# Chapter 1
# Geometry

## 1.1 Riemannian and Lorentzian Manifolds

### *1.1.1 Differential Geometry*

We collect here some basic facts and principles of differential geometry as the foundation for the sequel. For a more penetrating discussion and for the proofs of various results, we refer to [65]. Classical differential geometry as expressed through the tensor calculus is about coordinate representations of geometric objects and the transformations of those representations under coordinate changes. The geometric objects are invariantly defined, but their coordinate representations are not, and resolving this contradiction is the content of the tensor calculus.

We consider a $d$-dimensional differentiable manifold $M$ (assumed to be connected, oriented, paracompact and Hausdorff) and start with some conventions:

1. Einstein summation convention

$$a^i b_i := \sum_{i=1}^{d} a^i b_i. \tag{1.1.1}$$

   The content of this convention is that a summation sign is omitted when the same index occurs twice in a product, once as an upper and once as a lower index. This rule is not affected by the possible presence of other indices; for example,

$$\Lambda^i_j b^j = \sum_{j=1}^{d} \Lambda^i_j b^j. \tag{1.1.2}$$

   The conventions about when to place an index in an upper or lower position will be given subsequently. One aspect of this, however, is:
2. When $G = (g_{ij})_{i,j}$ is a metric tensor (a notion to be explained below) with indices $i, j$, the inverse metric tensor is written as $G^{-1} = (g^{ij})_{i,j}$, that is, by raising the indices. In particular

$$g^{ij} g_{jk} = \delta^i_k := \begin{cases} 1 & \text{when } i = k, \\ 0 & \text{when } i \neq k, \end{cases} \tag{1.1.3}$$

   the so-called Kronecker symbol.
3. Combining the previous rules, we obtain more generally

$$v^i = g^{ij} v_j \quad \text{and} \quad v_i = g_{ij} v^j. \tag{1.1.4}$$

J. Jost, *Geometry and Physics*,
DOI 10.1007/978-3-642-00541-1_1, 

4. For $d$-dimensional scalar quantities $(\phi^1, \dots, \phi^d)$, we can use the Euclidean metric $\delta_{ij}$ to freely raise or lower indices in order to conform to the summation convention, that is,

$$\phi_i = \delta_{ij}\phi^j = \phi^i. \tag{1.1.5}$$

A (finite-dimensional) manifold $M$ is locally modeled after $\mathbb{R}^d$. Thus, locally, it can be represented by coordinates $x = (x^1, \dots, x^d)$ taken from some open subset of $\mathbb{R}^d$. These coordinates, however, are not canonical, and we may as well choose other ones, $y = (y^1, \dots, y^d)$, with $x = f(y)$ for some homeomorphism $f$. When the manifold $M$ is differentiable—as always assumed here—we can cover it by local coordinates in such a manner that all such coordinate transitions are diffeomorphisms where defined. Again, the choice of coordinates is non-canonical. The basic content of classical differential geometry is to investigate how various expressions representing objects on $M$ like tangent vectors transform under coordinate changes. Here and in the sequel, all objects defined on a differentiable manifold will be assumed to be differentiable themselves. This is checked in local coordinates, but since coordinate transitions are diffeomorphic, the differentiability property does not depend on the choice of coordinates.

*Remark* For our purposes, it is often convenient, and in the literature, it is customary, to mean by "differentiability" smoothness of class $C^\infty$, that is, to assume that all objects are infinitely often differentiable. The ring of (infinitely often) differentiable functions on $M$ is denoted by $C^\infty(M)$. Nonetheless, at certain places where analysis is more important, we need to be more specific about the regularity classes of the objects involved. But for the moment, we shall happily assume that our manifold $M$ is of class $C^\infty$.

A tangent vector for $M$ at some point $p$ represented by $x_0$ in local coordinates[1] $x$ is an expression of the form

$$V = v^i \frac{\partial}{\partial x^i}. \tag{1.1.6}$$

This means that it operates on a function $\phi(x)$ in our local coordinates as

$$V(\phi)(x_0) = v^i \frac{\partial \phi}{\partial x^i}_{|x=x_0}. \tag{1.1.7}$$

The summation convention (1.1.1) applies to (1.1.7). The $i$ in $\frac{\partial}{\partial x^i}$ is considered to be a lower index since it appears in the denominator.

The tangent vectors at $p \in M$ form a vector space, called the tangent space $T_pM$ of $M$ at $p$. A basis of $T_pM$ is given by the $\frac{\partial}{\partial x^i}$, considered as derivative operators

[1] We shall not always be so careful in distinguishing a point $p$ as an invariant geometric object from its representation $x_0$ in some local coordinates, but frequently identify $p$ and $x_0$ without alerting the reader.

at the point $p$ represented by $x_0$ in the local coordinates, as in (1.1.7).[2] Whereas, as should become clear subsequently, this tangent space and its tangent vectors are defined independently of the choice of local coordinates, the representation of a tangent space does depend on those coordinates. The question then is how the same tangent vector is represented in different local coordinates $y$ with $x = f(y)$ as before. The answer comes from the requirement that the result of the operation of the tangent vector $V$ on a function $\phi$, $V(\phi)$, be independent of the choice of coordinates. Always applying the chain rule, here and in the sequel, this yields

$$V = v^i \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k}. \tag{1.1.8}$$

Thus, the coefficients of $V$ in the $y$-coordinates are $v^i \frac{\partial y^k}{\partial x^i}$. This is verified by the following computation:

$$v^i \frac{\partial y^k}{\partial x^i} \frac{\partial}{\partial y^k} \phi(f(y)) = v^i \frac{\partial y^k}{\partial x^i} \frac{\partial \phi}{\partial x^j} \frac{\partial x^j}{\partial y^k} = v^i \frac{\partial x^j}{\partial x^i} \frac{\partial \phi}{\partial x^j} = v^i \frac{\partial \phi}{\partial x^i} \tag{1.1.9}$$

as required.

More abstractly, changing coordinates by $f$ pulls a function $\phi$ defined in the $x$-coordinates back to $f^\star \phi$ defined for the $y$-coordinates, with $f^\star \phi(y) = \phi(f(y))$. If then $W = w^k \frac{\partial}{\partial y^k}$ is a tangent vector written in the $y$-coordinates, we need to push it forward as

$$f_\star W = w^k \frac{\partial x^i}{\partial y^k} \frac{\partial}{\partial x^i} \tag{1.1.10}$$

to the $x$-coordinates, to have the invariance

$$(f_\star W)(\phi) = W(f^\star \phi) \tag{1.1.11}$$

which is easily checked:

$$(f_\star W)\phi = w^k \frac{\partial x^i}{\partial y^k} \frac{\partial \phi}{\partial x^i} = w^k \frac{\partial}{\partial y^k} \phi(f(y)) = W(f^\star \phi). \tag{1.1.12}$$

In particular, there is some duality between functions and tangent vectors here. However, the situation is not entirely symmetric. We need to know the tangent vector only at the point $x_0$ where we want to apply it, but we need to know the function $\phi$ in some neighborhood of $x_0$ because we take its derivatives.

A vector field is then defined as $V(x) = v^i(x) \frac{\partial}{\partial x^i}$, that is, by having a tangent vector at each point of $M$. As indicated above, we assume here that the coefficients $v^i(x)$ are differentiable. The vector space of vector fields on $M$ is written as $\Gamma(TM)$. (In fact, $\Gamma(TM)$ is a module over the ring $C^\infty(M)$.)

[2] As here, we shall usually simply write $\frac{\partial}{\partial x^i}$ in place of $\frac{\partial}{\partial x^i}(p)$ or $\frac{\partial}{\partial x^i}(x_0)$, that is, we assume that the point where a derivative operator acts is clear from the context or the coefficient.

Later, we shall need the Lie bracket $[V, W] := VW - WV$ of two vector fields $V(x) = v^i(x)\frac{\partial}{\partial x^i}$, $W(x) = w^j(x)\frac{\partial}{\partial x^j}$; its operation on a function $\phi$ is

$$\begin{aligned}[V, W]\phi(x) &= v^i(x)\frac{\partial}{\partial x^i}\left(w^j(x)\frac{\partial}{\partial x^j}\phi(x)\right) - w^j(x)\frac{\partial}{\partial x^j}\left(v^i(x)\frac{\partial}{\partial x^i}\phi(x)\right) \\ &= \left(v^i(x)\frac{\partial w^j(x)}{\partial x^i} - w^i(x)\frac{\partial v^j(x)}{\partial x^i}\right)\frac{\partial \phi(x)}{\partial x^j}. \end{aligned} \tag{1.1.13}$$

In particular, for coordinate vector fields, we have

$$\left[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right] = 0. \tag{1.1.14}$$

Returning to a single tangent vector, $V = v^i \frac{\partial}{\partial x^i}$ at some point $x_0$, we consider a covector or cotangent vector $\omega = \omega_i dx^i$ at this point as an object dual to $V$, with the rule

$$dx^i\left(\frac{\partial}{\partial x^j}\right) = \delta^i_j \tag{1.1.15}$$

yielding

$$\omega_i dx^i\left(v^j\frac{\partial}{\partial x^j}\right) = \omega_i v^j \delta^i_j = \omega_i v^i. \tag{1.1.16}$$

This expression depends only on the coefficients $v^i$ and $\omega_i$ at the point under consideration and does not require any values in a neighborhood. We can write this as $\omega(V)$, the application of the covector $\omega$ to the vector $V$, or as $V(\omega)$, the application of $V$ to $\omega$.

The cotangent vectors at $p$ likewise constitute a vector space, the cotangent space $T^\star_p M$.

We have the transformation behavior

$$dx^i = \frac{\partial x^i}{\partial y_\alpha} dy^\alpha \tag{1.1.17}$$

required for the invariance of $\omega(V)$. Thus, the coefficients of $\omega$ in the $y$-coordinates are given by the identity

$$\omega_i dx^i = \omega_i \frac{\partial x^i}{\partial y_\alpha} dy^\alpha. \tag{1.1.18}$$

Again, a covector $\omega_i dx^i$ is pulled back under a map $f$:

$$f^\star(\omega_i dx^i) = \omega_i \frac{\partial x^i}{\partial y^\alpha} dy^\alpha. \tag{1.1.19}$$

The transformation rules (1.1.10), (1.1.19) apply to arbitrary maps $f : M \to N$ from $M$ into a possibly different manifold $N$, not only to coordinate changes or diffeomorphisms. So, we can always pull back a function or a covector and always push forward a vector under a map, but not always the other way around.

The transformation behavior of a tangent vector as in (1.1.8) is called contravariant, the opposite one of a covector as (1.1.18) covariant.

A 1-form then assigns a covector to every point in $M$, and thus, it is locally given as $\omega_i(x)dx^i$.

Having derived the transformation of vectors and covectors, we can then also determine the transformation rules for other tensors. **A lower index always indicates covariant, an upper one contravariant transformation.** For example, the metric tensor, written as $g_{ij}dx^i \otimes dx^j$,[3] with $g_{ij} = \langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \rangle$ being the inner product of those two basis vectors, operates on pairs of tangent vectors. It therefore transforms doubly covariantly, that is, becomes

$$g_{ij}(f(y))\frac{\partial x^i}{\partial y^\alpha}\frac{\partial x^j}{\partial y^\beta}dy^\alpha \otimes dy^\beta. \tag{1.1.20}$$

The purpose of the metric tensor is to provide a Euclidean product of tangent vectors,

$$\langle V, W\rangle = g_{ij}v^i w^j \tag{1.1.21}$$

for $V = v^i \frac{\partial}{\partial x^i}$, $W = w^i \frac{\partial}{\partial x^i}$. As a check, in this formula, $v^i$ and $w^i$ transform contravariantly, while $g_{ij}$ transforms doubly covariantly, so that the product as a scalar quantity remains invariant under coordinate transformations.

Similarly, we obtain the product of two covectors $\omega, \alpha \in T_x^\star M$ as

$$\langle \omega, \alpha\rangle = g^{ij}\omega_i\alpha_j. \tag{1.1.22}$$

We next introduce the concept of exterior $p$-forms and put

$$\Lambda^p := \Lambda^p(T_x^\star M) := \underbrace{T_x^\star M \wedge \cdots \wedge T_x^\star M}_{p \text{ times}} \quad \text{(exterior product)}. \tag{1.1.23}$$

On $\Lambda^p(T_x^\star M)$, we have the exterior product with $\eta \in T_x^\star M = \Lambda^1(T_x^\star M)$:

$$\begin{aligned} \Lambda^p(T_x^\star M) &\longrightarrow \Lambda^{p+1}(T_x^\star M) \\ \omega &\longmapsto \epsilon(\eta)\omega := \eta \wedge \omega. \end{aligned} \tag{1.1.24}$$

An exterior $p$-form is a sum of terms of the form

$$\omega(x) = \eta(x)dx^{i_1} \wedge \cdots \wedge dx^{i_p}$$

[3] Subsequently, we shall mostly leave out the symbol $\otimes$, that is, write simply $g_{ij}dx^i dx^j$ in place of $g_{ij}dx^i \otimes dx^j$.

where $\eta(x)$ is a smooth function and $(x^1, \dots, x^d)$ are local coordinates. That is, a $p$-form assigns an element of $\Lambda^p(T_x^\star M)$ to every $x \in M$. The space of exterior $p$-forms is denoted by $\Omega^p(M)$.

When $M$ carries a Riemannian metric $g_{ij} dx^i \otimes dx^j$, the scalar product on the cotangent spaces $T_x^\star M$ induces one on the spaces $\Lambda^p(T_x^\star M)$ by

$$\langle dx^{i_1} \wedge \cdots \wedge dx^{i_p}, dx^{j_1} \wedge \cdots \wedge dx^{j_p} \rangle := \det(\langle dx^{i_\mu}, dx^{j_\nu} \rangle) \tag{1.1.25}$$

and linear extension.

Given a Riemannian metric $g_{ij} dx^i \otimes dx^j$, also, in local coordinates, we can define the volume form

$$dvol_g := \sqrt{\det(g_{ij})} dx^1 \wedge \cdots \wedge dx^d. \tag{1.1.26}$$

This volume form depends on an ordering of the indices $1, 2, \dots, d$ of the local coordinates: since the exterior product is antisymmetric, $dx^i \wedge dx^j = -dx^j \wedge dx^i$, it changes its sign under an odd permutation of the indices. Thus, when we have a coordinate transformation $x = f(y)$ where the Jacobian determinant $\det(\frac{\partial x^i}{\partial y^\alpha})$ is negative, $dvol$ changes its sign; otherwise, it is invariant. Therefore, in order to have a globally defined volume form on the Riemannian manifold $M$, we need to exclude coordinate changes with negative Jacobian. The manifold $M$ is called *oriented* when it can be covered by coordinates such that all coordinate changes have a positive Jacobian. In that case, the volume form is well defined, and we can define the integral of a function $\phi$ on $M$ by

$$\int \phi(x)\, dvol_g(x). \tag{1.1.27}$$

We shall therefore assume the manifold $M$ to be oriented whenever we carry out such an integral. We can then also define the $L^2$-product of $p$-forms $\omega, \alpha \in \Omega^p(M)$:

$$(\omega, \alpha) := \int \langle \omega(x), \alpha(x) \rangle dvol_g(x). \tag{1.1.28}$$

We now assume that the dimension $d = 4$, the case of particular importance for the application of our geometric concepts to physics. Then when $\omega$ is a 2-form, $\omega \wedge \omega$ is a 4-form. We call $\omega$ self-dual or antiself-dual when the $+$ resp. $-$ sign holds in

$$\omega \wedge \omega = \pm \langle \omega, \omega \rangle dvol_g. \tag{1.1.29}$$

When $\omega_+$ is self-dual, and $\omega_-$ antiself-dual, we have

$$\langle \omega_+, \omega_- \rangle = 0 \tag{1.1.30}$$

that is, the spaces of self-dual and antiself-dual forms are orthogonal to each other. Every 2-form $\omega$ on a 4-manifold can be decomposed as the sum of a self-dual and an antiself-dual form,

$$\omega = \omega_+ + \omega_-. \tag{1.1.31}$$

We return to arbitrary dimension $d$.

**Definition 1.1** The *exterior derivative* $d : \Omega^p(M) \to \Omega^{p+1}(M)$ ($p = 0, \dots, \dim M$) is defined through the formula

$$d(\eta(x)dx^{i_1} \wedge \cdots \wedge dx^{i_p}) = \frac{\partial \eta(x)}{\partial x^j} dx^j \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_p} \tag{1.1.32}$$

and extended by linearity to all of $\Omega^p(M)$.

The exterior derivative enjoys the following product rule: If $\omega \in \Omega^p(M)$, $\vartheta \in \Omega^q(M)$, then

$$d(\omega \wedge \vartheta) = d\omega \wedge \vartheta + (-1)^p \omega \wedge d\vartheta, \tag{1.1.33}$$

from the formula $\omega \wedge \vartheta = (-1)^{pq} \vartheta \wedge \omega$ and (1.1.32).

Let $x = f(y)$ be a coordinate transformation,

$$\omega(x) = \eta(x) dx^{i_1} \wedge \cdots \wedge dx^{i_p} \in \Omega^p(M).$$

In the $y$-coordinates, we then have

$$f^*(\omega)(y) = \eta(f(y)) \frac{\partial x^{i_1}}{\partial y^{\alpha_1}} dy^{\alpha_1} \wedge \cdots \wedge \frac{\partial x^{i_p}}{\partial y^{\alpha_p}} dy^{\alpha_p} \tag{1.1.34}$$

which is the transformation formula for $p$-forms. The exterior derivative is compatible with this transformation rule:

$$d(f^*(\omega)) = f^*(d\omega), \tag{1.1.35}$$

which follows from the transformation invariance

$$\frac{\partial \eta(x)}{\partial x^j} dx^j = \frac{\partial \eta(f(y))}{\partial x^j} \frac{\partial f^j}{\partial y^\alpha} dy^\alpha = \frac{\partial \eta(f(y))}{\partial y^\alpha} dy^\alpha. \tag{1.1.36}$$

Thus, $d$ is independent of the choice of coordinates. $d$ satisfies the following important rule:

**Lemma 1.1**

$$d \circ d = 0. \tag{1.1.37}$$

*Proof* We check (1.1.37) for forms of the type

$$\omega(x) = f(x) dx^{i_1} \wedge \cdots \wedge dx^{i_p}$$

from which it extends by linearity to all $p$-forms. Now

$$\begin{aligned} d\circ d(\omega(x)) &= d\left(\frac{\partial f}{\partial x^j}dx^j\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_p}\right) \\ &= \frac{\partial^2 f}{\partial x^j\partial x^k}dx^k\wedge dx^j\wedge dx^{i_1}\wedge\cdots\wedge dx^{i_p}=0, \end{aligned}$$

since $\frac{\partial^2 f}{\partial x^j\partial x^k}=\frac{\partial^2 f}{\partial x^k\partial x^j}$ and $dx^j\wedge dx^k=-dx^k\wedge dx^j$. □

In the preceding, we have presented one possible way of conceptualizing transformations, the one employed by mathematicians: The same point $p$ is written in different coordinate systems $x$ and $y$, which are then functionally related by $x=x(y)$. Another view of transformations, often taken in the physics literature, is to move the point $p$ and consider the induced effect on tensors. Let us discuss the example of a 1-form $\omega(x)dx$. Within the fixed coordinates $x$, we vary the points represented by these coordinates by

$$x\mapsto x+\epsilon\xi(x)=:x+\epsilon\delta x \tag{1.1.38}$$

for some map $\xi$ and some small parameter $\epsilon$, and we want to take the limit $\epsilon\to 0$. We have the induced variation of our 1-form

$$\omega(x)dx\mapsto\omega(x+\epsilon\xi(x))d(x+\epsilon\xi(x))=:\omega(x)+\epsilon\delta\omega(x). \tag{1.1.39}$$

By Taylor expansion, we have

$$\begin{aligned} \omega(x+\epsilon\xi(x))d(x+\epsilon\xi(x)) &= \left(\omega_i(x)+\epsilon\frac{\partial\omega_i}{\partial x^k}\xi^k(x)\right)\left(dx^i+\epsilon\frac{\partial\xi^i}{\partial x^k}dx^k\right) \\ &\quad+\text{higher order terms} \end{aligned} \tag{1.1.40}$$

from which we conclude that for $\epsilon\to 0$

$$\delta\omega=\frac{\partial\omega_i}{\partial x^k}\xi^k dx^i+\omega_i\frac{\partial\xi^i}{\partial x^k}dx^k. \tag{1.1.41}$$

Of course, since $\frac{\partial\xi^i}{\partial x^k}dx^k=d\xi^i$, the last term in (1.1.41) agrees with the one required by (1.1.18).

To put the preceding into a slogan: *For setting up transformation rules in geometry, mathematicians keep the point fixed and change the coordinates, while physicists keep the same coordinates, but move the point around.* The first approach is well suited to identifying invariants, like the curvature tensor. The second one is convenient for computing variations, as in our discussion of actions below.

So far, we have computed derivatives of functions. We have also talked about vector fields $V(x)=v^i(x)\frac{\partial}{\partial x^i}$ as objects that depend differentiably on their arguments $x$. Of course, we can do the same for other tensors, like the metric $g_{ij}(x)dx^i\otimes dx^j$. This naturally raises the question about how to compute their derivatives. This encounters the problem, however, that in contrast to functions, the representation of such tensors depends on the choice of local coordinates, and we have described in some detail that and how they transform under coordinate changes. Precisely because of that transformation, they acquire a coordinate invariant meaning; for example, the operation of a vector on a function or the metric product between two vectors are both independent of the choice of coordinates.

It now turns out that on a differentiable manifold, there is in general no single canonical way of taking derivatives of vector fields or other tensors in an invariant manner. There are, in fact, many such possibilities, and they are called connections or covariant derivatives. Only when we have additional structures, like a Riemannian metric, can we single out a particular covariant derivative on the basis of its compatibility with the metric. For our purposes, however, we also need other covariant derivatives, and therefore, we now develop that notion. We shall treat this issue from a more abstract perspective in Sect. 1.2 below, and so the reader who wants to progress more rapidly can skip the discussion here.

Let $M$ be a differentiable manifold. We recall that $\Gamma(TM)$ denotes the space of vector fields on $M$. An (affine) connection or covariant derivative on $M$ is a linear map

$$\begin{aligned}\nabla : \Gamma(TM) \otimes_{\mathbb{R}} \Gamma(TM) &\to \Gamma(TM),\\ (V, W) &\mapsto \nabla_V W\end{aligned}$$

satisfying:

(i) $\nabla$ is tensorial in the first argument:

$$\begin{aligned}&\nabla_{V_1+V_2} W = \nabla_{V_1} W + \nabla_{V_2} W \quad \text{for all } V_1, V_2, W \in \Gamma(TM),\\ &\nabla_{fV} W = f\nabla_V W \quad \text{for all } f \in C^\infty(M),\ V, W \in \Gamma(TM);\end{aligned}$$

(ii) $\nabla$ is $\mathbb{R}$-linear in the second argument:

$$\nabla_V(W_1 + W_2) = \nabla_V W_1 + \nabla_V W_2 \quad \text{for all } V, W_1, W_2 \in \Gamma(TM)$$

and it satisfies the product rule

$$\nabla_V(fW) = V(f)W + f\nabla_V W \quad \text{for all } f \in C^\infty(M),\ V, W \in \Gamma(TM). \tag{1.1.42}$$

$\nabla_V W$ is called the covariant derivative of $W$ in the direction $V$. By (i), for any $x_0 \in M$, $(\nabla_V W)(x_0)$ only depends on the value of $V$ at $x_0$. By way of contrast, it also depends on the values of $W$ in some neighborhood of $x_0$, as it naturally should as a notion of a derivative of $W$. The example on which this is modeled is the Euclidean connection given by the standard derivatives, that is, for $V = V^i \frac{\partial}{\partial x^i}$, $W = W^j \frac{\partial}{\partial x^j}$,

$$\nabla_V^{eucl} W = V^i \frac{\partial W^j}{\partial x^i} \frac{\partial}{\partial x^j}.$$

However, this is not invariant under nonlinear coordinate changes, and since a general manifold cannot be covered by coordinates with only linear coordinate transformations, we need the above more general and abstract concept of a covariant derivative.

Let $U$ be a coordinate chart in $M$, with local coordinates $x$ and coordinate vector fields $\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^d}$ ($d = \dim M$). We then define the Christoffel symbols of the connection $\nabla$ via

$$\nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} =: \Gamma_{ij}^k \frac{\partial}{\partial x^k}. \tag{1.1.43}$$

Thus,

$$\nabla_V W = V^i \frac{\partial W^j}{\partial x^i} \frac{\partial}{\partial x^j} + V^i W^j \Gamma_{ij}^k \frac{\partial}{\partial x^k}. \tag{1.1.44}$$

In order to understand the nature of the objects involved, we can also leave out the vector field $V$ and consider the covariant derivative $\nabla W$ as a 1-form. In local coordinates

$$\nabla W = W^j_{;i} \frac{\partial}{\partial x^j} dx^i, \tag{1.1.45}$$

with

$$W^j_{;i} := \frac{\partial W^j}{\partial x^i} + W^k \Gamma^j_{ik}. \tag{1.1.46}$$

If we change our coordinates $x$ to coordinates $y$, then the new Christoffel symbols,

$$\nabla_{\frac{\partial}{\partial y^l}} \frac{\partial}{\partial y^m} =: \tilde{\Gamma}^n_{lm} \frac{\partial}{\partial y^n}, \tag{1.1.47}$$

are related to the old ones via

$$\tilde{\Gamma}^n_{lm}(y(x)) = \left\{ \Gamma^k_{ij}(x) \frac{\partial x^i}{\partial y^l} \frac{\partial x^j}{\partial y^m} + \frac{\partial^2 x^k}{\partial y^l \partial y^m} \right\} \frac{\partial y^n}{\partial x^k}. \tag{1.1.48}$$

In particular, due to the term $\frac{\partial^2 x^k}{\partial y^l \partial y^m}$, the Christoffel symbols do not transform as a tensor. However, if we have two connections ${}^1\nabla, {}^2\nabla$, with corresponding Christoffel symbols ${}^1\Gamma^k_{ij}, {}^2\Gamma^k_{ij}$, then the difference ${}^1\Gamma^k_{ij} - {}^2\Gamma^k_{ij}$ does transform as a tensor. Expressed more abstractly, this means that the space of connections on $M$ is an affine space.

For a connection $\nabla$, we define its torsion tensor via

$$T(V, W) := \nabla_V W - \nabla_W V - [V, W] \quad \text{for } V, W \in \Gamma(TM). \tag{1.1.49}$$

Inserting our coordinate vector fields $\frac{\partial}{\partial x^i}$ as before, we obtain

$$T_{ij} := T\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial x^j} - \nabla_{\frac{\partial}{\partial x^j}} \frac{\partial}{\partial x^i}$$

(since coordinate vector fields commute, i.e., $[\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}] = 0$)

$$(\Gamma^k_{ij} - \Gamma^k_{ji}) \frac{\partial}{\partial x^k}.$$

We call the connection $\nabla$ torsion-free or symmetric if $T \equiv 0$. By the preceding computation, this is equivalent to the symmetry

$$\Gamma^k_{ij} = \Gamma^k_{ji} \quad \text{for all } i, j, k. \tag{1.1.50}$$

Let $c(t)$ be a smooth curve in $M$, and let $V(t) := \dot{c}(t)$ $(= \dot{c}^i(t)\frac{\partial}{\partial x^i}(c(t))$ in local coordinates) be the tangent vector field of $c$. In fact, we should instead write $V(c(t))$ in place of $V(t)$, but we consider $t$ as the coordinate along the curve $c(t)$. Thus, in those coordinates $\frac{\partial}{\partial t} = \frac{\partial c^i}{\partial t}\frac{\partial}{\partial x^i}$, and in the sequel, we shall frequently and implicitly make this identification, that is, switch between the points $c(t)$ on the curve and the corresponding parameter values $t$. Let $W(t)$ be another vector field along $c$, i.e., $W(t) \in T_{c(t)}M$ for all $t$. We may then write $W(t) = \mu^i(t)\frac{\partial}{\partial x^i}(c(t))$ and form

$$\begin{aligned}\nabla_{\dot{c}(t)} W(t) &= \dot{\mu}^i(t)\frac{\partial}{\partial x^i} + \dot{c}^i(t)\mu^j(t)\nabla_{\frac{\partial}{\partial x^i}}\frac{\partial}{\partial x^j} \\ &= \dot{\mu}^i(t)\frac{\partial}{\partial x^i} + \dot{c}^i(t)\mu^j(t)\Gamma^k_{ij}(c(t))\frac{\partial}{\partial x^k}\end{aligned}$$

(the preceding computation is meaningful as we see that it depends only on the values of $W$ along the curve $c(t)$, but not on other values in a neighborhood of a point on that curve).

This represents a (nondegenerate) linear system of $d$ first-order differential operators for the $d$ coefficients $\mu^i(t)$ of $W(t)$. Therefore, for given initial values $\mu^i(0)$, there exists a unique solution $W(t)$ of

$$\nabla_{\dot{c}(t)} W(t) = 0.$$

This $W(t)$ is called the parallel transport of $W(0)$ along the curve $c(t)$. We also say that $W(t)$ is covariantly constant along the curve $c$.

Now, let $W$ be a vector field in a neighborhood $U$ of some point $x_0 \in M$. $W$ is called parallel if for any curve $c(t)$ in $U$, $W(t) := W(c(t))$ is parallel along $c$. This means that for all tangent vectors $V$ in $U$,

$$\nabla_V W = 0,$$

i.e.,

$$\frac{\partial}{\partial x^i} W^k + W^j \Gamma^k_{ij} = 0 \quad \text{identically in } U, \text{ for all } i, k,$$

$$\text{with } W = W^i \frac{\partial}{\partial x^i} \text{ in local coordinates.}$$

This now is a system of $d^2$ first-order differential equations for the $d$ coefficients of $W$, and so, it is overdetermined. Therefore, in general, such $W$ do not exist. Of course, they do exist for the Euclidean connection, because in Euclidean coordinates, the coordinate vector fields $\frac{\partial}{\partial x^i}$ are parallel.

We define the curvature tensor $R$ by

$$R(V, W)Z := \nabla_V \nabla_W Z - \nabla_W \nabla_V Z - \nabla_{[V,W]} Z, \tag{1.1.51}$$

or in local coordinates

$$R^k_{lij} \frac{\partial}{\partial x^k} := R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) \frac{\partial}{\partial x^l} \quad (i, j, l = 1, \dots, d). \tag{1.1.52}$$

The curvature tensor can be expressed in terms of the Christoffel symbols and their derivatives via

$$R^k_{lij} = \frac{\partial}{\partial x^i} \Gamma^k_{jl} - \frac{\partial}{\partial x^j} \Gamma^k_{il} + \Gamma^k_{im} \Gamma^m_{jl} - \Gamma^k_{jm} \Gamma^m_{il}. \tag{1.1.53}$$

We also note that, as the name indicates, the curvature tensor $R$ is, like the torsion tensor $T$, but in contrast to the connection $\nabla$ represented by the Christoffel symbols, a tensor. This means that when one of its arguments is multiplied by a smooth function, we may simply pull out that function without having to take a derivative of it. Equivalently, it transforms as a tensor under coordinate changes; here, the upper index $k$ stands for an argument that transforms as a vector, that is contravariantly, whereas the lower indices $l, i, j$ express a covariant transformation behavior. The curvature tensor will be discussed in more detail in Sect. 1.1.5.

A curve $c(t)$ in $M$ is called autoparallel or geodesic if

$$\nabla_{\dot{c}} \dot{c} = 0. \tag{1.1.54}$$

Geodesics will be discussed in detail and from a different perspective in Sect. 1.1.4. Here, we only display their equation and define the exponential map. In local coordinates, (1.1.54) becomes

$$\ddot{c}^k(t) + \Gamma^k_{ij}(c(t)) \dot{c}^i(t) \dot{c}^j(t) = 0 \quad \text{for } k = 1, \dots, d. \tag{1.1.55}$$

This constitutes a system of second-order ODEs, and given $x_0 \in M$, $V \in T_{x_0}M$, there exist a maximal interval $I_V \subset \mathbb{R}$ containing an open neighborhood of 0 and a geodesic

$$c_V : I_V \to M$$

with $c_V(0) = x_0$, $\dot{c}_V(0) = V$. We can then define the exponential map $\exp_{x_0}$ on some star-shaped neighborhood of $0 \in T_{x_0}M$:

$$\begin{aligned} \exp_{x_0} : \{V \in T_{x_0}M : 1 \in I_V\} &\to M, \\ V &\mapsto c_V(1). \end{aligned} \tag{1.1.56}$$

We then have $\exp_{x_0}(tV) = c_V(t)$ for $0 \leq t \leq 1$.

A submanifold $S$ of $M$ is called autoparallel or totally geodesic if for all $x_0 \in S$, $V \in T_{x_0}S$ for which $\exp_{x_0} V$ is defined, we have

$$\exp_{x_0} V \in S.$$

The infinitesimal condition needed for this property is that

$$\nabla_V W(x) \in T_x S$$

for any vector field $W(x)$ tangent to $S$ and $V \in T_x S$.

Now, let $M$ carry a Riemannian metric $g = \langle \cdot, \cdot \rangle$.

We say that $\nabla$ is a Riemannian connection if it satisfies the metric product rule

$$Z\langle V, W\rangle = \langle \nabla_Z V, W\rangle + \langle V, \nabla_Z W\rangle. \tag{1.1.57}$$

For any Riemannian metric $g$, there exists a unique torsion-free Riemannian connection, the so-called Levi-Civitá connection $\nabla^g$. It is given by

$$\begin{aligned}\langle \nabla^g_V W, Z\rangle = \frac{1}{2}\{&V\langle W, Z\rangle - Z\langle V, W\rangle + W\langle Z, V\rangle \\ &- \langle V, [W, Z]\rangle + \langle Z, [V, W]\rangle + \langle W, [Z, V]\rangle\}.\end{aligned} \tag{1.1.58}$$

The Christoffel symbols of $\nabla^g$ can be expressed through the metric; in local coordinates, with $g_{ij} = \langle \frac{\partial}{\partial x^i} \frac{\partial}{\partial x^j}\rangle$, we use the abbreviation

$$g_{ij,k} := \frac{\partial}{\partial x^k} g_{ij} \tag{1.1.59}$$

and have

$$\Gamma^k_{ij} = \frac{1}{2} g^{kl}(g_{il,j} + g_{jl,i} - g_{ij,l}), \tag{1.1.60}$$

or, equivalently,

$$g_{ij,k} = g_{jl}\Gamma^l_{ik} + g_{il}\Gamma^l_{jk} = \Gamma_{ikj} + \Gamma_{jki}. \tag{1.1.61}$$

The Levi-Civitá connection $\nabla^g$ respects the metric in the sense that if $V(t)$, $W(t)$ are parallel vector fields along a curve $c(t)$, then

$$\langle V(t), W(t)\rangle \equiv \text{const}, \tag{1.1.62}$$

that is, products between tangent vectors remain invariant under parallel transport.

### *1.1.2 Complex Manifolds*

We start with complex dimension 1. The Euclidean space $\mathbb{R}^2$ can be made into the complex vector space $\mathbb{C}^1$ on which multiplication by complex numbers of the form $a + ib$ is defined, with $i = \sqrt{-1}$. Conventions:

$$z = x + iy = x^1 + ix^2, \quad \bar{z} = x - iy. \tag{1.1.63}$$

In the physics literature, $z$ and $\bar{z}$ are formally viewed as independent coordinates. We define

$$\partial_z := \frac{\partial}{\partial z} = \frac{1}{2}(\partial_x - i\partial_y), \qquad \partial_{\bar{z}} = \frac{\partial}{\partial \bar{z}} = \frac{1}{2}(\partial_x + i\partial_y). \tag{1.1.64}$$

This is arranged so that

$$\partial_z z = 1, \qquad \partial_z \bar{z} = 0, \tag{1.1.65}$$

and so on. A function $f : \mathbb{C} \to \mathbb{C}$ is called holomorphic if

$$\partial_{\bar{z}} f = 0. \tag{1.1.66}$$

Mathematicians write $f(z)$ for any function of the complex variable $z$. Physicists instead write $f(z, \bar{z})$, reserving the notation $f(z)$ for a holomorphic function, that is, one satisfying (1.1.66) because that relation formally expresses independence of the coordinate $\bar{z}$. Similarly, $g : \mathbb{C} \to \mathbb{C}$ is antiholomorphic if

$$\partial_z g = 0. \tag{1.1.67}$$

Another reason for the physics convention is to consider the complexification $\mathbb{C}^2$ with coordinates $(z, z')$ of the Euclidean plane $\mathbb{C} = \mathbb{R}^2$. The slice defined by $\bar{z} = z'$ then yields the Euclidean plane, while $(z, z') = i(s + t, s - t)$ gives the Minkowski plane with metric $dt^2 - ds^2$.

When we use the conformal transformation $z = e^w$, with $w = \tau + i\sigma$, $-\infty < \tau < \infty$ and $0 \le \sigma < 2\pi$, and pass from $w = \tau + i\sigma$ to the light cone coordinates $\zeta^+ = \tau + \sigma$, $\zeta^- = \tau - \sigma$ (a so-called Wick rotation), we obtain the Minkowski metric in the form $d\zeta^+ d\zeta^-$.

In complex coordinates, the Laplace operator (see (1.1.103), (1.1.105) below) becomes

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} = 4\frac{\partial^2}{\partial z \partial \bar{z}}. \tag{1.1.68}$$

We next have the 1-forms

$$dz = dx + i dy, \qquad d\bar{z} = dx - i dy. \tag{1.1.69}$$

This is arranged so that

$$dz(\partial_z) = 1, \qquad dz(\partial_{\bar{z}}) = 0, \tag{1.1.70}$$

and so on, the analogs of (1.1.15). For a vector $v^1 \frac{\partial}{\partial x} + v^2 \frac{\partial}{\partial y}$, we write

$$v^z := v^1 + i v^2, \qquad v^{\bar{z}} := v^1 - i v^2, \tag{1.1.71}$$

and (in flat space)

$$v_z := \frac{1}{2}(v^1 - i v^2), \qquad v_{\bar{z}} := \frac{1}{2}(v^1 + i v^2). \tag{1.1.72}$$

In this notation, the Euclidean (flat) metric on $\mathbb{R}^2$, $g_{11} = g_{22} = 1$, $g_{12} = 0$, becomes

$$g_{z\bar{z}} = g_{\bar{z}z} = \frac{1}{2}, \qquad g_{zz} = g_{\bar{z}\bar{z}} = 0, \qquad g^{z\bar{z}} = g^{\bar{z}z} = 2, \qquad g^{zz} = g^{\bar{z}\bar{z}} = 0. \tag{1.1.73}$$

This is set up to be compatible with (1.1.4). Thus, (1.1.72) becomes a special case of

$$v_z = g_{zz}v^z + g_{z\bar{z}}v^{\bar{z}}. \tag{1.1.74}$$

The area form for this metric is

$$\frac{i}{2}dz \wedge d\bar{z} = dx \wedge dy. \tag{1.1.75}$$

The conventions become clearer when we observe

$$\sqrt{g_{11}g_{22} - g_{12}^2}\,dx \wedge dy = \sqrt{g_{zz}g_{\bar{z}\bar{z}} - g_{z\bar{z}}^2}\,dz \wedge d\bar{z}. \tag{1.1.76}$$

Also, for a twice covariant tensor,

$$\begin{aligned} V_{zz} &= \frac{1}{4}(V_{11} + 2iV_{12} - V_{22}), \qquad V_{\bar{z}\bar{z}} = \frac{1}{4}(V_{11} - 2iV_{12} - V_{22}), \\ V_{z\bar{z}} &= V_{\bar{z}z} = \frac{1}{4}(V_{11} + V_{22}) \end{aligned} \tag{1.1.77}$$

of which (1.1.73) is a special case.

The divergence is (in flat space)

$$\partial_{x^1}v^1 + \partial_{x^2}v^2 = \partial_z v^z + \partial_{\bar{z}}v^{\bar{z}}. \tag{1.1.78}$$

The divergence theorem (integration by parts, a special case of Stokes' theorem) is here

$$\int_\Omega (\partial_z v^z + \partial_{\bar{z}}v^{\bar{z}})\frac{i}{2}dz \wedge d\bar{z} = \frac{i}{2}\oint_{\partial\Omega}(v^z d\bar{z} - v^{\bar{z}}dz) \tag{1.1.79}$$

with a counterclockwise contour integral around $\Omega$.

We now turn to the higher-dimensional situation. The model space is now $\mathbb{C}^d$, the $d$-dimensional complex vector space. The preceding expressions defined for $d = 1$ then get equipped with coordinate indices:

$$z = (z^1, \ldots, z^d), \quad \text{with } z^j = x^j + iy^j \tag{1.1.80}$$

using $(x^1, y^1, \ldots, x^d, y^d)$ as Euclidean coordinates on $\mathbb{R}^{2d}$, and

$$z^{\bar{j}} := x^j - iy^j.$$

Likewise

$$\partial_{\bar{k}} := \frac{\partial}{\partial z^{\bar{k}}} := \frac{1}{2}\left(\frac{\partial}{\partial x^k} + i\frac{\partial}{\partial y^k}\right), \tag{1.1.81}$$

and so on. Then, a function $f : \mathbb{C}^d \to \mathbb{C}$ is holomorphic if

$$\partial_{\bar{k}} f = 0 \tag{1.1.82}$$

for $k = 1, \ldots, d$.

**Definition 1.2** A *complex manifold* of complex dimension $d$ ($\dim_{\mathbb{C}} M = d$) is a differentiable manifold of (real) dimension $2d$ ($\dim_{\mathbb{R}} M = 2d$) whose charts take values in open subsets of $\mathbb{C}^d$ with *holomorphic* coordinate transitions.

A one-dimensional complex manifold is also called a Riemann surface, but that subject will be taken up in more depth in Sect. 1.4.2 below.

Let $M$ again be a complex manifold of complex dimension $d$. Let $T_z^{\mathbb{R}} M := T_z M$ be the ordinary (real) tangent space of $M$ at $z$. We define the complexified tangent space

$$T_z^{\mathbb{C}} M := T_z^{\mathbb{R}} M \otimes_{\mathbb{R}} \mathbb{C} \tag{1.1.83}$$

which we then decompose as

$$T_z^{\mathbb{C}} M = \mathbb{C}\left\{ \frac{\partial}{\partial z^j}, \frac{\partial}{\partial z^{\bar{j}}} \right\} =: T_z' M \oplus T_z'' M, \tag{1.1.84}$$

where $T_z' M = \mathbb{C}\{\frac{\partial}{\partial z^j}\}$ is the holomorphic and $T_z'' M = \mathbb{C}\{\frac{\partial}{\partial z^{\bar{j}}}\}$ the antiholomorphic tangent space. In $T_z^{\mathbb{C}} M$, we have a conjugation mapping $\frac{\partial}{\partial z^j}$ to $\frac{\partial}{\partial z^{\bar{j}}}$, and so $T_z'' M = \overline{T_z' M}$. The same construction is possible for the cotangent space, and we have analogously

$$T_z^{\star\mathbb{C}} M = \mathbb{C}\{dz^j, dz^{\bar{j}}\} =: T_z^{\star'} M \oplus T_z^{\star''} M. \tag{1.1.85}$$

The important point is that these decompositions are invariant under coordinate changes because those coordinate changes are required to be holomorphic. In particular, we have the transformation rules

$$dz^j = \frac{\partial z^j}{\partial w^l} dw^l, \qquad dz^{\bar{k}} = \left( \overline{\frac{\partial z^k}{\partial w^m}} \right) dw^{\bar{m}} = \frac{\partial z^{\bar{k}}}{\partial w^{\bar{m}}} dw^{\bar{m}} \tag{1.1.86}$$

when $z = z(w)$.

The complexified space $\Omega^k(M; \mathbb{C})$ of $k$-forms can be decomposed into subspaces $\Omega^{p,q}(M)$ with $p + q = k$. $\Omega^{p,q}(M)$ is locally spanned by forms of the type

$$\omega(z) = \eta(z) dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q}. \tag{1.1.87}$$

Thus

$$\Omega^k(M) = \bigoplus_{p+q=k} \Omega^{p,q}(M). \tag{1.1.88}$$

We can then let the differential operators

$$\partial = \frac{1}{2}\left(\frac{\partial}{\partial x^j} - i\frac{\partial}{\partial y^j}\right)(dx^j + i dy^j) \quad \text{and}$$
$$\bar{\partial} = \frac{1}{2}\left(\frac{\partial}{\partial x^j} + i\frac{\partial}{\partial y^j}\right)(dx^j - i dy^j) \tag{1.1.89}$$

operate on such a form by

$$\partial\omega = \frac{\partial\eta}{\partial z^i} dz^i \wedge dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q} \tag{1.1.90}$$

and

$$\bar{\partial}\omega = \frac{\partial\eta}{\partial z^{\bar{j}}} dz^{\bar{j}} \wedge dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q}. \tag{1.1.91}$$

$\partial$ and $\bar{\partial}$ yield a decomposition of the exterior derivative $d$:

**Lemma 1.2**

$$d = \partial + \bar{\partial}. \tag{1.1.92}$$

*Moreover,*

$$\partial\partial = 0, \qquad \bar{\partial}\bar{\partial} = 0, \tag{1.1.93}$$
$$\partial\bar{\partial} = -\bar{\partial}\partial. \tag{1.1.94}$$

*Proof*

$$\begin{aligned}\partial + \bar{\partial} &= \frac{1}{2}\left(\frac{\partial}{\partial x^j} - i\frac{\partial}{\partial y^j}\right)(dx^j + i dy^j) + \frac{1}{2}\left(\frac{\partial}{\partial x^j} + i\frac{\partial}{\partial y^j}\right)(dx^j - i dy^j)\\ &= \frac{\partial}{\partial x^j} dx^j + \frac{\partial}{\partial y^j} dy^j = d.\end{aligned}$$

Consequently,

$$0 = d^2 = (\partial + \bar{\partial})(\partial + \bar{\partial}) = \partial^2 + \partial\bar{\partial} + \bar{\partial}\partial + \bar{\partial}^2$$

and decomposing this into types yields (1.1.93), (1.1.94). □

### *1.1.3 Riemannian and Lorentzian Metrics*

In local coordinates $x = (x^1, \ldots, x^d)$, a metric is represented by a nondegenerate, symmetric matrix

$$(g_{ij}(x))_{i,j=1,\ldots,d} \tag{1.1.95}$$

smoothly depending on $x$. Being symmetric, this matrix has $d$ real eigenvalues, and being nondegenerate, none of them is 0. When they are all positive, the metric is called Riemannian. When only one is positive, and therefore $d-1$ ones are negative, it is called Lorentzian.[4] The prototype of a Riemannian manifold is Euclidean space, $\mathbb{R}^d$ equipped with its Euclidean metric; the model for a Lorentz manifold is Minkowski space, namely $\mathbb{R}^d$ equipped with the inner product

$$\langle x, y\rangle = x^0 y^0 - x^1 y^1 - \cdots - x^{d-1} y^{d-1}$$

for $x = (x^0, x^1, \ldots, x^{d-1}),\ y = (y^0, y^1, \ldots, y^{d-1})$. (It is customary to use the indices $0, \ldots, d-1$ in place of $1, \ldots, d$ in the Lorentzian case, in order to better distinguish the time direction corresponding to 0 from the spatial ones.) This space is often denoted by $\mathbb{R}^{1,d-1}$.

The product of two tangent vectors $v, w \in T_p M$ with coordinate representations $(v^1, \ldots, v^d)$ and $(w^1, \ldots, w^d)$ (i.e. $v = v^i \frac{\partial}{\partial x^i}, w = w^j \frac{\partial}{\partial x^j}$) is then, as in (1.1.21),

$$\langle v, w\rangle := g_{ij}(x(p)) v^i w^j. \tag{1.1.96}$$

In particular, $\langle \frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\rangle = g_{ij}$. In a Lorentzian manifold, a vector $v$ with $\langle v, v\rangle > 0$ is called time-like, one with $\langle v, v\rangle < 0$ space-like, and a nontrivial one with $\|v\| = 0$ light-like.

A (smooth) curve $\gamma : [a, b] \to M$ ($[a, b]$ a closed interval in $\mathbb{R}$) is called time-like when $\langle \dot\gamma(t), \dot\gamma(t)\rangle > 0$ for all $t \in [a, b]$. Light- or space-like curves are defined analogously.

Similarly, the length or norm of $v$ is given by

$$\|v\| := \langle v, v\rangle^{\frac{1}{2}} \tag{1.1.97}$$

if $\langle v, w\rangle \geq 0$, and

$$\|v\| := -(-\langle v, v\rangle)^{\frac{1}{2}} \tag{1.1.98}$$

if $\langle v, w\rangle < 0$. On a Riemannian manifold, of course all vectors $v \neq 0$ have positive length.

Starting from the product (1.1.96), a metric then also induces products on other tensors. For example, for cotangent vectors $\omega = \omega_i dx^i, \lambda = \lambda_i dx^i \in T_p^* M$, we have

$$\langle \omega, \lambda\rangle = g^{ij}(x(p)) \omega_i \lambda_j, \tag{1.1.99}$$

[4] The conventions are not generally agreed upon in the literature (see [81] for a systematic survey of the older literature). The one employed here seems to be the one followed by the majority of physicists. Sometimes, however, for a Lorentzian metric, one requires $d-1$ positive and 1 negative eigenvalues. Of course, this simply changes the convention adopted here by a minus sign, without affecting the geometric or physical content. The latter convention looks natural when one wants to add a temporal dimension to already present spatial ones. The convention adopted here, in contrast, is natural when one starts with kinetics described by ordinary differential equations derived from a positive definite Lagrangian. Thus, the temporal dimension is the primary one and counted positively, whereas the additional spatial ones then lead to field theories.

that is, the induced product on the cotangent space is given by the inverse of the metric tensor. As a check, the reader should verify that this expression is invariant under coordinate transformations, with the transformation behavior of the metric now presented (or recalled from (1.1.20)).

Let $y = f(x)$. $v$ and $w$ then have representations $(\tilde{v}^1, \ldots, \tilde{v}^d)$ and $(\tilde{w}, \ldots, \tilde{w}^d)$ with $\tilde{v}^j = v^i \frac{\partial f^j}{\partial x^i}$, $\tilde{w}^j = w^i \frac{\partial f^j}{\partial x^i}$. The metric in the new coordinates, denoted by $h_{k\ell}(y)$, then satisfies

$$h_{k\ell}(f(x))\tilde{v}^k \tilde{w}^\ell = \langle v, w\rangle = g_{ij}(x) v^i w^j. \tag{1.1.100}$$

Therefore, the transformation rule is the one given in (1.1.20),

$$h_{k\ell}(f(x)) \frac{\partial f^k}{\partial x^i} \frac{\partial f^\ell}{\partial x^j} = g_{ij}(x). \tag{1.1.101}$$

Given a metric $(g_{ij}(x))_{i,j=1,\ldots,d}$, we put

$$\sqrt{g} := \sqrt{\det(g_{ij})} \tag{1.1.102}$$

and define the Laplace–Beltrami operator (Laplacian for short) acting on $C^\infty(M)$ as

$$\Delta := \Delta_g := \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^i}\left(\sqrt{g} g^{ij} \frac{\partial}{\partial x^j}\right). \tag{1.1.103}$$

We assume that our manifold $M$ is compact (and, as always, without boundary). We then have the integration by parts formula, using $\langle ., .\rangle$ for the product on 1-forms induced by the Riemannian metric $g$,

$$\begin{aligned} \int \langle df, dg\rangle \sqrt{g} dx^1 \cdots dx^d &= \int g^{ij} \frac{\partial f}{\partial x^i} \frac{\partial g}{\partial x^j} \sqrt{g} dx^1 \cdots dx^d \\ &= -\int f \Delta g \sqrt{g} dx^1 \cdots dx^d \end{aligned} \tag{1.1.104}$$

where $\sqrt{g} dx^1 \ldots dx^d$ is the volume form $dvol_g$ for the Riemannian metric as defined in (1.1.26). (Note that we are always assuming that our manifold $M$ is oriented. This avoids sign ambiguities in the volume form and permits global integration as in (1.1.104).)

In the Euclidean case, the Laplacian is simply the sum of the pure second derivatives,

$$\Delta = \sum_{i=1}^{d} \frac{\partial^2}{(\partial x^i)^2} \tag{1.1.105}$$

(cf. also (1.1.68) above). For the Minkowski metric, we have

$$\Delta = \frac{\partial^2}{\partial (x^0)^2} - \sum_{i=1}^{d-1} \frac{\partial^2}{\partial (x^i)^2}, \tag{1.1.106}$$

and this operator is often denoted by $\Box$ in the literature.

Generalizing (1.1.98), the metric $g$ induces a product $\langle \omega, \nu \rangle$ on $p$-forms, see (1.1.25), and we can then define the formal adjoint $d^*$ of the exterior derivative $d$ via

$$\int \langle d\mu, \nu \rangle dvol_g = \int \langle \mu, d^* \nu \rangle \, dvol_g \tag{1.1.107}$$

for a $(p-1)$-form $\mu$ and a $p$-form $\nu$. (Since $d : \Omega^p(M) \to \Omega^{p+1}(M)$, i.e., $d$ maps $p$-forms to $(p+1)$-forms, $d^* : \Omega^{p+1}(M) \to \Omega^p(M)$ maps $(p+1)$-forms to $p$-forms.) On functions, we then have

$$\Delta f = -d^* df. \tag{1.1.108}$$

More generally, one defines the Hodge Laplacian on $p$-forms by

$$dd^* + d^*d. \tag{1.1.109}$$

Since $d^* f = 0$ for functions, i.e, 0-forms $f$ (for the simple reason that there do not exist forms of degree $-1$), this is a generalization of (1.1.108)—up to the sign, and these differing sign conventions unfortunately cause a lot of confusion. We then have the general integration by parts formulae for $p$-forms

$$\int \langle d^* d\mu, \nu \rangle \, dvol_g = \int \langle d\mu, d\nu \rangle \, dvol_g = \int \langle \mu, d^* d\nu \rangle \, dvol_g \tag{1.1.110}$$

and

$$\begin{aligned} \int \langle (dd^* + d^*d)\mu, \nu \rangle \, dvol_g &= \int (\langle d\mu, d\nu \rangle + \langle d^*\mu, d^*\nu \rangle) \, dvol_g \\ &= \int \langle \mu, (dd^* + d^*d)\nu \rangle \, dvol_g. \end{aligned} \tag{1.1.111}$$

Let us briefly explain the relation with the cohomology of the (compact, oriented) manifold $M$. A $p$-form $\omega$ is called closed if

$$d\omega = 0, \tag{1.1.112}$$

and it is called exact if there exists some $(p-1)$-form $\eta$ with

$$\omega = d\eta. \tag{1.1.113}$$

Because of $d \circ d = 0$, see (1.1.37), any exact form is closed. Two closed $p$-forms $\omega^1, \omega^2$ are considered as cohomologically equivalent if their difference is exact, i.e., if there exists some $(p-1)$-form $\eta$ with

$$\omega^1 - \omega^2 = d\eta. \tag{1.1.114}$$

The equivalence classes of $p$-forms constitute a group, the $p$th (de Rham) cohomology group $H^p(M)$ of $M$. When $M$ carries a Riemannian metric $g$, one can identify a natural representative for each cohomology class as the unique form $\mu$ that minimizes

$$\int_M \langle \omega, \omega \rangle dvol_g \tag{1.1.115}$$

in its equivalence class. This minimizing form $\mu$ is then harmonic in the sense that

$$(dd^* + d^*d)\mu = 0, \tag{1.1.116}$$

or equivalently (as follows from (1.1.111) and the nonnegativity of the two terms in the middle integral)

$$d\mu = 0 \quad \text{and} \quad d^*\mu = 0. \tag{1.1.117}$$

Thus, a harmonic form is closed ($d\mu = 0$) and coclosed ($d^*\mu = 0$).

Since $M$ is compact, the dimension $b_p(M)$ (called the $p$th Betti number of $M$) of $H^p(M)$ is finite. This follows for instance from the fact that the elements of $H^p(M)$ are identified with the solutions of the elliptic differential equation (1.1.116). It is a general result in the theory of elliptic partial differential equations that their solution spaces satisfy a compactness principle.

### *1.1.4 Geodesics*

The length of a smooth (or at least rectifiable) curve $\gamma : [a, b] \to M$ is

$$L(\gamma) := \int_a^b \left\| \frac{d\gamma}{dt}(t) \right\| dt = \int_a^b \sqrt{g_{ij}(x(\gamma(t)))\dot{x}^i(t)\dot{x}^j(t)}dt \tag{1.1.118}$$

where we abbreviate $\dot{x}^i(t) := \frac{d}{dt}(x^i(\gamma(t)))$. Thus, time-, light-, or space-like curves have positive, vanishing, or negative length, respectively

The action of a time-like curve $\gamma$ is

$$S(\gamma) := \frac{1}{2}\int_a^b \left\| \frac{d\gamma}{dt}(t) \right\|^2 dt = \frac{1}{2}\int_a^b g_{ij}(x(\gamma(t)))\dot{x}^i(t)\dot{x}^j(t)dt. \tag{1.1.119}$$

Here, $\gamma$ is considered as the orbit of a mass point, which explains the name "action". In the mathematical literature, the action is often called energy, an unfortunate choice of terminology.

A massive particle in a Lorentzian manifold travels along a world line $x(\tau)$ with arclength

$$s = \int_{\tau_0}^{\tau_1} \left(g_{\alpha\beta}(x(\tau))\dot{x}^\alpha(\tau)\dot{x}^\beta(\tau)\right)^{\frac{1}{2}} d\tau,$$

where we assume $g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta > 0$ along the world line. Thus, the movement is time-like. When in place of $g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta > 0$, we have

$$g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0,$$

then the particle is massless, that is, a photon. $g_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta < 0$ would correspond to a movement with speed higher than that of light and is excluded.

By Hölder's inequality, for a time-like curve $\gamma$,

$$\int_a^b \left\| \frac{d\gamma}{dt} \right\| dt \leq (b-a)^{\frac{1}{2}} \left( \int_a^b \left\| \frac{d\gamma}{dt} \right\|^2 dt \right)^{\frac{1}{2}} \tag{1.1.120}$$

with equality precisely if $\| \frac{d\gamma}{dt} \| \equiv \text{const}$. This means that

$$L(\gamma)^2 \leq 2(b-a)S(\gamma), \tag{1.1.121}$$

again with equality only if $\gamma$ has constant norm.

The distance between $p, q \in M$ is

$$d(p,q) := \inf\{L(\gamma) : \gamma : [a,b] \to M \text{ with } \gamma(a) = p, \gamma(b) = q\}. \tag{1.1.122}$$

By the change of variables formula, if $\gamma : [a,b] \to M$ is a curve, and $\sigma : [a', b'] \to [a,b]$ is a change of parameter, then

$$L(\gamma \circ \sigma) = L(\gamma). \tag{1.1.123}$$

This is no longer so for the action, as follows with a little reflection on the equality discussion in (1.1.121). It is instructive to look at the stationary points of the action:

**Lemma 1.3** *The Euler–Lagrange equations* (*see Sect.* 2.3.1 *below*) *for the action S are*

$$\ddot{x}^i(t) + \Gamma^i_{jk}(x(t))\dot{x}^j(t)\dot{x}^k(t) = 0, \quad i = 1, \dots, d, \tag{1.1.124}$$

*where* $\Gamma^i_{jk}$ *are the Christoffel symbols* (1.1.60).

*Proof* As will be derived in Sect. 2.3.1 below, the Euler–Lagrange equations of a functional

$$I(x) = \int_a^b f(t, x(t), \dot{x}(t)) dt$$

are given by

$$\frac{d}{dt}\frac{\partial f}{\partial \dot{x}^i} - \frac{\partial f}{\partial x^i} = 0, \quad i = 1, \dots, d.$$

Thus, for our action $S$,

$$\frac{d}{dt}(g_{ik}(x(t))\dot{x}^k(t) + g_{ji}(x(t))\dot{x}^j(t)) - g_{jk,i}(x(t))\dot{x}^j(t)\dot{x}^k(t) = 0$$
$$\text{for } i = 1, \dots, d,$$

hence

$$g_{ik}\ddot{x}^k + g_{ji}\ddot{x}^j + g_{ik,\ell}\dot{x}^\ell \dot{x}^k + g_{ji,\ell}\dot{x}^\ell \dot{x}^j - g_{jk,i}\dot{x}^j \dot{x}^k = 0.$$

Renaming indices and using $g_{ik} = g_{ki}$, we get

$$2g_{\ell m}\ddot{x}^m + (g_{\ell k,j} + g_{j\ell,k} - g_{jk,\ell})\dot{x}^j \dot{x}^k = 0$$

and from this

$$g^{i\ell}g_{\ell m}\ddot{x}^m + \frac{1}{2}g^{i\ell}(g_{\ell k,j} + g_{j\ell,k} - g_{jk,\ell})\dot{x}^j \dot{x}^k = 0.$$

Because of $g^{i\ell}g_{\ell m} = \delta^i_m$ and thus $g^{i\ell}g_{\ell m}\ddot{x}^m = \ddot{x}^i$, (1.1.124) follows. □

**Definition 1.3** A geodesic is a curve $\gamma = [a, b] \to M$ that is a critical point of the action $S$, that is, satisfies (1.1.124).

Briefly interrupting our discussion, we point out that (1.1.124) is the same as (1.1.55). In other words, taking up the discussion at the end of Sect. 1.1.1, for the Levi-Cività connection, the two definitions of a geodesic, being autoparallel as in Sect. 1.1.1, or being a critical point of the action functional $S$ as defined here, are equivalent. In particular, we can also write the geodesic equation invariantly, as in (1.1.54), with a slight change of notation:

$$\nabla_{\frac{d}{dt}}\dot{x} = 0. \tag{1.1.125}$$

We now return to the discussion of geodesics as critical points of $S$. We say that a curve $\gamma$ is parametrized proportionally to arc length if $\langle \dot{x}, \dot{x} \rangle \equiv \text{const}$.

**Lemma 1.4** *Each geodesic is parametrized proportionally to arc length.*

*Proof* For a solution of (1.1.124),

$$\begin{aligned}
\frac{d}{dt}\langle \dot{x}, \dot{x} \rangle &= \frac{d}{dt}(g_{ij}(x(t))\dot{x}^i(t)\dot{x}^j(t)) \\
&= g_{ij}\ddot{x}^i\dot{x}^j + g_{ij}\dot{x}^i\ddot{x}^j + g_{ij,k}\dot{x}^i\dot{x}^j\dot{x}^k \\
&= -(g_{jk,\ell} + g_{\ell j,k} - g_{\ell k,j})\dot{x}^\ell\dot{x}^k\dot{x}^j + g_{\ell j,k}\dot{x}^k\dot{x}^\ell\dot{x}^j \\
&= 0, \quad \text{since } g_{jk,\ell}\dot{x}^\ell\dot{x}^k\dot{x}^j = g_{\ell k,j}\dot{x}^\ell\dot{x}^k\dot{x}^j.
\end{aligned}$$

Consequently, $\langle \dot{x}, \dot{x} \rangle \equiv \text{const.}$, and hence the curve is parametrized proportionally to arc length. □

As already discussed in Sect. 1.1.1, the next result follows from the Picard–Lindelöf theorem about the local existence and uniqueness of solutions of systems of ODEs.

**Lemma 1.5** *For each $p \in M$, $v \in T_pM$, there exist $\varepsilon > 0$ and precisely one geodesic*

$$c : [0, \varepsilon] \to M$$

*with $c(0) = p$ and $\dot{c}(0) = v$. This geodesic $c$ depends smoothly on $p$ and $v$.*

We now assume that the metric $g$ on $M$ is Riemannian, even though results corresponding to those stated below also hold in the case of other signatures, in particular for Lorentzian metrics.

If $x(t)$ is a solution of (1.1.124), so is $x(\lambda t)$ for any constant $\lambda \in \mathbb{R}$. Denoting the geodesic of Lemma 1.5 by $c_v$,

$$c_v(t) = c_{\lambda v}\left(\frac{t}{\lambda}\right) \quad \text{for } \lambda > 0,\ t \in [0, \varepsilon].$$

In particular, $c_{\lambda v}$ is defined on $[0, \frac{\varepsilon}{\lambda}]$.

Since $c_v$ depends smoothly on $v$, and $\{v \in T_pM : \|v\| = 1\}$ is compact, there exists $\varepsilon_0 > 0$ with the property that for $\|v\| = 1$, $c_v$ is defined at least on $[0, \varepsilon_0]$. Therefore, for any $w \in T_pM$ with $\|w\| \leq \varepsilon_0$, $c_w$ is defined at least on $[0, 1]$. Thus, as in (1.1.56):

**Definition 1.4** Let $p \in M$, $V_p := \{v \in T_pM : c_v \text{ is defined on } [0, 1]\}$.

$$\begin{aligned} \exp_p : V_p &\to M, \\ v &\mapsto c_v(1) \end{aligned} \tag{1.1.126}$$

is called the exponential map of $M$ at $p$.

One observes that the derivative of the exponential map $\exp_p$ at $0 \in T_pM$ is the identity. Therefore, with the help of the inverse function theorem, one checks that the exponential map $\exp_p$ maps a neighborhood of $0 \in T_pM$ diffeomorphically onto a neighborhood of $p \in M$. Since $T_pM$ is a vector space isomorphic to $\mathbb{R}^d$ (on which we choose a Euclidean orthonormal basis), we can consider the local inverse $\exp_p^{-1}$ as defining local coordinates in a neighborhood of $p$. These local coordinates are called normal coordinates with center $p$. In these coordinates, a basis of $T_pM$ that is orthonormal with respect to the Riemannian metric $g$ is identified with a Euclidean orthonormal basis of $\mathbb{R}^d$. This is the first part of the next lemma:

**Lemma 1.6** *In normal coordinates, the metric satisfies*

$$g_{ij}(0) = \delta_{ij}, \tag{1.1.127}$$

$$\Gamma^i_{jk}(0) = 0 \quad (\textit{and also } g_{ij,k}(0) = 0) \quad \textit{for all } i, j, k. \tag{1.1.128}$$

*Proof* (1.1.127) follows from the fact that the above identification $\Phi : T_pM \cong \mathbb{R}^d$ maps an orthonormal basis of $T_pM$ w.r.t. the metric $g$ (that is, a basis $e_1, \ldots e_d$ with

$\langle e_i, e_j \rangle = \delta_{ij}$ onto an orthonormal basis of $\mathbb{R}^d$. For (1.1.128), in normal coordinates, the straight lines through the origin of $\mathbb{R}^d$ are geodesic, as the line $tv$, $t \in \mathbb{R}$, $v \in \mathbb{R}^d$, is mapped onto $c_{tv}(1) = c_v(t)$, where $c_v(t)$ is the geodesic, parametrized by arc length, with $\dot{c}_v(0) = v$.

Inserting now $x(t) = tv$ into the geodesic equation (1.1.124), we obtain, using $\ddot{x}(t) = 0$,

$$\Gamma^i_{jk}(tv) v^j v^k = 0, \quad \text{for } i = 1, \dots, d.$$

In particular at 0, i.e., for $t = 0$,

$$\Gamma^i_{jk}(0) v^j v^k = 0 \quad \text{for all } v \in \mathbb{R}^d,\ i = 1, \dots, d.$$

Using the symmetry $\Gamma^i_{jk} = \Gamma^i_{kj}$, this implies

$$\Gamma^i_{\ell m}(0) = 0$$

for all $i$ and also for all $\ell, m$. By definition of $\Gamma^i_{jk}$, at $0 \in \mathbb{R}^d$, we obtain

$$g^{i\ell}(g_{j\ell,k} + g_{k\ell,j} - g_{jk,\ell}) = 0$$

for all free indices, hence also

$$g_{jm,k} + g_{km,j} - g_{jk,m} = 0.$$

Permuting the indices yields

$$g_{kj,m} + g_{mj,k} - g_{km,j} = 0,$$

which we add to obtain, for all indices,

$$g_{jm,k}(0) = 0. \qquad \square$$

This is a very useful result. When one has to check tensor equations, one can do this in arbitrary coordinates because by the definition of a tensor, results are coordinate independent. Now, it is often much easier to check such identities in normal coordinates at the point under consideration, making use of the vanishing of all first derivatives of the metric and all Christoffel symbols. We shall often employ this strategy in the sequel.

In fact, we can even achieve a little more: Let $c(s) : (-a, a) \to M$ be a geodesic parametrized by arclength, that is, $\langle \dot{c}(s), \dot{c}(s) \rangle = 1$ for $-a < s < a$ (see Lemma 1.4). Let $v^1(0), \dots, v^d(0)$ be an orthonormal basis of $T_{c(0)}M$ with $v^1 = \dot{c}(0)$, and let $v^i(t) \in T_{c(t)}M$ be the parallel transport of $v^i(0)$ along the geodesic $c(s)$. We define coordinates by mapping $(x^1, \dots, x^d)$ in some neighborhood of $0 \in \mathbb{R}^d$ to

$$(c(x^1), \exp_{c(x^1)}(x^2 v^2(x^1) + \cdots + x^d v^d(x^1))). \tag{1.1.129}$$

**Lemma 1.7** *The coordinates just described satisfy*

$$g_{ij}(x^1, 0, \dots, 0) = \delta_{ij}, \tag{1.1.130}$$

$$\Gamma^i_{jk}(x^1, 0, \dots, 0) = 0, \tag{1.1.131}$$

$$\text{(and also } g_{ij,k}(x^1, 0, \dots, 0) = 0) \tag{1.1.132}$$

*for all* $-a < x^1 < a, i, j, k$.

*Proof* By Lemma 1.4, $g_{11}(x^1, 0, \dots, 0)$ is constant, in fact $\equiv 1$ by our arclength assumption, as a function of $x^1$. Therefore, also $g_{11,1}(x^1, 0, \dots, 0) = 0$. Moreover, since the Levi-Cività connection $\nabla$ respects the metric (see (1.1.62)), $g_{jk}(x^1, 0, \dots, 0)$
$= \langle v^i(x^1), v^j(x^1)\rangle = \delta_{jk}$ for the other values of $j, k$. Therefore, also

$$g_{jk,1}(x^1, 0, \dots, 0) = 0 \quad \text{for all } j, k. \tag{1.1.133}$$

We continue to evaluate all expressions at $(x^1, 0, \dots, 0)$. All rays $tv$ for $v$ in the span of $v^2, \dots, v^d$ are mapped to geodesics, because the exponential map is applied to them. So, we obtain, as in the proof of Lemma 1.6, that

$$\Gamma^i_{\ell m}(x^1, 0, \dots, 0) = 0$$

for $i = 2, \dots, d$ and all $\ell, m$. By definition of $\Gamma^i_{jk}$, we obtain $g^{i\ell}(g_{j\ell,k} + g_{k\ell,j} - g_{jk,\ell}) = 0$ at $(x^1, 0, \dots, 0) \in \mathbb{R}^d$ for all free indices, hence also $g_{jm,k} + g_{km,j} - g_{jk,m} = 0$ for $m = 2, \dots, d$. Permuting the indices to get $g_{kj,m} + g_{mj,k} - g_{km,j} = 0$, adding these relations and combining them with (1.1.133) finally yields $g_{jm,k}(x^1, 0, \dots, 0) = 0$ for all indices. □

### *1.1.5 Curvature*

We now want to discuss the curvature tensor $R$ of the Levi-Cività connection $\nabla$. We recall (1.1.51):

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z. \tag{1.1.134}$$

In local coordinates (cf. (1.1.52)),

$$R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^\ell} = R^k_{\ell ij}\frac{\partial}{\partial x^k}. \tag{1.1.135}$$

We put

$$R_{k\ell ij} := g_{km} R^m_{\ell ij}, \tag{1.1.136}$$

i.e.

$$R_{k\ell ij} = \left\langle R\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^\ell}, \frac{\partial}{\partial x^k}\right\rangle. \tag{1.1.137}$$

There exist different sign conventions for the curvature tensor in the literature. We have adopted here a convention that hopefully minimizes conflict between them. As a consequence, the indices $k$ and $l$ appear in different orders at the two sides of (1.1.137).

The curvature tensor satisfies the following symmetries:

$$R(X,Y)Z = -R(Y,X)Z, \quad \text{i.e.} \quad R_{k\ell ij} = -R_{k\ell ji} \tag{1.1.138}$$

for vector fields $X, Y, Z, W$.

$$R(X,Y)Z + R(Y,Z)X + R(Z,X)Y = 0, \tag{1.1.139}$$

or with indices

$$R_{k\ell ij} + R_{kij\ell} + R_{kj\ell i} = 0 \tag{1.1.140}$$

(the first Bianchi identity).

$$\langle R(X,Y)Z, W\rangle = -\langle R(X,Y)W, Z\rangle, \tag{1.1.141}$$

with indices

$$R_{k\ell ij} = -R_{\ell kij}. \tag{1.1.142}$$

$$\langle R(X,Y)Z, W\rangle = \langle R(Z,W)X, Y\rangle, \tag{1.1.143}$$

with indices

$$R_{k\ell ij} = R_{ijk\ell}. \tag{1.1.144}$$

$$\frac{\partial}{\partial x^h} R_{k\ell ij} + \frac{\partial}{\partial x^k} R_{\ell hij} + \frac{\partial}{\partial x^\ell} R_{hkij} = 0 \tag{1.1.145}$$

(the second Bianchi identity). In order to practice tensor calculus, we give a proof of (1.1.145) in local coordinates. We recall (1.1.53):

$$R^k_{lij} = \frac{\partial}{\partial x^i}\Gamma^k_{jl} - \frac{\partial}{\partial x^j}\Gamma^k_{il} + \Gamma^k_{im}\Gamma^m_{jl} - \Gamma^k_{jm}\Gamma^m_{il}. \tag{1.1.146}$$

Since all expressions are tensors, we may choose normal coordinates around the point $x_0$ under consideration, i.e., for all indices

$$g_{ij}(x_0) = \delta_{ij}, \; g_{ij,k}(x_0) = 0 = \Gamma^k_{ij}(x_0) \tag{1.1.147}$$

(1.1.146) then becomes

$$R_{k\ell ij} = \frac{1}{2}(g_{jk,\ell i} + g_{\ell k,ij} - g_{j\ell,ki} - g_{ik,\ell j} - g_{\ell k,ij} + g_{i\ell,kj})$$

$$= \frac{1}{2}(g_{jk,\ell i} + g_{i\ell,kj} - g_{j\ell,ki} - g_{ik,\ell j}), \tag{1.1.148}$$

and also, differentiating (1.1.146) and using once more the vanishing of all terms containing first derivatives of $g_{ij}$ at $x_0$,

$$R_{k\ell ij,h} = \frac{1}{2}(g_{jk,\ell ih} + g_{i\ell,kjh} - g_{j\ell,kih} - g_{ik,\ell jh}). \tag{1.1.149}$$

This yields the second Bianchi identity:

$$\begin{aligned} R_{k\ell ij,h} + R_{\ell hij,k} + R_{hkij,\ell} &= \frac{1}{2}(g_{jk,\ell ih} + g_{i\ell,kjh} - g_{j\ell,kih} - g_{ik,\ell jh} \\ &\quad + g_{j\ell,hik} + g_{ih,\ell jk} - g_{jh,\ell ik} - g_{i\ell,hjk} \\ &\quad + g_{jh,ki\ell} + g_{ik,hj\ell} - g_{jk,hi\ell} - g_{ih,kj\ell}) \\ &= 0. \end{aligned}$$

The sectional curvature of the plane spanned by the (linearly independent) tangent vectors $X = \xi^i \frac{\partial}{\partial x^i}$, $Y = \eta^i \frac{\partial}{\partial x^i} \in T_x M$ is defined as

$$K(X \wedge Y) := \langle R(X, Y)Y, X\rangle \frac{1}{|X \wedge Y|^2}, \tag{1.1.150}$$

($|X \wedge Y|^2 = \langle X, X\rangle\langle Y, Y\rangle - \langle X, Y\rangle^2$), with indices

$$K(X \wedge Y) = \frac{R_{ijk\ell}\xi^i\eta^j\xi^k\eta^\ell}{g_{ik}g_{j\ell}(\xi^i\xi^k\eta^j\eta^\ell - \xi^i\xi^j\eta^k\eta^\ell)} = \frac{R_{ijk\ell}\xi^i\eta^j\xi^k\eta^\ell}{(g_{ik}g_{j\ell} - g_{ij}g_{k\ell})\xi^i\eta^j\xi^k\eta^\ell}. \tag{1.1.151}$$

The Ricci curvature in the direction $X = \xi^i \frac{\partial}{\partial x^i} \in T_x M$ is defined as the average of the sectional curvatures of all planes in $T_x M$ containing $X$,

$$\mathrm{Ric}(X, X) = g^{j\ell}\left\langle R\left(X, \frac{\partial}{\partial x^j}\right)\frac{\partial}{\partial x^\ell}, X\right\rangle, \tag{1.1.152}$$

and the Ricci tensor is thus the contraction of the curvature tensor,

$$R_{ik} = g^{j\ell}R_{ijk\ell} = R^j_{ijk}. \tag{1.1.153}$$

(1.1.144) implies the symmetry

$$R_{ik} = R_{ki}. \tag{1.1.154}$$

The scalar curvature is the contraction of the Ricci curvature,

$$R = g^{ik}R_{ik} = R^i_i. \tag{1.1.155}$$

For $d = \dim M = 2$, the curvature tensor is determined by the scalar curvature:

$$R_{ijk\ell} = R(g_{ik}g_{j\ell} - g_{ij}g_{k\ell}). \tag{1.1.156}$$

For $d = 3$, the curvature tensor is determined by the Ricci tensor. For $d > 3$, the part of the curvature tensor not yet determined by the Ricci tensor is given by the Weyl tensor

$$W_{ijk\ell} = R_{ijk\ell} + \frac{2}{d-2}(g_{i\ell}R_{kj} - g_{ik}R_{\ell j} + g_{jk}R_{\ell i} - g_{j\ell}R_{ki}) + \frac{2}{(d-1)(d-2)}R(g_{ik}g_{\ell j} - g_{i\ell}g_{kj}). \tag{1.1.157}$$

### *1.1.6 Principles of General Relativity*

General relativity describes the physical force of gravity and its relation with the structure of space–time. The fundamental physical insight behind the theory of general relativity is that the effects of acceleration cannot be distinguished from those of gravity. The presence of matter changes the geometry of space, and acceleration is experienced in relation to that geometry. In particular, the geometry of space and time is dynamically determined by the physical laws, and in contrast to other physical theories, is thus not assumed as independently given. These physical laws in turn are deduced from symmetry principles, more precisely from the principle of general covariance, that is, that the physics should be independent of its coordinate description. For this, Riemannian geometry has developed the appropriate formal tools.

Let $M$ be a Lorentz manifold with local coordinates $(x^0, x^1, x^2, x^3)$ and metric

$$(g_{\alpha\beta})_{\alpha,\beta=0,\dots,3}.$$

We recall the Christoffel symbols

$$\Gamma^\alpha_{\beta\gamma} = \frac{1}{2}g^{\alpha\delta}(g_{\beta\delta,\gamma} + g_{\gamma\delta,\beta} - g_{\beta\gamma,\delta}),$$

and those objects from which the essential invariants of a metric come, that is, the curvature tensor $R^\alpha_{\beta\gamma\delta} = \Gamma^\alpha_{\beta\delta,\gamma} - \Gamma^\alpha_{\beta\gamma,\delta} + \Gamma^\alpha_{\eta\gamma}\Gamma^\eta_{\beta\delta} - \Gamma^\alpha_{\eta\delta}\Gamma^\eta_{\beta\gamma}$, and its contractions, the Ricci tensor (1.1.154), $R_{\alpha\beta} = R^\gamma_{\alpha\gamma\beta}$, and the scalar curvature (1.1.156), $R = g^{\alpha\beta}R_{\alpha\beta}$.

The Einstein field equations couple the metric $g_{\alpha\beta}$ of the underlying differentiable manifold with the matter and fields on that manifold. These equations involve the Ricci curvature and are

$$R^{\alpha\beta} - \frac{1}{2}g^{\alpha\beta}R = \kappa T^{\alpha\beta}. \tag{1.1.158}$$

Here, $\kappa = \frac{8\pi g}{c^4}$ where $g$ is the gravitational constant. $(T^{\alpha\beta})_{\alpha,\beta}$ is the energy–momentum tensor. It describes the matter and fields present. When $T$ is given, the Einstein equations then determine the metric of space–time.[5] The presence of a nonvanishing energy–momentum tensor in the field equations makes space–time curved. The curvature in turn leads to gravity. (1.1.158) is equivalent to

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = \kappa T_{\alpha\beta}. \tag{1.1.159}$$

Taking the trace in (1.1.159) leads to $R - 2R = \kappa T$, that is,

$$R = -\kappa T. \tag{1.1.160}$$

($T = g^{\alpha\beta} T_{\alpha\beta}$; note that the dimension of $M$ is 4.) Using (1.1.160), (1.1.159) becomes equivalent to

$$R_{\alpha\beta} = \kappa \left( T_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} T \right). \tag{1.1.161}$$

In the special case where $(T_{\alpha\beta}) = 0$, that is, when neither matter nor fields are present, (1.1.161) becomes

$$R_{\alpha\beta} = 0, \tag{1.1.162}$$

i.e., the Ricci curvature of $M$ vanishes.

Hilbert discovered that the Einstein field equations can be derived from a variational principle. In fact, they are the Euler–Lagrange equations for the action functional

$$L_0(g) = \int_M R\sqrt{-g}\, dx = \int_M R\, d\mathrm{Vol}_M(x), \tag{1.1.163}$$

called the Einstein–Hilbert functional. To see this, we consider a family

$$g^t_{\alpha\beta} = g_{\alpha\beta} + t h_{\alpha\beta}$$

of metrics with $(h_{\alpha\beta})$ having compact support if $M$ is not compact itself. Quantities obtained from the metric $g^t_{\alpha\beta}$ will always carry a superscript $t$; for example,

$$R^t_{\alpha\beta}$$

is the Ricci tensor of $g^t_{\alpha\beta}$. We also put

$$\delta g_{\alpha\beta} = \frac{d}{dt} (g^t_{\alpha\beta})_{|t=0} = h_{\alpha\beta},$$

$$\delta R_{\alpha\beta} = \frac{d}{dt} (R^t_{\alpha\beta})_{|t=0}, \quad \text{etc.}$$

[5]Classically, the topology of $M$ is assumed fixed. However, it turns out that the equations may lead to space–time singularities, like black holes, which will then affect the underlying topology. Such singularities can occur and are sometimes even inevitable, even if suitable and physically natural restrictions are imposed on the energy–momentum tensor, like nonnegativity. We do not pursue that issue here, however, but refer to [56]. There, also the cosmological implications of such singularities are discussed.

Finally, we shall use the abbreviation

$$\gamma := \sqrt{-g}.$$

We then have

$$\frac{d}{dt} L_0(g^t)_{|t=0} = \int_M \delta(R\gamma)\, dx. \tag{1.1.164}$$

Now

$$\delta(R\gamma) = \delta(g^{\alpha\beta} R_{\alpha\beta}\gamma) = g^{\alpha\beta}\gamma\delta R_{\alpha\beta} + R_{\alpha\beta}\delta(g^{\alpha\beta}\gamma).$$

We now claim that

$$g^{\alpha\beta}\delta R_{\alpha\beta} = \operatorname{div} V \quad \left(= \frac{1}{\gamma}\frac{\partial}{\partial x^\alpha}(\gamma V^\alpha)\right) \tag{1.1.165}$$

for the vector field $V$ with components

$$V^\gamma = g^{\alpha\beta}\delta\Gamma^\gamma_{\alpha\beta} - g^{\gamma\alpha}\delta\Gamma^\beta_{\beta\alpha}. \tag{1.1.166}$$

Proof of (1.1.165): Let $p \in M$. We introduce normal coordinates near $p$; thus, at $p$, the metric tensor is diagonal and

$$g_{\alpha\beta,\gamma}(p) = 0 \quad \text{and} \quad \Gamma^\alpha_{\beta\gamma}(p) = 0 \quad \text{for all } \alpha, \beta, \gamma.$$

In particular, at $p$

$$\frac{\partial}{\partial x^\alpha}\gamma = 0 \quad \text{for all } \alpha.$$

In these coordinates, (1.1.165) then follows from the definition of the Ricci tensor. While the Christoffel symbols $\Gamma^\alpha_{\beta\gamma}$, as the components of a connection, do not transform tensorially, the $\delta\Gamma^\alpha_{\beta\gamma}$ do transform tensorially as derivatives, that is, as infinitesimal differences of connections. The right-hand side of (1.1.165) is thus a tensor, and so is the left-hand side. The equality of two tensors can be checked in arbitrary coordinates. Since we have just verified (1.1.165) in normal coordinates, (1.1.165) then also holds in arbitrary coordinates, and we have completed its proof.

We now get

$$\begin{aligned}\int_M \delta(R\gamma) &= \int g^{\alpha\beta}\gamma\delta R_{\alpha\beta}\, dx + \int R_{\alpha\beta}\delta(g^{\alpha\beta}\gamma)\, dx \\ &= \int \operatorname{div} V\gamma\, dx + \int R_{\alpha\beta}\delta(g^{\alpha\beta}\gamma)\, dx \\ &= \int R_{\alpha\beta}\delta(g^{\alpha\beta}\gamma)\, dx \end{aligned} \tag{1.1.167}$$

by Gauss's theorem, since $V$ has compact support.

Now

$$\delta\gamma = -\frac{1}{2}\gamma^{-1}\delta\det(g_{\alpha\beta}) = \frac{1}{2}\gamma g^{\alpha\beta}\delta g_{\alpha\beta}$$

and moreover

$$\delta g^{\alpha\beta} = -g^{\alpha\gamma} g^{\beta\gamma} \delta g_{\gamma\delta} \quad (\text{from } g^{\alpha\beta} g_{\beta\gamma} = \delta^{\alpha}_{\gamma})$$

and therefore

$$\delta(g^{\alpha\beta}\gamma) = \gamma\left(\frac{1}{2}\, g^{\alpha\beta} g^{\gamma\delta} - g^{\alpha\gamma} g^{\beta\delta}\right)\delta g_{\gamma\delta}. \tag{1.1.168}$$

(1.1.164), (1.1.167), (1.1.168) imply

$$\delta L_0 = \int_M \left(\frac{1}{2} g^{\alpha\beta} R - R^{\alpha\beta}\right)\gamma\,\delta g_{\alpha\beta}\, dx = 0. \tag{1.1.169}$$

If this holds for all variations $\delta g_{\alpha\beta}$ with compact support, we have

$$R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R = 0, \tag{1.1.170}$$

which implies, as in the derivation of (1.1.162), that

$$R_{\alpha\beta} = 0. \tag{1.1.171}$$

Einstein also tentatively introduced a cosmological constant $\Lambda$ that has the effect of changing the Einstein–Hilbert functional (1.1.163) to

$$L_\Lambda(g) = \int_M (R - 2\Lambda)\sqrt{-g}\, dx \tag{1.1.172}$$

and the Einstein field equations (1.1.158) to

$$R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R + \Lambda g^{\alpha\beta} = \kappa T^{\alpha\beta}. \tag{1.1.173}$$

While a nontrivial cosmological constant is presently appearing in some cosmological models, we put it to 0 for our present discussion. It is straightforward, however, to include a nontrivial $\Lambda$ in the subsequent formulas.

In the presence of some matter fields $\phi$, we assume a Lagrangian

$$L_1 = \int_M F(g, \phi, \nabla^g \phi)\sqrt{-g}\, dx \tag{1.1.174}$$

depending on the fields and their covariant derivatives w.r.t. the Levi-Cività connection, as well as possibly also directly on the metric $g$. When we consider a variation $\delta g_{\alpha\beta}$ of the metric that does not change the fields, we put

$$\delta L_1 = \frac{1}{2} \int_M T^{\alpha\beta} \delta g_{\alpha\beta} dx. \tag{1.1.175}$$

In other words, the energy–momentum tensor is defined as the variation of the matter Lagrangian w.r.t. the metric. In order to fully justify (1.1.175), we need to observe that all the variations of all metric dependent terms in $L_1$ are proportional to $\delta g_{\alpha\beta}$. For the volume form, this has been verified in (1.1.168). The covariant derivative $\nabla^g$ occurring in (1.1.175) also depends on the metric (see (1.1.60)). One easily computes, for example in normal coordinates, that the variation $\delta\Gamma^{\alpha}_{\beta\delta}$ of the Christoffel symbol is proportional to a combination of covariant derivatives of $\delta g_{\alpha\beta}$, and that the covariant derivatives can then be integrated away by parts in the computation of $\delta L_1$. When one then considers the full Lagrangian

$$L := L_0 + \kappa L_1, \tag{1.1.176}$$

with $\kappa$ as a coupling constant, we thus obtain from (1.1.169) and (1.1.175), for variations $\delta g_{\alpha\beta}$ of the metric,

$$\delta L = \int_M \left(\frac{1}{2} g^{\alpha\beta} R - R^{\alpha\beta} + \kappa T^{\alpha\beta}\right) \gamma\, \delta g_{\alpha\beta}\, dx. \tag{1.1.177}$$

Thus, when $\delta L = 0$ for all variations of the metric, we obtain (1.1.158).

For a more extended discussion of this variational principle, we refer to the presentation in [81] or [56].

Finally, we mention the so-called semiclassical Einstein equations

$$R_{\alpha\beta} - \frac{1}{2} g_{\alpha\beta} R = \kappa \langle \psi | \hat{T}_{\alpha\beta} | \psi \rangle \tag{1.1.178}$$

where the energy momentum tensor $T_{\alpha\beta}$ in (1.1.159) is replaced by the expectation value of the energy–momentum operator with respect to some quantum state $\psi$. This quantum state in turn depends on the metric $g$ through the Schrödinger equation. Here, we are invoking concepts that find their natural place in the second part of this book. Equation (1.1.178) arises in the context of quantum fields on an external space–time. The coupling of a quantum system to a classical one in (1.1.178) leads to questions of consistency which we do not enter here. We refer to the discussion in [74].

Variational principles will be taken up in more generality below in Sect. 2.3.1, and in Sect. 2.4, the energy–momentum tensor will appear again. Also, we shall see there that a consequence of Noether's theorem (see Sect. 2.3.2) is that under the fundamental assumption of general relativity, namely invariance of $L$ under coordinate transformations—that is, diffeomorphism invariance—the energy–momentum tensor is divergence free.

## 1.2 Bundles and Connections

### *1.2.1 Vector and Principal Bundles*

Let $M$ be a differentiable manifold. In this section, we present the basic aspects of the theory of vector and principal bundles. We point out that we have already studied one particular vector bundle over $M$ in Sect. 1.1.1, its tangent bundle $TM$.

A fiber bundle (or simply, a bundle) over $M$ consists of a total space $E$, a fiber $F$ (both of them also differentiable manifolds), and a projection $\pi : E \to M$ such that each $x \in M$ has a neighborhood $U$ for which $E_{|U} = \pi^{-1}(U)$ is diffeomorphic to $U \times F$ such that the fibers are preserved. This means that there exists a diffeomorphism

$$\varphi : \pi^{-1}(U) \to U \times F$$

with

$$\pi = p_1 \circ \varphi. \tag{1.2.1}$$

($p_1 : U \times F \to U$ is the projection onto the first factor.)

$\varphi$ is called a local trivialization of the bundle over $U$. Let $\{U_\alpha\}$ be an open covering of $M$ with local trivializations $\{\varphi_\alpha\}$. If

$$U_\alpha \cap U_\beta \neq \emptyset,$$

we obtain transition maps

$$\varphi_{\beta\alpha} : U_\alpha \cap U_\beta \to \mathrm{Diff}(F) \quad (= \text{group of diffeomorphisms of } F)$$

via

$$\varphi_\beta \circ \varphi_\alpha^{-1}(x, v) = (x, \varphi_{\beta\alpha}(x)v). \tag{1.2.2}$$

Omitting the base point, which is fixed by (1.2.1), from our notation, we shall usually simply write

$$\varphi_{\beta\alpha} = \varphi_\beta \circ \varphi_\alpha^{-1}.$$

We have

$$\text{for } x \in U_\alpha\colon \quad \varphi_{\alpha\alpha}(x) = \mathrm{id}_F, \tag{1.2.3}$$

$$\text{for } x \in U_\alpha \cap U_\beta\colon \quad \varphi_{\alpha\beta}(x)\varphi_{\beta\alpha}(x) = \mathrm{id}_F, \tag{1.2.4}$$

$$\text{for } x \in U_\alpha \cap U_\beta \cap U_\gamma\colon \quad \varphi_{\alpha\gamma}(x)\varphi_{\gamma\beta}(x)\varphi_{\beta\alpha}(x) = \mathrm{id}_F. \tag{1.2.5}$$

$E$ can be reconstructed from its transition maps:

$$E = \coprod_\alpha U_\alpha \times F / \sim \tag{1.2.6}$$

with

$$(x, v) \sim (y, w) :\Leftrightarrow x = y \quad \text{and} \quad w = \varphi_{\beta\alpha}(x)v$$
$$(x \in U_\alpha, y \in U_\beta, v, w \in F).$$

When we have some (differentiable) $f_\alpha : U_\alpha \to \mathrm{Diff}(F)$, we can replace the trivialization $\varphi_\alpha$ over $U_\alpha$ by

$$\varphi'_\alpha = f_\alpha \circ \varphi_\alpha, \tag{1.2.7}$$

and conversely, we can obtain any trivialization $\varphi'_\alpha$ over $U_\alpha$ in this manner via

$$f_\alpha := \varphi'_\alpha \circ \varphi_\alpha^{-1}$$
$$(\varphi_\alpha^{-1} \text{ assigns to each } x \text{ the diffeomorphism inverse to } \varphi_\alpha(x)). \tag{1.2.8}$$

If $f_\alpha, f_\beta$ are as above, the transition maps change according to

$$\varphi_{\beta\alpha}^{-1} = \varphi'_\beta \circ \varphi'^{\,-1}_\alpha = f_\beta \circ \varphi_{\beta\alpha} \circ f_\alpha^{-1}. \tag{1.2.9}$$

The special case where all transition maps take their values in an *Abelian* subgroup $A$ of $\mathrm{Diff}(F)$ yields some additional structure: The transition maps $\{\varphi_{\beta\alpha}\}$ then define a Čech cocycle on $M$ with values in $A$, because (1.2.4) and (1.2.5) imply

$$\delta(\{\varphi_{\beta\alpha}\}) = 0$$

for the boundary operator $\delta$. By (1.2.9), two such cocycles $\{\varphi_{\beta\alpha}\}$ and $\{\varphi'_{\beta\alpha}\}$ define the same bundle if $\{\varphi_{\beta\alpha}^{-1} \circ \varphi'_{\beta\alpha}\}$ is a coboundary. Thus, in this case, we can consider a bundle as a cohomology class in $H^1(M, A)$.[6]

A section of $E$ is a smooth map

$$s : M \to E$$

satisfying

$$\pi \circ s = \mathrm{id}.$$

We denote the space of sections by $C^\infty(E)$ or $\Gamma(E)$.

For our purposes, we shall only need two special (closely related) types of fiber bundles. The fiber $F$ will be either a vector space $V$ or a Lie group $G$. The important general principle here is to require that the transition maps respect the corresponding structure. Thus, they are not allowed to assume arbitrary values in $\mathrm{Diff}(F)$, but only in some fixed Lie group $G$. $G$ is called the structure group of the bundle.

According to this principle, the fiber of a vector bundle is a real or complex vector space $V$ of some real dimension $n$, and the structure group is $Gl(n, \mathbb{R})$ or some subgroup. A bundle whose fiber is a Lie group $G$ is called a principal bundle, and the total space is denoted by $P$. The structure group is $G$ or some subgroup, and it operates by left multiplication on the fiber $G$. Right multiplication on $G$ induces a right action of $G$ on $P$ via local trivializations:

$$P \times G \to P, \qquad (x, g) * h = (x, gh) \quad \text{for } p = (x, g) \in P,$$

with the composition rule $(p * g)h = p * gh$. This action is free, that is, $p * g = p \Leftrightarrow g = e$ (neutral element). The projection $\pi : P \to M$ is obtained by simply identifying $x \in M$ with an orbit of this action, that is,

$$\pi : P \to P/G = M.$$

The groups $Gl(n, \mathbb{R})$, $O(n)$, $SO(n)$, $U(n)$ and $SU(n)$ will be the ones of interest for us. Acting as linear groups on a vector space, they preserve linear, Euclidean, or Hermitian structures. For example, a Euclidean structure, that is, a (positive definite) scalar product, is an additional structure on a vector space. According to the general principle, if we have such a structure on our fiber, it has to be respected by the transition maps. As before, this restricts the transformations permitted. In our example, we thus allow only $O(n)$ in place of $Gl(n, \mathbb{R})$. Such a restriction of the admissible transformations by imposing an additional structure that has to be preserved is called a reduction of the structure group.

---

[6]We assume here that $M$ is connected; otherwise, in place of $A$ itself, we should utilize the locally constant sheaf of $A$.

Principal and vector bundles are closely related. Let $P \to M$ be a principal bundle with fiber $G$, and let the vector space $V$ carry a representation of $G$. We then construct a vector bundle $E$ with fiber $V$ using the following free right action of $G$ on $P \times V$:

$$P \times V \times G \to P \times V,$$
$$(p, v) * g = (p * g, g^{-1}v).$$

The projection

$$P \times V \to P \to M$$

is invariant under this action, and

$$E := P \times_G V := (P \times V)/G \to M$$

is a vector bundle with fiber

$$G \times_G V = (G \times V)/G = V$$

and structure group $G$. Via the left action of $G$ on $V$, the transition maps for $P$ yield transition maps for $E$. Conversely, if we have a vector bundle with structure group $G$, we construct a $G$-principal bundle $P$ by

$$\coprod_\alpha U_\alpha \times G/\sim$$

with

$$(x_\alpha, g_\alpha) \sim (x_\beta, g_\beta) :\Leftrightarrow x_\alpha = x_\beta \text{ in } U_\alpha \cap U_\beta \quad \text{and} \quad g_\beta = \varphi_{\beta\alpha}(x) g_\alpha.$$

(Here, $\{U_\alpha\}$ is a local trivialization of $E$ with transition maps $\varphi_{\beta\alpha}$; these transition maps are in $Gl(n, \mathbb{R})$. Since the elements $g_\alpha$ are in the structure group $G$ which is assumed to be a linear group, that is, a subgroup of $Gl(n, \mathbb{R})$, we can form the product $\varphi_{\beta\alpha}(x) g_\alpha$.) $P$ can be viewed as the bundle of admissible bases of $E$. In a local trivialization, each fiber of $E$ is identified with $\mathbb{R}^n$ or $\mathbb{C}^n$, and each admissible base is represented by a matrix with coefficients in $\mathbb{R}$ or $\mathbb{C}$. The transition maps then effect a base change. In each local trivialization, the action of $G$ on $P$ is given by matrix multiplication.

All standard operations on vector spaces extend to vector bundles. If we have a vector bundle $E$ with fiber $V_x$ over $x$, we can form the dual bundle $E^\star$ with fiber the dual space $V_x^\star$ of $V_x$. Applying this construction to the tangent bundle $TM$ yields the cotangent bundle $T^\star M$. If $E_1$ and $E_2$ are vector bundles, we can form the bundles $E_1 \oplus E_2$, $E_1 \otimes E_2$ and $E_1 \wedge E_2$ by performing the corresponding operations on the fibers. In particular, from the cotangent bundle $T^\star M$, we obtain the bundle $\Lambda^p(M)$ introduced in (1.1.23), whose sections are the exterior $p$-forms.

### 1.2.2 Covariant Derivatives

Let $E$ be a vector bundle over $M$. We may view $E$ as a family of vector spaces parametrized by $M$. A local trivialization $\varphi$ over $U$ identifies the fibers over $U$ with each other. Changing the local trivialization then also changes this identification of the fibers. The identification thus depends on the choice of a local trivialization and is therefore not canonical. Hence, while we can decide whether a section of $E$ is differentiable because all transition maps depend differentiably on $x$ and therefore do not affect the differentiability of a section in some local trivialization, there is no canonical way to specify the value of its derivative. In particular, we do not have a criterion for a section being constant along a curve in $M$.

Therefore, in order to be able to differentiate a section, we need to introduce and specify an additional structure on $E$, a so-called covariant derivative or connection. We point out that this includes and generalizes the concept of a covariant derivative developed in Sect. 1.1.1 for the tangent bundle.

A covariant derivative is an operator

$$D : \Gamma(E) \to \Gamma(E) \otimes_{C^\infty(M)} \Gamma(T^*M)$$

with the following properties: For $\sigma \in \Gamma(E)$, $V \in T_xM$, we write

$$D\sigma(V) =: D_V\sigma$$

and require (for all $x \in M$):

(i) $D$ is tensorial in $V$:

$$D_{V+W}\sigma = D_V\sigma + D_W\sigma \quad \forall V, W \in T_xM,\ \sigma \in \Gamma(E),$$

$$D_{fV}\sigma = fD_V\sigma \quad \forall V \in T_xM,\ f \in C^\infty(M),\ \sigma \in \Gamma(E).$$

(Remark: It does not really make sense to multiply an element $V \in T_xM$ by a function $f \in C^\infty(M)$. What the preceding rule means is that when we take a section $V \in \Gamma(TM)$ of the tangent bundle, the value $(D_V\sigma)(x)$ depends only on the value of $V$ at the point $x$, but not on the values at other points. This is not so for $\sigma$ as rule (iii) shows.)

(ii) $D$ is linear in $\sigma$:

$$D_V(\sigma + \tau) = D_V\sigma + D_V\tau \quad \forall V \in T_xM,\ \sigma, \tau \in \Gamma(E).$$

(iii) $D$ satisfies the product rule:

$$D_V(f\sigma) = V(f)\sigma + fD_V\sigma \quad \forall V \in T_xM,\ f \in C^\infty(M),\ \sigma \in \Gamma(E).$$

An example, which is not really typical, but in a certain sense a local model, is the trivial bundle $M \times \mathbb{R}$ over $M$, where we can put

$$D_V\sigma := d\sigma(V) = V(\sigma)$$

to obtain a covariant derivative. In the general case, let $\varphi$ be a trivialization of $E$ over $U$,

$$E_{|U} \cong U \times \mathbb{R}^n \quad (=\varphi(\pi^{-1}(U))).$$

Via this local identification, a base of $\mathbb{R}^n$ yields a base $\mu_1, \ldots, \mu_n$ of sections of $E_{|U}$. Any section $\sigma$ can then be written over $U$ as

$$\sigma(x) = a^k(x)\mu_k(x).$$

Then

$$D\sigma = (da^k)\mu_k + a^k D\mu_k. \tag{1.2.10}$$

Since $(\mu_j)_{j=1,\ldots,n}$ is a base of sections, we can write

$$D\mu_k = A_k^j \mu_j. \tag{1.2.11}$$

Here, for each $x$, $A(x) = (A_k^j(x))_{j,k=1,\ldots,n}$ is a $T_x^*M$-valued matrix, that is, an element of $\mathfrak{gl}(n, \mathbb{R}) \otimes T_x^*M$. In symbols,

$$A \in \Gamma\big(\mathfrak{gl}(n, \mathbb{R}) \otimes T^*M_{|U}\big).$$

In our trivialization, we write this as

$$D = d + A. \tag{1.2.12}$$

We now wish to determine the transformation behavior of $A$ under a change of the local trivialization. Let $\{U_\alpha\}$ be an open covering of $M$ that yields a local trivialization with transition maps

$$\varphi_{\beta\alpha} : U_\alpha \cap U_\beta \to Gl(n, \mathbb{R}).$$

$D$ defines a $T^*M$-valued matrix $A_\alpha$ on $U_\alpha$. Let the section $\mu$ be represented by $\mu_\alpha$ on $U_\alpha$. A Greek index $\alpha$ here is not a coordinate index, but refers to the chosen covering $\{U_\alpha\}$. Thus

$$\mu_\beta = \varphi_{\beta\alpha}\mu_\alpha \quad \text{on } U_\alpha \cap U_\beta.$$

This implies

$$\varphi_{\beta\alpha}(d + A_\alpha)\mu_\alpha = (d + A_\beta)\mu_\beta \quad \text{on } U_\alpha \cap U_\beta. \tag{1.2.13}$$

On the left-hand side, we have first computed $D\mu$ in the local trivialization determined by $U_\alpha$ and then transformed the result into the local trivialization determined by $U_\beta$, while on the right-hand side, we have directly expressed $D\mu$ in the latter. We conclude

$$A_\alpha = \varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha} + \varphi_{\beta\alpha}^{-1} A_\beta \varphi_{\beta\alpha}. \tag{1.2.14}$$

We have thus found the transformation behavior. $A_\alpha$ does not transform as a tensor, because of the term $\varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha}$. The difference of two connections, however, does transform as a tensor. The space of all connections on a given vector bundle is therefore an affine space. The difference of two connections is a $\mathfrak{gl}(n, \mathbb{R})$-valued 1-form.

Having a connection $D$ on a vector bundle $E$, it is now our aim to extend $D$ to associated bundles, requiring suitable compatibility conditions. We start with the dual bundle $E^*$. Let

$$(\cdot, \cdot) : E \otimes E^* \to \mathbb{R}$$

be the bilinear pairing between $E$ and $E^*$. The base dual to some base $\mu_1, \dots, \mu_n$ of $E$ is denoted by $\mu^{*1}, \dots, \mu^{*n}$, i.e.,

$$(\mu_i, \mu^{*j}) = \delta_i^j. \tag{1.2.15}$$

We then define the connection $D^*$ on $E^*$ by requiring

$$d(\mu, \nu^*) = (D\mu, \nu) + (\mu, D^*\nu) \tag{1.2.16}$$

for all $\mu \in \Gamma(E)$, $\nu \in \Gamma(E^*)$. In our above notation

$$D = d + A, \qquad D^* = d + A^*. \tag{1.2.17}$$

From (1.2.15) (cf. (1.2.11)) we then compute

$$\begin{aligned} 0 = d(\mu_i, \hat{\mu}^j) &= (A_i^k \mu_k, \mu^{*j}) + (\mu_i, A_l^{*j} \mu^{*l}) \\ &= A_i^j + A_i^{*j}, \end{aligned}$$

i.e.,

$$A^* = -A. \tag{1.2.18}$$

We now construct a connection on a product bundle from connections on the factors. If $E_1$ and $E_2$ are vector bundles over $M$ with connections $D_1$, $D_2$, we obtain a connection $D$ on $E := E_1 \otimes E_2$ by

$$\begin{aligned} D(\mu_1 \otimes \mu_2) &= D_1\mu_1 \otimes \mu_2 + \mu_1 \otimes D_2\mu_2 \\ &(\mu_i \in \Gamma(E_i),\ i = 1, 2). \end{aligned} \tag{1.2.19}$$

We apply this construction to $\mathrm{End}(E) = E \otimes E^*$ to obtain a connection that is again denoted by $D$. For a section $\sigma = \sigma_j^i \mu_i \otimes \mu^{*j}$, we then have

$$\begin{aligned} D(\sigma_j^i \mu_i \otimes \mu^{*j}) &= d\sigma_j^i \mu_i \otimes \mu^{*j} + \sigma_j^i A_i^k \mu_k \otimes \mu^{*j} - \sigma_j^i A_k^j \mu_i \otimes \mu^{*k} \\ &= d\sigma + [A, \sigma]. \end{aligned} \tag{1.2.20}$$

Thus, the connection induced on $\mathrm{End}(E)$ operates via the Lie bracket. In a slightly different interpretation, we can view a connection $D$ as a map

$$D : \Gamma(E) \to \Gamma(E) \otimes \Omega^1(M).$$

Using the notation

$$\Omega^p(E) := \Gamma(E) \otimes \Omega^p(M),$$

we extend $D$ to a map

$$D : \Omega^p(E) \to \Omega^{p+1}(E)$$

by

$$D(\mu\omega) = D\mu \wedge \omega + \mu d\omega \tag{1.2.21}$$

(where $\mu \in \Gamma(E)$, $\omega \in \Omega^p(M)$, and we have written $\mu\omega$ in place of $\mu \otimes_{C^\infty(M)} \omega$.[7]) The **curvature** of a connection $D$ is now defined as

$$F := D^2 : \Omega^0(E) \to \Omega^2(E).$$

$D$ is called flat if

$$F = 0.$$

Since the exterior derivative $d$ satisfies (1.1.37), i.e.,

$$d \circ d = 0$$

we obtain the de Rham complex

$$\Omega^0 \xrightarrow{d} \Omega^1 \xrightarrow{d} \Omega^2 \xrightarrow{d} \dots \quad (\Omega^p = \Omega^p(M)).$$

The sequence

$$\Omega^0(E) \xrightarrow{D} \Omega^1(E) \xrightarrow{D} \Omega^2(E) \xrightarrow{D} \dots$$

however, is not necessarily a complex, since in general $F \neq 0$. For $\mu \in \Gamma(E)$ $(= \Omega^0(E))$, we compute

$$\begin{aligned} F(\mu) &= (d + A) \circ (d + A)\mu \\ &= (d + A)(d\mu + A\mu) \\ &= (dA)\mu - A\, d\mu + A\, d\mu + A \wedge A\mu \end{aligned}$$

(the minus sign arises because $A$ takes values in 1-forms), that is,

$$F = dA + A \wedge A. \tag{1.2.22}$$

If we write

$$A = A_j\, dx^j$$

in local coordinates, with $A_j \in \Gamma(\mathfrak{gl}(n, \mathbb{R})) = \Gamma(\mathrm{End}(E))$, (1.2.22) becomes

$$F = \frac{1}{2}\left(\frac{\partial A_j}{\partial x^i} - \frac{\partial A_i}{\partial x^j} + [A_i, A_j]\right) dx^i \wedge dx^j. \tag{1.2.23}$$

$F$ is a map from $\Omega^0(E)$ to $\Omega^2(E)$, i.e.,

$$F \in \Omega^2(E) \otimes (\Omega^0(E))^* = \Omega^2\big(\mathrm{End}(E)\big).$$

Therefore, according to our rules (1.2.20) and (1.2.21),

$$\begin{aligned} DF &= dF + [A, F] \\ &= dA \wedge A - A \wedge dA + [A, dA + A \wedge A] \quad \text{(by (1.2.22))} \\ &= dA \wedge A - A \wedge dA + A \wedge dA - dA \wedge A + [A, A \wedge A] \end{aligned}$$

[7] We leave it to the reader to (easily) verify that (1.2.21) is well defined, even though the decomposition $\mu \otimes_{C^\infty(M)} \omega$ is not canonical.

$$
\begin{aligned}
&= [A, A \wedge A] \\
&= [A_i\, dx^i, A_j\, dx^j \wedge A_k\, dx^k] \\
&= A_i A_j A_k (dx^i \wedge dx^j \wedge dx^k - dx^j \wedge dx^k \wedge dx^i \\
&\quad - dx^i \wedge dx^k \wedge dx^j + dx^k \wedge dx^j \wedge dx^i) \\
&= 0.
\end{aligned}
$$

We thus obtain the Bianchi identity:

**Theorem 1.1** *The curvature of a connection $D$ satisfies*

$$DF = 0. \tag{1.2.24}$$

The Bianchi identity can also be derived in a conceptually more interesting manner from the equivariance of the curvature ($f^* F_D = F_{f^* D}$, $F_D =$ curvature of $D$) under bundle automorphisms $f$, that is, diffeomorphisms commuting with the group action (cf. [95]).

Using the notation of (1.2.13), we now wish to determine the transformation behavior of the curvature $F$ of a connection $D$. From (1.2.14),

$$
\begin{aligned}
F_\alpha &= dA_\alpha + A_\alpha \wedge A_\alpha \\
&= d(\varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha}) + d(\varphi_{\beta\alpha}^{-1} A_\beta \varphi_{\beta\alpha}) \\
&\quad + \varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha} \wedge \varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha} + \varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha} \wedge \varphi_{\beta\alpha}^{-1} A_\beta \varphi_{\beta\alpha} \\
&\quad + \varphi_{\beta\alpha}^{-1} A_\beta \wedge d\varphi_{\beta\alpha} + \varphi_{\beta\alpha}^{-1} A_\beta \wedge A_\beta \varphi_{\beta\alpha}.
\end{aligned}
$$

Because of

$$d(\varphi_{\beta\alpha}^{-1}) = -\varphi_{\beta\alpha}^{-1} d\varphi_{\beta\alpha} \varphi_{\beta\alpha}^{-1}$$

the derivatives of $\varphi_{\beta\alpha}$ cancel, and we obtain

$$F_\alpha = \varphi_{\beta\alpha}^{-1} F_\beta \varphi_{\beta\alpha}. \tag{1.2.25}$$

Thus, $F$ transforms as a tensor, in contrast to $A$.

### *1.2.3 Reduction of the Structure Group. The Yang–Mills Functional*

We now wish to implement the general principle formulated above that additional structures on the fibers of a bundle lead to restrictions on the admissible transformations. In the previous section, $Gl(n, \mathbb{R})$ was the structure group of our vector bundle. This reflected the fact that we only had a linear (vector space) structure on our fibers, but nothing else. We shall now consider vector spaces with a structure group $G \subset Gl(n, \mathbb{R})$. The group $G$ will then be interpreted as the invariance group of some structure on the fibers. Let $\mathfrak{g}$ be the Lie algebra of $G$. For a connection $D$ on the vector bundle $E$ with fiber $\mathbb{R}^n$, we then require compatibility with the $G$-structure. To make this more precise, we consider local trivializations

$$\varphi : \pi^{-1}(U) \to U \times \mathbb{R}^n$$

of $E$ whose transition functions preserve the $G$-structure, that is, ones that transform $G$-bases $\mu_1, \ldots, \mu_n$ (meaning that the matrix with the columns $\mu_1, \ldots, \mu_n$ is contained in $G$) into $G$-bases. Linear algebra (Gram-Schmidt) tells us that we can always construct such trivializations. In such a trivialization, we also require of

$$D = d + A$$

that

$$A \in \Gamma(\mathfrak{g} \otimes T^* M_{|U}). \tag{1.2.26}$$

Let us consider some examples. $G = O(n)$ means that each fiber of $E$ possesses a Euclidean scalar product $\langle \cdot, \cdot \rangle$. Via a corresponding local trivialization, for each $x \in U$, we then obtain an orthonormal base $e_1(x), \ldots, e_n(x)$ of the fiber $V_x$ over $x$ depending smoothly on $x$, namely $\varphi^{-1}(x, e_1, \ldots, e_n)$, where $e_1, \ldots, e_n$ is an orthonormal base of $\mathbb{R}^n$ w.r.t. the standard Euclidean scalar product. We then want that the Leibniz rule holds, i.e.,

$$d\langle \sigma, \tau \rangle = \langle D\sigma, \tau \rangle + \langle \sigma, D\tau \rangle, \tag{1.2.27}$$

that is, we require that $\langle \cdot, \cdot \rangle$ is covariantly constant. This implies in particular

$$0 = d\langle e_i, e_j \rangle = \langle A\, e_i, e_j \rangle + \langle e_i, A\, e_j \rangle, \tag{1.2.28}$$

that is, $A$ is skew symmetric, $A \in \mathfrak{o}(n)$. A connection $D$ satisfying the Leibniz rule is called a *metric connection.*

Analogously, for $G = U(n)$ we have a Hermitian product on the fibers, and the corresponding Leibniz rule implies

$$A \in \mathfrak{u}(n). \tag{1.2.29}$$

We then speak of a *Hermitian connection.*

Ad$E$ is defined to be the bundle with fibers $(\mathrm{Ad}E)_x \subset \mathrm{End}(V_x)$ consisting of those endomorphisms of $V_x$ that are contained in $G$. $\mathrm{Ad}E = P \times_G \mathfrak{g}$, where $P$ is the associated principal bundle $G$ acts on $\mathfrak{g}$ by the adjoint representation. Analogously, $\mathrm{Aut}(E)$ is the bundle with fiber $G$, now considered as the automorphism group of $V_x$, that is,

$$\mathrm{Aut}(E) = P \times_G G,$$

where $G$ acts by conjugation. (Thus, $\mathrm{Aut}(E)$ is not a principal bundle.) (The reason for this action is the compatibility with the action

$$P \times V \times G \to P \times V, \qquad (p, v) * g = (pg, g^{-1}v),$$

because with $(p,h) * g = (pg, g^{-1}hg)$, we obtain $g^{-1}hg(g^{-1}v) = g^{-1}(hv)$, since $G$ acts on $V$ from the left.) Sections of $\mathrm{Aut}(E)$ are called gauge transformations, and the group of gauge transformations is called the gauge group.

A section $s \in \Gamma(\mathrm{Aut}(E))$ operates on a connection $D$ by

$$s^* D = s^{-1} \circ D \circ s, \tag{1.2.30}$$

hence, for $\mu \in \Gamma(E)$

$$s^*(D)\mu = s^{-1} D(s\mu), \tag{1.2.31}$$

and, with $D = d + A$

$$s^*(A) = s^{-1} ds + s^{-1} A\, s. \tag{1.2.32}$$

In our present notation, the transformation rule (1.2.25) for the curvature $F$ of $D$ becomes

$$s^*(F) = s^{-1} \circ F \circ s. \tag{1.2.33}$$

Here, we consider $F$ as an element of $\Omega^2(\mathrm{Ad}E) = \Gamma(\mathrm{Ad}E \otimes \Lambda^2 T^*M)$, and $s$ acts trivially on the factor $\Lambda^2 T^*M$. The induced product on the fibers $\mathrm{Ad}E_x \otimes \Lambda^2 T_x^*M$ that comes from the bundle metric of $E$ and the Riemannian metric of $M$ will be denoted by $\langle . , . \rangle$.

**Definition 1.5** Let $M$ be a compact Riemannian manifold with metric $g$, $E$ a vector bundle with a bundle metric over $M$, $D$ a metric connection on $E$ with curvature $F_D \in \Omega^2(\mathrm{Ad}E)$. The *Yang–Mills functional* applied to $D$ is

$$YM(D) := \int_M \langle F_D, F_D \rangle dvol_g. \tag{1.2.34}$$

We now recall from Sect. 1.2.2 that the space of all connections on $E$ is an affine space; the difference of two connections is contained in $\Omega^1(\mathrm{End}E)$. Therefore, the space of all metric connections on $E$ is an affine space as well; the difference of two metric connections is contained in $\Omega^1(\mathrm{Ad}E)$. For deriving the Euler–Lagrange equations for the Yang–Mills functional, the variations to consider are therefore

$$D + tB \quad \text{with } B \in \Omega^1(\mathrm{Ad}E).$$

For $\sigma \in \Gamma(E) = \Omega^0(E)$,

$$\begin{aligned} F_{D+tB}(\sigma) &= (D + tB)(D + tB)\sigma \\ &= D^2\sigma + tD(B\sigma) + tB \wedge D\sigma + t^2(B \wedge B)\sigma \\ &= (F_D + t(DB) + t^2(B \wedge B))\sigma, \end{aligned} \tag{1.2.35}$$

as $D(B\sigma) = (DB)\sigma - B \wedge D\sigma$. Therefore,

$$\frac{d}{dt} YM(D + tB)_{|t=0} = \frac{d}{dt} \int \langle F_{D+tB}, F_{D+tB} \rangle_{|t=0}$$
$$= 2 \int \langle DB, F_D \rangle. \tag{1.2.36}$$

Using the definition of $D^*$ (1.2.17), this becomes

$$\frac{d}{dt} YM(D + tB)_{|t=0} = 2 \int \langle B, D^* F_D \rangle.$$

This vanishes for all variations $B$ if

$$D^* F_D \equiv 0. \tag{1.2.37}$$

**Definition 1.6** A metric connection $D$ on the vector bundle $E$ with a bundle metric over the Riemannian manifold $M$ satisfying (1.2.37) is called a *Yang–Mills connection*.

In tensor notation, $F_D = F_{ij} dx^i \wedge dx^j$, and we want to express (1.2.37) in local coordinates with the normalization $g_{ij}(x) = \delta_{ij}$. In such coordinates,

$$d^*(F_{ij} dx^i \wedge dx^j) = -\frac{\partial F_{ij}}{\partial x^i} dx^j,$$

and hence from (1.2.18)

$$D^* F_D = \left( -\frac{\partial F_{ij}}{\partial x^i} - [A_i, F_{ij}] \right) dx^j.$$

The Yang–Mills equation (1.2.37) in local coordinates thus reads

$$\frac{\partial F_{ij}}{\partial x^i} + [A_i, F_{ij}] = 0 \quad \text{for } j = 1, \ldots, d. \tag{1.2.38}$$

In the preceding, we have defined the Yang–Mills functional for metric connections, i.e., ones with structure group $G = O(n)$. Obviously, the same construction works for other compact structure groups, in particular for $U(m)$ and $SU(m)$. Those groups operate on the fibers of complex vector bundles. For a complex vector bundle, one has the structure group $Gl(m, \mathbb{C})$, that is, those of complex linear maps, and a Hermitian structure then, as explained, is a reduction of the structure group to $U(m)$. We now consider complex vector bundles, as for them, we can define important cohomology classes from the curvature of a connection, the so-called Chern classes, as we shall now explain. Thus, $E$ now is a *complex* vector bundle of Rank $m$, that is, with fiber $\mathbb{C}^m$, over the compact manifold $M$. $D$ is a connection on $E$ with curvature

$$F = D^2 : \Omega^0 \to \Omega^2(E). \tag{1.2.39}$$

$F$ satisfies the transformation rule (1.2.25):

$$F_\alpha = \varphi_{\beta\alpha}^{-1} F_\beta \varphi_{\alpha\beta}. \tag{1.2.40}$$

Therefore, we can consider $F$ as an element of $\mathrm{Ad}E$. Since $E$ is a complex vector bundle with structure group $\mathfrak{gl}(m,\mathbb{C})$, $\mathrm{Ad}E = \mathrm{End}E = \mathrm{Hom}_{\mathbb{C}}(E,E)$. Thus,

$$F \in \Omega^2(\mathrm{Ad}E), \tag{1.2.41}$$

that is, $F$ is a 2-form with values in the endomorphisms of $E$. Therefore, $\frac{i}{2\pi}F$ (the factor is simply chosen for convenient normalization) has eigenvalues $\lambda_k$, $k = 1,\dots,m$, which are 2-forms. We then define exterior forms $c_j(E) \in \Omega^{2j}(M)$, $j = 1,\dots,m$, on $M$ via

$$\sum_{j=0}^{m} c_j(E)t^j = \det\left(\frac{i}{2\pi}tF + \mathrm{Id}\right) = \prod_{k=1}^{m}(1+\lambda_k t). \tag{1.2.42}$$

From the Bianchi identity (1.2.24), i.e., $DF = 0$, one concludes that $dc_j(E) = 0$ for all $j$. Thus, the $c_j(E)$ are closed and therefore represent cohomology classes. One also verifies that these classes do not depend on the choice of the connection $D$ on $E$. These cohomology classes are called the *Chern classes* of the complex vector bundle $E$ over $M$. Thus, from an arbitrary Hermitian connection on the bundle $E$, we can compute topological invariants of $E$ and $M$.

For $j = 1, 2$, we get

$$c_1(E) = \frac{i}{2\pi}\,\mathrm{tr}\,F, \tag{1.2.43}$$

$$c_2(E) - \frac{m-1}{2m}c_1(E)\wedge c_1(E) = \frac{1}{8\pi^2}\,\mathrm{tr}(F_0\wedge F_0), \tag{1.2.44}$$

where

$$F_0 := F - \frac{1}{m}\,\mathrm{tr}\,F\cdot \mathrm{Id}_E \quad \text{is the trace-free part of } F. \tag{1.2.45}$$

We now consider a $U(m)$ vector bundle $E$ over a *four-dimensional* oriented Riemannian manifold $M$. We let $D$ be a Hermitian connection on $E$ with curvature $F = D^2$. As explained in (1.1.29), (1.1.31), we can decompose $F_0$ into its self-dual and antiself-dual components

$$F_0 = F_0^+ + F_0^-. \tag{1.2.46}$$

We recall (1.1.30), i.e., that the exterior product of a self-dual 2-form with an antiself-dual one vanishes, and obtain

$$\begin{aligned}\mathrm{tr}(F_0\wedge F_0) &= \mathrm{tr}(F_0^+\wedge F_0^+) + \mathrm{tr}(F_0^-\wedge F_0^-)\\ &= -|F_0^+|^2 + |F_0^-|^2\end{aligned} \tag{1.2.47}$$

by (1.1.29) (note that the trace is the negative of the Killing form of the Lie algebra $\mathfrak{u}(m)$, that is, $A\cdot B = -\,\mathrm{tr}(AB)$, which explains the difference in sign between (1.2.47) and (1.1.29)).

From (1.2.44), we obtain by integration over $M$

$$\left(c_2(E) - \frac{m-1}{2m}c_1(E)^2\right)[M] = -\frac{1}{8\pi^2}\int (|F_0^+|^2 - |F_0^-|^2)\sqrt{g}\,dx^1\wedge\cdots\wedge dx^d. \tag{1.2.48}$$

The Yang–Mills functional then can be decomposed as

$$
\begin{aligned}
YM(D) &= \int_M \left(\frac{1}{m}|\operatorname{tr} F|^2 + |F_0|^2\right)\sqrt{g}dx^1 \wedge \cdots \wedge dx^d \\
&= \int_M \left(\frac{1}{m}|\operatorname{tr} F|^2 + |F_0^+|^2 + |F_0^-|^2\right)\sqrt{g}\, dx^1 \wedge \cdots \wedge dx^d. \qquad (1.2.49)
\end{aligned}
$$

Since $\operatorname{tr} F$ represents the cohomology class $-2\pi i c_1(E)$, the cohomology class of $\operatorname{tr} F$ is fixed, and

$$
\int_M |\operatorname{tr} F|^2 \sqrt{g}\, dx^1 \wedge \cdots \wedge dx^d
$$

becomes minimal if $\operatorname{tr} F$ minimizes the $L^2$-norm in this class ($\operatorname{tr} F$ therefore has to be a harmonic 2-form, see (1.1.115), (1.1.117)). Next, because of the constraint (1.2.48), that is, because the difference of the two integrals of $F_0^+$ and $F_0^-$ is fixed by the topology of $M$ and $E$, and therefore, $\int |F_0|^2$ becomes minimal if one of them vanishes, i.e.,

$$
F_0^+ = 0 \quad \text{or} \quad F_0^- = 0, \qquad (1.2.50)
$$

i.e. if $F_0$ is antiself-dual or self-dual. Which of these two alternatives can hold depends on the sign of $(c_2(E) - \frac{m-1}{m} c_1(E)^2)[M]$.

We now assume that the structure group of the complex vector bundle $E$ is reduced to $SU(m)$. Thus, the fiber of $\mathrm{Ad}E$ is $\mathfrak{su}(m)$ which is trace-free. Therefore, if $D$ is an $SU(m)$ connection, its curvature $F \in \Omega^2(\mathrm{Ad}E)$ satisfies

$$
\operatorname{tr} F = 0. \qquad (1.2.51)
$$

Consequently, by (1.2.43)

$$
c_1(E) = 0,
$$

and by (1.2.44), (1.2.48)

$$
c_2(E)[M] = -\frac{1}{8\pi^2} \int_M (|F^+|^2 - |F^-|^2)\sqrt{g}dx^1 \wedge \cdots \wedge dx^d.
$$

Thus again, the difference of the two parts of the Yang–Mills functional is topologically fixed, and as in (1.2.49),

$$
YM(D) = \int_M (|F^+|^2 + |F^-|^2)\sqrt{g}dx^1 \wedge \cdots \wedge dx^d
$$

is therefore minimized if $F$ is (anti)self-dual; again, which of the two possibilities can hold depends on the sign of $c_2(E)[M]$. We conclude that:

**Theorem 1.2** *For a vector bundle $E$ with structure group $SU(m)$ over a compact oriented four-dimensional manifold $M$, an $SU(m)$ connection $D$ on $E$ yields an* *absolute minimum for the Yang–Mills functional if its curvature $F$ is self-dual or antiself-dual.*

For a systematic presentation of four-dimensional Yang–Mills theory, we refer to [31].

### *1.2.4 The Kaluza–Klein Construction*

Here, we take up the discussion of Sect. 1.1.6 and combine it with a bundle construction. The idea was first put forward by Theodor Kaluza in order to unify gravity with electromagnetism. Although this was not successful in its original form, the general idea is still important and alive today.

Kaluza's ansatz was to consider, in place of the Lorentz manifold $M$, a fiber bundle $\bar{M}$ over $M$. Kaluza took the real axis $\mathbb{R}$ as the fiber. This was then modified by Oscar Klein who chose the fiber $S^1$, that is, the compact Abelian Lie group $U(1)$, and this is also what we shall do here. Subsequently, we shall consider more general fibers. Following the physics literature, we shall always assume that $\bar{M}$ is a principal fiber bundle.

We obtain a metric

$$\bar{g} = \pi^* g + \bar{A} \otimes \bar{A} \tag{1.2.52}$$

on $\bar{M}$ where $\pi : \bar{M} \to M$ is the projection, $g$ is the Lorentz metric on $M$, and $\bar{A}$ is the 1-form for some $U(1)$ connection on $\bar{M}$. (More precisely, $\bar{A} = \pi^* A$ where $A$ is the connection form on $M$.) As in Sect. 1.1.6, we take the total scalar curvature as our action functional, that is,

$$\mathcal{L}(\bar{g}) = \int_{\bar{M}} \bar{R}\sqrt{-\bar{g}}\, dx^0 \cdots dx^3 d\xi, \tag{1.2.53}$$

where $\bar{R}$ is the scalar curvature of $\bar{g}$ and $\xi$ is the fiber coordinate. To rewrite this functional, we first give the formulae for the Ricci curvature of $\bar{g}$. Let $\bar{V}$ be a unit vector field in the fiber direction. Because of the form (1.2.52) of the metric, this simply means that $\bar{V}$ is dual to $\bar{A}$. For each tangent vector field $X$ on $M$, we consider the horizontal lift $\bar{X}_h$ determined by

$$\pi_* \bar{X}_h = X \quad \text{and} \quad \bar{g}(\bar{X}_h, \bar{V}) = 0.$$

Let $F$ be the curvature form for the connection $A$, i.e.,

$$F = dA \quad (\text{and } \pi^* F = d\bar{A})$$

(note that $A$ is a $U(1)$ connection, hence Abelian, and so, here we do not have an $A \wedge A$ term in the formula for the curvature).

We then have, for the Ricci tensor $\bar{R}(\cdot,\cdot)$ of $\bar{g}$,

$$\bar{R}(\bar{X}_h, \bar{Y}_h) = R(X,Y) + 2F \circ F(X,Y), \tag{1.2.54}$$

where $R(\cdot,\cdot)$ is the Ricci tensor of $M$, and in local coordinates with

$$F = F_{\alpha\beta}\, dx^\alpha \wedge dx^\beta,$$

we have

$$(F \circ F)_{\alpha\beta} = g^{\gamma\delta} F_{\alpha\gamma} F_{\delta\beta} = -g^{\gamma\delta} F_{\alpha\gamma} F_{\beta\delta} \tag{1.2.55}$$

and

$$\bar{R}(\bar{X}_h, \bar{V}) = -d^* F(X), \tag{1.2.56}$$

$$\bar{R}(\bar{V}, \bar{V}) = |F|^2 = F^{\alpha\beta} F_{\alpha\beta}. \tag{1.2.57}$$

In particular, the scalar curvature satisfies

$$\bar{R} = \operatorname{tr} \bar{R}(\cdot,\cdot) = R - |F|^2, \tag{1.2.58}$$

where, of course, $R$ is the scalar curvature of $g$. Upon integration over the fibers, (1.1.177) hence becomes

$$\mathcal{L}(\bar{g}) = \int_M (R - |F|^2)\sqrt{-g}\, dx, \tag{1.2.59}$$

that is, the sum of the Einstein–Hilbert functional of the base and the Yang–Mills functional of the fiber. If the Einstein field equations for the vacuum hold for such a metric on $\bar{M}$, then, by (1.1.162), $\bar{M}$ has to have vanishing Ricci curvature, and then by (1.2.57), $F$ has to vanish. Since $F$ is supposed to represent the electromagnetic field, this does not constitute a desirable physical consequence of this ansatz.[8]

We can extend this construction to principal fiber bundles $\pi : \bar{M} \to M$ with compact non-Abelian structure group $G$. For that purpose, let $g'$ be a $G$-invariant metric on the fiber, which we can then extend to all of $T\bar{M}$ with the help of a connection (given by a 1-form $A$). Let $g$ again be a metric on the base $M$. For tangent vectors $\bar{W}, \bar{Z}$ on $\bar{M}$, we put

$$\bar{g}(\bar{Z}, \bar{W}) = g(\pi_* \bar{Z}, \pi_* \bar{W}) + g'(Z, W)$$

(where $Z, W$ denote the projections of $\bar{Z}, \bar{W}$ onto the fibers), obtaining a metric on $\bar{M}$. If $\bar{U}$ and $\bar{V}$ are tangential to a fiber, we obtain, with notation analogous to that above,

$$\bar{R}(\bar{X}_h, \bar{Y}_h) = R(X, Y) - 2g^{\gamma\delta} g'(F_{\alpha\gamma}, F_{\beta\delta}) X^\alpha Y^\beta$$
$$\left(X = X^\alpha \frac{\partial}{\partial x^\alpha}, Y = Y^\alpha \frac{\partial}{\partial x^\alpha}\right), \tag{1.2.60}$$

[8]In an alternative interpretation, one might consider $\bar{g}$ as consisting of $g$ and $A$ and interpret the Euler–Lagrange equations for (1.2.59) as coupled Einstein–Maxwell equations for the metric $g$ and the potential $A$. In that case, the undesired consequence that $F$ has to vanish does not follow, but then we have a coupling rather than a unification of gravity and electromagnetism.

$$\bar{R}(\bar{X}_h, \bar{V}) = -g'(d^* F(X), \bar{V}) \tag{1.2.61}$$

$$\bar{R}(\bar{U}, \bar{V}) = \bar{R}'(\bar{U}, \bar{V}) + \det(g^{\gamma\delta})^{\frac{1}{2}} \sum_{\alpha,\beta} g'(F_{\alpha\beta}, \bar{U}) g'(F_{\alpha\beta}, \bar{V}). \tag{1.2.62}$$

The action functional becomes

$$\mathcal{L}(\bar{g}) = \int_{\bar{M}} (R + R' - |F|^2)\, dvol_{\bar{M}}. \tag{1.2.63}$$

(Here, $R$ is integrated on the base and the result is multiplied with the volume of the fiber, whereas $R'$ can be integrated on any fiber, by $G$-invariance, and the result is multiplied with the volume of $M$—assuming $M$ to be compact again.)

The Einstein field equations for the vacuum now no longer require the vanishing of $F$. $|F|^2$ has to be constant, however, when those equations hold, and base and fiber must have constant scalar curvature. In fact, taking the trace in (1.2.62) yields constant scalar curvature in the fiber direction when the field equations hold, and because the scalar curvature of the metric on the fiber bundle is constant, the scalar curvature in the fiber direction also has to be constant. Taking the trace in (1.2.60) then yields constant scalar curvature on the base.

## 1.3 Tensors and Spinors

### *1.3.1 Tensors*

We have already encountered the tangent bundle $TM$ of a manifold $M$; its dual bundle is the cotangent bundle $T^\star M$. The fiber of the tangent bundle over $p \in M$ is the tangent space $T_p M$, and the fiber of the cotangent bundle is the cotangent space $T_p^\star M$.

**Definition 1.7** A $p$ times *contravariant* and $q$ times *covariant* tensor (field) on a differentiable manifold $M$ is a section of

$$\underbrace{TM \otimes \cdots \otimes TM}_{p \text{ times}} \otimes \underbrace{T^\star M \otimes \cdots \otimes T^\star M}_{q \text{ times}}. \tag{1.3.1}$$

We recall that on a complex manifold, we have the decompositions

$$T^{\mathbb{C}} M = T' M \oplus T'' M, \qquad T^{\star\mathbb{C}} M = T^{\star'} M \oplus T^{\star''} M, \tag{1.3.2}$$

which are invariant under (holomorphic) coordinate changes, and the transformation rules (1.1.86),

$$dz^j = \frac{\partial z^j}{\partial w^l} dw^l, \qquad dz^{\bar{k}} = \frac{\partial z^{\bar{k}}}{\partial w^{\bar{m}}} dw^{\bar{m}} \tag{1.3.3}$$

when $z = z(w)$. We can therefore also speak of covariant tensors of type $(r, s)$, meaning sections of

$$\underbrace{T^{\star'}M \otimes \cdots \otimes T^{\star'}M}_{r \text{ times}} \otimes \underbrace{T^{\star''}M \otimes \cdots \otimes T^{\star''}M}_{s \text{ times}}. \tag{1.3.4}$$

(Contravariant tensors are defined analogously, with the tangent bundle in place of the cotangent bundle.)

For simplicity, we now consider the case of complex dimension 1, that is, of a Riemann surface, in order not to have to bother with too many indices. The reader will surely be able to transfer the subsequent considerations to the case of an arbitrary (finite) dimension. We return to the conceptualization of variations described in (1.1.22), (1.1.39), (1.1.41) and perform a variation

$$z \mapsto z + \epsilon f(z) =: z + \epsilon \delta z \tag{1.3.5}$$

with a *holomorphic* $f$. We want to determine the induced variation $\delta\omega$ of an $(r, s)$-form, that is, of an object of the type

$$\Omega(z, \bar{z}) = \omega(z, \bar{z})(dz)^r (d\bar{z})^s. \tag{1.3.6}$$

Here, $r$ and $s$ are called the conformal weights of $\omega$. Analogously to (1.1.41), we obtain the induced variation

$$\delta_{f,\bar{f}}\Omega(z, \bar{z}) = (r(\partial_z f) + s(\partial_{\bar{z}} \bar{f}) + f\partial_z + \bar{f}\partial_{\bar{z}})\Omega(z, \bar{z}). \tag{1.3.7}$$

$r + s$ is called the scaling dimension, because for $z \mapsto \lambda z$, $\lambda \in \mathbb{R}$,

$$\Omega = \omega(z, \bar{z})(dz)^r (d\bar{z})^s \mapsto \lambda^{r+s}\omega(\lambda z, \lambda\bar{z})(dz)^r (d\bar{z})^s. \tag{1.3.8}$$

$r - s$ is called the conformal spin, because for $z \mapsto e^{-i\vartheta}z$,

$$\Omega = \omega(z, \bar{z})(dz)^r (d\bar{z})^s \mapsto e^{-i(r-s)\vartheta}\omega(e^{-i\vartheta}z, e^{i\vartheta}\bar{z})(dz)^r (d\bar{z})^s. \tag{1.3.9}$$

### *1.3.2 Clifford Algebras and Spinors*

Let $V$ be a vector space of dimension $n$ over a field $F$, which we shall take to be $\mathbb{R}$ or $\mathbb{C}$ in the sequel, equipped with a quadratic form $Q : V \times V \to F$. We then form the Clifford algebra $Cl(Q)$ as the quotient of the tensor algebra $\bigoplus_{k\geq 0} \underbrace{V \otimes \cdots \otimes V}_{k \text{ times}}$ of $V$ by the two-sided ideal generated by all elements of the form

$$v \otimes v - Q(v, v). \tag{1.3.10}$$

In other words, the product in the Clifford algebra is

$$\{v, w\} := vw + wv = 2Q(v, w). \tag{1.3.11}$$

Let $e_1, \ldots, e_n$ be a basis of $V$. This basis then satisfies

$$e_i e_j + e_j e_i = 2Q(e_i, e_j). \tag{1.3.12}$$

The dimension of $Cl(Q)$ is $2^n$, a basis being given by

$$e_0 := 1,\ e_{\alpha_1} e_{\alpha_2} \cdots e_{\alpha_k}, \quad \text{with } 1 \le \alpha_1 < \cdots < \alpha_k \le n. \tag{1.3.13}$$

We define the degree of $e_{\alpha_1} \cdots e_{\alpha_k}$ to be $k$. The degree of $e_0$ is 0. We let $Cl^k(Q)$ be the vector space of elements of $Cl(Q)$ of degree $k$. We also let $Cl^{ev}(Q)$ and $Cl^{odd}(Q)$ denote the space of elements of even and odd degree, resp. We have

$$\begin{aligned} Cl^0(Q) &= \mathbb{R} \text{ or } \mathbb{C} \\ Cl^1(Q) &= V, \end{aligned} \tag{1.3.14}$$

whereas

$$Cl^2(Q) =: \mathfrak{spin}(Q) \tag{1.3.15}$$

is a Lie algebra with bracket

$$[a,b] := ab - ba. \tag{1.3.16}$$

It acts on $Cl^1(Q) = V$ via

$$\tau(a)v := [a,v] = av - va. \tag{1.3.17}$$

(Using (1.3.11), one verifies that for $a \in Cl^2(V)$, $v \in Cl^1(V)$, we have $av - va \in Cl^1(V)$.)

The simply connected Lie group with Lie algebra $\mathfrak{spin}(Q)$ is then denoted by $Spin(Q)$ and called the spin group. According to the general theory of representations of Lie groups (see e.g. [45]), representations of $\mathfrak{spin}(Q)$ lift to ones of $Spin(Q)$.

*Example*

1. $Q = 0$: This yields the so-called Grassmann algebra with multiplication rule

   $$\vartheta_i \vartheta_j + \vartheta_j \vartheta_i = 0,$$

   for some basis $\vartheta_1, \dots, \vartheta_n$.
2. For $F = \mathbb{R}$, consider the quadratic form $Q$ with

   $$Q(e_i, e_i) = \begin{cases} 1 & \text{for } i = 1, \dots p, \\ -1 & \text{for } i = p+1, \dots, n, \end{cases} \qquad Q(e_i, e_j) = 0 \quad \text{for } i \neq j$$

   for some basis $e_1, \dots, e_n$ of $V$. Putting $q := n - p$, we denote the corresponding Clifford algebra by

   $$Cl(p,q).$$

   $p = 0$ yields the Clifford algebra $Cl(0,n)$ usually considered in Riemannian geometry. Of course, for given $n$, the Clifford algebra $Cl(p,q) (p+q=n)$ depends on the choice of $p \in \{0, \dots, n\}$. This is no longer so for the complexification

   $$Cl^{\mathbb{C}}(n) := Cl(p,q) \otimes_{\mathbb{R}} \mathbb{C} \quad (p+q=n).$$

In fact, we have

$$Cl^{\mathbb{C}}(m) \cong \mathbb{C}^{2^n \times 2^n} \quad \text{for } m = 2n,$$
$$Cl^{\mathbb{C}}(m) \cong \mathbb{C}^{2^n \times 2^n} \oplus \mathbb{C}^{2^n \times 2^n} \quad \text{for } m = 2n+1.$$

We define the Pauli matrices

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{1.3.18}$$

They form a basis of the space of $2 \times 2$ Hermitian matrices. We have

$$\{\sigma_i, \sigma_j\} := \sigma_i \sigma_j + \sigma_j \sigma_i = 2\delta_{ij}\sigma_0 \quad \text{for } i, j = 1, 2, 3. \tag{1.3.19}$$

(Note the + sign here: $\{\sigma_i, \sigma_j\}$ is an anticommutator, not a commutator.)

The correspondence

$$e_0 \mapsto \sigma_0, \qquad e_1 \mapsto \sigma_1, \qquad e_2 \mapsto \sigma_3, \qquad e_1 e_2 \mapsto -i\sigma_2$$

thus yields a two-dimensional representation of $Cl(2,0)$, whereas mapping

$$e_1 \mapsto \sigma_1, \qquad e_2 \mapsto i\sigma_2, \qquad e_1 e_2 \mapsto -\sigma_3$$

yields one of $Cl(1,1)$ and

$$e_1 \mapsto i\sigma_1, \qquad e_2 \mapsto i\sigma_2, \qquad e_1 e_2 \mapsto -i\sigma_3$$

yields one of $Cl(0,2)$. The representations of $Cl(2,0)$ and $Cl(1,1)$ are both isomorphic to the algebra of real $2 \times 2$ matrices, whereas that of $Cl(0,2)$ is isomorphic to the quaternions $\mathbb{H}$. In particular, for later reference, we emphasize that we have displayed here real representations of $Cl(2,0)$ and $Cl(1,1)$.

Looking at $Cl(2,0)$, which will be of particular interest for us, and extending the representation to the complexification, we make the following observation which we will subsequently place in a general context. $ie_1e_2$ is represented by $\sigma_2$, and it anticommutes with both $e_1$ and $e_2$. Therefore, the representation of $Cl^2(2,0) = \mathfrak{spin}(2,0)$ leaves the eigenspaces of $ie_1e_2$ invariant. In contrast, $e_1$ and $e_2$, that is, the elements of $Cl^1(2,0)$, interchange them. (In particular, as a representation of $\mathfrak{spin}(2,0)$, the representation is reducible; the two parts themselves are irreducible, however. Here, this is trivial, because they are one-dimensional, but the pattern is general.) The eigenvalues of $ie_1e_2$ are $\pm 1$, and its eigenspaces are generated in our representation by the vectors

$$\begin{pmatrix} 1 \\ i \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ -i \end{pmatrix}.$$

The correspondence

$$e_0 \mapsto \sigma_0, \qquad \dots, \qquad e_3 \mapsto \sigma_3$$

yields a two-dimensional representation of $Cl(3,0)$.

We define the Dirac matrices

$$\gamma^0 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}, \qquad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad \text{for } j = 1, 2, 3,$$

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} 0 & \sigma_0 \\ \sigma_0 & 0 \end{pmatrix},$$

where each 0 represents a $2 \times 2$ block; i.e., the $\gamma^i$ are $4 \times 4$ matrices. The matrix $\gamma^0$ is Hermitian, while $\gamma^1, \gamma^2, \gamma^3$ are skew Hermitian. (This is expressed in the formula $\gamma^0\gamma^\mu\gamma^0 = \gamma^{\mu\dagger}$ for $\mu = 0, 1, 2, 3$.) They satisfy

$$\{\gamma^0, \gamma^0\} = 2I = \{\gamma^5, \gamma^5\},$$

$$\{\gamma^j, \gamma^j\} = -2I \quad \text{for } j = 1, 2, 3,$$

$$\{\gamma^i, \gamma^k\} = 0 \quad \text{for } i \neq k,$$

where $I$ is the $4 \times 4$ identity matrix. Thus, we obtain a four-dimensional representation of $Cl(1,3)$ and $Cl^{\mathbb{C}}(4)$, called the Dirac representation, by

$$e_i \mapsto \gamma^{i-1} \quad \text{for } i = 1, 2, 3, 4.$$

(Note: it might be better to denote the Dirac matrices by $\gamma^1, \dots, \gamma^4$ instead of $\gamma^0, \dots, \gamma^3$. Here, however we follow the convention in the physics literature; $\gamma^5$ will subsequently be denoted by $\Gamma$ when we consider arbitrary dimensions.) We also consider

$$\sigma^{\mu\nu} = \frac{1}{2}[\gamma^\mu, \gamma^\nu],$$

where $[.,.]$ is an ordinary commutator. (Note: in the physics literature, there is an additional factor $i$ in the definition of $\sigma^{\mu\nu}$.)

In the Dirac representation, we have

$$\sigma^{0i} = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix}, \qquad \sigma^{ij} = -\sum_k \varepsilon_{ijk} i \begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}$$

$$\left(\varepsilon_{ijk} := \begin{cases} 1 & \text{if } (i,j,k) \text{ is an even permutation of } (1,2,3) \\ -1 & \text{if } (i,j,k) \text{ is an odd permutation of } (1,2,3) \\ 0 & \text{otherwise.} \end{cases}\right)$$

In the Weyl representation, we instead define

$$\gamma^0 = \begin{pmatrix} 0 & -\sigma_0 \\ -\sigma_0 & 0 \end{pmatrix}, \qquad \gamma^j = \begin{pmatrix} 0 & \sigma_j \\ -\sigma_j & 0 \end{pmatrix}, \quad \text{for } j = 1, 2, 3,$$

$$\gamma^5 = i\gamma^0\gamma^1\gamma^2\gamma^3 = \begin{pmatrix} \sigma_0 & 0 \\ 0 & -\sigma_0 \end{pmatrix}.$$

In this case, we have

$$\sigma^{0i} = \begin{pmatrix} \sigma_i & 0 \\ 0 & -\sigma_i \end{pmatrix}, \qquad \sigma^{ij} = -\sum_k \varepsilon_{ijk} i \begin{pmatrix} \sigma_k & 0 \\ 0 & \sigma_k \end{pmatrix}.$$

Therefore, the action of the $\sigma^{\mu\nu}$ is reducible into two subspaces of (complex) dimension 2 each. Finally, we have the pseudo-Majorana representation, where all $\gamma^\mu$

are purely imaginary:

$$\gamma^0 = \begin{pmatrix} 0 & \sigma_2 \\ \sigma_2 & 0 \end{pmatrix}, \qquad \gamma^1 = \begin{pmatrix} i\sigma_3 & 0 \\ 0 & i\sigma_3 \end{pmatrix},$$
$$\gamma^2 = \begin{pmatrix} 0 & -\sigma_2 \\ \sigma_2 & 0 \end{pmatrix}, \qquad \gamma^3 = \begin{pmatrix} -i\sigma_1 & 0 \\ 0 & -i\sigma_1 \end{pmatrix}.$$

We now wish to consider representations of $Cl(2n, 0)$ and $Cl^{\mathbb{C}}(2n)$ more abstractly. We consider the algebra generated by a basis $\gamma_1, \ldots, \gamma_{2n}$ of $\mathbb{R}^{2n}$ satisfying

$$\{\gamma_\mu, \gamma_\nu\} = 2\delta_{\mu\nu}$$

and set

$$\begin{aligned} a_1 &:= \frac{1}{2}(\gamma_1 + i\gamma_2), & a_1^\dagger &:= \frac{1}{2}(\gamma_1 - i\gamma_2), \\ &\vdots & &\vdots \\ a_n &:= \frac{1}{2}(\gamma_{2n-1} + i\gamma_{2n}), & a_n^\dagger &:= \frac{1}{2}(\gamma_{2n-1} - i\gamma_{2n}) \end{aligned}$$

in $\mathbb{R}^{2n} \otimes \mathbb{C}$. In the physics literature, the $a_i, a_i^\dagger$ are called fermion annihilation and creation operators. We equip $\mathbb{C}^n$ with the coordinates $z^1 = x^1 + ix^2, \ldots, z^n = x^{2n-1} + ix^{2n}$. We let $\Lambda^{(0,q)}\mathbb{C}^n$ be the space of $(0, q)$-forms, i.e. the vector space of differential forms generated by

$$dz^{\bar{i}_1} \wedge \cdots \wedge dz^{\bar{i}_q}, \quad 1 \le i_1 < \cdots < i_q \le n \quad (dz^{\bar{1}} = dx^1 - idx^2, \text{ etc.})$$

We let $\varepsilon(dz^{\bar{j}})$ denote the exterior multiplication by $dz^{\bar{j}}$ from the left, i.e.,

$$\varepsilon(dz^{\bar{j}})(dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q}) = dz^{\bar{j}} \wedge dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q},$$

sending $(0, q)$-forms to $(0, q + 1)$-forms. Likewise, we let $\iota(dz^{\bar{j}})$ be the adjoint of $\varepsilon(dz^{\bar{j}})$ w.r.t. the natural metric on $\mathbb{C}^n$; thus

$$\begin{aligned} &\iota(dz^{\bar{j}})(dz^{\bar{j}_1} \wedge \cdots \wedge dz^{\bar{j}_q}) \\ &= \begin{cases} 0 & \text{if } j \notin \{j_1, \ldots, j_q\}, \\ (-1)^{\mu-1} dz^{\bar{j}_1} \wedge \cdots \wedge \widehat{dz^{\bar{j}_\mu}} \wedge \cdots \wedge dz^{\bar{j}_q}) & \text{if } j = j_\mu. \end{cases} \end{aligned}$$

We then obtain the desired representation by

$$a_j^\dagger \mapsto \varepsilon(dz^{\bar{j}}),$$
$$a_k \mapsto \iota(dz^{\bar{k}}).$$

Of course, one verifies that the formulae

$$\{a_i, a_j\} = 0, \qquad \{a_i^\dagger, a_j^\dagger\} = 0, \qquad \{a_i, a_j^\dagger\} = \delta_{ij}$$

are represented by

$$\{\varepsilon(dz^{\bar{j}}), \varepsilon(dz^{\bar{k}})\} = 0, \qquad \{\iota(dz^{\bar{j}}), \iota(dz^{\bar{k}})\} = 0, \qquad \{\varepsilon(dz^{\bar{j}}), \iota(dz^{\bar{k}})\} = \delta_{jk}.$$

The space

$$S := \Lambda^{(0,\cdot)}\mathbb{C}^n := \bigoplus_{q=0}^{n} \Lambda^{(0,q)}\mathbb{C}^n$$

on which $Cl^{\mathbb{C}}(2n)$ thus acts is called spinor space. The elements of $S$ are called ((complex) Dirac) spinors. Since by (1.3.14), $V$ is a subspace of its Clifford algebra, it therefore operates by multiplication on any representation of that Clifford algebra, in particular on $S$. This is called Clifford multiplication.

The representation $S$ is not irreducible as a representation of $\mathfrak{spin}(2n, 0)$, however. To see this, we consider the "chirality operator"

$$\Gamma := i^{\frac{n}{2}} \gamma_1 \cdots \gamma_{2n}$$

(for $2n = 4$, one often writes $\gamma_5$ in place of $\Gamma$ as explained above)

(with the usual exponential series, we can also write $\Gamma = \exp(i\pi N)$, with the "number operator" $N := \sum_{j=1}^{n} a_j^\dagger a_j$).

$$\{\Gamma, \gamma_\mu\} = 0 \quad \text{for } \mu = 1, \ldots, 2n,$$
$$\Gamma^2 = 1.$$

Thus, we may decompose $Cl^{\mathbb{C}}(n)$ into the eigenspaces $Cl^{\mathbb{C}}(n)^{\pm}$ of $\Gamma$ for the eigenvalues $\pm 1$, and these eigenspaces are interchanged by Clifford multiplication with any $v \in \mathbb{C}^n \backslash \{0\}$. Thus

$$P_{\pm} := \frac{1}{2}(1 \pm \Gamma)$$

project onto the eigenspaces of $\Gamma$, and we get a corresponding decomposition

$$S = S^+ \oplus S^-$$

into "positive and negative chirality spinors" (also called right- and left-handed spinors), or "Weyl spinors". If $p - q \equiv 0, 1, 2 \mod 8$, one may also find a real representation of $Cl(p, q)$. The corresponding spinors are called real or Majorana spinors. An important example is $n = 4$, $p = 3$, $q = 1$. Likewise, for $q - p \equiv 0, 1, 2 \mod 8$, there exist imaginary or pseudo-Majorana spinors.

The Lie algebra $\mathfrak{so}(n)$ consists of skew symmetric matrices. It is generated by the matrices $M^{ij}$ with coefficients

$$(M^{ij})_{ab} = \delta_a^i \delta_b^j - \delta_a^j \delta_b^i.$$

They satisfy the commutation rules

$$[M^{ij}, M^{kl}] = -\delta^{ik} M^{jl} + \delta^{jk} M^{il} + \delta^{il} M^{jk} - \delta^{jl} M^{ik}.$$

These rules are also satisfied by

$$\sigma_{ij} := -\frac{1}{4}(\gamma_i \gamma_j - \gamma_j \gamma_i)$$

where $\gamma_1, \ldots, \gamma_n$ are a basis of $\mathbb{R}^n$ with $\{\gamma_i, \gamma_j\} = -\delta_{ij}$. (Note that this differs by a factor $-\frac{1}{2}$ from the convention employed in the definition of the Dirac and Weyl representations above.)

Thus

$$M^{ij} \mapsto \sigma_{ij}$$

yields a representation of $\mathfrak{so}(n)$ on $\mathbb{R}^n$; in fact, we may identify $\mathfrak{so}(n)$ with $\mathfrak{spin}(0,n)$. Since $\mathfrak{spin}(0,n) = Cl^2(0,n)$ we thus get an induced representation of $\mathfrak{so}(n)$ on the spinor space $S$. This representation, however, does not lift to one of $SO(n)$, but only to one of $Spin(0,n)$, the two-sheeted cover of $SO(n)$.

In the case when $n$ is even, since each $\sigma_{ij}$ is a sum of products of two $\gamma_i$, and since Clifford multiplication with each $\gamma_i$ interchanges the eigenspaces of $S^{\pm}$ of $\Gamma$, $\sigma_{ij}$ leaves these eigenspaces invariant.

To summarize: We have established an isomorphism $\mathfrak{so}(n) \longleftrightarrow \mathfrak{spin}(n)$. Thus, $\mathfrak{so}(n)$ operates on $Cl^1(0,n)$, and each representation of the Clifford algebra $Cl(0,n)$ therefore induces a representation of $\mathfrak{so}(n)$. In particular, in this manner, we obtain the spinor representation of $\mathfrak{so}(n)$ (which induces a double valued representation of $SO(n)$).

*Remark* The presentation here partly follows that of [22]. The original reference for Clifford modules is [6].

### *1.3.3 The Dirac Operator*

As explained, since the vector space $V$ is a subspace of the Clifford algebra $Cl(Q)$, it operates on any representation of that Clifford algebra. We can thus multiply a vector, an element of $V$, with a spinor, an element of $S$, that is, we have Clifford multiplication

$$V \times S \to S. \tag{1.3.20}$$

In fact, since multiplication by an element of $V$ interchanges $S^+$ and $S^-$, we have an operation

$$V \times S^{\pm} \to S^{\mp}. \tag{1.3.21}$$

Denoting the representation by $\gamma$ and letting $\frac{\partial}{\partial x^i}$ be the partial derivative in the direction of $e_i$, we can define the Dirac operator

$$\not{D} := \gamma(e_i)\frac{\partial}{\partial x^i}, \tag{1.3.22}$$

which operates on spinor fields. The square $\not{D} \circ \not{D}$ of the Dirac operator is then a linear combination of second derivatives; that linear combination depends on the quadratic form $Q$ defining the Clifford algebra. If the quadratic form $Q$ is represented by the identity matrix, that is, if we consider the Clifford algebra $Cl(n,0)$, the square of the Dirac operator is the (negative definite) Laplace operator (see (1.1.103), (1.1.105))

$$\Delta = \sum_i \frac{\partial^2}{\partial x_i^2}. \tag{1.3.23}$$

In order to develop some structural insights, it is now useful to start with the complex case, or more precisely with a complex vector space $V$ with a nondegenerate quadratic form $Q$. As $Q$ is nondegenerate, it induces a nondegenerate bilinear form $\langle .,. \rangle$, w.r.t. which $V$ is self-dual. Moreover, on a representation $S$ of the Clifford algebra $Cl(Q)$, we can find a nondegenerate bilinear form $(.,.)$ that is invariant under multiplication by $v \in V = Cl^1(Q)$:

$$(vs, t) = (s, vt) \tag{1.3.24}$$

for all $s, t \in S$. We can then use $(.,.)$ to identify $S$ with its dual $S^\star$, and (1.3.20) then induces morphisms

$$\Gamma : S^\star \times S^\star \to V \tag{1.3.25}$$

and

$$\tilde{\Gamma} : S \times S \to V. \tag{1.3.26}$$

In (1.3.25), to any two elements of $S^\star$, we assign a vector $v \in V$ that operates on a pair $\sigma, \tau$ of elements of $S$ by $(v\sigma, \tau)$, cf. (1.3.20), (1.3.24).

Using bases $\{s_a\}$ and $\{e_\mu\}$ of $S^\star$ and $V$, we write (1.3.25) as

$$\Gamma(s_a, s_b) = \Gamma^\mu_{ab} e_\mu. \tag{1.3.27}$$

These morphisms are symmetric and equivariant w.r.t. the representation of $Cl(Q)$. Turning to the real case, the situation is not as convenient: we cannot always find real versions of these morphisms; they only exist in certain cases. This depends on the classification of Clifford algebras. They always exist for the Minkowski signature, that is, for the Clifford algebra $Cl(1, n-1)$, in any dimension $n$. They also exist for $Cl(2, 0)$, the case of particular interest for us.

### *1.3.4 The Lorentz Case*

Let us also exhibit the relation between the orthogonal group and the spin group in the Lorentz case. There exist many references on this topic, including the classic [81]. Let $x = (x^0, x^1, x^2, x^3) \in \mathbb{R}^{1,3}$. We put

$$\langle x, x \rangle = x^0x^0 - x^1x^1 - x^2x^2 - x^3x^3. \tag{1.3.28}$$

The subgroup of $Gl(4, \mathbb{R})$ that preserves $\langle x, x \rangle$ is the Lorentz group $O(1, 3)$. It consists of two components that are distinguished by the value of the determinant, $+1$ or $-1$, and have otherwise the same properties. Thus, we consider the identity component $SO(1, 3)$ where the determinant is $+1$, without essential loss of generality.

We shall see that the corresponding spin group is $Sl(2,\mathbb{C})$, the group of complex $2\times 2$-matrices with determinant 1. To $x$, we associate the Hermitian matrix

$$X := x^\mu \sigma_\mu = \begin{pmatrix} x^0+x^3 & x^1-ix^2 \\ x^1+ix^2 & x^0-x^3 \end{pmatrix}, \tag{1.3.29}$$

where $\sigma_0,\dots,\sigma_3$ are the Pauli matrices. We first note that

$$\langle x,x\rangle = \det X. \tag{1.3.30}$$

Since we have

$$\mathrm{Tr}(\sigma_\mu\sigma_\nu) = 2\delta_{\mu\nu}, \tag{1.3.31}$$

we obtain

$$x^\mu = \frac{1}{2}\mathrm{Tr}(X\sigma_\mu) \tag{1.3.32}$$

as the inverse of the equation expressing $X$ in terms of $x$.

In the physics literature, one writes the Hermitian matrix $X$ as

$$\begin{pmatrix} X^{1\dot{1}} & X^{1\dot{2}} \\ X^{2\dot{1}} & X^{2\dot{2}} \end{pmatrix}. \tag{1.3.33}$$

By the Hermitian condition, $X^{\alpha\dot{\beta}} = \overline{X^{\beta\dot{\alpha}}}$ so that $X^{1\dot{1}}$ and $X^{2\dot{2}}$ are real. In this notation, (1.3.32) becomes

$$x^0 = \frac{1}{2}(X^{1\dot{1}} + X^{2\dot{2}}), \qquad x^1 = \frac{1}{2}(X^{1\dot{2}} + X^{2\dot{1}}),$$
$$x^2 = \frac{i}{2}(X^{1\dot{2}} - X^{2\dot{1}}), \qquad x^3 = \frac{1}{2}(X^{1\dot{1}} - X^{2\dot{2}}).$$

We may use the relation (1.3.29) between a vector $x$ and a Hermitian matrix $X$ to define an operation of $Sl(2,\mathbb{C})$ on $\mathbb{R}^{1,3}$ as follows:

For $A\in Sl(2,\mathbb{C})$, we put

$$T(A)X := X' := AXA^\dagger. \tag{1.3.34}$$

With indices, this is written as

$$X'^{\sigma\dot{\tau}} = A^\sigma_\beta \bar{A}^{\dot{\tau}}_{\dot{\gamma}} X^{\beta\dot{\gamma}}. \tag{1.3.35}$$

Here, the dotted indices refer to the transformation according to the conjugate complex of $A$, and this then explains the convention employed in (1.3.33). The fact that two $A$s appear in (1.3.34) suggests that one consider this expression as a product: Instead of the 4-vector $X$, we take two spinors $\phi$, $\chi$ that transform according to

$$\phi'^\alpha = A^\alpha_\beta \phi^\beta, \qquad \chi'^{\dot{\gamma}} = A^{\dot{\gamma}}_{\dot{\delta}} \chi^{\dot{\delta}}. \tag{1.3.36}$$

Their product then transforms like $X$ in (1.3.35),

$$\phi'^\sigma \chi'^{\dot{\tau}} = A^\sigma_\beta \bar{A}^{\dot{\tau}}_{\dot{\gamma}} \phi^\beta \chi^{\dot{\gamma}}. \tag{1.3.37}$$

Using the above formulae, we can express (1.3.34) as a transformation of the vector $x$:

$$x'^{\mu} = \frac{1}{2}\operatorname{Tr}(X\sigma_{\prime\mu}) = \frac{1}{2}\operatorname{Tr}(AXA^{\dagger}\sigma_{\mu}) = \frac{1}{2}x^{\nu}\operatorname{Tr}(A\sigma_{\nu}A^{\dagger}\sigma_{\mu}). \tag{1.3.38}$$

Thus,

$$x' = Bx, \tag{1.3.39}$$

with

$$B^{\mu}_{\nu} = \frac{1}{2}\operatorname{Tr}(A\sigma_{\nu}A^{\dagger}\sigma_{\mu}) = \frac{1}{2}\operatorname{Tr}(\sigma_{\mu}A\sigma_{\nu}A^{\dagger}) \tag{1.3.40}$$

as the trace is invariant under cyclic permutations.

One may check from this that (1.3.34) induces a Lorentz transformation, but this can more easily be derived from the fact that (1.3.34) maps Hermitian matrices to Hermitian matrices and preserves the determinant (since $A \in Sl(2,\mathbb{C})$ has determinant 1), and (1.3.30) then implies that $\langle\cdot,\cdot\rangle = x^0x^0 - x^1x^1 - x^2x^2 - x^3x^3$ (see (1.3.28)) is preserved.

Also, this yields a homomorphism

$$T : Sl(2,\mathbb{C}) \to SO(1,3)$$

with kernel $\{\pm 1\}$ ($\pm$ identity in $Sl(2,\mathbb{C})$ leads to the identity in $SO(1,3)$ in (1.3.34)), and image the identity component of the Lorentz group.[9] $Sl(2,\mathbb{C})$ is the universal cover of the identity component of the Lorentz group, which is doubly connected. Therefore, in the physics literature, representations of $Sl(2,\mathbb{C})$ are usually considered as double-valued representations of $SO(1,3)$.

We also observe that the homomorphism $T$ in (1.3.34) maps $SU(2)$ to $SO(3)$. Namely we have, for $A \in SU(2)$,

$$\operatorname{Tr}(T(A)X) = \operatorname{Tr}(AXA^{\dagger}) = \operatorname{Tr}(AXA^{-1}) = \operatorname{Tr}(X) = 2x^0$$

in the notations of (1.3.29). Thus, $T(A)$ preserves $x^0$, and since $\langle x, x\rangle = x^0x^0 - x^1x^1 - x^2x^2 - x^3x^3$ is also preserved, it preserves

$$x^1x^1 + x^2x^2 + x^3x^3$$

and therefore yields an orthogonal transformation of the $x^1, x^2, x^3$ space. As before, this yields a twofold covering of $SO(3)$, and $SU(2) \cong Spin(3)$.

Since $SO(1,3)$ acts by automorphisms on $\mathbb{R}^{1,3}$ which can be considered as a group of translations, we can form the semidirect product $SO(1,3) \ltimes \mathbb{R}^{1,3}$ where $(B,b) \in SO(1,3) \ltimes \mathbb{R}^{1,3}$ operates on $\mathbb{R}^{1,3}$ via $x \mapsto Bx + b$ and where "semidirect product" refers to the obvious composition rule. The group of all isometries of Minkowski space is the semidirect product $O(1,3) \ltimes \mathbb{R}^{1,3}$, the Poincaré group, but it suffices for our purposes to consider its connected component containing the identity. Again, it is covered by $Sl(2,\mathbb{C}) \ltimes \mathbb{R}^{1,3}$.

---

[9] The Lorentz group has four connected components, all isomorphic to $SO(1,3)$, and we obtain the other components by space- and time-like reflections from the identity component.

The irreducible unitary representations of $Sl(2,\mathbb{C}) \ltimes \mathbb{R}^{1,3}$ were classified by Wigner. We sketch here those aspects of the representation theory that are directly relevant for elementary particle physics. A mathematical treatment to which we refer for further details and which emphasizes the applications in physics is given in [98], whereas a comprehensive presentation from the perspective of physics can be found in [103]. Since $\mathbb{R}^{1,3}$ is a normal subgroup of $Sl(2,\mathbb{C}) \ltimes \mathbb{R}^{1,3}$, the study of the representations proceeds by describing the orbits of the action of $Sl(2,\mathbb{C})$ on $\mathbb{R}^{1,3}$, identifying the isotropy group of a point on each orbit, called the "little group" in physics, and then finding the representations of those isotropy groups. We know from (1.3.28) that

$$m^2 := \langle x, x\rangle = x^0x^0 - x^1x^1 - x^2x^2 - x^3x^3 \tag{1.3.41}$$

is preserved by the action of $Sl(2,\mathbb{C})$ on $\mathbb{R}^{1,3}$. In particular, each orbit must be contained in a level set of $m^2$. Physically, $m$ is the mass of the particle defined by the representation, and it then suffices to consider the case $m \geq 0$. Using the identification (1.3.29) of $x \in \mathbb{R}^{1,3}$ with the matrix

$$X := x^\mu \sigma_\mu = \begin{pmatrix} x^0 + x^3 & x^1 - ix^2 \\ x^1 + ix^2 & x^0 - x^3 \end{pmatrix},$$

for $m^2 > 0$, we can select the point

$$\begin{pmatrix} \pm m & 0 \\ 0 & \pm m \end{pmatrix},$$

depending on whether $x^0 > 0$ or $< 0$. Since this is a multiple of the identity matrix, its isotropy group, that is, the group of matrices leaving it invariant under conjugation, see (1.3.34), is $SU(2)$. As described for instance in [45, 75, 98], the irreducible unitary representations of $SU(2)$ come in a discrete family, parametrized by a half integer

$$L = 0, \frac{1}{2}, 1, \frac{3}{2}, \ldots \tag{1.3.42}$$

which can be identified with the spin of the particle. Thus, the class of representations corresponding to an orbit with $m^2 > 0$ is described by the continuous parameter $m^2$ and the discrete parameter $L$ from (1.3.42).

A point on an orbit with $m^2 = 0$ is

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

and its isotropy group is defined by the invariance condition

$$A \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} A^\dagger = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

This implies that $A$ has to be of the form

$$A = \begin{pmatrix} e^{i\theta} & z \\ 0 & e^{i\theta} \end{pmatrix}$$

for some $z \in \mathbb{C}, \theta \in \mathbb{R}$. Looking at the conjugation

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{i\theta} \end{pmatrix} = \begin{pmatrix} 1 & ze^{2i\theta} \\ 0 & 1 \end{pmatrix},$$

we see that the isotropy group, denoted by $\tilde{E}(2)$, is a double cover (because of the angle $2\theta$) of the group of Euclidean motions $SO(2) \ltimes \mathbb{C}$. By the same strategy as before, for determining its representations, we should look at the orbits of the $SO(2)$ action on $\mathbb{C}$ which are the origin 0 and the concentric circles about 0. The representations corresponding to the latter do not occur in elementary particle physics. So, we are left with the origin whose isotropy group, the little group, is $SO(2)$. Its irreducible representations are all one-dimensional and labeled by

$$s = 0, \pm\frac{1}{2}, \pm 1, \dots \tag{1.3.43}$$

where the factor $\frac{1}{2}$ corresponds to the fact that the rotations were about an angle $2\theta$. The key for understanding the representations of $SU(2)$ is the following. The Lie algebra $\mathfrak{su}(2)$ is generated by

$$t_\mu := \frac{i}{2}\sigma_\mu, \quad \mu = 1, 2, 3 \tag{1.3.44}$$

with the Pauli matrices $\sigma_\mu$, see (1.3.18). They satisfy

$$[t_\mu, t_\nu] = \epsilon_{\mu\nu\sigma} t_\sigma \tag{1.3.45}$$

with the totally antisymmetric tensor $\epsilon_{\mu\nu\sigma}$. The real matrices

$$e_+ := -i(t_1 - it_2) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad e_+ := -i(t_1 + it_2) = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{1.3.46}$$

$$h := -it_3 = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.3.47}$$

yield a basis of the Lie algebra $\mathfrak{sl}(2, \mathbb{R})$ and satisfy

$$[h, t_+] = t_+, \qquad [h, t_-] = -t_-, \qquad [t_+, t_-] = 2h. \tag{1.3.48}$$

From this, one deduces that when $\rho$ is a representation of $\mathfrak{sl}(2, \mathbb{C})$ on a vector space $V$ and $v_\lambda$ is an eigenvector of $\rho(h)$ with eigenvalue $\lambda$, then $\rho(t_\pm)v_\lambda$ are eigenvectors of $\rho(h)$ with eigenvalues $\lambda \pm 1$. One then finds that the possible values of $\lambda$ are $L, L-1, \dots, -L$ for some half integer $L = 0, \frac{1}{2}, 1, \dots$, see e.g. [45, 75, 98]. Since the eigenvalues are nondegenerate, the dimension of this representation is then $2L + 1$.

### *1.3.5 Left- and Right-handed Spinors*

We now put the transformation rule (1.3.37) for the product of two spinors into a more general perspective that will be needed below in Sect. 2.2.1 for defining Lagrangians for spinors. According to our previous general discussion, in the present case of $\mathbb{R}^4 \cong \mathbb{C}^2$, the spinor space is a four-dimensional complex vector space, i.e., isomorphic to $\mathbb{C}^4$. We have already seen in Sect. 1.3.2 that the spinor representation is not irreducible as a representation of the spin group, but splits into the direct sum of two chiral representations, i.e., each spinor can be written as

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}. \tag{1.3.49}$$

$\psi_L$ is called a left-handed, $\psi_R$ a right-handed spinor.

$A \in Sl(2, \mathbb{C})$ then acts via

$$\begin{aligned} \psi_L &\mapsto A\psi_L, \\ \psi_R &\mapsto (A^\dagger)^{-1}\psi_R. \end{aligned} \tag{1.3.50}$$

With $A = (A_k^j)_{j,k=1,2}$, we have

$$\begin{aligned} \psi_L^i &\mapsto A_k^i \psi_L^k, \\ \psi_R^i &\mapsto \tilde{A}_k^i \psi_R^k \quad \text{with } \tilde{A}_k^i \bar{A}_k^j = \delta_{ij}. \end{aligned}$$

From (1.3.40), we also get with the help of (1.3.31)

$$B_\nu^\mu \sigma_\nu = A^\dagger \sigma_\mu A \tag{1.3.51}$$

(note that here the summation convention is used even though the position of the indices is not right—it would be better to write the $\sigma$s with upper indices, but we refrain here from changing an established convention). Putting

$$S(A) = \begin{pmatrix} A & 0 \\ 0 & (A^\dagger)^{-1} \end{pmatrix}, \qquad \psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}, \tag{1.3.52}$$

the action of $A$ is described by

$$\psi \mapsto S(A)\psi. \tag{1.3.53}$$

In the Weyl representation, with

$$\gamma^0 = \begin{pmatrix} 0 & -\sigma_0 \\ -\sigma_0 & 0 \end{pmatrix},$$

we then have

$$S^{-1} = \gamma^0 S^\dagger \gamma^0. \tag{1.3.54}$$

Finally (1.3.51) implies

$$S^{-1}\gamma^\mu S = B_\nu^\mu \gamma^\nu. \tag{1.3.55}$$

For two left-handed spinors (see (1.3.49)) $\phi, \chi$,

$$\phi\chi := \varepsilon_{\alpha\beta}\phi^\alpha \chi^\beta \tag{1.3.56}$$

transforms as a scalar under the spinor representation; namely

$$\begin{aligned}\varepsilon_{\alpha\beta}A^{\alpha}_{\gamma}\phi^{\gamma}A^{\beta}_{\delta}\chi^{\delta} &= A^1_{\gamma}A^2_{\delta}\phi^{\gamma}\chi^{\delta} - A^2_{\gamma}A^1_{\delta}\phi^{\gamma}\chi^{\delta} \\ &= (A^1_1A^2_2 - A^2_1A^1_2)\phi^1\chi^2 + (A^1_2A^2_1 - A^2_2A^1_1)\phi^2\chi^1 \\ &= \det A(\phi^1\chi^2 - \phi^2\chi^1) \\ &= \varepsilon_{\alpha\beta}\phi^{\alpha}\chi^{\beta} \quad (\text{since } \det A = 1 \text{ for } A \in Sl(2,\mathbb{C})).\end{aligned}$$

Similarly

$$\phi^{\alpha}\sigma_{\mu,\alpha\dot{\alpha}}\bar{\chi}^{\dot{\alpha}} \tag{1.3.57}$$

transforms as a vector, for $\mu = 0, 1, 2, 3$. This can be better understood by considering full spinors

$$\psi = \begin{pmatrix}\psi_L \\ \psi_R\end{pmatrix}. \tag{1.3.58}$$

Following the physics notation, in the Lorentzian case, we define $\psi^{\dagger}$ as the complex conjugate of $\psi$, and the Dirac-conjugated spinor as

$$\bar{\psi} := \psi^{\dagger}\gamma^0. \tag{1.3.59}$$

In the Riemannian case, the $\gamma^0$ is omitted, that is,

$$\bar{\psi} := \psi^{\dagger}. \tag{1.3.60}$$

Thus, returning to the Lorentzian case, $\bar{\psi}\omega = \bar{\psi}_L\omega_R + \bar{\psi}_R\omega_L$. Then in the Weyl representation,

$$\bar{\psi}\gamma^{\mu}\omega \tag{1.3.61}$$

transforms as a vector. In fact, applying a transformation $A \in Sl(2,\mathbb{C})$, we get

$$\begin{aligned}\psi^{\dagger}S(A)^{\dagger}\gamma^0\gamma^{\mu}S(A)\omega &= \bar{\psi}\gamma^0S(A)^{\dagger}\gamma^0\gamma^{\mu}S(A)\omega \\ &= \bar{\psi}S^{-1}\gamma^{\mu}S\omega \quad \text{by (1.3.54)} \\ &= B^{\mu}_{\nu}\bar{\psi}\gamma^{\nu}\omega \quad \text{by (1.3.55)},\end{aligned}$$

which is the required formula. Since $\bar{\psi}\gamma^{\mu}\omega = \psi^{\dagger}\gamma^0\gamma^{\mu}\omega = \psi_L^{\dagger}\sigma_{\mu}\omega_L - \psi_R^{\dagger}\sigma_{\mu}\omega_R$, we also see why (1.3.57) transforms as a vector.

## 1.4 Riemann Surfaces and Moduli Spaces

### *1.4.1 The General Idea of Moduli Spaces*

We start with some general principles; their meaning may become apparent only after reading the rest of this section, and the reader is advised to proceed when these principles are unclear and return to them later. It may be helpful, however, to try to understand the sequel in the light of these principles. In any case, the present section is more abstract and has more of a survey character than the preceding ones.

One is given a mathematical object with some varying structure. An example of such an object is a differentiable manifold $S$, and the structure could be a complex structure, a Riemannian metric—perhaps of a particular type—and so on. One wants to divide out all invariances; for example, one wants to identify all isometric metrics. The invariances usually constitute a (discrete or Lie) group. The resulting space of invariance classes is then a moduli space. This already suggests that there will be a problem (more precisely, singularities of the moduli space) caused by those particular instances of the structure that possess more invariances than the typical ones, for example, those Riemannian metrics that are highly symmetric. The reason is obvious, namely that for those instances, we need to divide out a larger group of invariances than for the other ones.

**Heuristic guiding principle**

*The moduli space for structures of some given type carries a structure of the same type.*

So, for example, we expect a moduli space of Riemannian metrics to carry a Riemannian metric itself, a moduli space of complex structures to be a complex space itself, a moduli space of algebraic varieties to be an algebraic variety itself.

Typically, the space of such structures is not compact, that is, these structures can degenerate. One then wishes to compactify the moduli space. The compactifying boundary then also contains (certain) degenerate versions of the structure. The choice of admissible degenerate structures—which need not be unique—can be subtle and should be carried out so that the resulting space is a Hausdorff space.

Often, one also wishes to get a fine moduli space $M_{fine}$. Let $p$ be a point in the (ordinary, or coarse) moduli space $M$ representing an instance $g$ of a structure. $M_{fine}$ then should be the fibration over $M$ with the fiber over $p$ being that $g$.

## *1.4.2 Riemann Surfaces and Their Moduli Spaces*

A Riemann surface can be defined in several different ways, that is, through different types of structures. While these notions turn out to be equivalent in the end, they lead to different approaches to the moduli space of Riemann surfaces and equip that moduli space with different structures, according to the above principle. We shall now explain these different structures and also illustrate why they are interesting, in particular how they lead to different mathematical constructions and applications. For more details and proofs, we refer to [64] and other references cited subsequently. A profound knowledge of Riemann surface theory is useful for understanding conformal field theory and string theory mathematically. Let $S$ be a compact differentiable orientable surface of genus $p$. If not explicitly stated otherwise, we assume $p > 1$.

The basic point is that one and the same such differentiable surface can carry a continuum of different Riemann surface structures. That is, there are many pairs $\Sigma_1, \Sigma_2$ of Riemann surfaces that are both diffeomorphic to $S$, but not equivalent as

Riemann surfaces. The moduli problem then consists of defining and understanding the space of all such Riemann surfaces (modulo holomorphic equivalence).

1. **A Riemann surface $\Sigma$ is a discrete (fixed point free, cocompact) faithful representation of the fundamental group $\pi_1(S)$ into $G := \mathrm{PSL}(2,\mathbb{R})$, determined up to conjugation by an element of $G$. The moduli space is the space of such representations modulo conjugation.**

More precisely: A Riemann surface $\Sigma$ is a quotient $H/\Gamma$, where $H = \{z = x+iy \in \mathbb{C} : y > 0\}$ is the Poincaré upper half plane and $\Gamma$ is a discrete group of isometries with respect to the hyperbolic metric

$$\frac{1}{y^2} dz \wedge d\bar{z}. \tag{1.4.1}$$

$\Gamma$ is a subgroup of the isometry group $\mathrm{PSL}(2,\mathbb{R})$ of $H$.[10] Here, $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\pm 1$, acting on $H$ via $z \mapsto \frac{az+b}{cz+d}$, with $a,b,c,d$ satisfying $ad-bc=1$ describing an element of $\mathrm{SL}(2,\mathbb{R})$. $\Gamma$ should operate properly discontinuously and freely. It thus should not contain elliptic elements. This excludes singularities of the quotient $H/\Gamma$ arising from fixed points of the action of $\Gamma$. In order to exclude cusps, that is, in order to ensure that $H/\Gamma$ is compact, parabolic elements (see insertion below) of $\Gamma$ also have to be excluded. Thus, all elements of $\Gamma$ different from the identity should be hyperbolic.

**Insertion:** Here, a transformation $z \mapsto \frac{az+b}{cz+d}$ of $H$ is called hyperbolic if it has two fixed points on the extended real axis $\bar{\mathbb{R}} = \partial H \cup \{\infty\}$, parabolic if it has one fixed point on $\bar{\mathbb{R}}$, and elliptic if it has a fixed point in $H$. Since the fixed points are computed to be $\frac{a-d}{2c} \pm \frac{1}{2c}\sqrt{(a+d)^2-4}$, the transformation is hyperbolic iff $|a+d| > 2$. The standard example of a hyperbolic transformation is $z \mapsto 2z$, with fixed points at 0 and $\infty$, and a parabolic one is given by $z \mapsto \frac{z}{z+1}$, which has its unique fixed point at 0. A hyperbolic transformation $\gamma$ maps the hyperbolic geodesic $l$ between its two fixed points $p_1, p_2$ (the semicircle through $p_1$ and $p_2$ orthogonal to the real axis) into itself, that is, it is a translation along the hyperbolic geodesic $l$. We can then easily visualize the operation of $\gamma$ on $H$; it simply maps each geodesic orthogonal to $l$ to another such geodesic orthogonal to $l$, with the shift already determined by the operation of $\gamma$ on $l$. When we consider the example $z \mapsto 2z$, the invariant geodesic is the imaginary axis. The invariant geodesic in $H$ becomes a closed geodesic on the surface $H/\Gamma$, with length given by the length of the shift. A parabolic transformation does not have a fixed geodesic, but instead rotates any geodesic through its fixed point into another such geodesic. Therefore, a parabolic transformation does not produce a closed geodesic in the quotient.

$\Gamma$ is isomorphic to the fundamental group $\pi_1(S)$. Thus, a Riemann surface is described by a faithful representation $\rho$ of $\pi_1(S)$ in $G := \mathrm{PSL}(2,\mathbb{R})$. This essentially leads to the approach of Ahlfors and Bers to Teichmüller theory. Here, we need to identify any two representations that only differ by a conjugation with an

[10] The isometries of $H$ are the same as the conformal automorphisms of $H$, because of the conformal invariance of the metric.

element of $G$. Thus, we consider the space of faithful representations up to conjugacy. A representation can be defined by the images of the generators, that is, by $2p$ elements of $G$, and this induces a natural topology on the moduli space. In particular, this allows us to compute the dimension of the moduli space: Each of the $2p$ generator images is described by three real degrees of freedom ($a, b, c, d$ satisfying the relation $ad - bc = 1$) which altogether yields $6p$ degrees of freedom. From this, we first need to subtract 3, the degrees of freedom for one generator, because the generators $a_1, b_1, \ldots, a_p, b_p$ of $\pi_1(\Sigma)$ are not independent, but satisfy the relation $a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_p b_p a_p^{-1} b_p^{-1} = 1$. We also need to subtract another 3 to account for the freedom of conjugating by an element $g$ of $\mathrm{PSL}(2, \mathbb{R})$. Thus, the (real) dimension of the moduli space of representations of $\pi_1(\Sigma)$ in $\mathrm{PSL}(2, \mathbb{R})$ modulo conjugations is $6p - 6$. This moduli space of representations of the fundamental group yields the Teichmüller space $T_p$. The moduli space $M_p$ is a branched quotient of that space.

Singularities of the moduli space arise when the image $\Gamma$ of $\rho$ has more automorphisms than such a generic subgroup of $G$ (whose only automorphisms are given by conjugations). Degenerations arise from limits of sequences of faithful, that is, injective representations $\rho_n$ that are no longer injective. Just as the Riemann surfaces are obtained as quotients $H/\Gamma$, the moduli space $M_p$ itself is likewise a quotient $T_p/C$ of the Teichmüller space $T_p$ by a discrete group, the so-called mapping class group. (This Teichmüller space $T_p$ is a complex space diffeomorphic—but not biholomorphic—to $\mathbb{C}^{3p-3}$. The complex structure was described by Bers through a holomorphic embedding into some complex Banach space. For recent results about this complex structure, we refer to [14]. $T_p$ parametrizes marked Riemann surfaces, that is, Riemann surfaces together with a choice of generators of the first homology group. Since all automorphisms of a hyperbolic Riemann surface act nontrivially on the first homology, Teichmüller space does not suffer from the problem of the moduli space, that Riemann surfaces with nontrivial automorphism groups can create singularities.)

This approach is also useful because it can be generalized to moduli spaces of representations of the fundamental group of a Kähler manifold in some linear algebraic group $G$. This is called non-Abelian Hodge theory and leads to profound insights into the structure of Kähler manifolds. In particular, because such representations can be shown to factor through holomorphic maps, this leads to the at present strongest approach to a general structure theory of Kähler manifolds via the Shafarevitch conjecture, see, e.g., [70–72].

2. **A Riemann surface $\Sigma$ is a 1-dimensional complex manifold. The moduli space is the semi-universal deformation space for such complex structures.**

More precisely: A Riemann surface $\Sigma$ is $S$ equipped with an (almost) complex structure. The relationship with 1 depends on the Poincaré uniformization theorem, which states that each compact Riemann surface of genus $p > 1$ can be represented as a quotient of $H$ as in 1. Conversely, each quotient $H/\Gamma$ as in 1 obviously inherits a complex structure from $H$, since $\Gamma$ operates by complex automorphisms on $H$.

The moduli space $M_p$ is then constructed as a universal space for variations of complex structures. This means that if $N$ is a complex space fibering over some base $B$ with the generic (=regular) fiber being a Riemann surface of genus $p$, we then obtain a holomorphic map $h : B_0 \to M_p$ where $B_0 \subset B$ are the points with regular fibers. In this manner, $M_p$, as a moduli space of complex structures, acquires a complex structure itself that is determined by the requirement that all these $h$ be holomorphic. Ideally, we would also like to have a holomorphic map $h_{fine} : N_0 \to M_{p,fine}$, $N_0$ being the space of regular fibers in $N$, mapping the fiber over $q \in B_0$ to the fiber over $h(q)$ in $M_{p,fine}$, but this is not always possible due to the difficulties with Riemann surfaces with nontrivial automorphisms. More precisely, $M_{p,fine}$ does not exist as such. A slight modification, however, leads to such a fine moduli space; namely, we only need to equip our Riemann surfaces additionally with some choice of a root of the canonical bundle in order to prevent nontrivial automorphisms. This is called a level structure. This gives a finite ramified cover of $M_p$. That cover is free of singularities and then yields a fine moduli space. (The Teichmüller space briefly described above is also a singularity-free cover of the moduli space, but, in contrast to the fine moduli space just introduced, it is an infinite cover and therefore not amenable to the constructions and techniques of algebraic geometry.) It is more subtle to understand what happens at the singular fibers. Here, we need a suitable compactification $\overline{M}_p$ of $M_p$ through certain singular Riemann surfaces. This, however, is better understood through the subsequent approaches to $M_p$ described below.

This construction is useful because, for example, it allows a geometric proof of the theorems of Arakelov-Parshin and Manin on the finiteness of the number of such fibrations of genus $p$ over a given compact base $B$ and the finiteness of the number of holomorphic sections of any given such fibration, see [69]. The idea is to show that because of the geometric properties of $M_p$ and $\overline{M}_p$, there can only be finitely many such holomorphic maps $h : B \to \overline{M}_p$ or (after taking care of the above need to take finite covers) from $N$ into a compactified fine moduli space.

3. **A Riemann surface is an algebraic curve, described by homogeneous polynomial equations. The moduli space is the space of coefficients of these polynomials modulo projective automorphisms.**

More precisely, a Riemann surface can be locally described as the common zero set of two homogeneous polynomials in three variables. The relationship with 2 depends on the Riemann–Roch theorem, which yields the existence of meromorphic functions.

**Insertion:** We briefly describe the relevant concepts. A line bundle $L$ on $\Sigma$ is given by an open covering $\{U_i\}_{i=1,\dots,m}$ of $\Sigma$ and transition functions $g_{ij} \in \mathcal{O}^*(U_i \cap U_j)$ ($\mathcal{O}^*$ denoting the nonvanishing holomorphic functions) satisfying

$$g_{ij} \cdot g_{ji} \equiv 1 \quad \text{on } U_i \cap U_j \text{ for all } i, j, \tag{1.4.2}$$

$$g_{ij} \cdot g_{jk} \cdot g_{ki} \equiv 1 \quad \text{on } U_i \cap U_j \cap U_k \text{ for all } i, j, k. \tag{1.4.3}$$

Two line bundles $L$, $L'$ with transition functions $g_{ij}$ and $g'_{ij}$, resp., are called isomorphic if there exist functions $\phi_i \in \mathcal{O}^*(U_i)$ for each $i$ with

$$g'_{ij} = \frac{\phi_i}{\phi_j} g_{ij} \quad \text{on each } U_i \cap U_j.$$

By multiplying transition functions we can define products of line bundles. The Abelian group of line bundles on $\Sigma$ is called the Picard group of $\Sigma$, $\mathrm{Pic}(\Sigma)$. The Picard group $\mathrm{Pic}(\Sigma)$ is isomorphic to the group of divisors $\mathrm{Div}(\Sigma)$ modulo linear equivalence. (Divisors are finite formal sums $\sum n_\alpha p_\alpha$ with $n_\alpha \in \mathbb{Z}$, $p_\alpha \in \Sigma$. The addition in $\mathbb{Z}$ induces a group structure on these divisors. Divisors are linearly equivalent when their difference is the divisor defined by a meromorphic function. This is verified when one expresses a divisor $D$ in terms of its local defining functions:

$$\left\{(U_i, f_i) : \frac{f_i}{f_j} \in \mathcal{O}^*(U_i \cap U_j)\right\}.$$

The function $f_i$ is meromorphic on $U_i$. When $p_\alpha \in U_i$, we require that $f_i$ has a zero (pole) of order $n_\alpha$ at $p_\alpha$ if $n_\alpha > 0 (< 0)$. At all other points, $f_i$ has to be holomorphic and nonzero.

We put

$$g_{ij} := \frac{f_i}{f_j} \quad \text{on } U_i \cap U_j$$

to define a line bundle, denoted by $[D]$.

Let $L$ be a line bundle with transition functions $g_{ij}$. A holomorphic section $h$ of $L$ is given by a collection $\{h_i \in \mathcal{O}(U_i)\}$ of holomorphic functions on $U_i$ satisfying

$$h_i = g_{ij} h_j \quad \text{on } U_i \cap U_j.$$

The zeros of a holomorphic section of a line bundle $L$ define an effective (i.e., all $n_i > 0$) divisor $E$, and when $L = [D]$, that divisor is linearly equivalent to $D$, that is, $E - D$ is the divisor of a meromorphic function. In general, the zeros and poles of a meromorphic section of $L$ define a divisor $D$ with $[D] = L$. The degree of a divisor is the sum of its coefficients, and from this one then defines also the degree of the line bundle $[D]$. Thus, the degree of a line bundle counts the zeros minus the poles of a meromorphic section.

The Riemann–Roch theorem for line bundles is then

**Theorem 1.3** *Let $L$ be a line bundle on the compact Riemann surface $\Sigma$ of genus $p$. Then the dimension of the space of holomorphic sections of $L$ satisfies the relation*

$$h^0(L) = \deg L - p + 1 + h^0(K \otimes L^{-1}) \tag{1.4.4}$$

*where $K$ is the canonical bundle of $\Sigma$, that is, the line bundle of holomorphic 1-forms.*

The equivalent formulation in terms of divisors replaces $h^0(L)$ by $h^0(D)$, the dimension of the space of effective divisors linearly equivalent to $D$.

Thus, the Riemann–Roch theorem can be viewed as an existence theorem for meromorphic functions, or, equivalently, for holomorphic sections of line bundles, whenever the right-hand side of (1.4.4) is positive. For example, $\deg K = 2p-2$ and $h^0(K) = p$, $\deg K^2 = 2\deg K = 4p-4$ and $h^0(K^2) = 3p-3$ for $p > 1$; $K^2$ is the line bundle whose sections are holomorphic quadratic differentials, that is, locally of the form $\varphi(z)dz^2$ with a holomorphic $\varphi$. Collections of holomorphic sections of a line bundle $L$ define mappings into projective spaces because a change of the local representation of $L$ multiplies them all by the same factor. One then needs sufficiently many independent sections to make such a map injective and thus to define an embedding of the Riemann surface into a projective space. In fact, one can show that every compact Riemann surface can be holomorphically embedded into $\mathbb{CP}^3$. Moreover, since by Chow's theorem every complex subvariety of $\mathbb{CP}^n$ is algebraic, our Riemann surface can then be represented by polynomial equations.

The relationship with 1 again goes via 2, that is, via the uniformization theorem. That theorem, however, is of a transcendental nature and thus outside the realm of algebraic geometry.

So, a Riemann surface becomes a (projective) algebraic variety in $\mathbb{CP}^3$, the zero set of algebraic equations. Such equations of a given degree can then be characterized by their coefficients. As automorphisms of $\mathbb{CP}^3$ lead to equivalent algebraic curves, one needs to divide these out. A difficulty emerges because the automorphism group of $\mathbb{CP}^3$ is not compact. Building upon the ideas of Hilbert, Mumford [83, 84] then developed geometric invariant theory to obtain the moduli space of algebraic curves. One then obtains the compactified Mumford–Deligne moduli space $\overline{M}_p$ as the moduli space of so-called stable curves, see [25]. As a moduli space of algebraic varieties, it is an algebraic variety itself, in agreement with the general principle.

4. **A Riemann surface is a collection of branch points on the Riemann sphere $S^2$ with branching orders satisfying the Riemann–Hurwitz formula. The moduli space is obtained from those collections by factoring out automorphisms of $S^2$.**

More precisely: Via some meromorphic function (whose existence again comes from the Riemann–Roch theorem), a Riemann surface is a branched cover of $S^2$, the Riemann sphere, which can also be identified with $\mathbb{CP}^1$. Again, we need to divide out automorphisms, this time those of $S^2$; they have the effect of moving the branch points around. This approach already led Riemann to count the number of moduli for Riemann surfaces of a given genus, that is, the dimension of the moduli space. This is explained in [51], for example.

5. **A Riemann surface is a finite algebraic extension of the field of rational functions $\mathbb{C}(x)$ in one variable over $\mathbb{C}$.**

From the algebraic representation in 3, one deduces that the field $k(\Sigma)$ of meromorphic functions on $\Sigma$ is a finite algebraic extension of the field of rational functions $\mathbb{C}(x)$ in one variable over $\mathbb{C}$. More precisely,

$$k(\Sigma) \sim \mathbb{C}(x)[y]/P(x, y) \tag{1.4.5}$$

for some irreducible polynomial $P$. For example, an elliptic curve, that is, a Riemann surface of genus 1, can be described by a cubic polynomial

$$y^2 - x(x-1)(x-\lambda) \tag{1.4.6}$$

for some $\lambda \in \mathbb{C} - \{0, 1\}$. For $z \in \Sigma$, we let $R_z$ be those meromorphic functions that are holomorphic at $z$. $R_z$ is then a subring of $k(\Sigma)$ and has a unique maximal ideal given by those functions that vanish at $z$. This means that, conversely, we can start with the field $k(\Sigma)$ and define the points of $\Sigma$ as the maximal ideals of local subrings of $k(\Sigma)$, and we may define a Riemann surface as a field of the form $\mathbb{C}(x)[y]/P(x,y)$ for some irreducible polynomial $P$. This encodes the functorial aspects: Let $\Sigma_1, \Sigma_2$ be compact Riemann surfaces, and let

$$\phi : k(\Sigma_2) \to k(\Sigma_1) \tag{1.4.7}$$

be a homomorphism whose restriction to $\mathbb{C}$ is the identity. Then there exists a unique holomorphic map

$$h : \Sigma_1 \to \Sigma_2 \tag{1.4.8}$$

with

$$\phi(f)(z) = f(h(z)) \tag{1.4.9}$$

for all $z \in \Sigma_1$ and all $f \in k(\Sigma_2)$.

This algebraic definition of a Riemann surface, which goes back to Dedekind and Weber, has the advantage that $\mathbb{C}$ can be replaced by any other algebraically closed field as the ground field. We may take finite fields $\mathbb{Z}_p$, and we can consider an algebraic equation $P(x,y)$ giving our Riemann surface as above, modulo $p$. Doing this for all prime numbers $p$ simultaneously yields important insights into the algebraic properties of such equations, see [36], and this was at the heart of Faltings' proof of the Mordell conjecture [35]. We can also take, instead of $\mathbb{C}$, a field of meromorphic functions on some variety $B$, in order to obtain an algebraic curve over a function field. In more elementary terminology, we now consider a polynomial $P(x,y)$ whose coefficients depend on the variable $w \in B$. We thus obtain a family of Riemann surfaces as in 2, but now from an algebraic point of view. The unification of those two possibilities of considering varying ground fields (depending on a prime number $p$ or on a variable $w$ in some algebraic variety) leads to arithmetic algebraic geometry.

6. **A Riemann surface $\Sigma$ is (defined by) an Abelian variety with a principal polarization, its Jacobian, that can be identified as the group of divisors of degree 0 on $\Sigma$ modulo linear equivalence or, equivalently, as the subgroup of the Picard group of line bundles of degree 0. Since not every principally polarized Abelian variety arises in this manner as the Jacobian of some Riemann surface, however, the moduli space of the latter is only a subvariety of the moduli space of principally polarized Abelian varieties.** By considering periods of holomorphic 1-forms, we can associate to a Riemann surface a principally polarized Abelian variety, its Jacobian. By Torelli's theorem, each Riemann surface is determined by its Jacobian. This means that we can identify a Riemann surface with this Abelian variety, and the space of Riemann surfaces becomes a subspace of the moduli space of principally polarized Abelian varieties. What is not so nice about this is that the solution of the Schottky problem, that is, the question of characterizing those Abelian varieties that are Jacobians of Riemann surfaces, is rather complicated [97].

**Insertion:** We explain the above concepts in some more detail. $H^0(\Sigma, \Omega^{1,0})$ is the space of holomorphic 1-forms on our compact Riemann surface $\Sigma$, that is, the holomorphic sections of the canonical bundle $K$. Thus, $h^0(\Omega^1) := h^0(K) = \dim_{\mathbb{C}} H^0(\Sigma, \Omega^{1,0}) = p$ by Riemann–Roch.

Let $\alpha_1, \dots, \alpha_p$ be a basis of $H^0(\Sigma, \Omega^{1,0})$, and $a_1, b_1, \dots, a_p, b_p$ a canonical homology basis for $\Sigma$. Then the period matrix of $\Sigma$ is defined as

$$\begin{pmatrix} \int_{a_1} \alpha_1 & \cdots & \int_{b_p} \alpha_1 \\ \vdots & & \vdots \\ \int_{a_1} \alpha_p & \cdots & \int_{b_p} \alpha_p \end{pmatrix}.$$

The column vectors of $\pi$,

$$P_i := \left( \int_{a_i} \alpha_1, \dots, \int_{a_i} \alpha_p \right) \quad \text{and} \quad P_{i+p} := \left( \int_{b_i} \alpha_1, \dots, \int_{b_i} \alpha_p \right), \quad i = 1, \dots, p,$$

are called the periods of $\Sigma$. $P_1, \dots, P_{2p}$ are linearly independent over $\mathbb{R}$ and thus generate a lattice

$$\Lambda := \left\{ n_1 P_1 + \cdots + n_{2p} P_{2p},\ n_j \in \mathbb{Z} \right\}$$

in $\mathbb{C}^p$.

**Definition 1.8** The Jacobian variety $J(\Sigma)$ of $\Sigma$ is the torus $\mathbb{C}^p/\Lambda$.

For each $z_0$ in $\Sigma$, we have the Abel map (a holomorphic embedding)

$$j : \Sigma \to J(\Sigma)$$

with

$$j(z) := \left( \int_{z_0}^{z} \alpha_1, \dots, \int_{z_0}^{z} \alpha_p \right) \bmod \Lambda.$$

Here $j(z)$ is independent of the choice of the path from $z_0$ to $z$, since a different choice changes the vector of integrals only by an element of $\Lambda$.

By the theorems of Abel and Jacobi, we obtain an isomorphism $\varphi$ from the group $\operatorname{Pic}^0(\Sigma)$ of line bundles of degree 0, that is, from the group $\operatorname{Div}^0(\Sigma)$ of divisors of degree 0 modulo linear equivalence, into the Jacobian $J(\Sigma)$ by writing a divisor $D$ of degree 0 as

$$D = \sum_{\nu} (z_\nu - w_\nu),$$

where $z_\nu, w_\nu \in \Sigma$ are not necessarily distinct, and putting

$$\varphi(D) := \left( \sum_{\nu} \int_{w_\nu}^{z_\nu} \alpha_1, \dots, \sum_{\nu} \int_{w_\nu}^{z_\nu} \alpha_p \right) \bmod \Lambda.$$

7. **A Riemann surface is a conformal structure on $S$, that is, a possibility to measure angles. Equivalently, it is an isometry class of Riemannian metrics modulo conformal factors. The moduli space is obtained by dividing the space of all Riemannian metrics on $S$ by isometries and conformal changes.** More

precisely: As already discovered by Gauss, a two-dimensional Riemannian manifold defines a conformal structure, that is, a Riemann surface. Different Riemannian metrics can lead to the same conformal structure, and so we need to divide out such equivalences. This is the approach of Tromba and Fischer, see [101]. Thus, we consider the space $R_p$ of all Riemannian metrics on $S$. As a space of Riemannian metrics, it carries itself a Riemannian metric. If $g$ is a Riemannian metric on $S$ and $h : S \to S$ is a diffeomorphism, $h^\star g$ is isometric to $g$ via $h$. Thus, we need to divide out the action of the diffeomorphism group $D_p$ of $S$. It acts isometrically on $R_p$ equipped with its Riemannian metric. Moreover, when we multiply a given metric $g$ by some positive function $\lambda$, the metric $\lambda g$ leads to the same conformal structure as $g$. Such multiplication by a positive function, however, does not induce an isometry of $R_p$ (and this is at the heart of the anomalies in string theory that ultimately force a particular dimension (26 in bosonic string theory)).

**Insertion:** Some details: Let $g \in R_p$ be some Riemannian metric on $S$. Suppressing the issues of the precise regularity class of the objects encountered, the tangent space $T_g R_p$ is given by symmetric $2 \times 2$ tensors $h = (h_{ij})$. Each such $h$ can be decomposed into its trace and trace-free parts:

$$h = \rho g + h' \qquad \rho : S \to \mathbb{R}, \tag{1.4.10}$$

$$h'_{ij} = h_{ij} - \frac{1}{2} g_{ij} g^{k\ell} h_{kl}. \tag{1.4.11}$$

The decomposition (1.4.10) is orthogonal w.r.t. the natural Riemannian structure on $T_g R_p$:

$$((h_{ij}), (\ell_{ij}))_{g,\kappa} := \int (g^{ijkm} + \kappa g^{ij} g^{km}) h_{ij} \ell_{km} \sqrt{\det g}\, dz^1\, dz^2 \tag{1.4.12}$$

with $\kappa > 0$ and

$$g^{ijkm} := \frac{1}{2}(g^{ik} g^{jm} + g^{im} g^{jk} - g^{ij} g^{km}).$$

Since the value of $\kappa$ will make no difference for us, we put $\kappa = \frac{1}{2}$ so that (1.4.12) becomes

$$((h_{ij}), (\ell_{ij}))_g := \int_S g^{ij} g^{km} h_{ik} \ell_{jm} \sqrt{\det g}\, dz^1\, dz^2. \tag{1.4.13}$$

As it stands, this is only a weak Riemannian metric on the infinite-dimensional space $R_p$, as (1.4.13) yields only an $L^2$-product, but Clarke [20] showed that it becomes a metric space with respect to the distance function induced by the Riemannian product of (1.4.13). (The completion of this metric space is identified in [19].)

From (1.4.13), we see that the Riemannian metric $(.,.)_g$ on $T_g R_p$ is invariant under the action of the diffeomorphism group, but not under conformal transformations.

In order to get rid of the ambiguity of the conformal factor, we need to find a suitable slice in $R_p$ transversal to the conformal changes. By Poincaré's theorem, any Riemannian metric on our surface $S$ of genus $p > 1$ is conformally equivalent to a unique hyperbolic metric, that is, $S$ becomes a quotient $H/\Gamma$ as above. This metric has constant curvature $-1$. With some differential geometry, one verifies that $-1$ is a regular value of the curvature functional, and so, by the implicit function theorem, the hyperbolic metrics yield a regular slice. Thus, we obtain the moduli space $M_p$ as the space $R_{p,-1}$ of metrics of curvature $-1$ divided by the action of $D_p$. In this way, the geometric structures on $R_p$ induce corresponding geometric structures on $M_p$ as described in Tromba's book [101]. $R_p$ is the space of symmetric, positive definite $2 \times 2$ tensors $(g_{ij})$ on $S$. As already explained, a tangent vector to $R_p$ is then a symmetric $2 \times 2$ tensor $(h_{ij})$, not necessarily positive definite. It is orthogonal to the conformal multiplications when it is trace-free, and it is orthogonal to the action of $D_p$ when it is divergence-free. Such a trace- and divergence-free symmetric tensor then can be identified with a holomorphic quadratic differential on the Riemann surface.

**Insertion:** Some details: We recall the decomposition

$$h = \rho g + h', \tag{1.4.14}$$

where $h'$ is trace-free. As we have seen, this decomposition is orthogonal w.r.t. the natural Riemannian metric on $T_g R_p$. In particular, since we only want to keep those directions that are orthogonal to conformal reparametrizations, we only need to consider the trace-free part $h'$. We next consider the infinitesimal action of the diffeomorphism group, with the aim of determining those $h'$ that are orthogonal to the action of that group as well. For that purpose, let $(\varphi_t) \subset D_p, \varphi_0 = id$, be a smooth family of diffeomorphisms, generated by the vector field

$$V(z) := \frac{d}{dt}\varphi_t(z)_{|t=0}. \tag{1.4.15}$$

The infinitesimal change of the metric $g$ under $(\varphi_t)$ is then given by

$$\frac{d}{dt}(\varphi_t^* g)_{|t=0}. \tag{1.4.16}$$

(This is the Lie derivative $L_V g$ of the metric in the direction of the vector field $V$.) With $\nabla$ denoting the covariant derivative for the metric $g$,

$$\begin{aligned} \frac{d}{dt}((\varphi_t^* g)_{|t=0})_{ij} &= g_{ik}(\nabla_{\frac{\partial}{\partial z^j}} V)^k + g_{jk}(\nabla_{\frac{\partial}{\partial z^i}})^k \\ &= g_{ij,k}V^k + g_{ik}V^k_{z^j} + g_{jk}V^k_{z^i}. \end{aligned} \tag{1.4.17}$$

In the above decomposition of $R_p$, the directions corresponding to conformal changes are given by the tensors $\rho g$, whereas those representing $D_p$ are of the form (1.4.17). It remains to identify the Teichmüller directions, i.e., those that are orthogonal to the preceding two types.

Our computations simplify considerably if we use conformal coordinates so that the metric $(g_{ij})$ is of the form

$$g_{ij}(z) = \lambda^2(z)\delta_{ij}. \tag{1.4.18}$$

If a symmetric tensor $h''$ is orthogonal to all multiples $\rho g$ of $g$, it has to be trace-free. If it is orthogonal to all tensors that arise from the infinitesimal action of the diffeomorphism group, that is, of type (1.4.16), we get, using the symmetry of $h''$

$$\begin{aligned} 0 &= \int g^{ij} g^{kl} h''_{ik} (g_{j\ell,m} V^m + 2 g_{jm} V^m_{z^\ell}) \sqrt{\det g}\, dz^1\, dz^2 \\ &= \int \frac{1}{\lambda^2} h''_{ik} \left( \delta_{ik} \left( \frac{\partial}{\partial z^m} \lambda^2 \right) V^m + 2\lambda^2 V^i_{z^k} \right) dz^1\, dz^2 \\ &= \int 2 h''_{ik} V^i_{z^k}\, dz^1\, dz^2, \quad \text{since } h'' \text{ is traceless.} \end{aligned}$$

If this holds for all vector fields $V$, we conclude

$$\frac{\partial}{\partial z^k} h''_{ik} = 0 \quad \text{for } i = 1, 2. \tag{1.4.19}$$

This means that $h''_{ik}$ is divergence free.

Thus, $h''$ is symmetric, trace-free, and divergence free. These conditions can be interpreted in a more concise manner as follows:

Being symmetric and trace-free, $h''$ is of the form

$$\begin{pmatrix} h''_{11} & h''_{12} \\ h''_{12} & h''_{22} \end{pmatrix} =: \begin{pmatrix} u & v \\ v & -u \end{pmatrix}.$$

Being divergence free, this tensor then has to satisfy

$$u_{z^1} = -v_{z^2}, \qquad u_{z^2} = v_{z^1}.$$

Thus, $u - iv$ is holomorphic, or, as a tensor,

$$\begin{aligned} h'' &= u(dz^1)^2 - u(dz^2)^2 + 2v\, dz^1\, dz^2 \\ &= \operatorname{Re}((u - iv)(dz^1 + i dz^2)^2) \end{aligned} \tag{1.4.20}$$

is the real part of a holomorphic quadratic differential

$$\phi\, dz^2 = (u - iv)\, dz^2.$$

Thus, we have identified the tangent directions of $R_p$ that correspond to nontrivial deformations of the complex structure as the (real parts of) holomorphic quadratic differentials on the Riemann surface defined by $(S, g)$.

Thus, the cotangent[11] space of $M_p$ at a point representing a Riemann surface $\Sigma$ is given by the holomorphic quadratic differentials on $\Sigma$. (This issue will be taken up again in Sect. 2.4 from a different point of view that also clarifies the relation between tangent and cotangent directions to the moduli space.) The complex dimension of this space is $3p - 3$ the Riemann–Roch theorem. $M_p$ then also inherits a Riemannian structure from that of $R_p$. The induced metric is the Petersson–Weil metric originally introduced in the context of approach 1. In a more abstract

[11] It is not very transparent from our preceding considerations that we have constructed the cotangent and not the tangent space, but a careful accounting of the transformation behaviors can clarify this issue.

framework, the so-called $L^2$-geometry of moduli spaces is investigated in [67]. Let $\Phi\,dz^2 = (u_1 - iv_1)\,dz^2$ and $\Psi\,dz^2 = (u_2 - iv_2)\,dz^2$ be two such differentials. Let $\rho^2(z)\,dz\,d\bar{z}$ be the hyperbolic metric. Then their Petersson–Weil product is

$$(\Phi\,dz^2, \Psi\,dz^2)_g = 2\int (u_1u_2 + v_1v_2)\cdot\frac{1}{\rho^2(z)}\,dz\,d\bar{z} = 2\,\mathrm{Re}\int \Phi\bar{\Psi}\frac{1}{\rho^2(z)}\,dz\,d\bar{z}. \tag{1.4.21}$$

We have now listed seven rather different approaches for defining what a Riemann surface is. It is a very remarkable and profound fact that all these approaches give fully compatible structures on the moduli space $M_p$. In each of them, one can construct a complex structure on $M_p$, and they all agree, and together with the Petersson–Weil metric, one then finds a Kähler structure on $M_p$.

Nevertheless, some remarks are in order here:

- From an algebraic point of view, the hyperbolic metric is a transcendental object and should be replaced by an algebraic one. There are also certain other natural metrics on a Riemann surface, like the Bergmann metric obtained from an $L^2$-orthonormal basis of holomorphic 1-forms, that is, the metric induced by embedding the Riemann surface into its Jacobian, or the Arakelov metric defined from an asymptotic expansion of the Green function, a rather natural object in string theory. One may replace the hyperbolic metric in (1.4.21) by another metric uniquely associated to each Riemann surface and still obtain a natural Riemannian metric on $M_p$. First steps in the direction of a systematic investigation have been done in [54, 55]. For more recent results in this direction, see [58, 59]. Let us briefly describe some of these constructions here. The Bergmann metric is given by

$$\rho_B^2 dz\wedge d\bar{z} := \sum_{i=1}^{p}\theta_i\wedge\bar{\theta}_i \tag{1.4.22}$$

where the $\theta_i$ are an $L^2$-orthonormal basis of the space of holomorphic 1-forms on $\Sigma$, that is,

$$\frac{i}{2}\int_\Sigma \theta_i\wedge\bar{\theta}_j = \delta_{ij}. \tag{1.4.23}$$

Equivalently, the metric is induced from the flat metric on the Jacobian $J(\Sigma)$ via the period map $j:\Sigma\to J(\Sigma)$. This latter description also shows that it does not depend on the choice of orthonormal basis—which, of course, is also readily checked directly. Moreover, the expression for the Bergmann metric is indeed positive definite, that is, it defines a metric, or equivalently, the derivative of the period map $j$ has maximal rank. This follows from the fact that there is no point on $\Sigma$ where all holomorphic 1-forms vanish simultaneously; this can be deduced from the Riemann–Roch theorem.

The Arakelov metric (references are [4, 18]) $\gamma^2 dz d\bar{z}$ is characterized by the property that its curvature is proportional to the Bergmann metric,

$$\frac{\partial^2}{\partial z \partial \bar{z}} \log \gamma = c_p \rho_B^2, \tag{1.4.24}$$

for some constant $c_p$ that depends only on $p$ and can, of course, be explicitly computed.[12] Alternatively, it is given in terms of an asymptotic expansion of the Green function of the Bergmann metric,

$$\log \gamma(z) = -\lim_{w \to z} (2\pi G(z, w) - \log |z - w|), \tag{1.4.25}$$

with $G$ satisfying

$$\frac{\partial^2}{\partial z \partial \bar{z}} G(z, w) = \frac{i}{2} \delta_w(z) + c_p \rho_B^2, \tag{1.4.26}$$

where $\delta_w$ is the Dirac functional supported at $w$, plus the normalization[13]

$$\int_\Sigma G(z, w) \frac{i}{2} \rho_B^2 dz \wedge d\bar{z} = 0. \tag{1.4.27}$$

The Green function is regular for $z \neq w$ and becomes $-\infty$ at $z = w$. $\exp 2\pi G(z, w)$ vanishes to first-order at $z = w$. The first term in the expansion of $\exp 2\pi G(z, w)$ is the universal term $|z - w|$, while the next one, $\gamma(z)$, encodes the geometry of the Riemann surface $\Sigma$.

If $\Delta_B$ is the Laplace operator for the Bergmann metric, and if $\phi_0, \phi_1, \ldots$ is an $L^2$-orthonormal basis of eigenfunctions with eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \ldots$, then the Green function is given by the expansion

$$G(z, w) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \phi_j(z) \overline{\phi_j(w)}. \tag{1.4.28}$$

In fact, one can perform this construction of the Green function and the associated metric on the basis of any conformal metric on $\Sigma$ in place of the Bergmann one. Arakelov discovered, however, that the Bergmann metric is distinguished here by the following property: When we use the Green function of a metric $g$ to define a metric on the canonical bundle $K$ by putting

$$\|dz\|(z_0) := \left( \lim_{z \to z_0} \frac{\exp 2\pi G(z, z_0)}{|z - z_0|} \right)^{-1}, \tag{1.4.29}$$

where the absolute value on the right-hand side is taken w.r.t. to local coordinates, that is, in $\mathbb{C}$, then the curvature of this metric on $K$ is a multiple of $g$ if and only

[12] In the sequel, $c_p$ will denote a generic such constant whose value can change between formulas.

[13] The characterization of the Arakelov metric in terms of its curvature likewise needs an additional normalization to fully determine it.

if we started with the Bergmann metric. In other words, we have the formula

$$\frac{1}{2\pi i}\frac{\partial^2}{\partial z\partial \bar{z}}\log \|s\|^2 dz\wedge d\bar{z} = c_p\rho_B^2 dz\wedge d\bar{z} \tag{1.4.30}$$

for any locally nonvanishing holomorphic section $s$ of $K$, and this is no longer valid for other metrics $g$ used to construct a Green function.

More generally, for a line bundle $L$ over $\Sigma$ with transition functions $g_{ij}$, a Hermitian metric $\lambda^2$ on $L$ is a collection of positive, smooth, real-valued functions $\lambda_i^2$ on $U_i$ with

$$\lambda_j^2 = \lambda_i^2 g_{ij}\overline{g_{ij}} \quad \text{on } U_i \cap U_j. \tag{1.4.31}$$

The norm of a section $h$ of $L$ given by the local collection $h_i$ is then defined via

$$\|h(z_0)\|^2 := \frac{|h_i(z_0)|^2}{\lambda_i^2(z_0)} \quad \text{for } z \in U_i. \tag{1.4.32}$$

The curvature or first Chern form is given by

$$c_1(L,\lambda^2) := \frac{1}{2\pi i}\frac{\partial^2}{\partial z\partial \bar{z}}\log \|h\|^2 dz\wedge d\bar{z} \tag{1.4.33}$$

for any meromorphic section $h$ and local coordinates $z$, and this is independent of the choices of $h$ and $z$. Arakelov called a Hermitian line bundle $L$ admissible w.r.t. a metric $\rho^2 dz\wedge d\bar{z}$ on $\Sigma$ if

$$c_1(L,\lambda^2) = \deg L\rho^2\, dz\wedge d\bar{z}. \tag{1.4.34}$$

Let $z_0 \in \Sigma$, and let $z$ be local coordinates mapping $z_0$ to 0. We can put a Hermitian metric on the line bundle $[z_0]$ by defining the norm of the local section $z$ in a neighborhood of $z_0$ as

$$|z|(z_1) = \exp G(z_1, z_0). \tag{1.4.35}$$

This metric is then admissible for the Bergmann metric. So, what is special about the Bergmann metric here is that if we start the construction of the Arakelov metric from the Green function of that metric then the curvature formula recovers that metric. This only holds for the Bergmann metric and not for any other one.

- We noted in 6 that we can inject the moduli space $M_p$ of Riemann surfaces of genus $p$ into the moduli space $A_p$ of principally polarized Abelian varieties of dimension $p$. The latter also carries a natural (locally Hermitian symmetric) metric. Since the map $j: M_p \to A_p$, while being injective by Torelli's theorem, is not of maximal rank everywhere, the pullback of that metric via $j$ has some singularities. Also, its behavior is qualitatively different from that of the Weil–Petersson metric, as will become clear below when we investigate degenerations of Riemann surfaces and their associated Jacobians.

### 1.4.3 Compactifications of Moduli Spaces

Some of the preceding approaches also naturally lead to compactifications of $M_p$.

1. We already mentioned the Mumford–Deligne compactification $\overline{M}_p$ as an algebraic variety. It consists of so-called stable curves, that is, possibly singular curves, but with a finite automorphism group. The sphere with no, one, or two punctures and the torus are thereby excluded. This is necessary for the Hausdorff property.

   The difficulty here can be seen from the following easy example: We consider annuli, and by the uniformization theorem, each annulus is characterized by a single modulus, a real number $0 < r < 1$; that is, it is conformally equivalent to an annulus

$$A_r := \{z \in \mathbb{C} : r < |z| < 1\}. \tag{1.4.36}$$

   Thus, the moduli space of annuli is $(0, 1)$. It seems obvious how to compactify it, namely by simply adding the boundary points $r = 0$ and $r = 1$. Now $r = 1$ does not correspond to a Riemann surface anymore, and so this is not a good limit. The annulus $A_r$, however, is conformally equivalent to the annulus

$$A'_r := \frac{1}{1-r} A_r = \left\{ z \in \mathbb{C} : \frac{r}{1-r} < |z| < \frac{1}{1-r} \right\}, \tag{1.4.37}$$

   which for $r \to 1$ converges to an infinite strip, that is, the limit can be identified with $\{x + iy \in \mathbb{C} : 0 < y < 1\}$. The boundary point $r = 0$ seems harmless because it simply corresponds to the punctured disk

$$D^* = \{z \in \mathbb{C} : 0 < |z| < 1\}. \tag{1.4.38}$$

   However, the annulus $A_r$ is also conformally equivalent to the annulus

$$A''_r := \frac{1}{\sqrt{r}} A_r = \left\{ z \in \mathbb{C} : \sqrt{r} < |z| < \frac{1}{\sqrt{r}} \right\}, \tag{1.4.39}$$

   and if we now let $r$ tend to 0, the limit is the punctured plane

$$\mathbb{C}^* = \{z \in \mathbb{C} : z \neq 0\}, \tag{1.4.40}$$

   which is not conformally equivalent to the punctured disk $D^*$. Thus, from the same limit $r \to 0$, we obtain two different limits, $D^*$ and $\mathbb{C}^*$, and therefore, we lose the Hausdorff property. Mumford's insight was that this problem essentially arises from the fact that the putative limit $\mathbb{C}^*$ has a noncompact automorphism group. In fact, its automorphism group contains all transformations of the form $z \to \lambda z$ for any $\lambda \in \mathbb{C}^*$. Mumford's theory then declared such limits as unstable and disallowed them. The problem of the noncompact automorphism group, however, will re-emerge later when we consider conformally invariant variational problems. The essential point is the following: We consider any Riemann surface $\Sigma$ and choose local coordinates $z$ in the open unit disk $U = \{z \in \mathbb{C} : |z| < 1\}$ around some point $p_0$, such that $p_0$ corresponds to 0. We then replace the coordinate $z$ by $z_\lambda := \lambda z \in \lambda U = \{z \in \mathbb{C} : |z| < \lambda\}$. When we let $\lambda \in \mathbb{R}$ tend to

$\infty$, we obtain $z_\infty \in \mathbb{C}$, but these are not coordinates for a local neighborhood of $p_0$ anymore because any fixed $z_\infty \in \mathbb{C}$ now corresponds to $p_0$ itself. In a sense to be made precise, they thus parametrize an infinitesimal neighborhood of $p_0$. We can compactify this infinitesimal coordinate patch $\mathbb{C}$ by adding the point at $\infty$ to obtain the sphere $S^2$. Thus, we have created a nontrivial Riemann surface, the sphere $S^2$, by blowing up a neighborhood of our point $p_0 \in \Sigma$. Again, if we allowed such processes in the construction of the moduli space, we would need to consider the union of $\Sigma$ and $S^2$ as a limit of the constant sequence $\Sigma$. (As this so-called "bubbling off" can be repeated, we should then even allow for infinitely many blown-up spheres.) At this point, as mentioned, this can simply be excluded by fiat, but the situation changes when these blown-up spheres carry some additional data, for example some part of the Lagrangian action in a variational problem.

2. We recall from 2 in Sect. 1.4.2 that if $N$ is a complex space fibering over some base $B$ with the generic (=regular) fiber being a Riemann surface of genus $p$, then we obtain a holomorphic map $h : B_0 \to M_p$ where $B_0 \subset B$ are the points with regular fibers. The fibers over $B_1 := B \backslash B_0$ are then singular, and we hope to extend $h$ across $B_1$, that is, obtain a holomorphic map $h : B \to \bar{M}_p$. Certain difficulties arise here from the possibility that not all such singular fibers in a holomorphic family need to be stable in the sense of Mumford. Thus, in particular, we cannot expect that the image of some point in $B_1$ is given by the complex structure of that singular fiber. Nevertheless, after lifting to finite covers so that the quotient singularities of $M_p$ disappear, one can extend $h$ to a holomorphic map $h : B \to \bar{M}_p$. This depends on certain hyperbolicity properties coming from the negative curvature of the Weil–Petersson metric on $M_p$ that lead to general extension properties for holomorphic maps, see [69].
3. While the preceding is a global aspect, one also has a convenient local model for degenerations of Riemann surfaces within 2. We consider two unit disks $D_1 = \{z \in \mathbb{C} : |z| < 1\}$ and $D_2 = \{w \in \mathbb{C} : |w| < 1\}$. For $t \in \mathbb{C}$, $|t| < 1$, we remove the interior disks $\{|z| \le |t|\}$, $\{|w| \le |t|\}$ and glue the rest by identifying $z$ with $w$ by the equation $zw = t$ to obtain an annular region $A_t$. For $t \to 0$, $A_t$ degenerates into the union of the two disks $D_1, D_2$ joined at the point $z = w = 0$. This is the local model for degeneration. The connection with the consideration of families as advocated in the preceding item of course comes from considering the smooth two-dimensional variety $N := \{(z, w, t) : zw - t = 0, |z|, |w|, |t| < 1\}$ for which $(z, w)$ yield global coordinates. $N$ fibers over the base $B := \{t : |t| < 1\}$, with a single singular fiber over $B_0 = \{0\}$.

   This local model is easily implemented in the context of compact Riemann surfaces as follows. We let $\Sigma_0$ be either a connected Riemann surface of genus $p - 1 > 0$ with two distinguished points $x_1, x_2$, called punctures, or the disjoint union of two Riemann surfaces $\Sigma^1, \Sigma^2$ of genera $p_1, p_2 > 0$ with $p_1 + p_2 = p$ and one puncture $x_i \in \Sigma^i$ each. We choose disjoint neighborhoods $U_1, U_2$ of the punctures and local coordinates $z : U_1 \to D_1$, $w : U_2 \to D_2$ with $z(x_1) = 0$, $w(x_2) = 0$. By performing the above grafting process on the coordinate disks $D_1$ and $D_2$, we obtain a Riemann surface $\Sigma_t$ of genus $p$ for $t \neq 0$. The correspondence

$$t \mapsto \Sigma_t \tag{1.4.41}$$

induces a map of $D^* = \{t \in \mathbb{C} : 0 < |t| < 1\}$ onto a complex curve in the moduli space $M_p$ which extends to a map from $D = \{t \in \mathbb{C} : |t| < 1\} = D^* \cup \{0\}$ into the compactification $\overline{M}_p$. (Because of the genus restrictions imposed, these degenerations all yield stable curves.)

4. Approaches 1 and 7 suggested looking at the moduli space of hyperbolic metrics. A hyperbolic metric on a compact surface can degenerate into a noncompact but complete hyperbolic metric of finite area with cusps. In the local model described in the previous item, this looks as follows. On the annulus $A_t = \{|t| < z < 1\}$, we have the hyperbolic metric

$$\frac{dz \wedge d\bar{z}}{|z|^2 \log^2 |z|} \left( \frac{\pi \frac{\log |z|}{\log |t|}}{\sin(\pi \frac{\log |z|}{\log |t|})} \right)^2. \tag{1.4.42}$$

For $|t| \to 0$, this converges to the hyperbolic metric on the punctured disk $\{z : 0 < |z| < 1\}$ given by

$$\frac{dz \wedge d\bar{z}}{|z|^2 \log^2 |z|}. \tag{1.4.43}$$

This metric is complete at 0, that is, the cusp 0 is at infinite distance from the points in the punctured disk. Also, the area of every punctured subdisk $\{z : 0 < |z| < \rho\}, 0 < \rho < 1$ is finite.

For the hyperbolic metric (1.4.42) on the annulus $A_t$, the middle curve $|z| = \sqrt{|t|}$ is the shortest of all the concentric circles, hence a closed geodesic, denoted by $c$. The reflection $z \mapsto \frac{t}{z}$ is then an isometry leaving $c$ fixed. Its length $l$ goes to 0 as $t \to 0$, while its distance from the boundary $|z| = 1$ goes to $\infty$. Thus, as $t$ goes to 0, the geodesic $c$ degenerates into a point curve at infinite distance from the interior. Therefore, in geometric terms, the degeneration is described by pinching a closed geodesic on some annulus inside our Riemann surface equipped with the hyperbolic metric. In fact, a hyperbolic metric on an annulus that is symmetric about a closed geodesic is uniquely determined by the length of that geodesic. That means that the hyperbolic on the annulus $A_t$ is induced by the metric of the Riemann surface $\Sigma_t$ as described above. Thus, even though we have presented it here as a local model, it captures the essential global aspects.

This consideration of varying hyperbolic metrics leads to the same compactification $\overline{M}_p$ of $M_p$ as a topological space, see [12]. The noncompact surfaces can be compactified as Riemann surfaces by adding a point at each cusp. We thus see that elements in the compactifying boundary of $\overline{M}_p$ correspond to surfaces of lower topological type with additional distinguished points, so-called punctures.

**Insertion:** The degeneration can also be described in terms of the generators of the discrete group $\Gamma$ considered in 1. Since hyperbolic elements are characterized by $|a+d| > 2$ and parabolic ones by $|a+d| = 2$, the relevant degeneration is one where we have a sequence $\Gamma_n$ of surface groups with hyperbolic elements $\gamma_n$ converging to a parabolic element $\gamma_0$. An example is the sequence of hyperbolic transformations

$$\gamma_n : z \mapsto \frac{(1+\frac{1}{n})z+\frac{1}{n}}{z+1} \tag{1.4.44}$$

converging to the parabolic transformation

$$\gamma_0 : z \mapsto \frac{z}{z+1}. \tag{1.4.45}$$

In the limit, the two fixed points of $\gamma_n$ merge into the single fixed point 0 of $\gamma_0$. Also, the length of the invariant geodesic for $\gamma_n$ approaches 0 as $n \to \infty$. Thus, again, the degeneration is described by pinching a closed geodesic on our Riemann surface equipped with the hyperbolic metric induced from $H$.

We now want to relate the geometric description of degeneration just established to the analytic model described previously. We first describe how to get from the analytic model to the geometric one. The behavior of the hyperbolic closed geodesic $|z| = \sqrt{|t|}$ for the hyperbolic metric (1.4.42) on the annulus $A_t$ translates into the following picture for hyperbolic isometries of $H$. We consider the hyperbolic isometry $\gamma_\lambda : z \mapsto \lambda z$ for some $\lambda > 1$. This leaves the imaginary axis in $H$ invariant, and so its image on the quotient $H/\Gamma$ by the group $\Gamma$ generated by $\gamma_\lambda$ is a closed geodesic of length $\int_1^\lambda \frac{dy}{y} = \log\lambda$. Via $z \mapsto \log\lambda \exp(\frac{2\pi i}{\log\lambda}(\log(-iz)+\log\lambda))$, $H/\Gamma$ is mapped onto $\mathbb{C}^*$, and the closed geodesic is mapped onto the circle $|w| = \log\lambda$.

In order to see how the geometric model can be translated into the analytic one, one uses the collar lemma, which says that if $\Sigma = H/\Gamma$ is a compact Riemann surface with a simple[14] closed geodesic $c$ of length $l$, then $\Sigma$ contains an annular region, called a collar, about $c$ isometric to $A_t$ with the hyperbolic metric $A_t$, $c$ corresponding to the middle curve $|z| = \sqrt{|t|}$. The boundary curves of the collar then are at a distance from $c$ of at least $\operatorname{arcsinh}(\frac{1}{\sinh(l/2)})$ which goes to $\infty$ as $l \to 0$. Thus, we are in the local situation described by the analytic model.

In fact, a theorem of Mumford says that pinching a simple closed geodesic is the only way a sequence of compact Riemann surfaces $\Sigma_n = H/\Gamma_n$ of fixed genus $p$ can degenerate. Namely, if the lengths of (simple) closed geodesics on $\Sigma_n$ are uniformly bounded below, then after selection of a subsequence, $\Gamma_n$ converges to a subgroup $\Gamma_0$ of $\mathrm{PSL}(2,\mathbb{R})$ for which $\Sigma_0 = H/\Gamma_0$ is a compact Riemann surface of the same genus $p$.

5. Since we have equipped $M_p$ in approach 7 with a Riemannian metric, the Petersson–Weil metric, we can study its compactification as a metric space. Again, as follows from the computations and estimates of Masur [80], this leads to the same $\overline{M}_p$ viewed as a topological space, see [107]. In particular, $M_p$ is not a complete metric space, that is, the boundary $\overline{M}_p \backslash M_p$ is at finite distance from the interior. Moreover, when we approach that boundary orthogonally

[14] That is, non-self-intersecting.

along some curve $c$, the tangent directions orthogonal to $c$ converge to boundary tangent directions. For a survey of some recent refinements of these results, see [108]. For the relation with the completion of the space $R_p$ of Riemannian metrics, see [19].

6. As explained in 6 of Sect. 1.4.1, by Torelli's theorem, the correspondence between a Riemann surface and its Jacobian leads to an injective mapping from $M_p$ into the moduli space $A_p$ of principally polarized Abelian varieties of dimension $p$. $A_p$ is a quotient $H_p/\Lambda_p$ of the Siegel upper half space by a discrete group ($H_1$ is simply the Poincaré upper half plane, and $H_p/\Lambda_p$ is then a higher-dimensional generalization of the modular curve $H/\mathrm{SL}(2,\mathbb{Z})$. $H_p$ is the space of symmetric complex $(p \times p)$ matrices with positive definite imaginary part. The discrete group $\Lambda_p$ is $\mathrm{Sp}(2p,\mathbb{Z})$, the group of real $(2p \times 2p)$ matrices $M$ with integer entries that satisfy $MJM^t = J$ for $J = \left(\begin{smallmatrix} 0 & Id \\ -Id & 0 \end{smallmatrix}\right)$). It admits a compactification first studied by Satake. Baily [8] then studied the induced compactification $\overline{\overline{M}}_p$. This is different from $\overline{M}_p$ and, in fact, highly singular. It can be obtained from $\overline{M}_p$ by forgetting the positions of the punctures or cusps of the limiting Riemann surfaces in $\overline{M}_p$. This is useful for the study of minimal surfaces of varying topological type, see [60, 61, 68], because the punctures would correspond to removable singularities. We shall discuss this issue briefly below in our study of the Dirichlet integral, our fundamental action functional, see Sect. 2.4.

   Also, in string theory, one ultimately wishes to extend the partition function over all possible genera, and one therefore needs some kind of universal moduli space that includes surfaces of all possible genera. The problem with the Mumford–Deligne compactification is that as the genus increases one gets surfaces with more and more punctures in the low boundary strata, in fact infinitely many in the limit of the genus going to infinity. This is avoided in the Satake–Baily compactification just described.

   There is another issue of interest here: We have described the degeneration of a family of Riemann surfaces by pinching a closed geodesic, that is, letting its length shrink to 0. These geodesics can be topologically of two different kinds. The first possibility is that it corresponds to a nontrivial homology class. When we pinch such a geodesic to a point and compactify the resulting surface by inserting two points, in the limit we still have a connected surface, but of lower genus, and therefore its space of holomorphic 1-forms has a smaller dimension. Therefore, the limiting surface also has a Jacobian of smaller dimension, and so we move into the boundary of $A_p$. The other possibility is that we pinch a curve that is homologically trivial, i.e., a commutator in the fundamental group $\pi_1(\Sigma)$. If we pinch such a curve and again compactify by inserting two points, the resulting surface is disconnected, but the genus $p$ is not lowered, that is, the sum $p_1 + p_2$ of the genera of the pieces $\Sigma_1$ and $\Sigma_2$ equals $p$. Therefore, also the dimension of the Jacobian is not lowered, and although we move to the boundary of the moduli space $M_p$, we stay inside the moduli space $A_p$. The Jacobian of our disconnected surface is simply the product of the Jacobians of the pieces $\Sigma_1$ and $\Sigma_2$. Of course, in order to substantiate these contemplations, we need to clarify in which sense the Jacobians of the family of degenerating surfaces converge to the Jacobian of the compactified limiting surface.

## 1.5 Supermanifolds

### *1.5.1 The Functorial Approach*

We present here the abstract mathematical setting of supermanifolds. We consider a super vector space (over a ground field of characteristic 0, like $\mathbb{R}$ or $\mathbb{C}$)

$$W = W_0 \oplus W_1$$

that is $\mathbb{Z}/2\mathbb{Z}$ graded. Elements $w$ of $W_0$ are called even, with parity $p(w) = 0$, those of $W_1$ odd, with parity $p(w) = 1$. Morphisms between super vector spaces are required to preserve the grading.

A super algebra $A$ is a super vector space together with a product $A \otimes A \to A$ which is a morphism in the above sense. It is also required to be associative and to have a unit, in the ordinary sense.

Now, the important point about super objects is that whenever an operation changes the order of two odd elements, a minus sign is introduced. In this sense, the super algebra $A$ is (super)commutative if for any two $a, b \in A$,

$$ab = (-1)^{p(a)p(b)} ba. \tag{1.5.1}$$

(Here and in the sequel, whenever the parity of an element enters a formula, that element is implicitly assumed to be of pure type, that is, either odd or even, but not a nontrivial sum of an odd and an even term. Generally, definitions are extended to inhomogeneous elements by linearity.)

The basic example of a commutative super algebra is a Grassmann algebra with generators $v_1, \dots, v_N$ satisfying

$$v_i v_j = -v_j v_i \quad \text{for all } i, j \tag{1.5.2}$$

and thus, in particular,

$$v_i^2 = 0 \quad \text{for all } i. \tag{1.5.3}$$

Hence, every element of this Grassmann algebra can be expanded as

$$v = a_0 + \sum_{i=1}^{N} a_i v_i + \cdots + a_{12\dots N} v_1 v_2 \cdots v_N. \tag{1.5.4}$$

Since the square of any generator vanishes by (1.5.4), the expansion terminates.

Similarly to (1.5.1), the rules defining Lie algebras pick up signs in the super context: The bracket of a super Lie algebra has to satisfy

$$[v, w] + (-1)^{p(v)p(w)}[w, v] = 0 \tag{1.5.5}$$

and the super Jacobi identity reads

$$[v,[w,u]]+(-1)^{p(v)(p(w)+p(u))}[w,[u,v]]+(-1)^{p(u)(p(v)+p(w))}[u,[v,w]]=0. \tag{1.5.6}$$

We now consider a complex super vector space $W$. A real structure on $W$ is given by a $\mathbb{C}$-antilinear automorphism

$$\kappa : W \to W$$

with

$$\kappa^2 w = (-1)^{p(w)} w. \tag{1.5.7}$$

This should be considered as complex conjugation. We point out, however, that on the odd part, we obtain a minus sign in (1.5.7). As an example, let us assume that over $\mathbb{R}$, the odd part $W_1$ has two generators $\vartheta_1$ and $\vartheta_2$; we may then put

$$\kappa(\vartheta_1) = \vartheta_2, \qquad \kappa(\vartheta_2) = -\vartheta_1. \tag{1.5.8}$$

A supersymmetric bilinear form[15] $(\cdot,\cdot)$ on $W$ is given by a symmetric form $(\cdot,\cdot)_0$ on $W_0$ and an alternating form $(\cdot,\cdot)_1$ on $W_1$ with

$$(\kappa v, \kappa w)_i = \overline{(v,w)_i} \quad \text{for } i = 0 \text{ and } 1. \tag{1.5.9}$$

This implies

$$\overline{(v,\kappa v)_i} = (\kappa v, \kappa^2 v)_i = (-1)^i (\kappa v, v)_i = (v, \kappa v)_i, \tag{1.5.10}$$

that is

$$(v,\kappa v) \quad \text{is real for all } v \in W. \tag{1.5.11}$$

We may thus call the form $(\cdot,\cdot)$ positive if

$$(v,\kappa v) > 0 \quad \text{for all } v \neq 0. \tag{1.5.12}$$

We can then define $(v,\kappa v)$ as a "norm" on a complex super vector space. The point is that $(v,v)=0$ if $v$ is odd. We therefore need $\kappa$ which is only meaningful if $W$ is defined over $\mathbb{C}$ so that each odd coordinate has two real components, as in our example. In that example, we could put

$$(\vartheta_1, \vartheta_2) = 1. \tag{1.5.13}$$

Then $\vartheta_1$ and $\vartheta_2$ would both have "norm" 1.

If we have a complex super algebra $A$, we could then require that $\kappa(ab) = \kappa(a)\,\kappa(b)$. If we wish to also include non-commutative algebras, like matrix algebras with their complex conjugation, it seems preferable to take as the basis object a star-operation, a $\mathbb{C}$-antilinear isomorphism from $A$ to the opposite algebra[16] satisfying

$$(ab)^* = (-1)^{p(a)p(b)} b^* a^*. \tag{1.5.14}$$

[15]Forms always take their values in $\mathbb{C}$.

[16]If the product in $A$ of $a$ and $b$ is $ab$, the product in the opposite algebra is defined as $(-1)^{p(a)p(b)}ba$.

A Hermitian form $\langle\cdot,\cdot\rangle$ on a complex super vector space is $\mathbb{C}$-antilinear in the first variable, $\mathbb{C}$-linear in the second one and satisfies

$$\overline{\langle v, w\rangle} = (-1)^{p(v)p(w)}\langle w, v\rangle. \tag{1.5.15}$$

We have, since $\langle\cdot,\cdot\rangle$ is assumed to be even, that

$$\langle v, w\rangle = 0 \quad \text{if } p(v) \neq p(w), \tag{1.5.16}$$

and also

$$\langle v, v\rangle \in \mathbb{R} \quad \text{for } v \text{ even}, \qquad \langle v, v\rangle \in i\mathbb{R} \quad \text{for } v \text{ odd}. \tag{1.5.17}$$

Then a super Hilbert space $H$ is a super vector space with a Hermitian form satisfying

$$\langle v, v\rangle > 0 \quad \text{for } v \text{ even}, \tag{1.5.18}$$

$$i^{-1}\langle v, v\rangle > 0 \quad \text{for } v \text{ odd} \tag{1.5.19}$$

and for which the ordinary Hilbert space structure defined by

$$\begin{aligned}
\langle\langle v, w\rangle\rangle &= \langle v, w\rangle && \text{for } v, w \text{ even},\\
\langle\langle v, w\rangle\rangle &= i^{-1}\langle v, w\rangle && \text{for } v, w \text{ odd},\\
\langle\langle v, w\rangle\rangle &= 0 && \text{for } v, w \text{ of different parities}
\end{aligned}$$

is complete. In the present treatise, we shall be concerned only with finite-dimensional super Hilbert spaces,[17] and the completeness is not an issue then because finite-dimensional Euclidean spaces are always complete.

### *1.5.2 Supermanifolds*

As for ordinary manifolds, there are several approaches to the definition of supermanifolds, and it is instructive to understand the relations between them. The standard model is $\mathbb{R}^{m|n}$ with even coordinates $(x^1, \dots, x^m)$ and odd coordinates $(\vartheta^1, \dots, \vartheta^n)$. Its sheaf of functions is $\mathcal{C}^\infty[\vartheta^1, \dots, \vartheta^n]$, the sheaf of commutative super algebras freely generated by odd quantities $\vartheta^1, \dots, \vartheta^n$ over the sheaf $\mathcal{C}^\infty$ of smooth functions on $\mathbb{R}^m$. Since the square of any $\vartheta^j$ vanishes, they generate a nilpotent ideal in this sheaf.

The functions in $\mathcal{C}^\infty[\vartheta^1, \dots, \vartheta^n]$ then admit expansions in the nilpotent variables. To explain this, we first consider $x = (x^1, \dots, x^m) \in U$ (open in $\mathbb{R}^m$) and $\xi = (\xi^1, \dots, \xi^m)$ where the $\xi^i$ are even nilpotent elements, i.e., of the form $\sum_{\alpha_1,\alpha_2} a_{\alpha_1,\alpha_2}\vartheta^{\alpha_1}\vartheta^{\alpha_2}+$ higher even-order terms, that is, the expansion starts with products of two $\vartheta^i$s. For a function that depends only on the even variables, we then require

$$\begin{aligned}
&F(x^1 + \xi^1, \dots, x^m + \xi^m)\\
&\quad = \sum_\gamma \frac{1}{\gamma_1! \cdots \gamma_m!}\partial_{x^1}^{\gamma_1} \cdots \partial_{x^m}^{\gamma_m} F(x^1, \dots, x^m)(\xi^1)^{\gamma_1} \cdots (\xi^m)^{\gamma_m}
\end{aligned} \tag{1.5.20}$$

[17] Perhaps, one should better speak of super Euclidean spaces in that case.

where $F$ as a function of $x = (x^1, \dots, x^m)$ is of class $C^\infty(U)$. Alternatively, we can view this as the rule for extending or pulling back a function of the ordinary coordinates $x = (x^1, \dots, x^m)$ to one of the coordinates $x + \xi = (x^1 + \xi^1, \dots, x^m + \xi^m)$. When the function is also allowed to depend on the odd variables, we have the expansion

$$F(x^1 + \xi^1, \dots, x^m + \xi^m, \vartheta^1, \dots, \vartheta^n) \\ = \sum_{\boldsymbol{\alpha}} \sum_{\boldsymbol{\gamma}} \frac{1}{\gamma_1! \cdots \gamma_m!} \partial_{x^1}^{\gamma_1} \cdots \partial_{x^m}^{\gamma_m} F^{\boldsymbol{\alpha}}(x^1, \dots, x^m)(\xi^1)^{\gamma_1} \cdots (\xi^m)^{\gamma_m} \vartheta^{\alpha_1} \dots \vartheta^{\alpha_k} \tag{1.5.21}$$

where the functions $F^{\boldsymbol{\alpha}}$ are of class $C^\infty(U)$. In these expansions, we may also allow for functions $F^{\boldsymbol{\alpha}}$ taking their values in a supercommutative algebra with unit in place of $\mathbb{R}$. Usually, these functions will then be even. We note that the expansions (1.5.20) and (1.5.21) contain a number of derivatives that depend on $n$. Since we want to reserve the flexibility to keep $n$ variable, we must work with $C^\infty$- instead of $C^k$-functions for some finite $k$.

There also exists a notion of (formal) integration, the Berezin integral, that inverts differentiation.

If $F$ is only a function of one odd variable $\vartheta$, we have

$$F(\vartheta) = a + b\vartheta \tag{1.5.22}$$

where $a, b$ are constants, i.e., independent of $\vartheta$. The integral of $F$ w.r.t. $\vartheta$ is then defined by linearity and the basic rules

$$\int d\vartheta = 0, \qquad \int \vartheta d\vartheta = 1. \tag{1.5.23}$$

This makes the integral translation invariant, i.e. for an odd $\varepsilon$,

$$\int F(\vartheta + \varepsilon) d\vartheta = \int (a + b\vartheta + b\varepsilon) d\vartheta = b \int \vartheta d\vartheta = \int F(\vartheta) d\vartheta. \tag{1.5.24}$$

Similarly, for a function $F$ of $n$ odd variables $\vartheta^1, \dots, \vartheta^n$,

$$F(\vartheta^1, \dots, \vartheta^n) = \sum_{\boldsymbol{\alpha}} b_{\boldsymbol{\alpha}} \vartheta^{\boldsymbol{\alpha}} \quad \begin{pmatrix} \text{with } \boldsymbol{\alpha} = 0 \text{ or } \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k) \\ 1 \le \alpha_1 < \alpha_2 \dots < \alpha_k \le n \end{pmatrix} \tag{1.5.25}$$

the integral is computed via the rules

$$\int d\vartheta^i = 0, \qquad \int \vartheta^i d\vartheta^j = \delta_{ij}. \tag{1.5.26}$$

Thus, we have

$$\int b_{\boldsymbol{\alpha}} \vartheta^{\boldsymbol{\alpha}} d\vartheta^{\alpha_k} \cdots d\vartheta^{\alpha_1} = b_{\boldsymbol{\alpha}} \quad \text{for } \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k). \tag{1.5.27}$$

A supermanifold of dimension $m|n$ can be defined by an atlas whose local charts are open domains of $\mathbb{R}^{m|n}$, that is, subsets with sheaf of functions $C^\infty(U_0)[\vartheta^1, \dots, \vartheta^n]$, where $U_0$ is an open subset of $\mathbb{R}^m$. In terms of functions, we are restricting the sheaf $\mathcal{C}^\infty[\vartheta^1, \dots, \vartheta^n]$ to $U_0$. This will of course be the general procedure for defining sub-supermanifolds. Note that we are restricting the even coordinates $x^1, \dots, x^m$ here, but not the odd ones. So, we have coordinate charts; coordinate transformations are then given by isomorphisms $f : U \to V$, $U, V$ open in $\mathbb{R}^{m|n}$. Such an isomorphism is given by even functions $f^1, \dots, f^m$ and odd functions $\phi^1, \dots, \phi^n$. To be an isomorphism, $f$ must be invertible, and the functions must be smooth, as always. And a morphism is invertible iff the underlying morphism defined by the $f^1, \dots, f^m$ is invertible; the odd functions do not play a role for invertibility.

Based on this, if $F$ is a function on the chart $U$, and if $f^1, \dots f^m, \phi^1, \dots, \phi^n$ are coordinate functions on our supermanifold, we can compute the values $F(f^1, \dots, f^m, \phi^1, \dots, \phi^n)$. We can therefore equivalently define a supermanifold as a topological space $M_0$ with a sheaf $\mathcal{O}_M$ of super ($\mathbb{R}$)-algebras that is locally isomorphic to $\mathbb{R}^{m|n}$. Functions on $M$ are then sections of the structure sheaf $\mathcal{O}_M$. Morphisms between supermanifolds $f : M \to N$ are then morphisms of ringed spaces, that is continuous maps $f_0 : M_0 \to N_0$ with a morphism of sheaves of super algebras from $f_0^* \mathcal{O}_N$ to $\mathcal{O}_M$. The odd functions generate a nilpotent ideal $J$ of $\mathcal{O}_M$, because the square of any odd coordinate is 0. The space $M_0$ with the sheaf $\mathcal{O}_M / J$ is then a smooth manifold of dimension $m$, called the reduced manifold $M_r$. A function $f$ on $M$ projects to a function $f_r$ on $M_r$, that is, a smooth function on $M_0$. The sheaf morphism determines the function. In particular, the evaluation of an odd function at a point of $M_0$ always yields 0. This also means that any map from an ordinary manifold, that is, a supermanifold of dimension $m|0$, into one of dimension $0|n$ vanishes identically. This can be remedied through the functor of points approach to supermanifolds. For a supermanifold $S$, an $S$-point of a supermanifold $M$ is a morphism $S \to M$. This construction is functorial in the sense that a morphism $\psi : T \to S$ induces a map from $M(S)$, the set of $S$-points of $M$, to $M(T)$ via $m \mapsto m \circ \psi$. Similarly, a morphism $f : M \to N$ induces $f_S : M(S) \to N(S)$, again functorially in $S$. In order to understand this more abstractly, we consider the so-called superpoints, the supermanifolds $\mathbb{R}^{0|n}$ defined as the space with structure sheaf $\mathbb{R}[\vartheta^1, \dots, \vartheta^n]$ (with anticommuting $\vartheta^j$, as always). Expressed differently, these are the supermanifolds $(\{\star\}, \Lambda_n)$ where $\{\star\}$ is an ordinary point and $\Lambda_n$ is a Grassmann algebra of $n$ generators. Then, see [93], these superpoints generate the category of finite-dimensional supermanifolds, that is, any such supermanifold is completely described by its superpoints. For supermanifolds $M, N$, one then defines (or, more precisely, shows the existence of) the inner Hom object $\underline{\mathrm{Hom}}(M, N)$ satisfying

$$\mathrm{Hom}(\mathbb{R}^{0|n}, \underline{\mathrm{Hom}}(M, N)) = \mathrm{Hom}(\mathbb{R}^{0|n} \times M, N) \tag{1.5.28}$$

for all $n \in \mathbb{N}$ and then also

$$\mathrm{Hom}(S, \underline{\mathrm{Hom}}(M, N)) = \mathrm{Hom}(S \times M, N) \tag{1.5.29}$$

for all supermanifolds $S$. In this way, the space of morphisms $M \to N$ also becomes a functor: For a supermanifold $S$, a morphism $M \times S \to N$, that is, a morphism $M \to N$ depending on a parameter in $S$, is then an $S$-point of $\underline{\mathrm{Hom}}(M, N)$. The morphisms $\mathbb{R}^{0|n} \times M \to N$ are then the superpoints of the supermanifold of morphisms $\underline{\mathrm{Hom}}(M, N)$ (in contrast to $\mathrm{Hom}(M, N)$ which is not a supermanifold, but rather the reduced space (see below) underlying the supermanifold $\underline{\mathrm{Hom}}(M, N)$).

In that way, we see that there also exist nontrivial odd functions on an ordinary manifold, say $\mathbb{R}$, even though their values vanish on all points of $\mathbb{R}$. To be concrete, consider $S = \mathbb{R}^{0|n}$. $\mathbb{R} \times S$ then has the sheaf $C^\infty(\mathbb{R}) \otimes \mathbb{R}[\vartheta^1, \ldots, \vartheta^n]$. Now take another space $T$ that is odd like $S$, with sheaf $\mathbb{R}[\eta^1, \ldots, \eta^m]$. We consider a map $\psi : \mathbb{R} \times S \to T$, that is $\psi : \mathbb{R}^{1|n} \to \mathbb{R}^{0|m}$, given by

$$\mathcal{C}^\infty(T) = \mathbb{R}[\eta^1, \ldots, \eta^m] \to C^\infty(\mathbb{R}) \otimes \mathbb{R}[\vartheta^1, \ldots, \vartheta^n],$$
$$\eta^j \mapsto a^j_k(t)\vartheta^k$$

which we can also write as

$$\eta^j(t) = a^j_k(t)\vartheta^k.$$

Of course, this vanishes at all points of $\mathbb{R} = (\mathbb{R} \times S)_0$, but nevertheless it is a nontrivial morphism.

In the converse direction, let us take $n = 1$, i.e., consider $S = \mathbb{R}^{0|1}$, the space with sheaf $\mathbb{R}[\vartheta]$ ($\vartheta^2 = 0$), and a morphism

$$S \to M_0$$

into some ordinary manifold $M_0$. This is given by an algebra homomorphism

$$C^\infty(M_0) \to \mathbb{R}[\vartheta],$$
$$f \mapsto a_0(f) + a_1(f)\vartheta.$$

The homomorphism condition implies first that

$$a_0 : C^\infty(M_0) \to \mathbb{R},$$
$$f \mapsto a_0(f)$$

is an algebra homomorphism, and, as is easily derived, it is therefore given by the evaluation at some point $x \in M_0$, that is $a_0(f) = f(x)$. Secondly we obtain, using the homomorphism condition,

$$a_0(fg) + a_1(fg)\vartheta = (a_0(f) + a_1(f)\vartheta)(a_0(g) + a_1(g)\vartheta)$$
$$= f(x)g(x) + (f(x)a_1(g) + g(x)a_1(f))\vartheta,$$

which means that $a_1$ is a derivation over functions, that is, the derivative in the direction of some tangent vector $v_x \in T_x M_0$,

$$a_1(f) = v_x f.$$

Thus, we could view the super point $S$ with its sheaf $\mathbb{R}[\vartheta]$ as an abstract (odd) tangent vector. The maps $S \to M_0$ correspond to points in the tangent bundle $TM_0$. If $M$ is a general supermanifold, the same applies, except that we get a sign from the odd functions, that is,

$$a_1(fg) = a_1(f)g(x) + (-1)^{p(f)} f(x)a_1(g).$$

Thus, $a_1$ is an odd homomorphism from the local ring at $x$ to $\mathbb{R}$.

In any case, we have a projection $M \to M_r$ from a supermanifold to its reduced manifold $M_r$. Conversely, by Batchelor's theorem, any smooth supermanifold is (non-canonically) isomorphic to one of the form $(M_r, \wedge^* V)$. Thus, we can obtain $M$ from the smooth ordinary manifold $M_r$ and a locally free module $V$ over the sheaf $C^\infty(M_r)$; namely, $M$ can be obtained as $M_r$ with the sheaf $\wedge^* V$ graded by the exterior degree mod 2, and the inclusion of $C^\infty(M_r)$ into $\wedge^* V$ defines a morphism $M \to M_r$ that retracts the embedding of $M_r$ into $M$. It is important to realize that these constructions are not canonical, since they are not invariant under automorphisms of $M$ if $m \geq 1, n \geq 2$. Namely, simply consider $\mathbb{R}^{1|2}$ with coordinates $(x, \vartheta^1, \vartheta^2)$ and the automorphism

$$(x, \vartheta^1, \vartheta^2) \mapsto (x + \vartheta^1\vartheta^2, \vartheta^1, \vartheta^2).$$

This example also shows us that decompositions of functions according to their degree are not invariant under automorphisms, and thus not invariant under coordinate transformations. Namely, if we have a function $f$ that, in the coordinates $(x, \vartheta^1, \vartheta^2)$, only depends on $x$, and if we denote its expression in the new coordinates $(x + \vartheta^1\vartheta^2, \vartheta^1, \vartheta^2)$ by $g$, we have

$$\begin{aligned} f(x) &= g(x + \vartheta^1\vartheta^2, \vartheta^1, \vartheta^2) \\ &= g_0(x) + g_0'(x)\vartheta^1\vartheta^2 + g_1(x)\vartheta^1\vartheta^2. \end{aligned}$$

Here, we have used the rule (1.5.20) for the Taylor expansion for a function $g_0$ of the even coordinates, and we then need to add a counter-term $g_1(x) = -g_0'(x)$ in order to compensate for the $\vartheta^1\vartheta^2$ term from the Taylor expansion of $g_0$. Note that here we work over a trivial base $S$.

If we are Taylor-expanding functions as explained, then if $M_0$ is an ordinary manifold, that is, a supermanifold with odd dimension 0, and if we consider a map $f_S : M_0 \times S \to N$, then the odd dimension of $S$ determines the maximal degree occurring in that expansion of $f_S$. In the physics literature, one expresses this by fixing the number $N$ of Grassmann generators. In the present framework, this corresponds to the odd dimension of $S$.

One should also note that a super vector space $W = W_0 \oplus W_1$ is not a supermanifold, unless the odd part $W_1$ is trivial. If the even and odd part have dimensions $m$ and $n$, resp., then $W$ has the underlying structure of an $m + n$-dimensional ordinary vector space, whereas the ordinary manifold $M_r$ underlying an $(m|n)$-dimensional supermanifold is only $m$-dimensional. Of course, one can canonically construct a supermanifold from a super vector space, but as such, the two structures of a super vector space and of a supermanifold are different.

A super Lie group is a supermanifold that is functorially characterized by the property that for all supermanifolds $S$, $\mathrm{Hom}(S, M)$ is a group such that the group operations are smooth morphisms of supermanifolds. It can be obtained by exponentiation from a super Lie algebra. More precisely, however, for that exponentiation, we also need to be able to multiply the elements of the super Lie algebra by the odd variables $\vartheta^j$, that is, on the super Lie algebra, we also need the structure of a left supermodule over the algebra spanned by the $\vartheta^j$.[18] For a super Lie group $H$, we can consider left multiplication by an element $h$

$$L_h : H \to H, \qquad L_h(k) = hk \quad \text{for } k \in H. \tag{1.5.30}$$

This induces a map $(L_h)_\star$ on the vector fields on $H$, given by

$$((L_h)_\star X)F := X(F \circ L_h) \quad \text{for functions } F. \tag{1.5.31}$$

When $(L_h)_\star X = X$ for all $h \in H$, the vector field is called left-invariant. The left-invariant vector fields then span a super Lie algebra (with the graded commutator of vector fields as the bracket) that is also a super module over the odd variables.

To see the principle, we consider $\mathbb{R}^{1|1}$ with coordinates $t, \vartheta$. This space carries a super Lie group structure given by

$$(t^1, \vartheta^1)(t^2, \vartheta^2) = (t^1 + t^2 + \vartheta^1\vartheta^2, \vartheta^1 + \vartheta^2). \tag{1.5.32}$$

The translation in the $t$-direction is generated by the vector field

$$\partial_t \left( := \frac{\partial}{\partial t} \right), \tag{1.5.33}$$

the one in the $\vartheta$-direction by

$$D := \partial_\vartheta - \vartheta \partial_t. \tag{1.5.34}$$

We note that $D$ does not induce a morphism in our sense as it changes the parity. We have the relation

$$[D, D] = 2D^2 = -2\partial_t. \tag{1.5.35}$$

$(D, \partial_t)$ constitute a basis of the left invariant vector fields on the super Lie group, while $(Q := \partial_\vartheta + \vartheta \partial_t, \partial_t)$ is a basis for the right invariant ones. We shall meet these vector fields when we consider supersymmetry transformations. The important point is that they generate diffeomorphisms of the superspace $\mathbb{R}^{1|1}$.

*Remark* The treatment of supermanifolds presented here has been developed by Leites [76], Manin [79], Bernstein, Deligne and Morgan [24], and Freed [39, 40]. Another reference is [102]. The comprehensive presentation of the subject is [11]. The superdiffeomorphism group is investigated in [94].

### *1.5.3 Super Riemann Surfaces*

As an example, we now consider super Riemann surfaces (SRSs). While above, we have defined supermanifolds over $\mathbb{R}$, it is straightforward to develop the same constructions over $\mathbb{C}$. An SRS then has one commuting complex coordinate $z$ and

[18] We only have a supermodule instead of a super vector space because that algebra is only a ring, but not a field.

one anticommuting one $\vartheta$. In addition, the coordinate transformations are required to be superconformal. To explain this, we start with the coordinate transformation formula for a single supercomplex manifold $M$ of complex dimension $(1|1)$, which has to be even, that is, of the form

$$\begin{aligned}\tilde{z} &= f(z),\\ \tilde{\vartheta} &= \vartheta h(z)\end{aligned} \tag{1.5.36}$$

with holomorphic functions $f$ and $h$ where $f$ is required to have a nonvanishing derivative, that is, to be conformal. The structure sheaf is thus of the form $\mathcal{O}_M = \mathcal{O}_{M,0} \oplus \mathcal{O}_{M,1}$ where $\mathcal{O}_{M,0}$ is the sheaf of holomorphic functions on the underlying Riemann surface $M_r$ and $\mathcal{O}_{M,1}$ is a sheaf of locally free modules of rank $0|1$ over $\mathcal{O}_{M,0}$. Up to a change of parity, this then defines a line bundle $L$ over $M_r$, and conversely, given such a line bundle $L$ over $M_r$, changing the parity of its sections from even to odd then defines the structure sheaf of supercomplex manifold of dimension $(1|1)$. Thus, such $(1|1)$-dimensional supercomplex manifolds and ordinary Riemann surfaces with a line bundle $L$ stand in bijective correspondence. When we look at families of such supercomplex manifolds, however, we may also take base spaces with odd directions, and we have the more general transformation formula

$$\begin{aligned}\tilde{z} &= f(z) + \vartheta k(z),\\ \tilde{\vartheta} &= g(z) + \vartheta h(z)\end{aligned} \tag{1.5.37}$$

with holomorphic functions $f, k, g, h$ and $f$ again conformal.

In order to define a super Riemann surface, we require in addition that the structure be superconformal. This means the following: We look at the derivative operators $\partial_z$ and $\tau := \partial_\vartheta + \vartheta\partial_z$; they satisfy

$$\frac{1}{2}[\tau,\tau] = \tau^2 = \partial_z. \tag{1.5.38}$$

We have the transformation rule

$$\tau = (\tau\tilde{\vartheta})\tilde{\tau} + (\tau\tilde{z} - \tilde{\vartheta}\tau\tilde{\vartheta})\tilde{\tau}^2. \tag{1.5.39}$$

(To see this, one computes

$$\partial_z = (f_z + \vartheta k_z)\partial_{\tilde{z}} + (g_z + \vartheta h_z)\partial_{\tilde{\vartheta}}, \tag{1.5.40}$$

$$\partial_\vartheta = h\partial_{\tilde{\vartheta}} + k\partial_{\tilde{z}}, \tag{1.5.41}$$

$$\tau\tilde{\vartheta} = h + \vartheta g_z, \tag{1.5.42}$$

$$\tau\tilde{z} = k + \vartheta f_z \tag{1.5.43}$$

from which

$$\begin{aligned}\tau &= \partial_\vartheta + \vartheta\partial_z = (h + \vartheta g_z)\partial_{\tilde{\vartheta}} + (k + \vartheta f_z)\partial_{\tilde{z}}\\ &= (h + \vartheta g_z)(\partial_{\tilde{\vartheta}} + \tilde{\vartheta}\partial_{\tilde{z}}) - (g + \vartheta h)(h + \vartheta g_z)\partial_{\tilde{z}} + (k + \vartheta f_z)\partial_{\tilde{z}}\\ &= (\tau\tilde{\vartheta})\tilde{\tau} + (\tau\tilde{z} - \tilde{\vartheta}\tau\tilde{\vartheta})\tilde{\tau}^2\end{aligned} \tag{1.5.44}$$

which is the required formula.) In the same manner as for an ordinary Riemann surface, that is, one with transition functions $\tilde{z} = f(z)$, the holomorphicity of $f$ implies that $\partial_z$ is a multiple of $\partial_{\tilde{z}}$, $\partial_z = \partial_z f \partial_{\tilde{z}}$. We now require for an SRS that $\tau$ transforms homogeneously, that is, $\tau$ is a multiple of $\tilde{\tau}$. In view of (1.5.39), for a family, this leads to the transformation law

$$\begin{aligned}\tilde{z} &= f(z) + \vartheta g(z) h(z),\\ \tilde{\vartheta} &= g(z) + \vartheta h(z)\end{aligned} \tag{1.5.45}$$

with

$$h^2(z) = \partial_z f(z) + g(z)\partial_z g(z). \tag{1.5.46}$$

Here, $f(z)$ is a commuting holomorphic function with $\frac{\partial}{\partial z} f \neq 0$, i.e., $f$ is conformal, and $g(z)$ is an anticommuting one.

These transformations then leave the line element $dz + \vartheta d\vartheta$ invariant up to conformal scaling. (The conformal factor is $\frac{\partial}{\partial z} f(z) + g(z)\frac{\partial}{\partial z} g(z)$, and one has to use (1.5.46).)

Given a single SRS $\Sigma$, we can put all the $g = 0$ and obtain the transformation rules

$$\begin{aligned}\tilde{z} &= f(z),\\ \tilde{\vartheta} &= \vartheta h(z)\end{aligned} \tag{1.5.47}$$

with $h^2(z) = \partial_z f(z)$. The holomorphic transformation functions $f$ of $z$ define an ordinary Riemann surface $\Sigma_r$, but the transformations of the odd coordinate $\vartheta$ additionally require the choice of a square root $h(z)$ of $\frac{\partial}{\partial z} f(z)$. In other words, they determine a spin structure on $\Sigma_r$. If $p$ is the genus of $\Sigma_r$, we have $2^{2p}$ different spin structures on $\Sigma_r$. In particular, we see that the super Teichmüller space of super Riemann surfaces of genus $p$ has at least $2^{2p}$ components (this does not hold for the super moduli space, because modular transformations can mix the spin structures). By the Riemann–Roch theorem (stated in 3 of Sect. 1.4.2 and recalled below), the number of even moduli (over $\mathbb{C}$) minus the number of conformal transformations of $\Sigma_r$ is $3p-3$ while the number of odd moduli minus the number of odd superconformal transformations is $2p-2$. (The even moduli here can be identified with sections of $K^2$, where $K$ is the canonical bundle of the underlying Riemann surface, while the odd ones correspond to sections of $K^{3/2}$. The Riemann–Roch theorem says that the space of sections of a line bundle $L$ over a Riemann surface $\Sigma$ of genus $p$ has dimension

$$h^0(\Sigma, L) = \deg L - p + 1 + h^0(\Sigma, K \otimes L^{-1}) \tag{1.5.48}$$

and the degree of the canonical bundle is $2p - 2$.)

On a sphere, we have no nontrivial spin structures and no super moduli, but, in agreement with the Riemann–Roch theorem, the superconformal transformations are of the form $f(z) = \frac{az+b}{cz+d}$ (with the normalization $ad - bc = 1$), $g(z) = \frac{\gamma z + \delta}{cz+d}$, that is, 3 even and 2 odd parameters.

More generally, on the supersphere, when, instead of (1.5.47), we allow for the general type of coordinate transformations (1.5.37), we obtain the orthosymplectic group OSp(1/2):

$$T = \begin{pmatrix} a & b & \alpha \\ c & d & \beta \\ \gamma & \delta & t \end{pmatrix},$$

$a, b, c, d, t$ even (commuting), $\alpha, \beta, \gamma, \delta$ odd (anticommuting).

$$T^{\mathrm{st}} K T = K \quad (T^{\mathrm{st}} \text{ supertransposed}),$$

with the orthosymplectic form

$$K = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The transformation

$$z \mapsto \frac{a\,z + b + \alpha\vartheta}{cz + d + \beta\vartheta}, \qquad \vartheta \mapsto \frac{\gamma z + \delta + t\vartheta}{cz + d + \beta\vartheta}$$

leaves the line element $dz + \vartheta\, d\vartheta$ invariant up to conformal scaling.

In that case, just to see some formulae,

$$dz \mapsto \frac{dz}{(cz + d + \beta\vartheta)^2},$$

$$z_{12} = z_1 - z_2 - \vartheta_1\vartheta_2 \mapsto \frac{z_{12}}{(cz_1 + d + \beta\vartheta_1)(cz_2 + d + \beta\vartheta_2)},$$

$$dz \wedge d\vartheta \mapsto \frac{dz \wedge d\vartheta}{cz + d + \beta\vartheta}.$$

Obviously, this extends the operation of $Sl(2, \mathbb{C})$ to the super case.

The supersphere can be covered by two coordinate patches, with transition

$$\tilde{z} = \frac{1}{z}, \qquad \tilde{\vartheta} = \frac{i\vartheta}{z}.$$

(Cf. (1.5.45): Here, $h = \sqrt{\frac{\partial}{\partial z} f}$.)

Genus 1 is next. A torus with a spin structure is described by the rigid super conformal transformation

$$(z, \vartheta) \cong (z + 1, \eta_1\vartheta) \cong (z + \tau, \eta_2\vartheta),$$

where $\tau$ is taken from the usual period domain, and the $\eta_i = \pm 1$ determine the spin structure. Since the only holomorphic functions on a torus are the constants, we obtain a nontrivial supermodulus $\nu$ only in the case of a trivial spin structure, that is, $\eta_1 = 1 = \eta_2$. In that case, the periodicities are

$$(z, \vartheta) \cong (z + 1, \vartheta) \cong (z + \tau + \vartheta\nu, \vartheta + \nu).$$

In agreement with Riemann–Roch, we then also have the odd superconformal transformation given infinitesimally as

$$(z, \vartheta) \mapsto (z + \vartheta\varepsilon, \vartheta + \varepsilon).$$

In particular, we see that there can exist nontrivial odd moduli, and the supermoduli space is bigger than just the moduli space of ordinary Riemann surfaces with spin structures. We observe, however, that $\varepsilon \mapsto -\varepsilon$ is a superconformal transformation, and so the supertori corresponding to $\varepsilon$ and $-\varepsilon$ are equivalent. Thus, the corresponding component of the supermoduli space is a $\mathbb{Z}_2$ super orbifold, with a singularity at $\varepsilon = 0$.

When we look at functions on this super torus, we obtain the periodicity condition

$$f(z, \vartheta) = f(z + \tau + \vartheta\nu, \vartheta + \nu),$$

that is, after Taylor expanding,

$$f_0(z) + f_1(z)\vartheta = f_0(z + \tau) + f_0'(z + \tau)\vartheta\nu + f_1(z + \tau)(\vartheta + \nu)$$

which implies that $f_0'$ vanishes when $\nu \neq 0$, that is, $f_0$ is constant (over a trivial base $S$ again). $f_1$ is less trivial. The situation becomes richer when we look at mappings between two such supertori, with moduli $(\tau, \nu)$ and $(\tilde{\tau}, \tilde{\nu})$, resp. We then expand to obtain

$$f_0(z) + \tilde{\tau} + f_1(z)\vartheta\tilde{\nu} = f_0(z + \tau) + f_0'(z + \tau)\vartheta\nu$$

and

$$f_1(z)\vartheta + \tilde{\nu} = f_1(z + \tau)(\vartheta + \nu).$$

The first equation expresses $f_0'$ in terms of $f_1$ or conversely, while the second one restricts $f_1$. However, we should be careful here as $f_0$ need not be holomorphic, and so $f_0'$ stands for a $(2 \times 2)$-matrix.

*Remark* For a treatment of super Riemann surfaces as needed for superstring theory, we refer to Crane and Rabin [21, 89] and Polchinski [88]. A general mathematical perspective is developed by Leites and his coauthors in [32]. A very lucid discussion, which we have also partly utilized here, can be found in [93].

We should note that the above definition is not the only possible for an SRS. In fact, there are several superextensions of the conformal algebra, and each of them could be taken as the basis for the definition of an SRS. The one used here corresponds to the superconformal algebra $\mathfrak{k}^L(1|1)$ and yields the $N = 1$ worldsheets of superstring theory and $2D$ supergravity.

### *1.5.4 Super Minkowski Space*

Now assume that we have a vector space $V$ with a quadratic form $Q$ and a representation of the Clifford algebra $Cl(Q)$ for which a symmetric equivariant morphism $\Gamma$

as in (1.3.25) exists. Following the presentation in [22], we may then construct an object that captures deeper aspects of the physical concept of supersymmetry than just a supermanifold, namely a space that incorporates the symmetry between vectors and spinors as representations of bosons and fermions, resp. For that purpose, we consider the vector space $V$ as the Lie algebra of its translations, and construct the super Lie algebra

$$\mathfrak{l} := V \oplus S^\star \tag{1.5.49}$$

and the bracket [., .]. This bracket is trivial on $V$ (that is, $V$ is central) and is given by

$$[s, t] = -2\Gamma(s, t) \in V \tag{1.5.50}$$

on $S^\star$. Super Minkowski space $M$ is then defined as the supermanifold underlying the Lie group $\exp(\mathfrak{l})$; its reduced space is thus given by the affine space $V$, and its odd directions are given by $S^\star$.

In the Minkowski case, the super Lie algebra (1.5.49) leads to the super Poincaré algebra

$$(V \oplus \mathfrak{so}(V)) \oplus S^\star. \tag{1.5.51}$$