathfrak{X}(\gamma)$ such that $X' = 0$ and $X(t_0) = \xi$.*

Exercise 56. Prove Proposition 55.

Due to Proposition 55 we are in a position to define parallel translation along a curve. Given assumptions as in the statement of Proposition 55, let $t_0, t_1 \in I$. Then there is a map

$$P : T_{\gamma(t_0)}M \to T_{\gamma(t_1)}M$$

defined as follows. Given $\xi \in T_{\gamma(t_0)}M$, let $X \in \mathfrak{X}(\gamma)$ be such that $X' = 0$ and $X(t_0) = \xi$. Define $P(\xi) = X(t_1)$. Here P depends (only) on γ, t_0 and t_1. In some situations, it may be useful to indicate this dependence, but if these objects are clear from the context, it is convenient to simply write P. The map P is called *parallel translation along* γ from $\gamma(t_0)$ to $\gamma(t_1)$. Parallel translation has the following property.

Proposition 57. *Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. Finally, let $t_0, t_1 \in I$ and $p_i = \gamma(t_i)$, $i = 0, 1$. Then parallel translation along γ from $\gamma(t_0)$ to $\gamma(t_1)$ is a linear isometry from $T_{p_0}M$ to $T_{p_1}M$.*

Proof. We leave it to the reader to prove that parallel translation is a vector space isomorphism. In order to prove that it is an isometry, let $v, w \in T_{p_0}M$ and $V, W \in \mathfrak{X}(\gamma)$ be such that $V' = W' = 0$, $V(t_0) = v$ and $W(t_0) = w$. Then $P(v) = V(t_1)$ and $P(w) = W(t_1)$. Compute

$$\begin{aligned}\langle P(v), P(w)\rangle =& \langle V(t_1), W(t_1)\rangle = \langle V(t_0), W(t_0)\rangle + \int_{t_0}^{t_1} \frac{d}{dt}\langle V, W\rangle dt \\ =& \langle v, w\rangle + \int_{t_0}^{t_1} (\langle V', W\rangle + \langle V, W'\rangle)\, dt = \langle v, w\rangle,\end{aligned}$$

where we have used property (3.9) of the derivative operator $'$, as well as the fact that $V' = W' = 0$. The proposition follows. □

Exercise 58. Prove that parallel translation is a vector space isomorphism.

Let us analyze what parallel translation means in the case of Euclidean space and Minkowski space. Let I and γ be as in the statement of Proposition 55, where (M, g) is either Euclidean space or Minkowski space. Note that the Christoffel symbols of g_{E} and g_{M} vanish with respect to standard coordinates on $\mathbb{R}^n$ and $\mathbb{R}^{n+1}$ respectively. An element $X \in \mathfrak{X}(\gamma)$ is therefore parallel if and only if the components of X with respect to the standard coordinate vectorfields are constant (just as we stated in the introduction). In particular, the result of the parallel translation does not depend on the curve. It is of importance to note that, even though this is true in the case of Euclidean space and Minkowski space, it is not true in general.

3.3 Geodesics

A notion which is extremely important both in Riemannian geometry and in Lorentz geometry is that of a geodesic. In Riemannian geometry, geodesics are locally length minimizing curves. In the case of general relativity (Lorentz geometry), geodesics are related to the trajectories of freely falling test particles, as well as the trajectories of light.

Definition 59. Let (M, g) be a semi-Riemannian manifold, $I \subseteq \mathbb{R}$ be an open interval and $\gamma : I \to M$ be a smooth curve. Then γ is said to be a *geodesic* if $\gamma'' = 0$; in other words, if $\gamma' \in \mathfrak{X}(\gamma)$ is parallel.

Keeping (3.12) in mind, geodesics are curves which with respect to local coordinates (x^i) satisfy the equation

$$\ddot{\gamma}^k + \left(\Gamma^k_{ij} \circ \gamma\right) \dot{\gamma}^i \dot{\gamma}^j = 0, \tag{3.13}$$

where we use the notation

$$\gamma^i = x^i \circ \gamma, \quad \dot{\gamma}^i = \frac{d\gamma^i}{dt}, \quad \ddot{\gamma}^i = \frac{d^2\gamma^i}{dt^2}.$$

It is important to note that, even though (3.13) is an ODE, it is (in contrast to the equation $X' = 0$ for a fixed curve γ) a *non-linear* ODE. Due to the fact that (3.13) is an autonomous ODE for γ and the fact that the Christoffel symbols are smooth functions, it is clear that geodesics are smooth curves. Due to local existence and uniqueness results for ODE's, we have the following proposition.

Proposition 60. *Let (M, g) be a semi-Riemannian manifold, $p \in M$ and $v \in T_pM$. Then there is a unique geodesic $\gamma : I \to M$ with the properties that*

- *$I \subseteq \mathbb{R}$ is an open interval such that $0 \in I$,*
- *$\gamma'(0) = v$,*
- *I is maximal in the sense that if $\alpha : J \to M$ is a geodesic (with J an open interval, $0 \in J$ and $\alpha'(0) = v$), then $J \subseteq I$ and $\alpha = \gamma|_J$.*

Proof. Since the uniqueness is clear from the definition, let us focus on existence.

Local existence and uniqueness. To begin with, note that there is an open interval I_0 containing 0 and a unique geodesic $\beta : I_0 \to M$ such that $\beta'(0) = v$; this is an immediate consequence of applying standard results concerning ODE's to the equation (3.13). In other words, local existence and uniqueness holds.

Global uniqueness. In order to proceed, we need to prove global uniqueness. In other words, we need to prove that if I_i, $i = 0, 1$, are open intervals containing 0 and $\beta_i : I_i \to M$ are geodesics such that $\beta_i'(0) = v$, then $\beta_0 = \beta_1$ on $I_0 \cap I_1$. In order to prove this statement, let $\mathcal{A}$ be the set of $t \in I_0 \cap I_1$ such that $\beta_0'(t) = \beta_1'(t)$. Note that if $t \in \mathcal{A}$, then $\beta_0(t) = \pi \circ \beta_0'(t) = \pi \circ \beta_1'(t) = \beta_1(t)$, where $\pi : TM \to M$ is the projection taking a tangent vector to its base point. In other words, if we can prove that $\mathcal{A} = I_0 \cap I_1$, it then follows that $\beta_0 = \beta_1$ on $I_0 \cap I_1$. Since $0 \in \mathcal{A}$, it is clear that $\mathcal{A}$ is non-empty. Due to local uniqueness, $\mathcal{A}$ is open. In order to prove that $\mathcal{A}$ is closed, let $t_1 \in I_0 \cap I_1$ belong to the closure of $\mathcal{A}$. Then there is a sequence $s_j \in \mathcal{A}$ such that $s_j \to t_1$. Since $\beta_i' : I_i \to TM$ are smooth maps, it is clear that

$$\beta_1'(t_1) = \lim_{j \to \infty} \beta_1'(s_j) = \lim_{j \to \infty} \beta_0'(s_j) = \beta_0'(t_1).$$

Thus $t_1 \in \mathcal{A}$. Summing up, $\mathcal{A}$ is an open, closed and non-empty subset of $I_0 \cap I_1$. Thus $\mathcal{A} = I_0 \cap I_1$. In other words, global uniqueness holds.

Existence. Let I_a, $a \in A$, be the collection of open intervals $I_a \subseteq \mathbb{R}$ such that

- $0 \in I_a$,
- there is a geodesic $\gamma_a : I_a \to M$ such that $\gamma_a'(0) = v$.

Due to local existence, we know that this collection of intervals is non-empty. Define

$$I = \bigcup_{a \in A} I_a.$$

Then $I \subseteq \mathbb{R}$ is an open interval containing 0. Moreover, due to global uniqueness, we can define a geodesic $\gamma : I \to M$ such that $\gamma'(0) = v$; simply let $\gamma(t) = \gamma_a(t)$ for $t \in I_a$. Finally, it is clear, by definition, that I is maximal. ☐

The geodesic constructed in Proposition 60 is called the *maximal geodesic* with initial data given by $v \in T_pM$.

Exercise 61. Prove that the maximal geodesics in Euclidean space and in Minkowski space are the straight lines.

Let γ be a geodesic. Then

$$\frac{d}{dt}\langle\gamma',\gamma'\rangle = \langle\gamma'',\gamma'\rangle + \langle\gamma',\gamma''\rangle = 0.$$

In other words, $\langle\gamma',\gamma'\rangle$ is constant so that the following definition makes sense.

Definition 62. Let (M,g) be a semi-Riemannian manifold and γ be a geodesic on (M,g). Then γ is said to be *spacelike* if $\langle\gamma',\gamma'\rangle > 0$ or $\gamma' = 0$; γ is said to be *timelike* if $\langle\gamma',\gamma'\rangle < 0$; and γ is said to be *lightlike* or *null* if $\langle\gamma',\gamma'\rangle = 0$, $\gamma' \neq 0$. A geodesic which is either timelike or null is said to be *causal*

Remark 63. In a spacetime, we can also speak of future oriented timelike, null and causal curves.

In general relativity, timelike geodesics are interpreted as the trajectories of freely falling test particles and null geodesics are interpreted as the trajectories of light. In particular, in Lorentz geometry, we can think of the timelike geodesics as freely falling observers. Moreover, if $\gamma : I \to M$ is a future oriented timelike geodesic in a spacetime and $t_0 < t_1$ are elements of I, then

$$\int_{t_0}^{t_1} \left(-\langle\gamma'(t),\gamma'(t)\rangle\right)^{1/2} dt$$

is the proper time between t_0 and t_1 as measured by the observer γ. Since the integrand is constant, it is clear that if $I = (t_-, t_+)$ and $t_+ < \infty$, then the amount of proper time the observer can measure to the future is finite (there is an analogous statement concerning the past if $t_- > -\infty$). This can be thought of as saying that the observer leaves the spacetime in finite proper time. One way to interpret this is that there is a singularity in the spacetime. It is therefore of interest to analyze under what circumstances $I \neq \mathbb{R}$. To begin with, let us introduce the following terminology.

Definition 64. Let (M,g) be a semi-Riemannian manifold and $\gamma : I \to M$ be a maximal geodesic in (M,g). Then γ is said to be a *complete geodesic* if $I = \mathbb{R}$. A semi-Riemannian manifold, all of whose maximal geodesics are complete is said to be *complete*.

Euclidean space and Minkowski space are both examples of complete semi-Riemannian manifolds. On the other hand, removing one single point from Euclidean space or Minkowski space yields an incomplete semi-Riemannian manifold. In other words, the notion of completeness is very sensitive. Moreover, it is clear that in order to interpret the presence of an incomplete causal geodesic as the existence of a singularity (as is sometimes done), it is necessary to ensure that the spacetime under consideration is maximal in some natural sense. Nevertheless, trying to sort out conditions ensuring that a spacetime (which is maximal in some natural sense) is causally geodesically incomplete is a fundamental problem. Due to the work of Hawking and Penrose, spacetimes are causally geodesically incomplete under quite general circumstances. The relevant results, which are known under the name of "the singularity theorems", are discussed, e.g., in [2, Chapter 14].

3.4 Variational characterization of geodesics

Another perspective on geodesics is obtained by considering the variation of the length of curves that are close to a fixed curve. To be more precise, let (M,g) be a semi-Riemannian manifold, $t_0 < t_1$ and $\epsilon > 0$ be real numbers, and

$$\nu : [t_0,t_1] \times (-\epsilon,\epsilon) \to M. \tag{3.14}$$

The function ν should be thought of as a variation of the curve $\gamma(t) = \nu(t, 0)$. Let

$$L(s) = \int_{t_0}^{t_1} |\langle \partial_t \nu(t, s), \partial_t \nu(t, s)\rangle|^{1/2} dt.$$

A natural question to ask is: what are the curves γ such that for every variation ν (as above, with an appropriate degree of regularity and fixing the endpoints t_0 and t_1), $L'(0) = 0$? Roughly speaking, it turns out to be possible to characterize geodesics as the curves for which $L'(0)$ for all such variations. In particular, geodesics in Riemannian geometry are the locally length minimizing curves.

Here we shall not pursue this perspective further, but rather refer the interested reader to, e.g., [2, Chapter 10].

Chapter 4

Curvature

The notion of curvature arose over a long period of time; cf. [4] for some of the history. However, in the interest of brevity, we here proceed in a more formal way. As indicated at the beginning of the previous chapter, one way to define curvature is through parallel translation along a closed curve. Here we define the curvature tensor via an "infinitesimal" version of this idea.

4.1 The curvature tensor

Proposition 65. *Let (M, g) be a semi-Riemannian manifold and ∇ denote the associated Levi-Civita connection. Then the function $R : \mathfrak{X}(M)^3 \to \mathfrak{X}(M)$ defined by*

$$R_{XY}Z = \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z \tag{4.1}$$

is linear over $C^\infty(M)$. In particular, it can thus be interpreted as a tensor field, and it is referred to as the Riemannian curvature tensor of (M, g).

Remark 66. Even though the notation $R(X, Y, Z)$ may seem more reasonable, the convention (4.1) is the one commonly used. Since R does not take its values in $C^\infty(M)$, the statement that it is a tensor field requires some justification. The reason for the terminology is that we can easily consider R to be a map from $\mathfrak{X}(M)^3 \times \mathfrak{X}^*(M) \to C^\infty(M)$ according to

$$(X, Y, Z, \eta) \mapsto \eta(R_{XY}Z).$$

That the corresponding map is linear over the functions in η is obvious. If it is linear over the functions in the other arguments, the tensor characterization lemma [3, Lemma 12.24, p. 318] thus yields the conclusion that we can think of R as of a tensor field.

Proof. That $\mathbb{R}$ is linear over the real numbers is clear. The only thing we need to prove is thus that

$$R_{(fX)Y}Z =fR_{XY}Z, \tag{4.2}$$
$$R_{X(fY)}Z =fR_{XY}Z, \tag{4.3}$$
$$R_{XY}(fZ) =fR_{XY}Z. \tag{4.4}$$

We prove one of these equalities and leave the other two as exercises. Compute

$$\begin{aligned}R_{XY}(fZ) =&\nabla_X[f\nabla_Y Z + Y(f)Z] - \nabla_Y[X(f)Z + f\nabla_X Z] - [X, Y](f)Z - f\nabla_{[X,Y]}Z\\ =&X(f)\nabla_Y Z + XY(f)Z + Y(f)\nabla_X Z + f\nabla_X\nabla_Y Z - YX(f)Z - X(f)\nabla_Y Z\\ &- Y(f)\nabla_X Z - f\nabla_Y\nabla_X Z - [X, Y](f)Z - f\nabla_{[X,Y]}Z\\ =&fR_{XY}Z.\end{aligned}$$

Due to Exercise 67, the proposition follows. □

Exercise 67. Prove (4.2) and (4.3).

By an argument similar to the proof of the tensor characterization lemma, cf. [3, pp. 318–319], Proposition 65 implies that it is possible to make sense of $R_{xy}z$ for $x, y, z \in T_pM$. In fact, choosing any vector fields $X, Y, Z \in \mathfrak{X}(M)$ such that $X_p = x$, $Y_p = y$ and $Z_p = z$, we can define $R_{xy}z$ by

$$R_{xy}z = (R_{XY}Z)(p);$$

the right hand side is independent of the choice of vectorfields X, Y, Z satisfying $X_p = x$, $Y_p = y$ and $Z_p = z$. Moreover, we can think of R_{xy} as a linear map from T_pM to T_pM.

The curvature tensor has several symmetries.

Proposition 68. *Let (M, g) be a semi-Riemannian manifold and $x, y, z, v, w \in T_pM$, where $p \in M$. Then*

$$\begin{aligned} R_{xy} &= - R_{yx}, && (4.5)\\ \langle R_{xy}v, w\rangle &= - \langle v, R_{xy}w\rangle, && (4.6)\\ R_{xy}z + R_{yz}x + R_{zx}y &= 0, && (4.7)\\ \langle R_{xy}v, w\rangle &= \langle R_{vw}x, y\rangle. && (4.8) \end{aligned}$$

Proof. Choose vectorfields X, Y, Z, V, W so that $X_p = x$ etc. Without loss of generality, we may assume the Lie brackets of any pairs of vectorfields in $\{X, Y, Z, V, W\}$ to vanish; simply choose these vectorfields to have constant coefficients relative to coordinate vectorfields (it is sufficient to carry out the computations locally).

That (4.5) holds is an immediate consequence of the definition. Note that (4.6) is equivalent to

$$\langle R_{xy}z, z\rangle = 0 \tag{4.9}$$

for all $x, y, z \in T_pM$. In order to prove (4.9), compute (using (3.2) and the fact that $[X, Y] = 0$)

$$\begin{aligned} \langle R_{XY}Z, Z\rangle &= \langle \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z, Z\rangle \\ &= X\langle \nabla_Y Z, Z\rangle - \langle \nabla_Y Z, \nabla_X Z\rangle - Y\langle \nabla_X Z, Z\rangle + \langle \nabla_X Z, \nabla_Y Z\rangle \\ &= \frac{1}{2}XY\langle Z, Z\rangle - \frac{1}{2}YX\langle Z, Z\rangle = \frac{1}{2}[X, Y]\langle Z, Z\rangle = 0. \end{aligned}$$

Thus (4.6) holds. In order to prove (4.7), compute

$$\begin{aligned} R_{XY}Z + R_{YZ}X + R_{ZX}Y &= \nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z + \nabla_Y\nabla_Z X - \nabla_Z\nabla_Y X \\ &\quad + \nabla_Z\nabla_X Y - \nabla_X\nabla_Z Y \\ &= \nabla_X[Y, Z] + \nabla_Z[X, Y] + \nabla_Y[Z, X] = 0. \end{aligned}$$

In order to prove (4.8), note that

$$\langle R_{xy}v + R_{yv}x + R_{vx}y, w\rangle = 0$$

due to (4.7). Adding up the four cyclic permutations of this equation and using (4.5) and (4.6) yields (4.8). We leave the details to the reader. □

Exercise 69. Prove (4.8).

4.2 Calculating the curvature tensor

It is of interest to derive a formula for the curvature in terms of a frame. To begin with, it is convenient to define the so-called connection coefficients.

Definition 70. Let (M,g) be a semi-Riemannian manifold, ∇ be the associated Levi-Civita connection and $\{e_i\}$ be a local frame. Then the *connection coefficients* associated with the frame $\{e_i\}$, denoted Γ^i_{jk}, are defined by

$$\nabla_{e_j} e_k = \Gamma^i_{jk} e_i. \tag{4.10}$$

Remark 71. In case $e_i = \partial_i$, the connection coefficients are the Christoffel symbols given by (3.5). However, for a general frame, the relation $\Gamma^k_{ij} = \Gamma^k_{ji}$ does typically not hold. This is due to the fact that the Lie bracket $[e_i, e_j]$ typically does not vanish.

In order to calculate the connection coefficients associated with a frame, it is useful to appeal to the Koszul formula (3.4). In this formula, the Lie brackets of the elements of the frame appear. It is of interest to note that the information concerning the Lie brackets is contained in the functions γ^k_{ij} defined by

$$[e_i, e_j] = \gamma^k_{ij} e_k. \tag{4.11}$$

Given Γ^i_{jk} and γ^i_{jk}, the curvature can be calculated according to the following formula.

Lemma 72. *Let (M,g) be a semi-Riemannian manifold, ∇ be the associated Levi-Civita connection and $\{e_i\}$ be a local frame. Let Γ^i_{jk} and γ^i_{jk} be defined by (4.10) and (4.11) respectively. Then*

$$R_{e_i e_j} e_k = -R_{ijk}{}^m e_m, \tag{4.12}$$

where

$$R_{ijk}{}^m = e_j(\Gamma^m_{ik}) - e_i(\Gamma^m_{jk}) + \Gamma^l_{ik}\Gamma^m_{jl} - \Gamma^l_{jk}\Gamma^m_{il} + \gamma^l_{ij}\Gamma^m_{lk}. \tag{4.13}$$

Remark 73. The motivation for including a minus sign in (4.12) is perhaps not so clear. There are several different conventions, but we have included the minus sign to obtain consistency with some of the standard references. The symbol $R_{ijk}{}^m$ should be thought of as the components of the curvature tensor (which is a $(1,3)$ tensor field) with respect to the frame $\{e_i\}$.

Proof. Let us compute

$$\begin{aligned} R_{e_i e_j} e_k =& \nabla_{e_i}\nabla_{e_j} e_k - \nabla_{e_j}\nabla_{e_i} e_k - \nabla_{[e_i,e_j]} e_k \\ =& \nabla_{e_i}(\Gamma^l_{jk} e_l) - \nabla_{e_j}(\Gamma^l_{ik} e_l) - \gamma^l_{ij}\nabla_{e_l} e_k \\ =& e_i(\Gamma^l_{jk}) e_l + \Gamma^l_{jk}\nabla_{e_i} e_l - e_j(\Gamma^l_{ik}) e_l - \Gamma^l_{ik}\nabla_{e_j} e_l - \gamma^l_{ij}\Gamma^m_{lk} e_m \\ =& e_i(\Gamma^m_{jk}) e_m + \Gamma^l_{jk}\Gamma^m_{il} e_m - e_j(\Gamma^m_{ik}) e_m - \Gamma^l_{ik}\Gamma^m_{jl} e_m - \gamma^l_{ij}\Gamma^m_{lk} e_m. \end{aligned} \tag{4.14}$$

The lemma follows. □

The curvature tensor with respect to local coordinates. In case the frame in Lemma 72 is given by $e_i = \partial_i$, the γ^k_{ij}'s vanish, and we obtain the formula

$$R_{ijk}{}^m = \partial_j \Gamma^m_{ik} - \partial_i \Gamma^m_{jk} + \Gamma^l_{ik}\Gamma^m_{jl} - \Gamma^l_{jk}\Gamma^m_{il}.$$

Moreover, in this case, the Γ^k_{ij}'s are given by (3.5). With respect to the standard coordinates, the Christoffel symbols of the Euclidean metric and the Minkowski metric vanish. In particular, the associated curvature tensors thus vanish. Moreover, this property (essentially) characterizes Euclidean space and Minkowski space. To prove this statement is, however, non-trivial.

Strategy for computing the curvature. The general strategy for computing the components of the curvature tensor is the following. First, choose a suitable local frame. Which frame is most appropriate depends on the context. Sometimes it is convenient to use a coordinate frame, but sometimes it is easier to carry out the computations with respect to an orthonormal frame. Once a choice of frame has been made, one first calculates the functions γ_{ij}^k determined by the Lie bracket. Then, one calculates the coefficients Γ_{ij}^k using the Koszul formula, (3.4). After this has been done, the components of the curvature can be calculated using (4.13). Needless to say, this is a cumbersome process in most cases.

4.3 The Ricci tensor and scalar curvature

Since the curvature tensor is a $(1,3)$ tensor field, we can contract two of the indices in order to obtain a covariant 2-tensor field. In fact, we define the *Ricci tensor* to be the covariant 2-tensor field whose components are given by

$$R_{ik} = R_{ijk}{}^{j}. \tag{4.15}$$

In terms of local coordinates, the components of the Ricci tensor are given by

$$R_{ik} = \partial_j \Gamma_{ik}^j - \partial_i \Gamma_{jk}^j + \Gamma_{ik}^l \Gamma_{jl}^j - \Gamma_{jk}^l \Gamma_{il}^j.$$

Again, the Ricci tensor of Euclidean space and Minkowski space vanish. In what follows, we denote the tensor field whose components are given by (4.15) by Ric. In other words, if $R_{ijk}{}^{m}$ are the components of the curvature tensor relative to a frame $\{e_i\}$, then

$$\mathrm{Ric}(e_i, e_k) = R_{ijk}{}^{j}.$$

The Ricci tensor is an extremely important object in semi-Riemannian geometry in general, and in general relativity in particular. It is of interest to derive alternate formulae for the Ricci tensor.

Lemma 74. *Let (M, g) be a semi-Riemannian manifold and $\{e_i\}$ be a a local orthonormal frame such that $\langle e_i, e_i \rangle = \epsilon_i$ (no summation on i). Then, for all $X, Y \in \mathfrak{X}(M)$,*

$$\mathrm{Ric}(X, Y) = \sum_j \epsilon_j \langle R_{e_j X} Y, e_j \rangle \tag{4.16}$$

on the domain of definition of the frame.

Proof. Note that with respect to a local frame $\{e_i\}$,

$$\langle R_{e_i e_j} e_k, e_l \rangle = -R_{ijk}{}^{m} \langle e_m, e_l \rangle = -R_{ijk}{}^{m} g_{ml}, \tag{4.17}$$

where all the components are calculated with respect to the frame $\{e_i\}$. Assume now that the frame is orthonormal so that $\langle e_i, e_j \rangle = 0$ if $i \neq j$ and $\langle e_i, e_i \rangle = \epsilon_i$ (no summation on i), where $\epsilon_i = \pm 1$. Letting $l = j$ in (4.17) then yields

$$-\epsilon_j R_{ijk}{}^{j} = \langle R_{e_i e_j} e_k, e_j \rangle$$

(no summation on j). Thus

$$R_{ijk}{}^{j} = -\epsilon_j \langle R_{e_i e_j} e_k, e_j \rangle = \epsilon_j \langle R_{e_j e_i} e_k, e_j \rangle$$

(no summation on j), where we have appealed to (4.5). Summing over j now yields

$$\mathrm{Ric}(e_i, e_k) = \sum_j \epsilon_j \langle R_{e_j e_i} e_k, e_j \rangle.$$

If $X = X^i e_i$ and $Y = Y^i e_i$ are elements of $\mathfrak{X}(M)$, we then obtain

$$\mathrm{Ric}(X, Y) = X^i Y^k \mathrm{Ric}(e_i, e_k) = X^i Y^k \sum_j \epsilon_j \langle R_{e_j e_i} e_k, e_j \rangle = \sum_j \epsilon_j \langle R_{e_j X} Y, e_j \rangle.$$

The lemma follows. □

The Ricci tensor is symmetric. It is of interest to note that, as a consequence of (4.16), (4.5), (4.6) and (4.8),

$$\mathrm{Ric}(X,Y) = \sum_j \epsilon_j \langle R_{e_j X} Y, e_j\rangle = \sum_j \epsilon_j \langle R_{Y e_j} e_j, X\rangle = \sum_j \epsilon_j \langle R_{e_j Y} X, e_j\rangle = \mathrm{Ric}(Y,X).$$

In other words, the Ricci tensor is a symmetric covariant 2-tensor field.

The scalar curvature. Finally, we define the *scalar curvature* S of a semi-Riemannian manifold by the formula

$$S = g^{ij} R_{ij}.$$

4.4 The divergence, the gradient and the Laplacian

The divergence of a vector field. Let (M, g) be a semi-Riemannian manifold with associated Levi-Civita connection ∇. If $X \in \mathfrak{X}(M)$, we can think of ∇X as $(1,1)$-tensor field according to

$$(Y, \eta) \mapsto \eta(\nabla_Y X);$$

note that this map is bilinear over the smooth functions and thus defines a $(1,1)$-tensor field due to the tensor characterization lemma. The components of this tensor field with respect to coordinates would in physics notation be written $\nabla_i X^j$. They are given by

$$\begin{aligned}\nabla_i X^j =& dx^j(\nabla_{\partial_i} X) = dx^j[\nabla_{\partial_i}(X^k \partial_k)] = dx^j[(\partial_i X^k)\partial_k + X^k \nabla_{\partial_i}\partial_k] \\ =& dx^j[(\partial_i X^k)\partial_k + X^k \Gamma^l_{ik}\partial_l] = \partial_i X^j + \Gamma^j_{ik} X^k.\end{aligned}$$

Contracting this tensor field yields a smooth function. We define the *divergence* of $X \in \mathfrak{X}(M)$, written $\mathrm{div}X$, to be the function which in local coordinates is given by

$$\mathrm{div}X = \nabla_i X^i = \partial_i X^i + \Gamma^i_{ik} X^k.$$

In Euclidean space, this gives the familiar formula, since $\Gamma^k_{ij} = 0$.

The gradient of a function. If $f \in C^\infty(M)$, then $df \in \mathfrak{X}^*(M)$. Applying the isomorphism $\sharp$ to df, we thus obtain a vectorfield referred to as the *gradient* of f:

$$\mathrm{grad}f = (df)^\sharp.$$

In local coordinates,

$$\mathrm{grad}f = g^{ij}(\partial_i f)\partial_j.$$

The Laplacian of a function. Finally, taking the divergence of the gradient yields the *Laplacian*

$$\Delta f = \mathrm{div}(\mathrm{grad}f).$$

In the case of Euclidean space, this definition yields the ordinary Laplacian. However, in the case of Minkowski space, it yields the wave operator.

4.5 Computing the covariant derivative of tensor fields

So far, we have only applied the Levi-Civita connection to vectorfields. However, it is also possible to apply it to tensor fields. To begin with, let us apply it to a one-form. To this end, let $\eta \in \mathfrak{X}^*(M)$ and $X, Y \in \mathfrak{X}(M)$. Then we define

$$(\nabla_X \eta)(Y) = X[\eta(Y)] - \eta(\nabla_X Y). \tag{4.18}$$

Exercise 75. Prove that $(\nabla_X\eta)(Y)$ defined by (4.18) is linear over the smooth functions in the argument Y (so that $\nabla_X\eta$ is a one-form due to the tensor characterization lemma). Prove, moreover, that

- $\nabla_X\eta$ is linear over $C^\infty(M)$ in X,
- $\nabla_X\eta$ is linear over $\mathbb{R}$ in η,
- $\nabla_X(f\eta) = X(f)\eta + f\nabla_X\eta$ for all $X \in \mathfrak{X}(M)$, $\eta \in \mathfrak{X}^*(M)$ and all $f \in C^\infty(M)$.

Finally, prove that

$$\nabla_X\eta = (\nabla_X\eta^\sharp)^\flat.$$

The components of the covariant derivative of a one-form. Note that $\nabla\eta$ can be thought of as a covariant 2-tensor field according to

$$(X, Y) \mapsto (\nabla_X\eta)(Y).$$

The components of this tensor field with respect to local coordinates are written $\nabla_i\eta_j$ in physics notation. They are given by

$$\nabla_i\eta_j = (\nabla_{\partial_i}\eta)(\partial_j) = \partial_i[\eta(\partial_j)] - \eta(\nabla_{\partial_i}\partial_j) = \partial_i\eta_j - \eta(\Gamma^k_{ij}\partial_k) = \partial_i\eta_j - \Gamma^k_{ij}\eta_k,$$

where Γ^k_{ij} are the Christoffel symbols associated with the coordinates (x^i).

The covariant derivative of tensorfields. In order to generalize the above construction to tensorfields, let T be a tensorfield of type (k, l). We can then think of T as a map from $\mathfrak{X}^*(M) \times \cdots \times \mathfrak{X}^*(M) \times \mathfrak{X}(M) \times \cdots \times \mathfrak{X}(M)$ (k copies of $\mathfrak{X}^*(M)$ and l copies of $\mathfrak{X}(M)$) to $C^\infty(M)$ which is multilinear over the smooth functions. If $\eta_1, \dots, \eta_k \in \mathfrak{X}^*(M)$ and $X, X_1, \dots, X_l \in \mathfrak{X}(M)$, then $\nabla_X T$ is defined by the relation

$$\begin{aligned}
&(\nabla_X T)(\eta_1, \dots, \eta_k, X_1, \dots, X_l)\\
=&X[T(\eta_1, \dots, \eta_k, X_1, \dots, X_l)]\\
&- T(\nabla_X\eta_1, \eta_2, \dots, \eta_k, X_1, \dots, X_l) - \cdots - T(\eta_1, \dots, \eta_{k-1}, \nabla_X\eta_k, X_1, \dots, X_l)\\
&- T(\eta_1, \dots, \eta_k, \nabla_X X_1, X_2, \dots, X_l) - \cdots - T(\eta_1, \dots, \eta_k, X_1, \dots, X_{l-1}, \nabla_X X_l).
\end{aligned} \tag{4.19}$$

Exercise 76. Prove that $(\nabla_X T)(\eta_1, \dots, \eta_k, X_1, \dots, X_l)$ defined by the formula (4.19) is linear over the smooth functions in $\eta_1, \dots, \eta_k, X_1, \dots, X_l$. Due to the tensor characterization lemma, this implies that $\nabla_X T$ is a tensorfield of type (k, l).

It is of interest to calculate ∇g, where g is the metric.

Exercise 77. Let (M, g) be a semi-Riemannian manifold and let ∇ be the associated Levi-Civita connection. Prove that $\nabla g = 0$.

4.5.1 Divergence of a covariant 2-tensor field

In the context of Einstein's equations, it is of interest to calculate the divergence of symmetric covariant 2-tensor fields. For that reason, we here wish to define the divergence and to derive a convenient formula for calculating it.

Let T be a symmetric covariant 2-tensor field. Then we can think of ∇T as a covariant 3-tensor field according to

$$(X, Y, Z) \mapsto (\nabla_X T)(Y, Z). \tag{4.20}$$

It is of interest to calculate the components of this tensor field with respect to a frame.

Lemma 78. *Let (M, g) be a semi-Riemannian manifold, ∇ be the associated Levi-Civita connection and $\{e_i\}$ be a local frame. Let Γ^i_{jk} and γ^i_{jk} be defined by (4.10) and (4.11) respectively. Finally, let T be a covariant 2-tensor field on M and $T_{ij} = T(e_i, e_j)$. Then*

$$(\nabla_{e_i} T)(e_j, e_k) = e_i(T_{jk}) - \Gamma^l_{ij} T_{lk} - \Gamma^l_{ik} T_{jl}. \tag{4.21}$$

Proof. Compute

$$\begin{aligned}(\nabla_{e_i} T)(e_j, e_k) =& e_i(T_{jk}) - T(\nabla_{e_i} e_j, e_k) - T(e_j, \nabla_{e_i} e_k)\\ =& e_i(T_{jk}) - \Gamma^l_{ij} T_{lk} - \Gamma^l_{ik} T_{jl}.\end{aligned}$$

The lemma follows. □

In physics notation, the components of the tensor field defined by (4.20) would be written $\nabla_i T_{jk}$, a convention we follow here. Moreover, we use this notation also in the case that the components are calculated with respect to a frame as opposed to only coordinate frames. However, which frame we use should be clear from the context. Note that $\nabla_i T_{jk} = \nabla_i T_{kj}$ when T is symmetric.

Definition 79. Let (M, g) be a semi-Riemannian manifold, ∇ be the associated Levi-Civita connection and T be a symmetric covariant 2-tensor field on M. Then we define $\mathrm{div} T$ to be the one-form whose components are given by

$$(\mathrm{div} T)_k = g^{ij} \nabla_i T_{jk}.$$

Remark 80. In physics notation, the definition of $\mathrm{div} T$ would be written

$$(\mathrm{div} T)_k = \nabla^i T_{ik};$$

first you raise the first index and then you contract with the second index.

Lemma 81. *Let (M, g) be a semi-Riemannian manifold, ∇ be the associated Levi-Civita connection and $\{e_i\}$ be a local orthonormal frame. Let Γ^i_{jk} and γ^i_{jk} be defined by (4.10) and (4.11) respectively. Finally, let T be a covariant 2-tensor field on M and $T_{ij} = T(e_i, e_j)$. Then*

$$(\mathrm{div} T)(X) = \sum_i \epsilon_i (\nabla_{e_i} T)(e_i, X) \tag{4.22}$$

for every $X \in \mathfrak{X}(M)$, where $\epsilon_i = \langle e_i, e_i \rangle$. Moreover,

$$(\mathrm{div} T)(e_j) = \sum_i \epsilon_i [e_i(T_{ij}) - \Gamma^l_{ii} T_{lj} - \Gamma^l_{ij} T_{il}]. \tag{4.23}$$

Proof. With respect to the orthonormal frame $\{e_i\}$,

$$(\mathrm{div} T)_k = g^{ij} \nabla_i T_{jk} = \sum_i \epsilon_i \nabla_i T_{ik} = \sum_i \epsilon_i (\nabla_{e_i} T)(e_i, e_k),$$

where it is taken for granted that all the indices are calculated with respect to the frame $\{e_i\}$. Say that $X = X^i e_i$ is a smooth vector field. Then

$$(\mathrm{div} T)(X) = X^k (\mathrm{div} T)(e_k) = X^k (\mathrm{div} T)_k = X^k \sum_i \epsilon_i (\nabla_{e_i} T)(e_i, e_k) = \sum_i \epsilon_i (\nabla_{e_i} T)(e_i, X).$$

Thus (4.22) holds. Combining this observation with (4.21) yields (4.23). □

4.6 An example of a curvature calculation

In the present section, we calculate the Ricci tensor of one specific metric; cf. (4.24) below. Our motivation for doing so is that the calculations illustrate the theory. However, the particular metric we have chosen is such that n-dimensional hyperbolic space and the 2-sphere are two special cases. Moreover, the Lorentz manifolds used by physicists to model the universe nowadays usually have a metric of the form (4.24).

The metric. Define the metric g by the formula

$$g = \epsilon dt \otimes dt + f^2(t) \sum_{i=1}^{n} dx^i \otimes dx^i \tag{4.24}$$

on $I \times U$, where I is an open interval and U is an open subset of $\mathbb{R}^n$. Moreover, t is the coordinate on the interval I and x^i are the standard coordinates on $\mathbb{R}^n$. Finally, f is a strictly positive smooth function on I and $\epsilon = \pm 1$. If $\epsilon = 1$, the metric is Riemannian, and if $\epsilon = -1$, g is a Lorentz metric.

The orthonormal frame. The curvature calculations can be carried out in many different ways. Here we shall use an orthonormal frame, denoted $\{e_\alpha\}$, $\alpha = 0, \dots, n$, and defined by

$$e_0 = \partial_t, \quad e_i = \frac{1}{f}\partial_i, \tag{4.25}$$

where $i = 1, \dots, n$; we shall here use the convention that Greek indices range from 0 to n and that Latin indices range from 1 to n.

The coefficients of the Lie bracket. Our goal is to compute the curvature of the metric (4.24). The strategy is to first compute the $\gamma^\lambda_{\alpha\beta}$, defined by the relation

$$[e_\alpha, e_\beta] = \gamma^\lambda_{\alpha\beta} e_\lambda. \tag{4.26}$$

Then the idea is to use the Koszul formula (3.4) to calculate the connection coefficients, defined by

$$\nabla_{e_\alpha} e_\beta = \Gamma^\lambda_{\alpha\beta} e_\lambda. \tag{4.27}$$

Since $\gamma^\lambda_{\alpha\beta} = -\gamma^\lambda_{\beta\alpha}$, it is sufficient to calculate $\gamma^\lambda_{\alpha\beta}$ for $\alpha < \beta$. Compute

$$[e_0, e_i] = -\frac{\dot{f}}{f^2}\partial_i = He_i,$$

where H is the function defined by

$$H = -\frac{\dot{f}}{f}.$$

Thus $\gamma^\alpha_{0i} = 0$ unless $\alpha = i$ and

$$\gamma^i_{0i} = H.$$

Since $[e_i, e_j] = 0$ for all $i, j = 1, \dots, n$, we have

$$\gamma^\alpha_{ij} = 0$$

for all $\alpha = 0, \dots, n$. To conclude, the only $\gamma^\lambda_{\alpha\beta}$'s that do not vanish are

$$\gamma^i_{0i} = H, \quad \gamma^i_{i0} = -H, \tag{4.28}$$

where $i = 1, \dots, n$ and we do not sum over i.

4.6.1 Computing the connection coefficients

The next step is to compute the connection coefficients. Let us first derive a general formula for the connection coefficients of an orthonormal frame.

Lemma 82. *Let (M, g) be a semi-Riemannian manifold and let $\{e_\alpha\}$, $\alpha = 0, \dots, n$, be an orthonormal frame on an open subset U of M. Define the connection coefficients $\Gamma^\lambda_{\alpha\beta}$ by the formula (4.27) and $\gamma^\lambda_{\alpha\beta}$ by (4.26). Then*

$$\Gamma^\lambda_{\alpha\beta} = \frac{1}{2}\left(-\epsilon_\lambda\epsilon_\alpha\gamma^\alpha_{\beta\lambda} + \epsilon_\lambda\epsilon_\beta\gamma^\beta_{\lambda\alpha} + \gamma^\lambda_{\alpha\beta}\right) \tag{4.29}$$

(no summation on any index), where $\epsilon_\alpha = g(e_\alpha, e_\alpha)$.

Proof. Due to the Koszul formula, (3.4),

$$\begin{aligned} 2\langle\nabla_{e_\alpha}e_\beta, e_\mu\rangle =& e_\alpha\langle e_\beta, e_\mu\rangle + e_\beta\langle e_\mu, e_\alpha\rangle - e_\mu\langle e_\alpha, e_\beta\rangle \\ & - \langle e_\alpha, [e_\beta, e_\mu]\rangle + \langle e_\beta, [e_\mu, e_\alpha]\rangle + \langle e_\mu, [e_\alpha, e_\beta]\rangle \\ =& - \langle e_\alpha, [e_\beta, e_\mu]\rangle + \langle e_\beta, [e_\mu, e_\alpha]\rangle + \langle e_\mu, [e_\alpha, e_\beta]\rangle \\ =& - \langle e_\alpha, \gamma^\nu_{\beta\mu}e_\nu\rangle + \langle e_\beta, \gamma^\nu_{\mu\alpha}e_\nu\rangle + \langle e_\mu, \gamma^\nu_{\alpha\beta}e_\nu\rangle \\ =& - \gamma^\nu_{\beta\mu}g_{\alpha\nu} + \gamma^\nu_{\mu\alpha}g_{\beta\nu} + \gamma^\nu_{\alpha\beta}g_{\mu\nu}, \end{aligned}$$

where we used the fact that the frame is orthonormal in the second step and $g_{\alpha\beta} = \langle e_\alpha, e_\beta\rangle$. On the other hand,

$$2\langle\nabla_{e_\alpha}e_\beta, e_\mu\rangle = 2\langle\Gamma^\nu_{\alpha\beta}e_\nu, e_\mu\rangle = 2\Gamma^\nu_{\alpha\beta}g_{\mu\nu}.$$

Combining these two equations yields

$$\Gamma^\nu_{\alpha\beta}g_{\mu\nu} = \frac{1}{2}\left(-\gamma^\nu_{\beta\mu}g_{\alpha\nu} + \gamma^\nu_{\mu\alpha}g_{\beta\nu} + \gamma^\nu_{\alpha\beta}g_{\mu\nu}\right).$$

Multiplying this equation with $g^{\lambda\mu}$ and summing over μ yields

$$\begin{aligned} \Gamma^\lambda_{\alpha\beta} =& \frac{1}{2}\left(-\gamma^\nu_{\beta\mu}g^{\lambda\mu}g_{\alpha\nu} + \gamma^\nu_{\mu\alpha}g^{\lambda\mu}g_{\beta\nu} + \gamma^\nu_{\alpha\beta}g^{\lambda\mu}g_{\mu\nu}\right) \\ =& \frac{1}{2}\left(-\epsilon_\lambda\epsilon_\alpha\gamma^\alpha_{\beta\lambda} + \epsilon_\lambda\epsilon_\beta\gamma^\beta_{\lambda\alpha} + \gamma^\lambda_{\alpha\beta}\right) \end{aligned}$$

(no summation on any index), where we have used the fact that $g_{\alpha\beta} = \epsilon_\alpha\delta_{\alpha\beta}$. The lemma follows. □

Calculating the connection coefficients in the case of the metric (4.24). Let us now return to the metric (4.24). Considering the formula (4.29), it is of interest to note the following. In all the terms on the right hand side, the indices in the $\gamma^\mu_{\nu\kappa}$'s are simply permutations of the indices in $\Gamma^\lambda_{\alpha\beta}$. In our particular setting, the only combination of indices in $\gamma^\mu_{\alpha\beta}$ that (may) give a non-zero result is if one of the indices is 0 and the other two are equal and belong to $\{1, \dots, n\}$. Let us compute, using (4.28) and (4.29),

$$\begin{aligned} \Gamma^i_{0i} =& \frac{1}{2}\left(-\epsilon\gamma^0_{ii} + \gamma^i_{i0} + \gamma^i_{0i}\right) = 0, \\ \Gamma^i_{i0} =& \frac{1}{2}\left(-\gamma^i_{0i} + \epsilon\gamma^0_{ii} + \gamma^i_{i0}\right) = -\gamma^i_{0i} = -H, \\ \Gamma^0_{ii} =& \frac{1}{2}\left(-\epsilon\gamma^i_{i0} + \epsilon\gamma^i_{0i} + \gamma^0_{ii}\right) = \epsilon\gamma^i_{0i} = \epsilon H \end{aligned}$$

(no summation on i). To conclude, the only connection coefficients which are non-zero are

$$\Gamma^i_{i0} = -H, \quad \Gamma^0_{ii} = \epsilon H \tag{4.30}$$

(no summation on i).

4.6.2 Calculating the components of the Ricci tensor

Let us now turn to the problem of calculating the Ricci tensor. It is useful to derive a general expression for the the components of the Ricci tensor with respect to an orthonormal frame.

Lemma 83. *Let (M, g) be a semi-Riemannian manifold and let $\{e_\alpha\}$, $\alpha = 0, \dots, n$, be an orthonormal frame on an open subset U of M. Define the connection coefficients $\Gamma^\lambda_{\alpha\beta}$ by the formula (4.27) and $\gamma^\lambda_{\alpha\beta}$ by (4.26). Then*

$$\mathrm{Ric}(e_\mu, e_\nu) = e_\alpha(\Gamma^\alpha_{\mu\nu}) + \Gamma^\lambda_{\mu\nu}\Gamma^\alpha_{\alpha\lambda} - e_\mu(\Gamma^\alpha_{\alpha\nu}) - \Gamma^\lambda_{\alpha\nu}\Gamma^\alpha_{\mu\lambda} - \gamma^\lambda_{\alpha\mu}\Gamma^\alpha_{\lambda\nu}, \tag{4.31}$$

where Einstein's summation convention applies.

Proof. Due to (4.16),

$$\mathrm{Ric}(e_\mu, e_\nu) = \sum_\alpha \epsilon_\alpha \langle R_{e_\alpha e_\mu} e_\nu, e_\alpha\rangle. \tag{4.32}$$

On the other hand, (4.14) yields

$$\begin{aligned}\langle R_{e_\alpha e_\mu} e_\nu, e_\alpha\rangle =& \langle e_\alpha(\Gamma^\beta_{\mu\nu})e_\beta + \Gamma^\lambda_{\mu\nu}\Gamma^\beta_{\alpha\lambda}e_\beta - e_\mu(\Gamma^\beta_{\alpha\nu})e_\beta - \Gamma^\lambda_{\alpha\nu}\Gamma^\beta_{\mu\lambda}e_\beta - \gamma^\lambda_{\alpha\mu}\Gamma^\beta_{\lambda\nu}e_\beta, e_\alpha\rangle \\ =& \epsilon_\alpha\left(e_\alpha(\Gamma^\alpha_{\mu\nu}) + \Gamma^\lambda_{\mu\nu}\Gamma^\alpha_{\alpha\lambda} - e_\mu(\Gamma^\alpha_{\alpha\nu}) - \Gamma^\lambda_{\alpha\nu}\Gamma^\alpha_{\mu\lambda} - \gamma^\lambda_{\alpha\mu}\Gamma^\alpha_{\lambda\nu}\right)\end{aligned}$$

(no summation on α), where we have used the fact that $\langle e_\alpha, e_\beta\rangle = \epsilon_\alpha\delta_{\alpha\beta}$. Combining this observation with (4.32) yields (4.31). □

The components of the Ricci tensor of the metric (4.24). Let us now compute the components of the Ricci tensor of the metric (4.24). Since the Ricci tensor is symmetric, it is sufficient to compute $\mathrm{Ric}(e_\mu, e_\nu)$ for $\mu \le \nu$. Before computing the individual components, let us make the following observations. Since the $\Gamma^\alpha_{\mu\nu}$'s only depend on t, $e_\lambda(\Gamma^\alpha_{\mu\nu}) = 0$ unless $\lambda = 0$. Keeping in mind that the only non-zero connection coefficients are given by (4.30), we conclude that

$$e_\alpha(\Gamma^\alpha_{\mu\nu}) = e_0(\Gamma^0_{\mu\nu}) = 0$$

unless $\mu = \nu = i$. Moreover,

$$e_\alpha(\Gamma^\alpha_{ii}) = e_0(\Gamma^0_{ii}) = \epsilon\dot{H}.$$

The 00-component of the Ricci tensor. Compute, using the above observations as well as the fact that the only non-zero connection coefficients are given by (4.30),

$$\begin{aligned}\mathrm{Ric}(e_0, e_0) =& e_\alpha(\Gamma^\alpha_{00}) + \Gamma^\lambda_{00}\Gamma^\alpha_{\alpha\lambda} - e_0(\Gamma^\alpha_{\alpha0}) - \Gamma^\lambda_{\alpha0}\Gamma^\alpha_{0\lambda} - \gamma^\lambda_{\alpha0}\Gamma^\alpha_{\lambda0} \\ =& -\sum_i e_0(\Gamma^i_{i0}) + \sum_i \gamma^i_{0i}\Gamma^i_{i0} = n\dot{H} - nH^2,\end{aligned}$$

where we have used the fact that the only non-zero $\gamma^\alpha_{\mu\nu}$'s are given by (4.28).

The $0i$-components of the Ricci tensor. Compute

$$\mathrm{Ric}(e_0, e_i) = e_\alpha(\Gamma^\alpha_{0i}) + \Gamma^\lambda_{0i}\Gamma^\alpha_{\alpha\lambda} - e_0(\Gamma^\alpha_{\alpha i}) - \Gamma^\lambda_{\alpha i}\Gamma^\alpha_{0\lambda} - \gamma^\lambda_{\alpha0}\Gamma^\alpha_{\lambda i} = 0;$$

since $\Gamma^\lambda_{0\beta} = 0$ regardless of what λ and β are, the first, second and fourth terms on the right hand side vanish; since $\Gamma^0_{0i} = 0$ and $\Gamma^j_{ji} = 0$ regardless of the values of i and j, it is clear that $\Gamma^\alpha_{\alpha i} = 0$ (so that the third term on the right hand side vanishes); in order for the first factor in the fifth term to be non-vanishing, we have to have $\lambda = \alpha = j$ for some $j = 1, \dots, n$, cf. (4.28), but if $\lambda = \alpha = j$, then the second factor in the fifth term vanishes.

The ij-components of the Ricci tensor, $i \ne j$. If $i \ne j$, then

$$\mathrm{Ric}(e_i, e_j) = e_\alpha(\Gamma^\alpha_{ij}) + \Gamma^\lambda_{ij}\Gamma^\alpha_{\alpha\lambda} - e_i(\Gamma^\alpha_{\alpha j}) - \Gamma^\lambda_{\alpha j}\Gamma^\alpha_{i\lambda} - \gamma^\lambda_{\alpha i}\Gamma^\alpha_{\lambda j} = 0. \tag{4.33}$$

To justify this calculation, note that in order for $\Gamma^{\alpha}_{\mu\nu}$ or $\gamma^{\alpha}_{\mu\nu}$ to be non-zero, two of the indices have to be non-zero and equal, and the remaining index has to be zero. For this reason, the first three terms on the right hand side of (4.33) vanish. Turning to the last two terms, there are indices such that one of the factors appearing in these terms are non-vanishing. However, then the other factor has to vanish.

The ij-components of the Ricci tensor, $i = j$. By arguments similar to ones given above,

$$\begin{aligned}\mathrm{Ric}(e_i, e_i) =& e_\alpha(\Gamma^{\alpha}_{ii}) + \Gamma^{\lambda}_{ii}\Gamma^{\alpha}_{\alpha\lambda} - e_i(\Gamma^{\alpha}_{\alpha i}) - \Gamma^{\lambda}_{\alpha i}\Gamma^{\alpha}_{i\lambda} - \gamma^{\lambda}_{\alpha i}\Gamma^{\alpha}_{\lambda i}\\ =& e_0(\Gamma^{0}_{ii}) + \Gamma^{0}_{ii}\Gamma^{j}_{j0} - \Gamma^{0}_{ii}\Gamma^{i}_{i0} - \gamma^{i}_{0i}\Gamma^{0}_{ii}\\ =& \epsilon\dot{H} - \epsilon n H^2 + \epsilon H^2 - \epsilon H^2 = \epsilon\dot{H} - \epsilon n H^2\end{aligned}$$

(no summation on i).

Summing up, we obtain the following lemma.

Lemma 84. *Let I be an open interval, U be an open subset of $\mathbb{R}^n$ and g be defined by (4.24), where t is the coordinate on the interval I and x^i are the standard coordinates on $\mathbb{R}^n$. Moreover, f is a strictly positive smooth function on I and $\epsilon = \pm 1$. Let $\{e_\alpha\}$, $\alpha = 0, \dots, n$, be the orthonormal frame defined by (4.25). Then*

$$\mathrm{Ric}(e_\alpha, e_\beta) = 0$$

if $\alpha \neq \beta$. Moreover,

$$\begin{aligned}\mathrm{Ric}(e_0, e_0) =& n\dot{H} - nH^2,\\ \mathrm{Ric}(e_i, e_i) =& \epsilon\dot{H} - \epsilon n H^2\end{aligned}$$

(no summation on i), where $i = 1, \dots, n$.

4.7 Calculating the Ricci curvature of the 2-sphere and of the n-dimensional hyperbolic space

It is of interest to apply the calculations of the previous section to two special cases, namely that of the 2-sphere and that of the n-dimensional hyperbolic space. Let us begin with the 2-sphere.

4.7.1 The Ricci curvature of the 2-sphere

Recall that the n-sphere and the metric on the n-sphere were defined in Example 27. Here we calculate the Ricci curvature in case $n = 2$.

Proposition 85. *Let $\mathbb{S}^2$ denote the 2-sphere and $g_{\mathbb{S}^2}$ denote the metric on the 2-sphere. If $\mathrm{Ric}[g_{\mathbb{S}^2}]$ denotes the Ricci curvature of $g_{\mathbb{S}^2}$, then*

$$\mathrm{Ric}[g_{\mathbb{S}^2}] = g_{\mathbb{S}^2}.$$

Remark 86. Note that the Ricci curvature of the 2-sphere is positive definite. There is a general connection between the positive definiteness of the Ricci tensor and the compactness of the manifold. In fact, the so-called Myers' theorem implies the following: If (M, g) is a geodesically complete and connected Riemannian manifold such that

$$\mathrm{Ric}[g](v, v) \geq c_0 g(v, v) \tag{4.34}$$

for some constant $c_0 > 0$, all $v \in T_pM$ and all $p \in M$, then M is compact and $\pi_1(M)$ is finite. We shall not prove this theorem here but refer the interested reader to [2, Theorem 24, p. 279] for a proof.

Remark 87. In the case of 3-dimensions, there is an even deeper connection between the positive definiteness of the Ricci tensor and the topology of the manifold. In fact, if (M, g) is a connected, simply connected and geodesically complete Riemannian manifold of dimension 3 with positive definite Ricci curvature (in the sense that (4.34) holds), then M is diffeomorphic to the 3-sphere. This result is due to Richard Hamilton and we shall not prove it here.

Proof. Let

$$\psi : (0, \pi) \times (0, 2\pi) \to \mathbb{S}^2$$

be defined by

$$\psi(\theta, \phi) = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta). \tag{4.35}$$

Then ψ is a diffeomorphism onto its image, the image of ψ is dense in $\mathbb{S}^2$ and

$$\psi^* g_{\mathbb{S}^2} = d\theta \otimes d\theta + \sin^2\theta \; d\phi \otimes d\phi. \tag{4.36}$$

We leave the verification of these statements as an exercise. Due to these facts (and the smoothness of the Ricci tensor and the metric), it is sufficient to verify that

$$g = d\theta \otimes d\theta + \sin^2\theta \; d\phi \otimes d\phi$$

satisfies $\mathrm{Ric} = g$ for $0 < \theta < \pi$ and $0 < \phi < 2\pi$.

The metric g is such that we are in the situation considered in Section 4.6; replace t with θ; x^1 with ϕ; n with 1; $f(t)$ with $\sin\theta$; and ϵ with 1. As in Section 4.6, we also introduce the frame

$$e_0 = \partial_\theta, \quad e_1 = \frac{1}{\sin\theta}\partial_\phi.$$

Note that

$$H = -\frac{1}{\sin\theta}\partial_\theta \sin\theta = -\cot\theta.$$

Moreover,

$$\dot{H} = 1 + \frac{\cos^2\theta}{\sin^2\theta} = \frac{1}{\sin^2\theta}.$$

Appealing to Lemma 84 then yields

$$\mathrm{Ric}(\partial_\theta, \partial_\theta) = \mathrm{Ric}(e_0, e_0) = \frac{1}{\sin^2\theta} - \frac{\cos^2\theta}{\sin^2\theta} = 1.$$

Similarly, $\mathrm{Ric}(e_1, e_1) = 1$ and $\mathrm{Ric}(e_0, e_1) = 0$. The proposition follows. □

Exercise 88. Let ψ be defined by (4.35). Prove that ψ is a diffeomorphism onto its image, that the image of ψ is dense in $\mathbb{S}^2$ and that (4.36) holds.

4.7.2 The curvature of the upper half space model of hyperbolic space

Let us define $\mathbb{U}^n$ by

$$\mathbb{U}^n = \{x = (x^1, \dots, x^n) \in \mathbb{R}^n : x^n > 0\}.$$

Moreover, define

$$g_{\mathbb{U}^n} = \frac{1}{(x^n)^2}\sum_{i=1}^n dx^i \otimes dx^i.$$

Then $(\mathbb{U}^n, g_{\mathbb{U}^n})$ is called the *upper half space model* of n-dimensional hyperbolic space. Here, we do not sort out the relation between this model and the metric defined in Example 27, but we calculate the Ricci tensor of $g_{\mathbb{U}^n}$.

Lemma 89. *Let* $1 \leq n \in \mathbb{Z}$ *and let* $\mathbb{U}^{n+1}$ *and* $g_{\mathbb{U}^{n+1}}$ *be defined as above. Then*

$$\mathrm{Ric}[g_{\mathbb{U}^{n+1}}] = -n g_{\mathbb{U}^{n+1}},$$

where $\mathrm{Ric}[g_{\mathbb{U}^{n+1}}]$ *denotes the Ricci curvature of* $g_{\mathbb{U}^{n+1}}$.

Proof. Let $\psi : \mathbb{R}^{n+1} \to \mathbb{U}^{n+1}$ be defined by

$$\psi(x^0, \dots, x^n) = [x^1, \dots, x^n, \exp(x^0)].$$

Then ψ is a diffeomorphism from $\mathbb{U}^{n+1}$ to $\mathbb{R}^{n+1}$ and

$$g = \psi^* g_{\mathbb{U}^{n+1}} = dx^0 \otimes dx^0 + e^{-2x^0} \sum_{i=1}^{n} dx^i \otimes dx^i.$$

Denoting x^0 by t, introducing f by $f(t) = e^{-t}$, and letting $\epsilon = 1$, we are exactly in the situation considered in Section 4.6. Compute

$$H = -\frac{\dot{f}}{f} = 1.$$

Lemma 84 then yields $\mathrm{Ric} = -ng$. The lemma follows. □

Chapter 5

Einstein's equations

Given the terminology introduced in the previous chapters, it is straightforward to write down Einstein's equations. Let (M, g) be a spacetime, Ric be the associated Ricci tensor and S be the associated scalar curvature. Then the *Einstein tensor*, denoted G, is defined by

$$G = \mathrm{Ric} - \frac{1}{2} S g.$$

Note that this is a symmetric covariant 2-tensor field. Einstein's equations relate the Einstein tensor to the matter content. The matter content is described by the so-called stress energy tensor T, which is a symmetric covariant 2-tensor field. The exact form of T depends on the matter model used. There is a very large number of choices of matter models. For that reason, we do not discuss the form of T here in any detail; we shall simply give some examples. Given that a matter model and a T have been specified, *Einstein's equations* take the form

$$G + \Lambda g = T, \tag{5.1}$$

where Λ is a constant, referred to as the *cosmological constant*. If there is matter present, these equations should be complemented with equations for the matter fields. Specializing (5.1) to the case of $T = 0$ corresponds to *Einstein's vacuum equations* (with a cosmological constant).

Even though we have developed the necessary mathematical structures needed in order to write down Einstein's equations, it is not so clear why (5.1) should constitute a physical theory of gravitational interaction. For this reason, we spend the next section giving a heuristic motivation for why the Lorentz geometry setting is natural, and why the equations take the specific form they do. Readers only interested in the mathematical aspects of the theory (and prepared to accept the equations as given) can proceed immediately to Section 5.3 (Section 5.2 contains a description of two classes of solutions of interest in physics, a section which can also be skipped by those only interested in the mathematical aspects).

5.1 A brief motivation of Einstein's equations

It is natural to divide the motivation for the equations into two parts. First, we wish to motivate why the theory should be geometric in nature, and, second, we wish to motivate the particular form of the equations.

5.1.1 Motivation for the geometric nature of the theory

In pre-relativity physics, the following two assumptions were typically made

- The notions of time, length and acceleration are absolute (as a consequence, the non-accelerated, or so-called inertial, frames are well defined).
- The laws of physics take the same form regardless of the choice of inertial frame.

The first assumption leads to the Galilean transformations relating the measurements made in one inertial frame with the measurements made in another frame. However, as we already pointed out in Section 2.7, Maxwell's equations do not transform well under these transformations. This led Einstein to the following assumptions:

- Acceleration is absolute (as a consequence, the non-accelerated, or so-called inertial frames are well defined).
- The speed of light is independent of inertial frame.
- The laws of physics take the same form regardless of the choice of inertial frame.

This then led to the Lorentz transformations. Since the Lorentz transformations are the isometries of the Minkowski metric, these assumptions can also be said to have led to the introduction of Minkowski space.

Even though Maxwell's equations transform well under the Lorentz transformations, Newtonian gravity does not. It is therefore clear that the classical theory of gravity has to be modified. In order to arrive at a new theory of gravity, there were several (in part quite philosophical) ideas that influenced the thinking of Einstein. The three perhaps most important ones were the following.

Mach's principle: The matter content of the universe should contribute to the local definition of the notion of what it means for a frame to be non-accelerating and non-rotating. In a universe devoid of matter, these concepts should not make sense. In short: *the concept of acceleration lacks meaning in the absence of matter.*

The principle of equivalence. The ratio of the masses of two bodies can be defined in two ways:

- the reciprocal ratio of the acceleration a given force imparts to them (*inert mass*).
- the ratio of the forces which act upon them in the same gravitational field (*gravitational mass*).

The principle of equivalence then states that the inert mass of a body equals the gravitational mass. Another way to express this statement is to say that a coordinate system at rest in a uniform gravitational field is equivalent to a coordinate system in uniform acceleration far away from all matter.

The principle of general covariance. Roughly speaking, this principle states that there are no preferred coordinate frames, and that the equations of physics should be independent of coordinates. A more precise formulation would be to say that all physical laws should be expressible as tensor equations on manifolds (thereby being independent of the particular coordinate representation). Moreover, the laws should reduce to those of special relativity in a frame which is in free fall (this notion can be given a precise meaning in the case of Lorentz geometry, but we shall refrain from doing so here).

Note that the principle of equivalence is quite remarkable. There is no reason why the inert mass should be the same as the gravitational mass. It is therefore highly desirable to develop a theory in which gravitational forces and acceleration are practically the same.

Geometric nature of the theory. In order to justify that it is natural for the theory to be geometric in nature, let us (following Einstein) carry out the following thought experiment. Let K be a frame in $\mathbb{R}^3$ in which ordinary Euclidean geometry holds. Give the axes of K the names

x, y and z. Let K' similarly be a frame (with x', y' and z'-axes) such that the origin of K and K' coincide and the z-axis coincides with the z'-axis. Assume however, that K' is rotating relative to K. Say now that we have a circle which is at rest relative to the the $x'y'$-plane. If O is the circumference of the circle and D is the diameter (as measured by K), then $O/D = \pi$. Assume now that O' and D' are the circumference and diameter of the circle, as measured with respect to K'. Then D' should equal D. The reason for this is that Lorentz transformations give rise to length contraction, but only in directions that are parallel to the motion. In this case the motion (rotation) is perpendicular to the diameter, so that there should be no length contraction in the direction of the diameter. On the other hand, $O < O'$, since the circumference is parallel to the motion. Thus

$$\frac{O'}{D'} > \frac{O}{D} = \pi.$$

In other words, with respect to K', the geometry is not Euclidean. It would thus seem that the fact that the frame K' is accelerated distorts the geometry. On the other hand, due to the principle of equivalence, acceleration should correspond to gravitation. It is therefore natural to expect that gravitation should distort the geometry. In order to determine what type of geometry is the most natural, it is useful to keep in mind that in special relativity, the natural geometry is that of Minkowski space. In general, we can expect the scalar product to change from point to point. This naturally leads to Lorentz geometry. In short, a natural way to model the spacetime is by a Lorentz manifold (or, in fact, a spacetime in the sense of Definition 36).

Interpretation of curvature. In order to connect gravitation (or, equivalently, acceleration) with the geometry, it is natural to consider particles in free fall. Since particles in free fall are ones upon which no forces act, they travel along straight lines in special relativity. Generalizing the notion of a straight line to the Lorentz geometry setting, the principle of general covariance leads to the postulate that in general relativity, freely falling particles follow timelike geodesics. In order to detect the influence of gravity, it is natural to consider a family of timelike geodesics. On the one hand, the gravitational tidal forces should lead to freely falling particles to either converge or diverge. In other words, the relative motion of members of a family of geodesics should correspond to the influence of the gravitational field. On the other hand, since it should be possible to carry out a geometric analysis of the relative motion, it should be possible to analyze which geometric objects correspond to the gravitational field. Let ν be as in (3.14). Let γ be defined by $\gamma(t) = \nu(t, 0)$ and assume that for every $s \in (-\epsilon, \epsilon)$, the curve α_s, defined by $\alpha_s(t) = \nu(t, s)$ is a timelike geodesic. Furthermore, let $V \in \mathfrak{X}(\gamma)$ be defined by the condition that $V(t)$ is the tangent vector of the curve $s \mapsto \nu(t, s)$ at the point $s = 0$ (we also write this $V(t) = \nu_s(t, 0)$). Then V can be thought of as an infinitesimal version of the displacement of the geodesics α_s (i.e., the freely falling test particles) relative to the curve γ. A computation shows that

$$V'' = R_{\gamma' V}\gamma' \tag{5.2}$$

(we refer the reader interested in a justification of this identity to [2, Lemma 3, p. 216]). Since V is the relative displacement, V'' is the relative acceleration of freely falling test particles. Thus (5.2) gives a relation between the relative acceleration of the freely falling test particles and the curvature tensor of the underlying Lorentz manifold. In other words, the gravitation should correspond to the curvature of the Lorentz manifold.

5.1.2 A motivation for the form of the equation

In the Newtonian picture, the starting point is a matter distribution ρ; at each spacetime point, the value of ρ is the amount of matter per unit volume at that point. The matter distribution gives rise to a gravitational field in the following way. First, a gravitational potential is obtained by solving Poisson's equation

$$\Delta\phi = 4\pi G\rho, \tag{5.3}$$

where G is the gravitational constant. The gravitational field is then obtained as $-\text{grad}\phi$. In the case of a point mass, ρ is a multiple of the Dirac delta function, and the corresponding gravitational field is the one giving rise to the standard formula for the gravitational force between two point masses.

In the case of general relativity, spacetime is described by a Lorentz manifold. Moreover, we should think of the metric as corresponding to the gravitational potential in the classical picture. Naively, it would thus seem natural to replace the left hand side of (5.3) by something which involves at most second order derivatives of the metric. Moreover, due to the principle of general covariance, it should be a tensor field (independent of the coordinates). Consider the right hand side of (5.3). Clearly, it should also be replaced by a tensorial expression. Moreover, due to the principle of general covariance, what this tensor is should be indicated by special relativity. However, in special relativity, there is a canonical way to collect all the matter into one tensor. This tensor is called the stress energy tensor, and it is a symmetric covariant 2-tensor field. Let us call it T. Due to the conservation laws for matter, it turns out that T has to be divergence free. Returning now to the left hand side of (5.3), and summing up the requirements: the replacement for the left hand side of (5.3) should be a symmetric covariant 2-tensor field constructed from the metric and its derivatives (up to order at most 2). Moreover, it should be divergence free. However, it then turns out that the only possibility for the left hand side (up to constant multiples) is $G + \Lambda g$, where G is the Einstein tensor and Λ is the cosmological constant. This leads to (5.1), where we have set the constant multiplying T to 1.

5.2 Modeling the universe and isolated systems

Studying Einstein's equations in all generality is very difficult. It is therefore natural to start by considering some special cases. In physics there are two natural situations of interest. First, one would like to model an isolated object, such as a planet, a star or a black hole. Second, one would like to model the entire universe.

5.2.1 Isolated systems

When modeling an isolated system, it is natural to start by assuming that the object under consideration is spherically symmetric and static (roughly speaking meaning that there is no "time dependence"). Naively, this leads to the assumption that the metric is of the form

$$g = N(r)dt \otimes dt + R(r)dr \times dr + A(r)g_{\mathbb{S}^2}$$

on $M = \mathbb{R} \times (r_0, \infty) \times \mathbb{S}^2$, where $r_0 \in \mathbb{R}$, N, R and A are functions on (r_0, ∞) and $g_{\mathbb{S}^2}$ is the standard round metric on $\mathbb{S}^2$. Imposing Einstein's vacuum equations on an ansatz of this type leads to the so-called Schwarzschild solutions, which are discussed at greater length in [1] (the reason for imposing Einstein's vacuum equations is that the solution should be thought of as describing the exterior of a planet, star, etc.). We refer the reader interested in a more detailed discussion of this case to [1].

5.2.2 Cosmology

When modeling the entire universe, the standard starting point is to assume that the universe is spatially homogeneous and isotropic. In practice, this means that the metric can be written in the form

$$g = -dt^2 + f^2(t)g_N$$

on $M = I \times N$, where I is an open interval and (N, g_N) is a Riemannian manifold. Moreover, the isometry group of (N, g_N) should be *transitive*. This means that for any pairs of points $p, q \in N$, there should be an isometry ϕ of (N, g_N) such that $\phi(p) = q$. This requirement correspond to the assumption of spatial homogeneity and in practice means that you cannot distinguish between different points on N; geometrically they are equivalent. The second condition is that for every $p \in M$ and every pair of vectors $v, w \in T_pM$ such that $g_N(v,v) = g_N(w,w)$, there is an isometry ϕ of (N, g_N) such that $d\phi(v) = w$. This requirement corresponds to the assumption of spatial isotropy and means that you cannot distinguish between different directions on N; geometrically they are equivalent.

Collectively, the assumptions of spatial homogeneity and isotropy are referred to as the *cosmological principle*. They are motivated by observations (particularly of the cosmic microwave background radiation) as well as the philosophical idea that we do not occupy a privileged position in the universe.

A natural question now arises: are there any Riemannian manifolds satisfying the assumptions we make concerning (N, g_N) above? It turns out that there are three possible geometries: Euclidean, hyperbolic and spherical. At present, physicists prefer the Euclidean geometry. In other words, the metrics of interest are of the form (4.24). We discuss this class of metrics next.

5.3 A cosmological model

Most of the current models of the universe are such that the relevant Lorentz metric is given by (4.24) with $n = 3$ and $\epsilon = -1$. The metric is defined on $M = I \times \mathbb{R}^3$, where the size of the interval I depends on the context. The standard model also includes matter of so-called dust type. In the present context, the relevant form of the corresponding stress energy tensor is

$$T = \rho dt \otimes dt, \tag{5.4}$$

where ρ is a function of t only. The equations that the geometry and the matter should satisfy consist of Einstein's equations (5.1) as well the equation for ρ implied by the requirement that $\text{div}T = 0$. It is of interest to write down what these equations mean in terms of f and ρ.

Lemma 90. *Let I be an open interval, $f > 0$ and $\rho \geq 0$ be smooth functions on I and let g be the metric on $M = I \times \mathbb{R}^3$ defined by (4.24). Finally, let Λ be a constant and T be defined by (5.4). Then the equations*

$$G + \Lambda g = T, \tag{5.5}$$

$$\text{div}T = 0 \tag{5.6}$$

are equivalent to

$$3H^2 - \Lambda = \rho, \tag{5.7}$$

$$2\dot{H} - 3H^2 + \Lambda = 0, \tag{5.8}$$

$$-\dot{\rho} + 3H\rho = 0. \tag{5.9}$$

Remark 91. Note that (5.7) and (5.8) imply (5.9).

Proof. To begin with, let us compute G. Due to Lemma 84, we know that

$$\text{Ric}(e_0, e_0) = 3\dot{H} - 3H^2, \quad \text{Ric}(e_i, e_i) = -\dot{H} + 3H^2$$

(no summation on i), $i = 1, \ldots, n$, and the remaining components of the Ricci tensor vanish; here the frame $\{e_\alpha\}$ is defined in (4.25). In particular, this means that the scalar curvature is given by

$$S = -\text{Ric}(e_0, e_0) + \sum_{i=1}^{3} \text{Ric}(e_i, e_i) = -6\dot{H} + 12H^2. \tag{5.10}$$

Thus

$$\begin{aligned}G(e_0,e_0) =&\mathrm{Ric}(e_0,e_0)-\frac{1}{2}Sg(e_0,e_0)=3\dot{H}-3H^2-3\dot{H}+6H^2=3H^2,\\ G(e_i,e_i) =&\mathrm{Ric}(e_i,e_i)-\frac{1}{2}Sg(e_i,e_i)=-\dot{H}+3H^2+3\dot{H}-6H^2=2\dot{H}-3H^2.\end{aligned}$$

Moreover, if $\alpha\neq\beta$, then $G(e_\alpha,e_\beta)=0$. In particular, we thus have

$$\begin{aligned}G(e_0,e_0)+\Lambda g(e_0,e_0) =&3H^2-\Lambda,\\ G(e_i,e_i)+\Lambda g(e_i,e_i) =&2\dot{H}-3H^2+\Lambda\end{aligned}$$

and

$$G(e_\alpha,e_\beta)+\Lambda g(e_\alpha,e_\beta)=0$$

if $\alpha\neq\beta$. Since the only component of T which is non-zero is the 00-component, we conclude that (5.5) is equivalent to the two equations

$$\begin{aligned}3H^2-\Lambda =&\rho,\\ 2\dot{H}-3H^2+\Lambda =&0.\end{aligned}$$

Turning to (5.6), note that (4.23) yields

$$(\mathrm{div}T)(e_\beta)=\sum_\alpha \epsilon_\alpha[e_\alpha(T_{\alpha\beta})-\Gamma^\lambda_{\alpha\alpha}T_{\lambda\beta}-\Gamma^\lambda_{\alpha\beta}T_{\alpha\lambda}].$$

Keeping in mind that the only non-zero connection coefficients are given by (4.30) and that the only component of T which is non-zero is $T_{00}=\rho$, this yields

$$(\mathrm{div}T)(e_0)=-\dot{\rho}+3H\rho. \tag{5.11}$$

Moreover,

$$(\mathrm{div}T)(e_k)=0. \tag{5.12}$$

The lemma follows. □

Exercise 92. Prove that (5.11) and (5.12) hold.

Note that (5.8) is a second order ODE for f and that (5.9) is a first order ODE for ρ. However, they can be combined to yield a system of first order equations for f, H and ρ:

$$\begin{aligned}\dot{f} =&-Hf, &(5.13)\\ \dot{H} =&\frac{3}{2}H^2-\frac{1}{2}\Lambda, &(5.14)\\ \dot{\rho} =&3H\rho. &(5.15)\end{aligned}$$

Given initial data for f, H and ρ, this system of equations has a unique corresponding solution. Note also that if f and ρ are strictly positive initially, then they are always strictly positive. This is a consequence of the following observation (we leave the proof as an exercise).

Lemma 93. *Let I be an open interval, $h\in C^1(I)$ and $g\in C^0(I)$. If $\dot{h}=gh$ on I, then*

$$h(t)=h(t_0)\exp\left(\int_{t_0}^t g(s)ds\right)$$

for all $t,t_0\in I$.

Since solutions to (5.13)–(5.15) are uniquely determined by initial data, it is not so clear that it is possible to combine these equations with (5.7). That this is nevertheless possible is a consequence of the following lemma.

Lemma 94. *Let $\Lambda \in \mathbb{R}$. Moreover, let $f_0 > 0$, $\rho_0 \geq 0$ and H_0 be real numbers and assume that they satisfy*

$$3H_0^2 - \Lambda = \rho_0. \tag{5.16}$$

Let f, H and ρ be the solution to (5.13)–(5.15) corresponding to the initial data $f(0) = f_0$, $H(0) = H_0$ and $\rho(0) = \rho_0$, and let I be the maximal interval of existence for the solution. Then (5.7) holds for all $t \in I$. In particular (5.7)–(5.9) are satisfied for all $t \in I$.

Proof. Define

$$\psi = 3H^2 - \Lambda - \rho.$$

By assumption, $\psi(0) = 0$. Compute

$$\dot{\psi} = 6H\dot{H} - \dot{\rho} = 3H(3H^2 - \Lambda) - 3H\rho = 3H\psi,$$

where we have used (5.14) and (5.15) in the second step. Due to Lemma 93, we conclude that $\psi(t) = 0$ for all $t \in I$. □

Due to this lemma, a natural way to construct solutions to (5.7)–(5.9) is the following. First specify initial data $f_0 > 0$, $\rho_0 \geq 0$ and H_0 to (5.13)–(5.15), satisfying (5.16). Then solve (5.13)–(5.15). The corresponding solution will then be a solution to (5.7)–(5.9). Let us now analyze how the corresponding solutions behave.

Lemma 95. *Let $\Lambda > 0$ and fix real numbers $f_0 > 0$, $\rho_0 \geq 0$ and $H_0 < 0$ satisfying (5.16). Let f, ρ and H denote the solution to (5.13)–(5.15) corresponding to these initial data, and let I denote the maximal interval of existence. There are two cases to consider. If $\rho_0 = 0$, then $I = \mathbb{R}$ and*

$$\rho(t) = 0, \quad H(t) = -\alpha_0, \quad f(t) = f_0 e^{\alpha_0 t}$$

for all $t \in \mathbb{R}$, where

$$\alpha_0 = (\Lambda/3)^{1/2}. \tag{5.17}$$

If $\rho_0 > 0$, then $I = (t_-, \infty)$, where $t_- > -\infty$. In fact,

$$H(t) = \alpha_0 \frac{1 + c_H e^{3\alpha_0 t}}{1 - c_H e^{3\alpha_0 t}}, \tag{5.18}$$

$$\rho(t) = 12\alpha_0^2 \frac{c_H e^{3\alpha_0 t}}{(1 - c_H e^{3\alpha_0 t})^2}, \tag{5.19}$$

$$f(t) = f_0 \left(\frac{c_H e^{3\alpha_0 t} - 1}{c_H - 1} \right)^{2/3} e^{-\alpha_0 t} \tag{5.20}$$

where

$$c_H = \frac{H(0) - \alpha_0}{H(0) + \alpha_0} \tag{5.21}$$

and α_0 is given by (5.17). Moreover, $c_H > 1$ and

$$t_- = -\frac{1}{3\alpha_0} \ln c_H. \tag{5.22}$$

Remark 96. Due to the assumptions concerning ρ_0 and Λ, the equation (5.16) implies that H_0 has to be non-zero. However, it could be either positive or negative. The choice corresponds to a choice of time orientation. Demanding that $H_0 < 0$ implies that increasing t corresponds to increasing f.

Remark 97. The reason we focus on the case $\Lambda > 0$ is that this is the case of greatest interest in cosmology at present.

Proof. In case $\rho_0 = 0$, we know that $3H^2 = \Lambda$. Since $H_0 < 0$, we also know that $H(t) < 0$ for all $t \in I$. Consequently, $H(t) = -\alpha_0$, where α_0 is given by (5.17). Due to (5.13), we conclude that $f(t) = f_0 e^{\alpha_0 t}$. The statements of the lemma concerning the case $\rho_0 = 0$ follow.

Let us now assume that $\rho_0 > 0$. Then $\rho(t) > 0$ for all $t \in I$. Due to (5.7), we know that

$$3H^2 = \rho + \Lambda > \Lambda.$$

Combining this observation with the fact that $H(0) < 0$, we conclude that $H(t) < -\alpha_0$ for all $t \in I$. On the other hand, (5.14) implies that

$$\dot{H} = \frac{3}{2}(H - \alpha_0)(H + \alpha_0). \tag{5.23}$$

Since $H(t) - \alpha_0 < 0$ and $H(t) + \alpha_0 < 0$ for all t, it is clear that H is a strictly increasing function. Moreover, this is a separable equation which can be solved explicitly. The solution is given by (5.18), where c_H is given by (5.21). Note that $c_H > 1$, since $H(0) < -\alpha_0$. It is of interest to analyze for which t the right hand side of (5.18) is well defined. The only problem that occurs is when $1 - c_H e^{3\alpha_0 t}$ equals zero. This happens when $t = t_-$, where t_- is defined by (5.22). Note that $t_- < 0$. Moreover, the right hand side of (5.18) is well defined for all $t > t_-$. As long as H is a well defined smooth function, it is clear from (5.13) and (5.15) that f and ρ are well defined. To conclude, we have a solution to (5.13)–(5.15) on $I = (t_-, \infty)$. Clearly, this interval cannot be extended to the right. Moreover, since $H(t) \to -\infty$ as $t \to t_-$, is is clear that it cannot be extended to the left. Thus the maximal interval of existence is given by $I = (t_-, \infty)$. Due to (5.7),

$$\rho(t) = 3H^2(t) - \Lambda = 3[H^2(t) - \alpha_0^2] = 12\alpha_0^2 \frac{c_H e^{3\alpha_0 t}}{(1 - c_H e^{3\alpha_0 t})^2}.$$

Thus (5.19) holds. Finally, given the formula (5.18) for H, (5.13) can be integrated to yield (5.20). ☐

It is of interest to note the following consequences of (5.18)–(5.20):

$$\lim_{t\to t_-} H(t) = -\infty, \tag{5.24}$$

$$\lim_{t\to t_-} \rho(t) = \infty, \tag{5.25}$$

$$\lim_{t\to t_-} f(t) = 0. \tag{5.26}$$

Since ρ is the energy density of the matter, (5.25) means that the energy density becomes unbounded as $t \to t_-$. Turning to the curvature $S(t)$, note that (5.10) and (5.14) imply that

$$S = -6\dot{H} + 12H^2 = 3H^2 + 3\Lambda.$$

In particular,

$$\lim_{t\to t_-} S(t) = \infty.$$

In this sense, the curvature becomes unbounded as $t \to t_-$. Clearly, something extreme happens as $t \to t_-$. These observations justify referring to the $t = t_-$ hypersurface as the big bang.

A natural next question to ask is whether timelike geodesics are complete or not. In order to be able to answer this question, it is, however, necessary to quote the following result concerning maximal existence intervals of solutions to autonomous systems of ODE's (we omit the proof).

Lemma 98. *Let $U \subseteq \mathbb{R}^n$ be an open set and let $F : U \to \mathbb{R}^n$ be a smooth function. Let $\xi_0 \in U$ and consider the initial value problem*

$$\frac{d\xi}{dt} = F \circ \xi, \tag{5.27}$$

$$\xi(0) = \xi_0. \tag{5.28}$$

Let $I = (t_-, t_+)$ be the maximal interval of existence of the solution to (5.27) and (5.28). If $t_+ < \infty$, then there is a sequence $t_k \in I$ such that $t_k \to t_+$ as $k \to \infty$, and such that either $|\xi(t_k)| \to \infty$ or $\xi(t_k)$ converges to a point of the boundary of U. Analogously, if $t_- > -\infty$, then there is a sequence $t_k \in I$ such that $t_k \to t_-$ as $k \to \infty$, and such that either $|\xi(t_k)| \to \infty$ or $\xi(t_k)$ converges to a point of the boundary of U.

We are now in a position to prove the desired statement concerning timelike geodesics.

Proposition 99. *Let (M, g) be a spacetime of the type constructed in Lemma 95 corresponding to initial data with $\rho_0 > 0$. Then $M = I \times \mathbb{R}^3$, where $I = (t_-, \infty)$ and $t_- > -\infty$. Let $\gamma : J \to M$ be a future directed maximal timelike geodesic in (M, g). Define $s_\pm \in \mathbb{R}$ by the requirement that $J = (s_-, s_+)$. Then $s_- > -\infty$ and $s_+ = \infty$, so that $J = (s_-, \infty)$. Moreover, the t-coordinate of γ converges to t_- as $s \to s_-$ and to ∞ as $s \to \infty$.*

Remark 100. The time orientation of (M, g) is determined by the requirement that $e_0 = \partial_t$ be future oriented. To say that γ is future directed thus means that

$$\langle e_0|_{\gamma(s)}, \dot{\gamma}(s)\rangle < 0$$

for all $s \in J$.

Remark 101. The physical interpretation of the statement is the following. A freely falling test particle (observer) follows a timelike geodesic, say γ. The parameter range of the geodesic corresponds (up to a constant factor) to the proper time as measured by the observer. The statement of the lemma implies that the observer has only "lived" for a finite time. Moreover, tracing the trajectory of the observer back towards the past, one reaches $t = t_-$ in finite parameter time. Due to Lemma 95 and the statements made after the proof of this lemma, we know that the energy density ρ and the scalar curvature S blow up as $t \to t_-$. This extreme behaviour will thus be experienced by the observer γ a finite proper time to the past.

Proof. Define the functions $v^\alpha : J \to \mathbb{R}$, $\alpha = 0, \dots, 3$, by the requirement that

$$\dot{\gamma}(s) = v^\alpha(s) e_\alpha|_{\gamma(s)}.$$

Here, a dot refers to a derivative with respect to the parameter of the curve γ. In case we compute a derivative with respect to t, we denote it by a prime. Note that

$$v^0 = -\langle \dot{\gamma}, e_0\rangle, \quad v^i = \langle \dot{\gamma}, e_i\rangle.$$

It is of interest to relate these expressions to the coordinate formulae for the curve. Define γ^α, $\alpha = 0, \dots, 3$, by

$$\gamma = (\gamma^0, \gamma^1, \gamma^2, \gamma^3).$$

Then

$$\dot{\gamma} = \dot{\gamma}^\alpha \partial_\alpha.$$

Thus

$$\dot{\gamma}^0(s) = v^0(s), \quad \dot{\gamma}^i(s) = \frac{1}{f[\gamma^0(s)]} v^i(s). \tag{5.29}$$

Reformulating the equation for the geodesic. Using the fact that $\ddot{\gamma} = 0$ and Proposition 51,

$$\begin{aligned} 0 =& \ddot{\gamma} = \dot{v}^\alpha(s) e_\alpha|_{\gamma(s)} + v^\alpha(s)\nabla_{\dot{\gamma}(s)} e_\alpha|_{\gamma(s)} = \dot{v}^\alpha(s) e_\alpha|_{\gamma(s)} + v^\alpha(s) v^\beta(s) \nabla_{e_\beta|_{\gamma(s)}} e_\alpha|_{\gamma(s)} \\ =& \dot{v}^\lambda(s) e_\lambda|_{\gamma(s)} + v^\alpha(s) v^\beta(s) \Gamma^\lambda_{\beta\alpha}[\gamma(s)] e_\lambda|_{\gamma(s)}. \end{aligned}$$

In other words, the equation $\ddot{\gamma} = 0$ is equivalent to the system of equations

$$\dot{v}^\lambda(s) + \Gamma^\lambda_{\beta\alpha}[\gamma(s)] v^\alpha(s) v^\beta(s) = 0,$$

$\lambda = 0, \dots, 3$. In case $\lambda = 0$,

$$\dot{v}^0(s) - H[\gamma^0(s)] \sum_i v^i(s) v^i(s) = 0, \tag{5.30}$$

where we have used the fact that the only non-zero connection coefficients are given by (4.30); note that $\epsilon = -1$ in the case of interest here. In case $\lambda = i$,

$$\dot{v}^i(s) - H[\gamma^0(s)] v^0(s) v^i(s) = 0. \tag{5.31}$$

Note that this equation, in combination with Lemma 93, yields the conclusion that v^i is either always non-zero or always zero.

The autonomous ODE. Note that the equations (5.29), (5.30) and (5.31) constitute an autonomous ODE for

$$\xi = (\xi^1, \dots, \xi^8) = (\gamma^0, \dots, \gamma^3, v^0, \dots, v^3).$$

The relevant ODE is of the form $\dot{\xi} = F \circ \xi$, where F is determined by (5.29), (5.30) and (5.31). Note that $H \circ \gamma^0$ is a smooth function as long as $\gamma^0 \in (t_-, \infty)$ and that $f \circ \gamma^0$ is a smooth and strictly positive function as long as $\gamma^0 \in (t_-, \infty)$. Since the remaining dependence of F on ξ is polynomial, it is clear that F is well defined on $U = (t_-, \infty) \times \mathbb{R}^7$. We can thus apply Lemma 98 in order to conclude that in order for s_+ to be strictly less than ∞, there has to be a sequence $s_k \to s_+$ such that either $|\xi(s_k)| \to \infty$ or $\gamma^0(s_k) \to t_-$. The statement concerning s_- is analogous.

Timelike geodesics. Let us now focus on timelike geodesics. We then have

$$\langle \dot{\gamma}, \dot{\gamma} \rangle = -\lambda_0,$$

where $\lambda_0 > 0$ is a constant. Note also that, due to the fact that γ is future oriented, $\langle \dot{\gamma}, e_0 \rangle < 0$. Thus $v^0 > 0$. Since $\dot{\gamma}^0 = v^0$, this means that the t-coordinate of the curve γ is strictly increasing. In fact, since

$$\langle \dot{\gamma}, \dot{\gamma} \rangle = -(v^0)^2 + \sum_i (v^i)^2,$$

we know that

$$\dot{\gamma}^0 = v^0 \geq \lambda_0^{1/2}.$$

Proving completeness/incompleteness. If $v^i(s) \neq 0$, (5.31) implies

$$\ln \frac{v^i(s)}{v^i(s_0)} = \int_{s_0}^s \frac{\dot{v}^i(\sigma)}{v^i(\sigma)} d\sigma = \int_{s_0}^s H[\gamma^0(\sigma)] v^0(\sigma) d\sigma.$$

On the other hand, (5.13) and (5.29) imply that

$$H \circ \gamma^0 \cdot v^0 = -\frac{f' \circ \gamma^0}{f \circ \gamma^0} \dot{\gamma}^0 = -\frac{d}{ds} \ln f \circ \gamma^0.$$

Combining the last two equations yields

$$\ln \frac{v^i(s)}{v^i(s_0)} = -\ln \frac{f[\gamma^0(s)]}{f[\gamma^0(s_0)]}.$$

In particular,

$$v^i(s) = v^i(s_0)\frac{f[\gamma^0(s_0)]}{f[\gamma^0(s)]}$$

(note that this relation also holds in case $v^i(s) = 0$). As long as $\gamma^0(s)$ belongs to (t_-, ∞), $v^i(s)$ is thus finite. Combining this observation with (5.30) and the fact that $H[\gamma^0(s)]$ is well defined and finite for s such that $\gamma^0(s)$ belongs to (t_-, ∞), we conclude that $v^0(s)$ is finite as long as $\gamma^0(s)$ belongs to (t_-, ∞). Finally, (5.29) yields a similar conclusion concerning γ^i. By the above observations concerning the maximal existence interval of solutions to $\dot{\xi} = F \circ \xi$, we conclude that if $J = (s_-, s_+)$ is the maximal interval of existence of the geodesic, then the only possibility for s_+ to be strictly less than ∞ is if

$$\lim_{s \to s_+} \gamma^0(s) = \infty. \tag{5.32}$$

The reason for this is the following. Since γ^0 is monotonically increasing, it converges to a limit as $s \to s_+$. If this limit is finite, say $t_+ < \infty$, this means that γ^0 is bounded on the interval $[s_0, s_+)$. Moreover, it is bounded away from t_-, since γ^0 is monotonically increasing. By the above observations, this means that all the other components of ξ are bounded on $[s_0, s_+)$. These observations contradict the statement that there is a sequence $s_k \to s_+$ such that either $|\xi(s_k)| \to \infty$ or $\gamma^0(s_k) \to t_-$. By a similar argument, the only possibility for s_- to be strictly larger than $-\infty$ is if

$$\lim_{s \to s_-} \gamma^0(s) = t_-.$$

Proving completeness to the future. In order to prove that the geodesic is future complete, assume that $s_+ < \infty$. Then (5.32) holds. On the other hand, due to (5.30), it is clear that v^0 is decreasing. Fixing s_0 in the existence interval for γ, we conclude that for $s \geq s_0$,

$$v^0(s_0) \geq v^0(s).$$

In particular, v^0 is bounded to the future. Combining this observation with

$$\gamma^0(s) = \gamma^0(s_0) + \int_{s_0}^{s} \dot{\gamma}^0(\sigma)d\sigma = \gamma^0(s_0) + \int_{s_0}^{s} v^0(\sigma)d\sigma \leq \gamma^0(s_0) + v^0(s_0)(s - s_0),$$

it is clear that $\gamma^0(s)$ cannot become infinite in finite parameter time to the future. This contradicts (5.32).

Proving incompleteness to the past. Assume that $s_- = -\infty$. Since $v^0(s) \geq v^0(s_0)$ for $s \leq s_0$, this implies that

$$\gamma^0(s) = \gamma^0(s_0) + \int_{s_0}^{s} \dot{\gamma}^0(\sigma)d\sigma = \gamma^0(s_0) + \int_{s_0}^{s} v^0(\sigma)d\sigma \leq \gamma^0(s_0) - v^0(s_0)(s_0 - s).$$

As $s \to s_- = -\infty$, this implies that $\gamma^0(s) \to -\infty$. However, we know that $\gamma^0(s) > t_- > -\infty$ for all $s \in (s_-, s_+)$. This leads to a contradiction, so that the geodesic is past incomplete. The lemma follows. □

Exercise 102. Prove the same statement for future directed null geodesics.

Finally, let us make the following observation concerning the causal structure of the spacetime at late times.

Proposition 103. *Let (M, g) be a spacetime of the type constructed in Lemma 95. Then $M = I \times \mathbb{R}^3$, where $I = (t_-, \infty)$. Fix $\bar{x}_i \in \mathbb{R}^3$, $i = 0, 1$, with $\bar{x}_0 \neq \bar{x}_1$. For T large enough, there is then no future directed causal curve $\gamma : (s_-, s_+) \to M$ with the property that $\gamma(s_0) = (t_0, \bar{x}_0)$, $t_0 \geq T$, and $\gamma(s_1) = (t_1, \bar{x}_1)$ for some $s_1 > s_0$.*

Remark 104. The physics interpretation of this statement is the following. The trajectories of galaxies in the universe are, roughly speaking, curves of the form $\gamma(s) = [\gamma^0(s), \bar{x}]$, where $\bar{x} \in \mathbb{R}^3$ is independent of s. Fixing the points $\bar{x}_i$ thus, in some sense, corresponds to fixing two separate galaxies in the universe. One fundamental question to ask is then: is it possible for observers in the galaxy corresponding to $\bar{x}_0$ to send information to observers in the galaxy corresponding to $\bar{x}_1$? Since information has to travel along causal curves, this is equivalent to the question if there is a future directed causal curve γ starting at $(t_0, \bar{x}_0)$ and ending at $(t_1, \bar{x}_1)$. The statement of the proposition is that if t_0 is large enough, there is no such causal curve. In other words, it is not possible for observers in the different galaxies to communicate. An alternate formulation of the proposition is thus the following. Given two distinct galaxies, there is a time (distance from the big bang) such that after that time, it is not possible for observers in the two galaxies to communicate (how large this time is of course depends on how close the galaxies are initially). The reason for this inability to communicate is the presence of the positive cosmological constant.

Proof. Let $\gamma : (s_-, s_+) \to M$ be a future directed causal curve such that $\gamma(s_0) = (t_0, \bar{x}_0)$. Let us analyze how far the curve can travel in the spatial direction. In other words, let

$$\bar{\gamma}(s) = [\gamma^1(s), \gamma^2(s), \gamma^3(s)].$$

The question is then: what is the maximal Euclidean length of the curve $\bar{\gamma}|_{[s_0,s_+)}$? Due to the causality of the curve

$$0 \geq \langle \dot{\gamma}, \dot{\gamma} \rangle = -(\dot{\gamma}^0)^2 + f^2 \circ \gamma^0 |\dot{\bar{\gamma}}|^2.$$

Since γ is future oriented, $\dot{\gamma}^0 > 0$, so that this inequality implies that

$$|\dot{\bar{\gamma}}| \leq \frac{1}{f \circ \gamma^0} \dot{\gamma}^0. \tag{5.33}$$

On the other hand, due to (5.13),

$$f' = -Hf \geq \alpha_0 f,$$

where we have used (5.7) in the last step. This means that

$$f(t) \geq f(t_0) \exp[\alpha_0 (t - t_0)]$$

for $t \geq t_0$. Combining this estimate with (5.33) yields

$$|\dot{\bar{\gamma}}| \leq \frac{1}{f(t_0)} e^{\alpha_0 t_0} e^{-\alpha_0 \gamma^0} \dot{\gamma}^0.$$

Since $t_0 = \gamma^0(s_0)$, this inequality yields

$$\int_{s_0}^{s_+} |\dot{\bar{\gamma}}(s)| ds \leq \frac{1}{f(t_0)} e^{\alpha_0 t_0} \int_{s_0}^{s_+} e^{-\alpha_0 \gamma^0(s)} \dot{\gamma}^0(s) ds \leq \frac{1}{f(t_0)} e^{\alpha_0 t_0} \int_{t_0}^{\infty} e^{-\alpha_0 t} dt = \frac{1}{\alpha_0 f(t_0)}.$$

The length of the curve $\bar{\gamma}|_{[s_0,s_+)}$ is thus bounded by the right hand side of this inequality. Since the right hand side tends to zero as $t_0 \to \infty$, it is clear that for t_0 large enough, the length of the curve is strictly smaller than $|\bar{x}_0 - \bar{x}_1|$. Thus there is no $s \in [s_0, s_+)$ such that $\bar{\gamma}(s) = \bar{x}_1$. Since this conclusion is independent of the causal curve γ satisfying $\gamma(s_0) = (t_0, \bar{x}_0)$, the lemma follows. □

Bibliography

[1] Bär, C.: Theory of Relativity, lecture notes, Summer term 2013.

[2] O'Neill, B.: Semi Riemannian Geometry. Academic Press, Orlando (1983)

[3] Lee, J. M.: Introduction to Smooth Manifolds (Second Edition). Springer, New York (2013)

[4] Spivak, M.: A comprehensive introduction to differential geometry, Volume two. Publish or Perish, Inc., Houston, Texas, USA (1979)