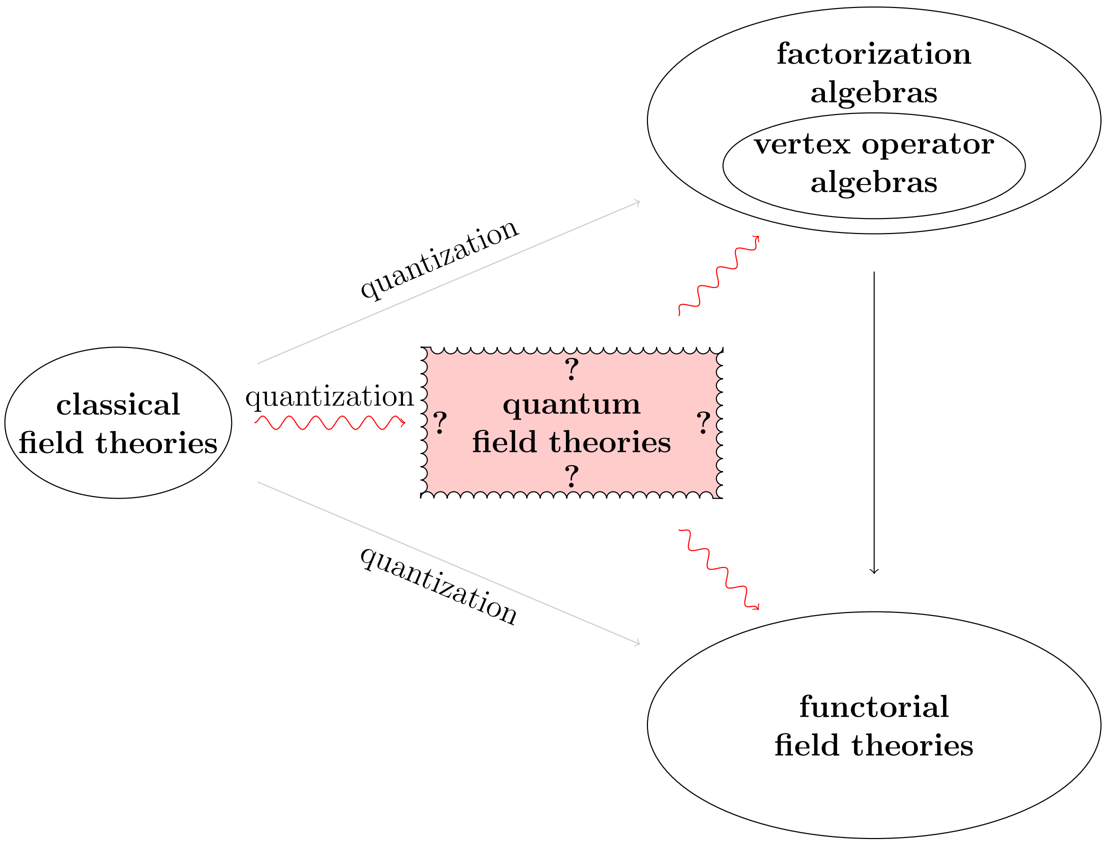
Functorial Field Theories and Factorization Algebras

September 3, 2014

Contents

1 Introduction

Here is a cartoon map of the area of mathematics this course will be exploring.



As the picture shows, the elephant in the middle is *quantum field theory*, a theory that so far has evaded a mathematical rigorous definition/construction, a fact that is indicated pictorially by squiggly lines. There are many competing mathematical approaches to quantum field theories, among them

- Functorial field theories, pioneered by Segal [Se2], Atiyah [At] and Kontsevich in the 1980's;
- Vertex operator algebras, first introduced by Borcherds in 1986 [Bo];
- Factorization algebras, defined by Beilinson and Drinfeld in 2004 [BD], is a more geometric and much broader notion than vertex operator algebras.

Factorization algebras and functorial field theories should be thought of as mathematical approximations to what a physicist might mean by a "quantum field theory"; in particular, such a field theory should give rise to a factorization algebra and a functorial field theory as indicated by the unlabeled squiggly arrows. Moreover, there are mathematically well-defined objects called *classical field theories* which are expected to lead to quantum field theories by a process called *quantization*, a process that is even less understood than the formal definition of quantum field theories. One important benchmark for the quality of a mathematical theory that hopes to be a contender for a mathematical definition of quantum field theory is the success it has in constructing objects in that theory that should be considered quantizations of classical field theories. This has only been partially done for factorization algebras and functorial field theories which is indicated by the gray arrows originating from classical field theories.

In this course we start out discussing various classical field theories. Then we will motivate the definition of functorial field theories by the heuristic discussion of a particular classical field theory known as the *non-linear σ-model*. Next will be the discussion of functorial field theories that will be based on the survey paper [ST]. Then we will be talking about factorization algebras, drawing from [Gw] and [CG]. The final topic will be a construction producing a functorial field theory from a factorization algebra. This is based on ongoing joint work with Bill Dwyer and Peter Teichner.

2 Classical field theories

Definition 2.1. A *classical field theory* is given by the following data:

- a manifold $\mathcal{M}$, typically infinite dimensional, which is called the *space of fields*;
- a smooth map $\mathcal{S}\colon \mathcal{M} \to \mathbb{R}$ called the *action*.

This is a first, very preliminary definition of what a classical field theory is. As we will see in the examples we will discuss, the space space of fields $\mathcal{M}$ is typically a space of smooth maps on a manifold M or, more generally, the space of smooth sections of a bundle over M. Moreover, for $\phi \in \mathcal{M}$, the action $\mathcal{S}(\phi)$ is given as an integral over M. This implies that the condition for ϕ to be a critical point of $\mathcal{S}$ amounts to a differential equation for ϕ, which is called the *Euler-Lagrange equation* for the action functional $\mathcal{S}$.

In classical physics, we often describe physical quantities in terms of functions which satisfy differential equations, for example

- the path of a classical point particle satisfies Newton's equation $F = ma$, where F is the net force acting on the particle, m is the mass of the particle, and a is its acceleration.
- the electromagnetic field satisfies Maxwell's equation.
- the gravitational field satisfies Einstein's equation.

To go from these differential equations to an action function $\mathcal{S}\colon \mathcal{M} \to \mathbb{R}$, we need to do a bit of reverse engineering: in each of these cases, we need to construct a suitable action function such that the associated Euler-Lagrange equation characterizing the critical points of $\mathcal{S}$ is the differential equation we start with.

2.1 Classical mechanics

For simplicity we restrict to a particular simple classical mechanical system, namely a point particle moving in $\mathbb{R}^3$, or more generally in a Riemannian manifold X, under the influence of a force field F. Now, physicists don't typically call classical mechanics a "field theory", but from a mathematical perspective it seems ok to do that, the same way that a scalar can be thought of as a 1×1-matrix.

Mathematically, the movement of a point particle in the manifold X is described by a smooth map $\phi\colon [a, b] \to X$ with $\phi(t) \in X$ giving the position of the particle at time $t \in [a, b]$. According to Newton's law it satisfies the differential equation $F = ma$, where F is the net force acting on the particle, m is the mass of the particle, and a is its acceleration. Written more carefully, Newton's law is the second order differential equation

$$F(\phi(t)) = m\ddot{\phi}(t). \tag{2.2}$$

Here both sides are vectors in the tangent space $T_{\phi(t)}X$. It is clear how to define the acceleration vector $\ddot{\phi}(t) \in T_{\phi(t)}X$ if ϕ is a path in $\mathbb{R}^n$. The definition of the acceleration vector $\ddot{\phi}(t)$ for a path in a general Riemannian manifold X requires a discussion of connections, which we will do in the next subsection.

That's fine, but what is the space of fields $\mathcal{M}$ and the action functional $\mathcal{S}\colon \mathcal{M} \to \mathbb{R}$? The first question is easy to answer:

$$\mathcal{M} := C^\infty([a, b], X) := \{\phi\colon [a, b] \to X \mid \phi \text{ is smooth}\}.$$

The action functional $\mathcal{S}\colon \mathcal{M} \to \mathbb{R}$ emerges in a reformulation of Newton's law called the *Lagrangian formulation* of classical mechanics. This requires the assumption that the force vector field F can be written in the form $F = -\operatorname{grad} V$ for some function $V\colon X \to \mathbb{R}$ that physicists call the *potential*. Here $\operatorname{grad} V$ is the *gradient vector field* of V (see Definition 2.12 for the definition of the gradient vector field of a function on a general Riemannian manifold). The Riemannian metric and the potential allow us to define the function

$$L\colon TX \longrightarrow \mathbb{R} \qquad \text{by} \qquad L(x, v) := \frac{m}{2}||v||^2 - V(x)$$

that physicists call the *Lagrangian*. Here $x \in X$, $v \in T_xX$, and $||v||^2 = \langle v, v\rangle_x$ is the norm-squared of the tangent vector v as measured in the Riemannian metric on X (see Definition 2.10 for the definition of a (pseudo) Riemannian metric). Physicists call $\frac{m}{2}||v||^2$ the *kinetic* *energy* and $V(x)$ the *potential energy*. The Lagrangian in turn can be used to define an action functional $\mathcal{S}\colon \mathcal{M} \to \mathbb{R}$ on the space $\mathcal{M}$ of all smooth maps $\phi\colon [a,b] \to X$ by setting

$$\mathcal{S}(\phi) := \int_a^b L(\phi(t), \dot{\phi}(t))dt. \tag{2.3}$$

According to the main result of the Lagrangian formulation of classical mechanics, solutions of Newton's differential equation (2.2) correspond essentially to critical points of the action functional $\mathcal{S}$. However, we need *boundary conditions* to obtain the correct statement (without them, the functional $\mathcal{S}$ on $\mathcal{M}$ has *no critical points*): fix points $x, y \in X$ and let $\mathcal{M}_{x,y}$ be the subspace of $\mathcal{M}$ consisting of smooth maps $\phi\colon [a,b] \to X$ with $\phi(a) = x$, $\phi(b) = y$. Then the precise statement is the following.

Theorem 2.4. *A path $\phi_0 \in \mathcal{M}_{x,y}$ is a critical point of the action functional $\mathcal{S}\colon \mathcal{M}_{x,y} \to \mathbb{R}$ if and only if it is a solution of Newton's equation* (2.2)*.*

If $\mathcal{M}_{x,y}$ were a finite dimensional smooth manifold, we would know what it means for the function $\mathcal{S}$ to be smooth, and what it means that $\phi \in \mathcal{M}_{x,y}$ is a critical point of $\mathcal{S}$. Due to the infinite dimensional nature of $\mathcal{M}_{x,y}$ we first need to *define* what this means in order to be provide a proof of the above theorem rather than just a heuristic argument.

Let us first define when a map $\mathcal{M}_{x,y} \to \mathbb{R}$ is smooth. More generally, if M, X are finite dimensional manifolds, possibly with boundary, and $\mathcal{M}'$ is a subset of the set $\mathcal{M} := C^\infty(M, X)$ of smooth maps $M \to X$, we want to define what it means for maps $S \to \mathcal{M}'$ and $\mathcal{M}' \to N$ to be smooth, where S, N are finite-dimensional smooth manifolds.

Definition 2.5. Let S, M, X, N be finite dimensional manifolds. A map

$$\phi\colon S \longrightarrow \mathcal{M}' \subset \mathcal{M} := C^\infty(M, X)$$

is *smooth* if the corresponding map $\widehat{\phi}\colon S \times M \to X$ given by $\widehat{\phi}(s,x) = (\phi(s))(x)$ is smooth. A map $f\colon \mathcal{M}' \to N$ is *smooth* if for all smooth maps $\phi\colon S \to \mathcal{M}'$ the composition

$$S \xrightarrow{\phi} \mathcal{M}' \xrightarrow{f} N$$

is smooth.

Next we want to define what it means to say that $\phi_0 \in \mathcal{M}' \subset \mathcal{M} = C^\infty(M, X)$ is a critical point of a smooth function $f\colon \mathcal{M}' \to \mathbb{R}$. We note that a point $\phi_0 \in W$ of a finite dimensional smooth manifold W is a critical point of a smooth function $f\colon W \to \mathbb{R}$ if and only if for all smooth paths $\phi\colon \mathbb{R} \to W$ with $\phi(0) = \phi_0$ the point $0 \in \mathbb{R}$ is a critical point of the composition $\mathbb{R} \xrightarrow{\phi} W \xrightarrow{f} \mathbb{R}$. This motivates the following definition.

Definition 2.6. A point $\phi_0 \in \mathcal{M}' \subset \mathcal{M} = C^\infty(M, X)$ is a *critical point* of a smooth function $f\colon \mathcal{M}' \to \mathbb{R}$ if for all smooth maps $\phi\colon \mathbb{R} \to \mathcal{M}'$ with $\phi(0) = \phi_0$ the point $0 \in \mathbb{R}$ is a critical point of the composition $\mathbb{R} \xrightarrow{\phi} \mathcal{M}' \xrightarrow{f} \mathbb{R}$.

Remark 2.7. The above abstract definition of smooth functions on subsets $\mathcal{M}'$ of mapping spaces and their critical points works for much more general objects $\mathcal{M}'$, namely contravariant functors $\mathcal{M}'$ from the category of smooth manifolds to the category of sets (we note that a subset $\mathcal{M}'$ of the set $C^\infty(M, X)$ of smooth maps determines such a functor by sending a manifold S to the set of smooth maps from S to $\mathcal{M}'$). This general nonsense of talking about functors from some category to the category of sets is surprisingly useful and goes under the name *functor of point formalism.*

Proof. Let us first show that the action functional $\mathcal{S}\colon \mathcal{M}_{x,y} \to \mathbb{R}$ is in fact smooth. By Definition 2.5 we need to show that for every smooth map $\phi\colon S \to \mathcal{M}_{x,y}$ the composition

$$S \xrightarrow{\phi} \mathcal{M}_{x,y} \xrightarrow{\mathcal{S}} \mathbb{R}$$

is smooth. The smoothness of ϕ is equivalent to the smoothness of the corresponding map

$$\widehat{\phi}\colon S \times [a, b] \longrightarrow X \qquad (s, t) \mapsto (\phi(s))(t).$$

In terms of the map $\widehat{\phi}$ the composition $\mathcal{S} \circ \phi$ is given by

$$\mathcal{S}(\phi(s)) = \int_a^b L(\widehat{\phi}(s, t), \frac{\partial}{\partial t}\widehat{\phi}(s, t))dt \tag{2.8}$$

which in fact is a *smooth* function of $s \in S$.

Next we want to discuss when a point $\phi_0 \in \mathcal{M}_{x,y}$ is a critical point of the action functional $\mathcal{S}$. So in view of Definition 2.6 let $\phi\colon \mathbb{R} \to \mathcal{M}_{x,y}$ be a smooth path with $\phi(0) = \phi_0$. We note that the corresponding map $\widehat{\phi}\colon \mathbb{R} \times [a, b] \longrightarrow X$ is smooth, and has the properties

$$\widehat{\phi}(0, t) = \phi_0(t) \qquad \widehat{\phi}(s, a) = x \qquad \widehat{\phi}(s, b) = y.$$

To determine the condition on ϕ_0 guaranteeing that 0 is a critical point of the composition (2.8), we differentiate at $s = 0$ and obtain

$$\begin{aligned}
\frac{\partial}{\partial s}_{|s=0}(\mathcal{S}(\phi(s)) &= \frac{\partial}{\partial s}_{|s=0} \int_a^b L(\widehat{\phi}(s, t), \frac{\partial \widehat{\phi}}{\partial t}(s, t))dt \\
&= \frac{\partial}{\partial s}_{|s=0} \int_a^b \left(\frac{m}{2}\langle \frac{\partial \widehat{\phi}}{\partial t}(s, t), \frac{\partial \widehat{\phi}}{\partial t}(s, t)\rangle - V(\widehat{\phi}(s, t)) \right) \, dt
\end{aligned}$$

Moving the s-derivative inside the integral, and using the product (resp. chain rule) for the first (resp. second) term, this is equal to

$$\int_a^b \left(m \left\langle \frac{\partial \widehat{\phi}}{\partial t}(0,t), \frac{\partial^2 \widehat{\phi}}{\partial t \partial s}(0,t) \right\rangle - dV(\frac{\partial \widehat{\phi}}{\partial s}(0,t) \right) dt$$

We note that $\frac{\partial \widehat{\phi}}{\partial t}(0,t) = \dot{\phi}_0(t) \in T_{\phi_0(t)}X$ is the tangent vector of the path $\phi(t)$. The map $t \mapsto \frac{\partial \widehat{\phi}}{\partial s}(0,t)$ is a vector field along the path ϕ_0, which geometrically can be interpreted as the tangent vector of the path $\phi\colon \mathbb{R} \to \mathcal{M}_{x,y}$ at the point $\phi(0) = \phi_0 \in \mathcal{M}_{x,y}$. This is often called the *infinitesimal variation* of the path ϕ_0, and the notation $\delta\phi$ is common for it. Using this notation, and rewriting the second summand in terms of the gradient of V we obtain

$$\int_a^b \left(m\langle \dot{\phi}_0(t), \frac{\partial}{\partial t}(\delta\phi)(t)\rangle - \langle \operatorname{grad} V(\phi_0(t)), (\delta\phi)(t)\rangle \right) dt$$

The key step in this calculation is the next one, namely integration by parts in the first summand. We note that there are no boundary terms since $(\delta\phi)(t) = \frac{\partial \widehat{\phi}}{\partial s}(0,t)$ vanishes for $t = a, b$ due to $\widehat{\phi}(s,a) = x$ and $\widehat{\phi}(s,b) = y$ for all $s \in \mathbb{R}$. Using $F = -\operatorname{grad} V$ in the second summand, and combining both summands and putting all these steps together, we obtain

$$\frac{\partial}{\partial s}_{|s=0}(\mathcal{S}(\phi(s)) = \int_a^b \left\langle -m\ddot{\phi}_0(t) + F(\phi_0(t)), (\delta\phi)(t) \right\rangle dt. \tag{2.9}$$

This shows that if $\phi_0 \in \mathcal{M}_{x,y}$ satisfies Newton's equation (2.2), then ϕ_0 is a critical point of $\mathcal{S}$.

To prove the converse statement let us assume that ϕ_0 is a critical point of $\mathcal{S}$. It is not hard to show that *every* section $\delta\phi \in \Gamma([a,b], \phi_0^*TX)$ with $(\delta\phi)(a) = (\delta\phi)(b) = 0$ is obtained as $\frac{\partial \widehat{\phi}}{\partial s}(0,t)$ for some smooth map $\widehat{\phi}\colon \mathbb{R} \times [a,b] \to X$ with $\widehat{\phi}(0,t) = \phi_0(t)$ and $\phi(s,a) = x$, $\phi(s,b) = y$. It follows that for the corresponding path $\phi\colon \mathbb{R} \to \mathcal{M}_{x,y}$ with $\phi(0) = \phi_0$ the formula (2.9) holds. Hence our assumption that ϕ_0 is a critical point of $\mathcal{S}$ implies that the right hand side of (2.9) vanishes for *every* $\delta\phi \in \Gamma([a,b]; \phi_0^*TX)$ with $\delta\phi(a) = \delta\phi(b) = 0$.

We claim that this implies that $\alpha := -m\ddot{\phi} + F(\phi_0) \in \Gamma([a,b], \phi^*TX)$ vanishes. To see this, assume $\alpha(t_0) \neq 0$ for some $t_0 \in (a,b)$. Then choose $\delta\phi$ to be the product of α with a bump function that is $\equiv 1$ in a neighborhood of t_0 and which vanishes at a and b. Then the integrand above is non-negative and positive for $t = t_0$, leading to the desired contradiction to our assumption that the integral is trivial. □

2.1.1 Digression: Riemannian manifolds and connections

The goal of this section is twofold:

- extend the statement of Theorem 2.4 to paths $\phi\colon [a,b] \to X$ in a general (pseudo) Riemannian manifold by making sense of
  - the gradient vector field $\operatorname{grad} V$ of a smooth function $V\colon X \to \mathbb{R}$;
  - the acceleration vector $\ddot{\phi}(t) \in T_{\phi(t)}X$.
- explain the basic notion of *connection* on a vector bundle that will be used throughout these lectures.

Definition 2.10. A *pseudo Riemannian metric* on a smooth manifold X is a family

$$\langle -,- \rangle_x \colon T_xX \times T_xX \longrightarrow \mathbb{R}$$

of non-degenerate symmetric bilinear forms on the tangent spaces T_xX of X depending smoothly on $x \in X$. Here *non-degenerate* means that the map

$$T_xX \longrightarrow T_x^*X := \operatorname{Hom}(T_xX, \mathbb{R}) \qquad v \mapsto (w \mapsto \langle v, w \rangle_x) \tag{2.11}$$

from the tangent space T_xX to the cotangent space T_x^*X is injective. This of course implies that this map is an isomorphism since the finite dimensional vector space T_xM and its dual space T_x^*M have the same dimension. Later, we'll be talking about scalar The informal statement that $\langle -,- \rangle_x$ *depends smoothly on* x means that the section of the vector bundle $(TX \otimes TX)^\vee$ (the dual of $TX \otimes TX$) given by

$$X \ni x \;\mapsto\; (v \otimes w \mapsto \langle v, w \rangle) \in (TX \otimes TX)^\vee_x = \operatorname{Hom}(T_xX \otimes T_xX, \mathbb{R})$$

is smooth.

Since "non-degenerate symmetric bilinear form" is quite a mouthful, we'll refer to $\langle -,- \rangle_x$ as *scalar product*. We should emphasize that in general we *do not* require that $\langle -,- \rangle_x$ is positive definite. If it is, then $\langle -,- \rangle$ is a *Riemannian metric* on X. In general, we can find a basis $\{v_1, \dots, v_n\}$ of T_xM such that the matrix $\langle v_i, v_j \rangle_x$ is diagonal. Let p be the number of positive diagonal entries, and let q be the number of negative diagonal entries of this matrix (which turn out to be independent of the choice of basis). The pair (p,q) is called the *signature* of $\langle -,- \rangle_x$. We note that a pseudo metric is a metric if and only if $q = 0$.

Definition 2.12. Let $V \in C^\infty(X)$ be a smooth function on a smooth manifold X, and let $dV \in \Omega^1(X) = \Gamma(X, T^*X)$ be the differential of V. If $\langle -,- \rangle$ is a (pseudo) Riemannian metric on X, the *gradient vector field* $\operatorname{grad} V \in \Gamma(X, TX)$ is the section corresponding to dV via the isomorphism

$$\Gamma(X, TX) \cong \Gamma(X, T^*X)$$

induced by the vector bundle isomorphism $TX \xrightarrow{\cong} T^*X$ given on the fibers over x by the isomorphism (2.11).

Our next goal is to define the acceleration $\ddot{\phi}(t) \in T_{\phi(t)}X$ of a smooth path $\phi\colon [a,b] \to X$. This will be done in two steps:

Construction of $\dot{\phi}$. First we want to discuss $\dot{\phi}$, and we start by asking: What kind of object is $\dot{\phi}$? We know what we mean by $\dot{\phi}(t)$ for $t \in [a,b]$, namely the tangent vector of the curve ϕ at the point $\phi(t) \in X$. So whatever $\dot{\phi}$ is, we can evaluate $\dot{\phi}$ at any $t \in [a,b]$ to obtain the vector $\dot{\phi}(t) \in T_{\phi(t)}X$. Restating this, $\dot{\phi}$ is a section of the vector bundle over $[a,b]$ whose fiber over $t \in [a,b]$ is the vector space $T_{\phi(t)}X$. In other words, $\dot{\phi}$ is a section of the pull-back vector bundle $\phi^*TX \to [a,b]$.

Construction of $\ddot{\phi}$. We note that a section of the trivial vector $M \times V \to M$ with fiber V over a smooth manifold M is the same thing as a smooth function $M \to V$ to the vector space V. Hence a section $s \in \Gamma(M,E)$ of a vector bundle $E \to M$ can be thought of as a generalization of a vector-valued function on M. So we should try to interpret $\ddot{\phi} = \frac{\partial}{\partial t}\dot{\phi}$ as some kind of derivative of the section $\dot{\phi} \in \Gamma([a,b], \phi^*TX)$ in the direction of the standard vector field $\frac{\partial}{\partial t}$ on $[a,b]$. This leads us to introduce *covariant derivatives* in Definition 2.15 below, which is exactly the kind of gadget we are looking for: a covariant derivative ∇ on a vector bundle $E \to M$ allows us to differentiate any section $s \in \Gamma(M,E)$ in the direction of a vector field $X \in \Gamma(M,TM)$ to obtain a new section denoted $\nabla_X s \in \Gamma(M,E)$ which should be thought of as the "derivative of s in the direction of the vector field X".

Before giving the formal definition of a connection ∇ on a vector bundle $E \to M$, let us first motivate the definition by discussing derivatives of sections $\Gamma(M,E)$ of the trivial vector bundle $E = M \times V \to M$ which we identify with $C^\infty(M,V)$, the smooth V-valued functions on M. Given a vector field $X \in \Gamma(M,TM)$ and a smooth function $f \in C^\infty(M)$, we can form $Xf \in C^\infty(M)$, the derivative of f in the direction of the vector field X. More generally, if $s \in C^\infty(M,V) = \Gamma(M,E)$ is a V-valued function, we can form $Xs \in C^\infty(M)$, the derivative of s in the direction of the vector field X. For pedagogical reasons, we will use the notation $\nabla_X s$ instead of Xs in this paragraph. We can ask ourselves how the section $\nabla_X s \in \Gamma(M,E)$ depends on the section s and the vector field X. In particular, we can ask about compatibility of the construction $\nabla_X s$ with respect to multiplication with functions $f \in C^\infty(M)$, noting that this multiplication makes the space of sections of *any* vector bundle over M a *module* over the algebra $C^\infty(M)$. Here is the well-known answer:

$$\nabla_X s \text{ is } C^\infty(M)\text{-linear in the } X\text{-variable} \tag{2.13}$$

$$\nabla_X s \text{ is a } C^\infty(M)\text{-derivation in the } s\text{-variable} \tag{2.14}$$

The second property means that $\nabla_X s$ is $\mathbb{R}$-linear in the s-variable and that it satisfies the Leibniz rule

$$\nabla_X(fs) = (Xf)s + f\nabla_X s$$

for all $f \in C^\infty(M)$.

Definition 2.15. A *covariant derivative* or *connection* on a vector bundle $E \to M$ is a map

$$\nabla \colon \Gamma(M, TM) \times \Gamma(M, E) \longrightarrow \Gamma(M, E) \qquad (X, s) \mapsto \nabla_X s$$

with properties 2.13 and 2.14.

Example 2.16. Here are some examples of connections.

1. If $E = M \times V \to M$ is the trivial vector bundle, then we have the "tautological" connection ∇^{taut} given by $\nabla^{\text{taut}}_X s := Xs$ as discussed in the paragraph before Definition 2.15.

2. Any vector bundle $E \to M$ has a connection.

   Homework 2.17. Prove this statement. Hint: Use local trivializations and the tautological connection on trivial bundles to construct connections on the restrictions of E to open subsets that cover M. Then show that these connections can be "glued" to a connection on E using partitions of unity.

3. There are *many* connections on any vector bundle E, as can seen as follows. Suppose ∇ is a connection on E and

   $$A \in \Omega^1(M, \text{End}(E)) = \Gamma(M, T^*M \otimes \text{End}(E))$$

   is an endomorphism-valued 1-form. Then $\nabla + A$ is again a connection, defined by

   $$(\nabla + A)_X s := \nabla_X s + A_X s \qquad \text{for } X \in \Gamma(M, TM),\ s \in \Gamma(M, E).$$

   Here $A_X \in \Gamma(M, \text{End}(E))$ is the bundle endomorphism obtained by evaluating the 1-form A on the vector field X.

   Homework 2.18. Show that $\nabla + A$ is in fact a connection. Moreover, show that any connection ∇' is of the form $\nabla + A$ for some $A \in \Omega^1(M, \text{End}(E))$.

   We note that these two statements imply that the space of connections is an *affine vector space* for the vector space $\Omega^1(M, \text{End}(E))$.

4. A (pseudo) Riemannian metric on a manifold M (see Definition 2.10) determines a unique connection ∇ on the tangent bundle TM, called the *Levi-Civita connection*. It is characterized by the following two properties:

   metric property $X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle$ for all $X, Y, Z \in \Gamma(TM)$; in other words, we can differentiate the product $\langle Y, Z\rangle \in C^\infty(M)$ using the "product rule".

torsion free property $\nabla_X Y - \nabla_Y X = [X,Y]$ for all $X, Y \in \Gamma(TM)$, where $[X,Y] \in \Gamma(TM)$ is the Lie bracket of the vector fields X, Y.

Homework 2.19. Show that if M is the Euclidean space $\mathbb{R}^n$ with its standard metric, then the Levi-Civita connection agrees with the tautological connection on $TM = M \times \mathbb{R}^n$.

5. If $f\colon N \to M$ is a smooth map and $E \to N$ is a vector bundle, then a connection ∇ on E induces a pullback connection on the pullback bundle f^*E.

Remark 2.20. Given the uniqueness of the Levi-Civita connection ∇ on the tangent bundle TM determined by a Riemannian metric $\langle -, - \rangle$, it might be tempting to think that a (pseudo) metric on *any* vector bundle $E \to M$ determines uniquely a connection ∇ on E. Here the notion of a (pseudo) metric on E is a slight generalization of that of a Riemannian metric as described in Definition 2.10 by simply replacing the tangent bundle TM by the vector bundle E.

We note that of the two properties characterizing the Levi-Civita connection, the "metric property" makes sense for $Y, Z \in \Gamma(E)$ and $X \in \Gamma(TM)$, while the "torsion-free property" does not make sense: if ∇ is a connection on E, then the term $\nabla_X Y$ requires X to be a vector field, and Y to be a section of E, while for the term $\nabla_Y X$ it is the other way around.

So we can insist that a connection on E is metric, but that condition does *not* determine the connection uniquely.

Finally we ready to define the acceleration $\ddot{\phi}$ of a path ϕ in a Riemannian manifold X, thus making sense of Newton's equation (2.2) in that general case.

Definition 2.21. Let $\phi\colon [a,b] \to X$ be a smooth path in a Riemannian manifold X, and let $\dot{\phi} \in \Gamma([a,b], \phi^*TX)$ be its tangent vector field. Then the *acceleration vector field* $\ddot{\phi} \in \Gamma([a,b], \phi^*TX)$ is defined by

$$\ddot{\phi} := \nabla_{\frac{\partial}{\partial t}} \dot{\phi} \in \Gamma([a,b], \phi^*TX).$$

Here the connection ∇ on ϕ^*TX is the pull-back of the Levi-Civita connection on TX, and $\frac{\partial}{\partial t}$ is the coordinate vector field on $[a,b]$.

If the Riemannian manifold X is the Euclidean space $\mathbb{R}^n$, then by homework problem 2.19 the Levi-Civita connection on TX is the tautological connection on $TX = X \times \mathbb{R}^n$. This implies in particular that for the pull-back connection ∇ on ϕ^*TX we have

$$\ddot{\phi} = \nabla_{\frac{\partial}{\partial t}} \dot{\phi} = \nabla^{\mathrm{taut}}_{\frac{\partial}{\partial t}} \dot{\phi} = \frac{\partial}{\partial t} \dot{\phi},$$

which is the usual definition for acceleration of a curve ϕ in Euclidean space.

Homework 2.22. Prove Theorem 2.4 in the general case where X is a (pseudo) Riemannian manifold.

2.2 Electromagnetism

The *electric field* E is a time dependent vector field on $\mathbb{R}^3$ (or some open subset U we are focussing on). The same is true for the *magnetic field* B. These satisfy a system of first order differential equations known as the *Maxwell equations*:

$$\operatorname{div} B = 0 \qquad \operatorname{curl} E = -\frac{\partial B}{\partial t} \tag{2.23}$$

$$\operatorname{div} E = 0 \qquad \operatorname{curl} B = \frac{\partial E}{\partial t} \tag{2.24}$$

Here $\operatorname{div} E$ (resp. $\operatorname{curl} E$) is the *divergence* (resp. *curl*) of the vector field $E = (E_1, E_2, E_3)$, which can be conveniently defined in terms of the vector $\nabla := (\frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3})$:

$$\begin{aligned} \operatorname{div} E = \nabla \cdot E &= \frac{\partial E_1}{\partial x_1} + \frac{\partial E_2}{\partial x_2} + \frac{\partial E_3}{\partial x_3} \\ \operatorname{curl} E = \nabla \times E &= \left(\frac{\partial E_3}{\partial x_2} - \frac{\partial E_2}{\partial x_3}, \frac{\partial E_1}{\partial x_3} - \frac{\partial E_3}{\partial x_1}, \frac{\partial E_2}{\partial x_1} - \frac{\partial E_1}{\partial x_2} \right) \end{aligned}$$

More precisely, these are the equations in a region not involving any electrical charges or currents, for example in the vacuum. More generally, in the presence of charges (resp. currents), the right hand side of the two equations (2.24) is modified by adding the charge (resp. current) to it. For that reason, we will refer to equations (2.23) as the *homogeneous Maxwell equations* and to equations (2.24) as the *inhomogeneous Maxwell equations*, despite the fact that we will deal only with the homogenous form of the Maxwell equations (2.24).

In his 1985 paper entitled *On some recent interactions between mathematics and physics* [Bo] that I highly recommend, Raoul Bott says this about this form of the Maxwell equations:

> This sort of no-nonsense description has many virtues but it completely obscures the geometro-topological aspects of these equations.

The key to a *geometric* understanding of Maxwell's equations is to express them in a coordinate independent way: this shouldn't come as a surprise since often there is a dichotomy between calculational power and geometric understanding. For example, for *calculations* with orthogonal maps $\mathbb{R}^3 \to \mathbb{R}^3$ with positive determinant, matrices are the tool of choice, while a better *geometric understanding* of the maps is provided by describing them as rotations about some axis.

A first step towards a coordinate free description of Maxwell's equations is to provide a coordinate free version of curl and div in terms of the de Rham complex on $\mathbb{R}^3$.

2.2.1 Digression: The de Rham complex

We recall that for any manifold M of dimension n we have the *de Rham complex*

$$\Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \dots \xrightarrow{d} \Omega^n(M). \tag{2.25}$$

Here $\Omega^k(M) := \Gamma(M, \Lambda^k T^*M)$ is the vector space of *k-forms on M*, the space of sections of the k-th exterior power of the cotangent bundle of M. In particular, $\Omega^0(M)$ is the same as the space $C^\infty(M)$ of smooth functions on M. The linear map

$$d\colon \Omega^k(M) \longrightarrow \Omega^{k+1}(M) \tag{2.26}$$

is the *de Rham differential*. It is uniquely characterized by the following properties:

- for $f \in \Omega^0(M) = C^\infty(M)$ the 1-form $df \in \Omega^1(M) = \Gamma(M, T^*M)$ is the differential of f, a section of the cotangent bundle.
- $d^2 = 0$
- d is a graded derivation with respect to the wedge product

  $$\Omega^k(M) \otimes \Omega^\ell(M) \xrightarrow{\wedge} \Omega^{k+\ell}(M), \tag{2.27}$$

  which means that for $\alpha \in \Omega^k(M)$, $\beta \in \Omega^\ell(M)$, we have

  $$d(\alpha \wedge \beta) = (d\alpha) \wedge \beta + (-1)^{k\ell} \alpha \wedge (d\beta). \tag{2.28}$$

The wedge product gives $\Omega^*(M) := \bigoplus_{k=0}^n \Omega^k(M)$ the structure of a $\mathbb{Z}$-graded algebra; it is *graded commutative* in the sense that for differential forms $\alpha \in \Omega^k(M)$, $\beta \in \Omega^\ell(M)$ we have

$$\alpha \wedge \beta = (-1)^{k\ell} \beta \wedge \alpha.$$

Let us illustrate how the three properties of the de Rham differential d allow us to determine d explicitly for $M = \mathbb{R}^n$. Let $x^1, \ldots, x^n \in C^\infty(\mathbb{R}^n)$ be the coordinate functions, also know as projection maps $x^k\colon \mathbb{R}^n \to \mathbb{R}$, given by $(x_1, \ldots, x_n) \mapsto x_k$. Any differential k-form on $\mathbb{R}^n$ can be written as a sum of terms of the form $\omega = f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ with $f \in C^\infty(\mathbb{R}^n)$ and $i_1 < i_2 < \cdots < i_k$. The differential of this k-form can be calculated using the derivation property (2.28) by thinking of ω as the wedge product of $f \in C^\infty(\mathbb{R}^n) = \Omega^0(\mathbb{R}^n)$ and the 1-forms $dx^{i_j} \in \Omega^1(M)$. This gives a sum of $k+1$ terms, each of which involves the differential d on one of the $k+1$ factors. Noting that $d(dx^{i_j}) = 0$ due to $d^2 = 0$, we obtain

$$d\left(f\, dx^{i_1} \wedge \cdots \wedge dx^{i_k}\right) = df \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx^i \wedge dx^{i_1} \wedge \cdots \wedge dx^{i_k}. \tag{2.29}$$

With these preparations, we can give a coordinate free interpretation of div, curl and grad in terms of the de Rham complex on $\mathbb{R}^3$, or more generally, some open subset $U \subset \mathbb{R}^3$.

Homework 2.30. Show that the exterior derivative for differential forms on $\mathbb{R}^3$ corresponds to the classical operations of *gradient* resp. *curl* resp. *divergence*. More precisely, show that there is a commutative diagram

$$\begin{tikzcd} C^\infty(U) \arrow[r, "\mathrm{grad}"] \arrow[d, equal] & \Gamma(U,TU) \arrow[r, "\mathrm{curl}"] \arrow[d, "\Phi_1"', "\cong"] & \Gamma(U,TU) \arrow[r, "\mathrm{div}"] \arrow[d, "\Phi_2"', "\cong"] & C^\infty(U) \arrow[d, "\Phi_3"', "\cong"] \\ \Omega^0(U) \arrow[r, "d"] & \Omega^1(U) \arrow[r, "d"] & \Omega^2(U) \arrow[r, "d"] & \Omega^3(U) \end{tikzcd} \tag{2.31}$$

Here we identify a vector field $E \in \Gamma(U,TU)$ with a triple (E_1, E_2, E_3) of smooth functions on U. The vertical isomorphisms are given by

$$\begin{aligned} \Phi_1(E_1,E_2,E_3) &= E_1 dx_1 + E_2 dx_2 + E_3 dx_3 \\ \Phi_2(E_1,E_2,E_3) &= E_1 dx_2 \wedge dx_3 + E_2 dx_3 \wedge dx_1 + E_3 dx_1 \wedge dx_2 \\ \Phi_3(f) &= f dx_1 \wedge dx_2 \wedge dx_3 \end{aligned}$$

From a topologist's point of view the condition $d^2 = 0$ means that the sequence of vector spaces and linear maps (2.25) is a *cochain complex*; the prefix "co" here simply means that the differential d has degree $+1$ in the sense that d goes from the vector space indexed by k to that indexed by $k+1$. This makes it possible to define the *de Rham cohomology* of the manifold M as

$$H^k_{dR}(M) := \frac{\ker\big(d\colon \Omega^k(M) \to \Omega^{k+1}(M)\big)}{\mathrm{im}\,(d\colon \Omega^{k-1}(M) \to \Omega^k(M))} = \frac{\{\text{closed } k\text{-forms}\}}{\{\text{exact } k\text{-forms}\}} \tag{2.32}$$

It turns out that there is an isomorphism

$$H^k_{dR}(M) \cong H^k(M;\mathbb{R}) \tag{2.33}$$

between the de Rham cohomology group $H^k_{dR}(M)$ and the singular cohomology group $H^k(M;\mathbb{R})$ with coefficients in $\mathbb{R}$. This statement is known as *de Rham's Theorem*. We want to stress that despite notational similarities, there is a big difference in flavor of the two sides related by the de Rham isomorphism (2.33). This singular cohomology group $H^k(M;\mathbb{R})$ can be defined for any topological space M, whereas the de Rham cohomology group $H^k_{dR}(M)$ uses the fact that M is a smooth manifold in an essential way: e.g., k-forms ω belonging to the kernel of d are solutions of the *differential equation* $d\omega = 0$ (the relation between the cohomology groups $H^k(M;\mathbb{R})$ and solutions of PDE's is even more strikingly expressed by the Hodge Theorem 2.42).

Long exact sequences provide a very effective tool to calculate cohomology groups. For example, the punctured 3-space $\mathbb{R}^3 \setminus \{0\}$ is homotopy equivalent to the 2-sphere S^2 and hence

$$H^2_{dR}(\mathbb{R}^3 \setminus \{0\}) \cong H^2(\mathbb{R}^3 \setminus \{0\};\mathbb{R}) \cong H^2(S^2;\mathbb{R}) \cong \mathbb{R}. \tag{2.34}$$

2.2.2 The homogeneous Maxwell equations

We recall that by homogeneous Maxwell equations we mean the two equations (2.23). We will think of the components E_i, B_i of the time dependent vector fields $E = (E_1, E_2, E_3)$, $B = (B_1, B_2, B_3)$ as smooth functions of $(x_0, x_1, x_2, x_3) \in \mathbb{R}^4 = \mathbb{R}\times\mathbb{R}^3$. Here $(x_1, x_2, x_3) \in \mathbb{R}^3$ is a point in "space" and $x_0 = t \in \mathbb{R}$ is a point in "time". Out of E and B we can manufacture the following 2-form $F \in \Omega^2(\mathbb{R}^4)$:

$$F := \big(E_1 dx^1 + E_2 dx^2 + E_3 dx^3\big) \wedge dx^0 + B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2. \quad (2.35)$$

We note that 2-forms on $\mathbb{R}^4$ are sections of the vector bundle $\Lambda^2 T^*\mathbb{R}^4$ of dimension $\binom{4}{2} = 6$. The above construction provides an isomorphism between the space of pairs (E, B) of time-dependent vector fields on $\mathbb{R}^3$ and 2-forms on $\mathbb{R}^4$ (it can be thought of as a particular trivialization of the 2-form bundle $\Lambda^2 T^*\mathbb{R}^4$).

Question. How can Maxwell's equations be expressed in terms of a differential equation for the 2-form F?

As discussed in the last section, there is a *canonical* first order differential equation for any differential form F, namely $dF = 0$. Let us determine what this equation corresponds to in terms of E and B. For this calculation we observe that equation (2.29) can be interpreted as the equation

$$d = \sum dx^i \frac{\partial}{\partial x^i},$$

of operators acting on $\Omega^*(\mathbb{R}^n)$, where $\frac{\partial}{\partial x_i}$ differentiates a form $f dx^{i_1} \wedge \cdots \wedge dx^{i_k}$ by differentiating the coefficient function, and dx^i acts on form via wedge product. In particular for $\mathbb{R}^4$ with coordinates (x_0, x_1, x_2, x_3) we can decompose the de Rham differential as

$$d = \sum_{i=0}^{3} dx^i \frac{\partial}{\partial x^i} = dx^0 \frac{\partial}{\partial x_0} + d^{\mathbb{R}^3} \qquad \text{with} \qquad d^{\mathbb{R}^3} = \sum_{i=1}^{3} dx^i \frac{\partial}{\partial x^i}.$$

For our calculation, we note that we can rewrite F as

$$F = \Phi_1(E) \wedge dx^0 + \Phi_2(B).$$

Hence

$$\begin{aligned} dF =& d(\Phi_1(E) \wedge dx^0 + \Phi_2(B)) = (dx^0 \frac{\partial}{\partial x_0} + d^{\mathbb{R}^3})(\Phi_1(E) \wedge dx^0 + \Phi_2(B)) \\ =& dx^0 \wedge \frac{\partial}{\partial x_0} \Phi_1(E) \wedge dx^0 + d^{\mathbb{R}^3} \Phi_1(E) \wedge dx^0 + dx^0 \wedge \frac{\partial}{\partial x_0} \Phi_2(B) + d^{\mathbb{R}^3} \Phi_2(B) \end{aligned}$$

We conclude:

Lemma 2.36. *The homogeneous Maxwell equations* (2.23) *are equivalent to* $dF = 0$.

How can we express the other two equations in terms of the electromagnetic 2-form F? As we've emphasized, the differential equation $dF = 0$ is the *only* equation we can write down if we are just given a 2-form F on some manifold M. So we conclude that we need more structure on M in order to be able to write down other differential equations. It will turn out that the second pair of Maxwell's equations is equivalent to the condition

$$d(\star F) = 0.$$

Here $\star\colon \Omega^k(M) \xrightarrow{\cong} \Omega^{n-k}(M)$ is the *Hodge-star isomorphism* for an n-dimensional manifold M. Its construction requires a (pseudo) metric on M.

2.2.3 Digression: Hodge theory

This digression is about Hodge theory, which is basically asking the question what additional structures on the de Rham complex of a manifold M are obtained from a (pseudo) Riemannian metric on M. Here is a list of these.

1. A (pseudo) Riemannian metric on a manifold M induces a metric on the vector bundle $\Lambda^k TM$. More generally, a metric on vector bundle $\pi\colon E \to M$ (see Remark 2.20) induces a metric on vector bundles that can be constructed out of E, like the k-fold tensor product $E^{\otimes k}$, its k-th symmetric power $S^k E$ and its k-th exterior power $\Lambda^k E$. Let us describe the metric on the exterior power bundle $\Lambda^k E$ which we are particularly interested in since the space of k-forms $\Omega^k(M)$ is the space $\Gamma(M, \Lambda^k T^*M)$ of sections of the k-th exterior power of the cotangent bundle.

   Let $x \in M$ and let $v_1, \dots, v_k$, $w_1, \dots, w_k$ be vectors in the fiber $E_x = \pi^{-1}(x)$. Then we define a scalar product (i.e., a non-degenerate symmetric bilinear form) $\langle -, - \rangle_x$ on $\Lambda^k E_x$ by declaring

   $$\langle v_1 \wedge \cdots \wedge v_k, w_1 \wedge \cdots \wedge w_k \rangle_x := \det \begin{pmatrix} \langle v_1, w_1 \rangle_x & \dots & \langle v_1, w_2 \rangle_x \\ \vdots & & \vdots \\ \langle v_2, w_1 \rangle_x & \dots & \langle v_k, w_k \rangle_x \end{pmatrix}.$$

   Here $\langle v_i, w_j \rangle_x$ is the given metric on E applied to the vectors $v_i, w_j \in E_x$ (it seems unnecessary to clutter the notation for the metric by superscripts indicating the bundle it lives on; the bundle is evident from the vectors we apply the metric to). It is not hard to show that this depends smoothly on $x \in M$ and hence defines a (pseudo) metric on the exterior power bundle $\Lambda^k E$. We note that if the pseudo metric on E we start with is positive definite, i.e., it is a *metric*, then so is $\langle -, - \rangle_x^{\Lambda^k E}$.

2. If M is a manifold of dimension n and $x \in M$, the vector space $\Lambda^n T_x^* M$ is 1-dimensional and hence there are precisely *two* vectors $\omega_x \in \Lambda^n T_x^* M$ with

$$\langle \omega_x, \omega_x \rangle = \pm 1. \tag{2.37}$$

We note that if ω_x is one of the two, then $-\omega_x$ is the other one, and both have the same norm-squared. (If you are given the signature of the pseudo metric, i.e., the number of positive and negative entries in a diagonal matrix representing the metric on $T_x M$, can you figure out whether the form $\langle -, - \rangle_x^{\Lambda^n T^* M}$ is positive or negative definite?)

The additional structure on M needed to have a *distinguished* vector ω_x with the property (2.37) is an *orientation* on M. We recall that an *orientation* on M can be defined as a choice

$$\sigma_x \in \pi_0 \left(\Lambda^n T_x^* M \setminus \{0\} \right) := \text{set of components of the space } \Lambda^n T_x^* M \setminus \{0\}$$

for every $x \in M$. The choice σ_x is required to "depend continuously" on x. This can be made precise by arguing that we can define a topology on the disjoint union

$$\widetilde{M} := \coprod_{x \in M} \pi_0(\Lambda^n T_x^* M \setminus \{0\})$$

such that the obvious projection map $\widetilde{M} \to M$ becomes a double covering called the *orientation covering* of M. An orientation is a section $\sigma \colon M \to \widetilde{M}$ of the orientation covering.

From the discussion above we see that pseudo metric on M plus an orientation σ determine uniquely an n-form $\mathrm{vol} := \omega \in \Omega^n(M) = \Gamma(M; \Lambda^n T^* M)$ called *volume form* which is uniquely characterized by the property (2.37) and the requirement $\omega_x \in \sigma_x$.

3. A (pseudo) metric plus an orientation on a compact manifold M induces a non-degenerate symmetric bilinear form $(-,-)$ on $\Omega^k(M)$ for all k defined by

$$(\alpha, \beta) := \int_M \langle \alpha, \beta \rangle \, \mathrm{vol} \qquad \text{for } \alpha, \beta \in \Omega^k(M).$$

Here $\langle \alpha, \beta \rangle$ is the smooth function on M whose value at $x \in M$ is given by $\langle \alpha_x, \beta_x \rangle_x$.

4. A (pseudo) metric and an orientation on a compact n-manifold M induces a vector bundle isomorphism

$$\star \colon \Lambda^k T^* M \xrightarrow{\cong} \Lambda^{n-k} T^* M,$$

called the *Hodge star operator*, which is characterized by

$$\alpha \wedge \star \beta = \langle \alpha, \beta \rangle \, \mathrm{vol}_x \qquad \text{for all } \alpha, \beta \in \Lambda^k T_x^* M.$$

Abusing language, the induced isomorphism on sections

$$\star\colon \Omega^k(M) \xrightarrow{\cong} \Omega^{n-k}(M)$$

is also called the Hodge star operator. It is characterized by $\alpha \wedge \star\beta = (\alpha, \beta)\,\mathrm{vol}$ for all $\alpha, \beta \in \Omega^k(M)$.

5. The de Rham differential $d\colon \Omega^k(M) \to \Omega^{k+1}(M)$ has an *adjoint*

$$d^*\colon \Omega^{k+1}(M) \to \Omega^k(M).$$

This means that

$$(d\alpha, \beta) = (\alpha, d^*\beta) \qquad \text{for all } \alpha \in \Omega^k(M),\ \beta \in \Omega^{k+1}(M). \tag{2.38}$$

Homework 2.39. Prove this statement by showing that the operator

$$d^* := (-1)^{k+1} \star^{-1} d\star\colon \Omega^{k+1}(M) \to \Omega^k(M)$$

satisfies equation (2.77). Hint: use Stoke's Theorem.

6. The following operators build from the de Rham operator d and its adjoint d^* are particularly interesting since they are *elliptic* differential operators on *Riemannian* manifolds (this does require the positivity condition on the metric on TM). Ellipticity is a relatively easy to check condition on a differential operator that guarantees that the kernel of an elliptic differential operator on a compact manifold is finite dimensional.

de Rham operator	$d + d^*\colon \Omega^*(M) \longrightarrow \Omega^*(M)$	(2.40)
Laplace-Beltrami operator	$\Delta = (d + d^*)^2 = dd^* + d^*d\colon \Omega^*(M) \longrightarrow \Omega^*(M)$	(2.41)

We note that unlike the de Rham operator, the Laplace operator preserves the degree of differential forms, and restricts to an operator

$$\Delta\colon \Omega^k(M) \longrightarrow \Omega^k(M)$$

for each k. Elements of $\mathcal{H}^k(M) := \ker\big(\Delta\colon \Omega^k(M) \longrightarrow \Omega^k(M)\big)$ are called *harmonic* k-forms (or harmonic functions for $k = 0$). You may be familiar with the fact that any harmonic function on a compact Riemannian manifold is locally constant; in particular, the space of harmonic functions is a finite dimensional vectorspace whose dimension is equal to the number of connected components of M. This illustrates a very special case of the general statement that the kernel of an elliptic differential operator on a compact manifold is finite dimensional.

We note that the discussion above can be recast as saying that for compact manifolds Riemannian M the map

$$\mathcal{H}^k(M) \longrightarrow H^k_{dR}(M) \quad \text{given by} \quad \alpha \mapsto [\alpha] \tag{2.42}$$

is a vector space isomorphism for $k=0$. According to the *Hodge Theorem* this is true for any k, yielding in particular the cool fact the dimension of the space of solutions of the partial differential equation $\Delta\alpha = 0$ for $\alpha \in \Omega^k(M)$ is a *topological invariant* of M, namely its k-th Betti-number, the dimension of $H^k(M;\mathbb{R})$.

For the map above to make sense we need to know that a harmonic form is closed.

Homework 2.43. Show that a k-form α is harmonic if and only if $d\alpha = 0$ and $d^*\alpha = 0$.

2.2.4 The inhomogeneous Maxwell equations

Our first goal is to equip $\mathbb{R}^4$ with a pseudo metric such that the inhomogeneous Maxwell equations (2.24) correspond to $d(*F) = 0$. It turns out that the right kind of metric is not Riemannian, but rather *Lorentzian*, that is a pseudo metric of signature $(3,1)$, see Definition 2.10; in other words, the scalar product on each tangent space with respect to a suitable basis is a diagonal matrix with one negative entry; all other entries are positive.

The standard Lorentz metric $\langle -, \rangle$ on $\mathbb{R}^4$ is given by stipulating that the tangent vectors $\frac{\partial}{\partial x_i}$ are mutually perpendicular, and that

$$\langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_i} \rangle = \begin{cases} -1 & i = 0 \\ +1 & i = 1,2,3 \end{cases}.$$

Equipped with the standard Lorentz metric, $\mathbb{R}^4$ is referred to as *Minkowski space* or *Minkowski space-time.*

Homework 2.44. Show that the Hodge star $\star\colon \Omega^2(\mathbb{R}^4) \to \Omega^2(\mathbb{R}^4)$ on Minkowski space is given by

$$\star(dx_i \wedge dx_j) = \begin{cases} -\operatorname{sign}(i,j,k,\ell) & 0 \in \{i,j\} \\ \operatorname{sign}(i,j,k,\ell) & 0 \ni \{i,j\} \end{cases}$$

Here k, ℓ are the “missing” indices, that is, $\{i,j,k,\ell\} = \{0,1,2,3\}$, and $\operatorname{sign}(i,j,k,\ell)$ is the sign of the permutation (i,j,k,ℓ). Also, we use the standard orientation of $\mathbb{R}^4$, given by the ordered basis $\{\frac{\partial}{\partial x^0}, \frac{\partial}{\partial x^1}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\}$ of the tangent space of $\mathbb{R}^4$. In particular, the volume form $\mathrm{vol} \in \Omega^4(\mathbb{R}^4)$ determined by the Lorentz metric and this orientation is given by

$$\mathrm{vol} = dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3.$$

Applying this to

$$
\begin{aligned}
F = & \left(E_1 dx^1 + E_2 dx^2 + E_3 dx^3\right) \wedge dx^0 \\
& + B_1 dx^2 \wedge dx^3 + B_2 dx^3 \wedge dx^1 + B_3 dx^1 \wedge dx^2
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\star F = & E_1 dx^2 \wedge dx^3 + E_2 dx^3 \wedge dx^1 + E_3 dx^1 \wedge dx^2 \\
& - B_1 dx^1 \wedge dx^0 - B_2 dx^2 \wedge dx^0 - B_3 dx_3 \wedge dx_0 .
\end{aligned}
$$

In other words, applying the Hodge star to F corresponds to the interchange

$$(E, B) \mapsto (-B, E).$$

Hence the equation $d(\star F) = 0$ is equivalent to the homogeneous Maxwell equations (2.23) with E, B exchanged in this way. These are precisely the inhomogeneous Maxwell equations (2.24), which proves the following statement.

Lemma 2.45. *The inhomogeneous Maxwell equations* (2.24) *are equivalent to* $d(\star F) = 0$.

2.2.5 The Maxwell equations as Euler-Lagrange equations

After succeeding to rewrite the Maxwell equations in the coordinate free form

$$dF = 0 \qquad d(\star F) = 0 \tag{2.46}$$

for the electromagnetic 2-form F, we are now ready to reformulate these equations as the Euler-Lagrange equations of a suitable action functional

$$\mathcal{S} \colon \mathcal{M} \longrightarrow \mathbb{R}.$$

Let $\mathcal{M} = \Omega^1(M)$ be the vector space of 1-forms on some compact space-time M equipped with a Lorentzian metric. For $A \in \Omega^1(M)$, let

$$\mathcal{S}(A) := \frac{1}{2}(dA, dA),$$

where $(-,-)$ is the scalar product determined on $\Omega^2(M)$ by the Lorentz metric (see item (3) in section 2.2.3).

Proposition 2.47. *A is a critical point of this functional if and only if $F = dA$ satisfies the Maxwell equations* (2.46).

Proof. Let $A_s \in \Omega^1(M)$ be a path of 1-forms parametrized by $s \in \mathbb{R}$. We calculate

$$\frac{\partial}{\partial s}\mathcal{S}(A_s) = \frac{1}{2}\frac{\partial}{\partial s}(dA_s, dA_s) = (dA_s, d\frac{\partial}{\partial s}A_s) = (d^*dA_s, \frac{\partial}{\partial s}A_s).$$

Letting as usual $\delta A := \frac{\partial}{\partial s}_{|s=0} A_s \in \Omega^1(M)$ be the tangent of the curve A_s at $s = 0$, we obtain:

$$\frac{\partial}{\partial s}_{|s=0}\mathcal{S}(A_s) = (d^*dA_0, \delta A).$$

In particular, A_0 is a critical point of $\mathcal{S}$ if and only if $(d^*dA_0, \delta A) = 0$ for all $\delta A \in \Omega^1(M)$. Since $(-,-)$ is non-degenerate, this is the case if and only if $d^*dA = d^*F = \pm \star^{-1} d \star F$ vanishes. This in turn is equivalent to the second Maxwell equation $d(\star F) = 0$. We note that the first Maxwell equation $dF = 0$ is automatic due to $d^2 = 0$. □

In physics parlance the 1-form A with $dA = F$ is called a *vector potential* for the electromagnetic field F. This terminology is motivated by the analogous relationship between the potential energy function V and its differential, which corresponds to the force field.

Two aspects of the above result are a little worrisome, namely:

- Not every electromagnetic field $F \in \Omega^2(M)$ satisfying the Maxwell equations (2.46) is of the form $F = dA$: the first Maxwell equation $dF = 0$ guarantees that F is a closed form, but not every closed form is exact; in fact, the quotient of closed forms modulo exact forms is the de Rham cohomology group $H^2_{dR}(M)$ (see equation (2.32)). This is trivial for $M = \mathbb{R}^4$, but non-trivial for example for $M = \mathbb{R} \times (\mathbb{R}^3 \setminus 0)$.

- Many vector potentials A lead to the same electromagnetic field F, since for $A + df \in \Omega^1(M)$ for any function $f \in C^\infty(M)$ we have

  $$d(A + df) = dA + d^2 f = dA = F.$$

  In particular, all critical points of $\mathcal{S}$ are highly degenerate: if A is a critical point of $\mathcal{S}$, then so is $A + df$ for every function f.

Question. Can the Maxwell equations (2.46) for the electromagnetic field $F \in \Omega^2(M)$ on a Lorentz manifold M be interpreted as the Euler-Lagrange equations of an action functional if the de Rham cohomology class $[F] \in H^2_{dR}(M)$ is non-zero?

The key for doing this is to interpret the symbols A and d in the equation

$$dA = F \in \Omega^2(M)$$

differently, namely to think of A as a connection on a Hermitian line bundle L, and to think of dA as its curvature. If L is the trivial line bundle, Hermitian connections can be identified with 1-forms, and the curvature of this connection is given by the differential of the 1-form.

2.2.6 Digression: curvature

Let ∇ be a connection on a vector bundle $E \to M$ of rank k. Given two vector fields $X, Y \in \Gamma(M; TM)$ we may ask whether the two linear maps

$$\nabla_X, \nabla_Y \colon \Gamma(M; E) \longrightarrow \Gamma(M; E)$$

commute. We note that if ∇ is the tautological connection on the trivial vector bundle $E = M \times V \to M$ (see item (1) in Example 2.16), then for $s \in \Gamma(M; E) = C^\infty(M, V)$ we have

$$(\nabla_X \nabla_Y - \nabla_Y \nabla_X)s = (XY - YX)s = [X, Y]x = \nabla_{[X,Y]}s,$$

where $[X, Y]$ is the commutator of X and Y, which is again a vector field on M (homework: check that this commutator is again a derivation of $C^\infty(M)$). For a general connection ∇, the section $R^\nabla(X, Y)s \in \Gamma(M; E)$ given by the formula

$$R^\nabla(X, Y)s := \nabla_X \nabla_Y s - \nabla_Y \nabla_X s - \nabla_{[X,Y]} s \tag{2.48}$$

is in general non-zero.

Homework 2.49. Show that the map

$$\begin{aligned} R^\nabla \colon \Gamma(M; TM) \times \Gamma(M; TM) \times \Gamma(M; E) &\longrightarrow \Gamma(M; E) \\ (X, Y, s) &\mapsto R^\nabla(X, Y)s \end{aligned}$$

is $C^\infty(M)$-linear in each variable.

This implies that the map R^∇ is induced by a vector bundle homomorphism

$$TM \otimes TM \otimes E \longrightarrow E.$$

Abusing notation, we'll write again R^∇ for this vector bundle homomorphism. This vector bundle map corresponds to vector bundle homomorphisms

$$TM \otimes TM \longrightarrow E^\vee \otimes E = \mathrm{End}(E) \qquad \text{or} \qquad \underline{\mathbb{R}} \to T^*M \otimes T^*M \otimes \mathrm{End}(E),$$

all of which we will denote by R^∇. We note that the last vector bundle map from the trivial line bundle $\underline{\mathbb{R}}$ to $T^*M \otimes T^*M \otimes \mathrm{End}(E)$ can be interpreted as a section of this bundle. Moreover, the skew-symmetry

$$R^\nabla(X, Y)s = -R^\nabla(Y, X)s$$

implies that R^∇ is in fact a section of the bundle $\Lambda^2 T^*M \otimes \mathrm{End}(E)$.

Definition 2.50. If ∇ is a connection on a vector bundle $E \to M$, its *curvature 2-form* is the section

$$R = R^\nabla \in \Gamma(M; \Lambda^2 T^*M \otimes \mathrm{End}(E)) = \Omega^2(M; \mathrm{End}(E))$$

determined by this construction.

More geometrically, given a point $p \in M$, tangent vectors $x, y \in T_pM$ and a vector $e \in E_p$, then the endomorphism

$$R(x, y)\colon E_p \longrightarrow E_p$$

sends e to $(R(X, Y)s)(p) \in E_p$, that is, the section $R(X, Y)s \in \Gamma(M; E)$ evaluated at the point $p \in M$. Here $X, Y \in \Gamma(M; TM)$ are vector fields with $X(p) = x$, $Y(p) = y$, and s is a section of E with $s(p) = e$. The $C^\infty(M)$-linearity of $R(X, Y)s$ in each variable guarantees that $(R(X, Y)s)(p) \in E_p$ is independent of how we extend the given tangent vectors x, y to vector fields X, Y and how the vector $e \in E_p$ is extended to a section s.

Now let us consider the case where L is a line bundle, that is, a vector bundle of rank 1. In this case the endomorphism bundle $\mathrm{End}(L)$ is isomorphic to the trivial line line bundle $\underline{\mathbb{R}} = M \times \mathbb{R}$. An explicit isomorphism is given by

$$\begin{aligned} \underline{\mathbb{R}} = M \times \mathbb{R} &\longrightarrow \mathrm{End}(L) \\ (x, r) &\mapsto (x, r\,\mathrm{id}_{L_x}), \end{aligned}$$

where $\mathrm{id}_{L_x} \in \mathrm{End}(L)_x = \mathrm{End}(L_x)$ is the identity endomorphism of the fiber L_x. In particular, the curvature 2-form

$$R \in \Omega^2(M; \mathrm{End}(L)) = \Omega^2(M; \underline{\mathbb{R}}) = \Omega^2(M)$$

is an ordinary 2-form.

Proposition 2.51. *Let $L \to M$ be a real line bundle with connection ∇.*

1. *The curvature 2-form $R = R^\nabla \in \Omega^2(M)$ is closed, that is, $dR = 0 \in \Omega^3(M)$.*

2. *Let $\nabla' = \nabla + A$ be the connection on L obtained from ∇ by adding a 1-form $A \in \Omega^1(M)$. Then $R^{\nabla'} = R^\nabla + dA \in \Omega^2(M)$.*

3. *The de Rham cohomology class $[R^\nabla] \in H^2_{dR}(M)$ is independent of the connection ∇.*

4. *If ∇ is a metric connection, then $R^\nabla = 0$.*

5. *A closed 2-form $F \in \Omega^2(M)$ is the curvature 2-form R^∇ of some connection ∇ on a real line bundle L if and only if F is exact.*

Homework 2.52. Prove parts (1) and (2) of this proposition. Hint: Show first part (2), and deduce (1) from (2) using the fact that for a trivial line bundle L every connection is of the form $\nabla = \nabla^{\mathrm{taut}} + A$ for some A, where ∇^{taut} is the tautological connection on the trivial bundle.

Proof. To prove part (3), let ∇, ∇' be two connections on L. Then by Homework 2.18 ∇' can be written as $\nabla' = \nabla + A$ for some $A \in \Omega^1(M)$. By part (2) $R^{\nabla'} = R^{\nabla} + dA$, and hence the closed forms $R^{\nabla'}$, R^{∇} differ by an exact form and hence represent the same de Rham cohomology class.

To prove (4), consider the *sphere bundle* $S(L) = \{(p,v) \in L \mid ||v|| = 1\} \subset L$. The projection map $\pi\colon L \to M$ restricts to a double covering $S(L) \to M$. In particular, for each point $p \in M$ we can find a section s of L with $||s(p)|| = 1$ defined in an open neighborhood U of p. Using this section to trivialize $L_{|U}$, we can write $\nabla = \nabla^{\mathrm{taut}} + A$ for some 1-form A defined on U. We note that by construction $\nabla^{\mathrm{taut}} s = 0$, since via the trivialization $\underline{\mathbb{R}} \cong L$ the constant function 1 corresponds to the section s. We also note that the derivative of the smooth function $||s||^2 = 1$ in the direction of all vector fields X is zero. Hence we have

$$0 = X||s||^2 = X\langle s,s\rangle = 2\langle \nabla_X s, s\rangle = 2\langle \nabla_X^{\mathrm{taut}} s + A(X)s, s\rangle = 2\langle A(X)s, s\rangle.$$

This implies that $A \in \Omega^1(U)$ vanishes, and so hence does $R^{\nabla}_{|U} = dA \in \Omega^2(U)$. Since M can be covered by open set U as above, it follows that $R^{\nabla} \in \Omega^2(M)$ vanishes.

Concerning part (5), if F is exact, that is of the form $F = dA$ for some $A \in \Omega^1(M)$, then the connection $\nabla := \nabla^{\mathrm{taut}} + A$ on the trivial line bundle $\underline{\mathbb{R}}$ has curvature $R^{\nabla} = R^{\nabla^{\mathrm{taut}}} + dA = dA$ by part (2).

To prove the converse, we first argue that we can construct a metric and a metric connection on any vector bundle E. This is certainly true for the trivial vector bundle $E = M \times V \to M$ by picking an inner product on V and taking the tautological connection ∇^{taut}. In particular, for any vector bundle E we can use local trivializations to construct metrics $\langle -,-\rangle^i$ and metric connections ∇^i on the restrictions $E_{|U_i}$ to open subsets U_i that form an open cover of M. Using a partition of unity $\varphi_i \in C^\infty(M)$ for this cover we can construct

the metric $\qquad \langle v,w\rangle_p := \sum_i \varphi_i(p)\langle v,w\rangle_p^i \qquad v,w \in E_p$

and the metric connection $\qquad \nabla_X s := \sum_i \varphi^i \nabla_X^i s \qquad X \in \Gamma(M;TM),\ s \in \Gamma(M;E).$

Now to prove that the curvature 2-form of any connection is exact, let ∇ be a metric connection on the line bundle L, and let ∇' be an arbitrary connection. As before we write ∇' in the form $\nabla' = \nabla + A$ for some $A \in \Omega^1(M)$. It follows that

$$R^{\nabla'} = R^{\nabla} + dA = dA.$$

Here the first equality is by part (2), and the second by part (4). □

Corollary 2.53. *The de Rham cohomology class $[R^\nabla] \in H^2_{dR}(M)$ represented by the curvature 2-form R^∇ of a connection ∇ is* independent *of the connection ∇.*

These results allow us to generalize our action functional

$$\mathcal{M} := \Omega^1(M) \xrightarrow{\mathcal{S}} \mathbb{R} \qquad A \mapsto \frac{1}{2}(dA, dA) \tag{2.54}$$

from the beginning of section 2.2.5 as follows. Let $L \to M$ be a fixed real line bundle on the Lorentz manifold M. Define the action functional

$$\mathcal{M}_L := \{\text{connections on } L\} \xrightarrow{\mathcal{S}} \mathbb{R} \qquad \text{by} \qquad \nabla \mapsto \frac{1}{2}(R^\nabla, R^\nabla) \tag{2.55}$$

We observe that if L is the trivial line bundle, then there is a bijection

$$\mathcal{M} \longleftrightarrow \mathcal{M}_L \qquad \text{given by} \qquad A \mapsto \nabla^{\text{taut}} + A,$$

where ∇^{taut} is the tautological connection on the trivial line bundle. The two action maps correspond to each other under this bijection by part (2) of Proposition 2.51. In other words, the new space of fields $\mathcal{M}_L$ and its action functional (2.55) is a generalization of our previous action functional (2.54).

We recall that the point of our current discussion is to address the issue that not every $F \in \Omega^2(M)$ satisfying the Maxwell equations is of the form $F = dA$ for some $A \in \Omega^1(M)$. So the hope is that there are *more* F's satisfying the Maxwell equations which are the curvature 2-forms of some connection on a real line bundle over M. Alas, this is not the case as part (5) of Proposition 2.51 shows.

The reader might wonder why we spend so much time discussing a generalization which then turns out to be useless for our purposes. The reason is that the above construction can be "tweaked" to obtain the Maxwell equations as as Euler-Lagrange equations for electromagnetic fields $F \in \Omega^2(M)$ which are not necessarily exact. The idea is to look at connections on *complex line bundles*. This requires the following digression.

2.2.7 Digression: Connections and their curvature on complex vector bundles

Let $\pi\colon E \to M$ be a complex vector bundle over a smooth manifold M. This means in particular that each fiber $E_p = \pi^{-1}(p)$ over a point $p \in M$ is a complex vector space, and that the space of section $\Gamma(M; E)$ is a complex vector space: given a section $s \in \Gamma(M; E)$ and $z \in \mathbb{C}$ the section $zs \in \Gamma(M; E)$ is determined by $(zs)(p) = z(s(p)) \in E_p$. More generally, we can multiply a section s by a complex valued function $f \in C^\infty(M, \mathbb{C})$ by setting $(fs)(p) := f(p)s(p)$. This makes the space of sections $C^\infty(M;E)$ a module over the algebra $C^\infty(M,\mathbb{C})$.

Just to avoid misunderstandings, there is nothing complex about the manifold M here: it's just a smooth manifold; in particular its tangent bundle TM is a real vector bundle which generally is not a complex vector bundle. We can obtain a complex vector bundle $TM_\mathbb{C} := TM \otimes_\mathbb{R} \mathbb{C}$, the *complexified tangent bundle*, whose fiber $T_pM_\mathbb{C}$ over $p \in M$ is the complexification $T_pM \otimes_\mathbb{R} \mathbb{C}$ of the tangent space T_pM. We will refer to sections of $TM_\mathbb{C}$ as *complex vector fields*. For example, for $M = \mathbb{R}^2 = \mathbb{C}$, the coordinate vector fields $\frac{\partial}{\partial x}$ and $\frac{\partial}{\partial y}$ are sections of TM, while

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \qquad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

are complex vector fields. Now we can define connections and their curvature on complex vector bundles by copying Definition 2.15 of connections on real vector bundles. The motto is to replace the algebra $C^\infty(M)$ of real-valued function by the bigger algebra $C^\infty(M,\mathbb{C})$ of complex-valued functions on M.

Definition 2.56. Let $E \to M$ be a complex vector bundle. A *connection* on E is a map

$$\nabla \colon \Gamma(M;TM_\mathbb{C}) \times \Gamma(M;E) \longrightarrow \Gamma(M;E) \qquad (X,s) \mapsto \nabla_X s$$

which is $C^\infty(M,\mathbb{C})$-linear in the X-variable, and a $C^\infty(M,\mathbb{C})$-derivation in the s-variable.

The *curvature* of a connection ∇ is given by the formula

$$R^\nabla(X,Y)s := \nabla_X\nabla_Y s - \nabla_Y\nabla_X s - \nabla_{[X,Y]}s \in \Gamma(M;E)$$

for $X, Y \in \Gamma(M;TM_\mathbb{C})$, $s \in \Gamma(M;E)$. The section $R(X,Y)s$ is $\mathbb{C}^\infty(M,C)$-linear in the X, Y, and s slot, and hence is induced by a map of complex vector bundles

$$R^\nabla \colon TM_\mathbb{C} \otimes TM_\mathbb{C} \otimes E \longrightarrow E.$$

As in the real case, this map of vector bundles can be interpreted as an endomorphism valued 2-form $R^\nabla \in \Omega^2(M;\mathrm{End}(E))$, in which case we refer to $R = R^\nabla$ as the *curvature 2-form* of the connection ∇.

Let $L \to M$ be a complex line bundle, that is, a complex vector bundle of rank one. Then the endomorphism bundle $\mathrm{End}(L)$ is also a line bundle, and the identity section of $\mathrm{End}(L)$ is a nowhere vanishing section that allows us to identify $\mathrm{End}(L)$ with the trivial complex line bundle $\underline{\mathbb{C}} = M \times \mathbb{C} \to M$. If ∇ is a connection on L, then its curvature

$$R^\nabla \in \Omega^2(M;\mathrm{End}(L)) = \Omega^2(M;\underline{\mathbb{C}}) =: \Omega^2(M;\mathbb{C})$$

is a complex valued 2-form.

We recall that our interest in curvature 2-forms stems from the desire to interpret the electromagnetic 2-form $F \in \Omega^2(M)$ as some kind of curvature 2-form. So need to come to grips with the fact that F is a *real* 2-form, while the curvature 2-form R^∇ of complex line bundle is *complex*. Examples show that R^∇ is typically not a real form (see Homework 2.59). In fact, it will turn out that R^∇ is *purely imaginary* if the connection ∇ is compatible with a Hermitian metric on L. Then dividing by $i = \sqrt{-1}$ produces the desired real 2-form.

Definition 2.57. A *Hermitian inner product* on a complex vector space V is a map

$$\langle -, - \rangle \colon V \times V \longrightarrow \mathbb{C}$$

with the following properties:

1. $\langle -, - \rangle$ is $\mathbb{C}$-linear in the first slot and $\mathbb{C}$-antilinear in the second slot.
2. $\langle v, w \rangle = \overline{\langle w, v \rangle}$, the complex conjugate of $\langle w, v \rangle$.
3. $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0$ only for $v = 0$.

A *Hermitian metric* on a complex vector bundle $E \to M$ is a family $\langle -, - \rangle_p$, $p \in M$, of Hermitian inner products on the fibers E_p of E which depend smoothly on p. A *Hermitian vector bundle* is a complex vector bundle equipped with a Hermitian metric. A *Hermitian connection* on a Hermitian vector bundle E is a connection which is compatible with the Hermitian metric $\langle -, - \rangle$ in the sense that

$$X\langle s, s' \rangle = \langle \nabla_X s, s' \rangle + \langle s, \nabla_X s' \rangle \qquad \text{for all } X \in \Gamma(M; TM_{\mathbb{C}}),\ s, s' \in \Gamma(M; E).$$

Example 2.58. Let $F \to M$ be a Hermitian vector bundle, and let $E \subset F$ be a sub bundle. Then a Hermitian connection ∇^F on F induces a Hermitian connection ∇^E on E by defining

$$\nabla^E_X s := \mathrm{proj}^E \nabla^F_X s \qquad \text{for } X \in \Gamma(M; TM_{\mathbb{C}}),\ s \in \Gamma(M; E).$$

Here $\mathrm{proj}^E \colon F \to E$ is the vector bundle map given by orthogonal projection in each fiber. In particular, the tautological connection on the trivial bundle $M \times V \to M$ for a Hermitian inner product space V induces a Hermitian connection on any sub bundle $E \subset M \times V$. A particular example is provided by the *tautological complex line bundle* γ over the complex projective space $\mathbb{CP}^n$. We recall that

$$\mathbb{CP}^n := \{\text{1-dimensional subspaces } V \subset \mathbb{C}^{n+1}\} = \left(\mathbb{C}^{n+1} \setminus \{0\}\right)/\mathbb{C}^\times,$$

where $z \in \mathbb{C}^\times = \mathbb{C} \setminus \{0\}$ act on $(z_0, \dots, z_n) \in \mathbb{C}^{n+1} \setminus \{0\}$ by multiplying each component by z. The line bundle γ

$$\gamma := \{(V, v) \mid V \in \mathbb{CP}^n,\ v \in V\} \subset \mathbb{CP}^n \times \mathbb{C}^{n+1}$$

is by construction a sub bundle of the trivial bundle with fiber $\mathbb{C}^{n+1}$. Hence the standard Hermitian inner product on $\mathbb{C}^{n+1}$ induces a Hermitian metric on γ as well as a Hermitian connection ∇.

Homework 2.59. Let $\iota\colon \mathbb{C} \to \mathbb{CP}^1$ be the embedding given by $z \mapsto [1, z]$ (the equivalence class of $(1, z) \in \mathbb{C}^2 \setminus \{0\}$ in $\mathbb{CP}^1$). We remark that this extends to a homemorphism from S^2, the one-point-compactification of $\mathbb{C}$, to $\mathbb{CP}^1$. For explicit calculations of the Hermitian connection ∇ on $\iota^*\gamma$ and its curvature, it is useful to take advantage of the nowhere vanishing section

$$s_0\colon \mathbb{C} \longrightarrow \iota^*L \subset \mathbb{C} \times \mathbb{C}^2 \qquad \text{defined by} \qquad s_0(z) := (z, (1, z)).$$

1. Show that
$$\nabla_{\frac{\partial}{\partial z}} s_0 = \frac{\bar{z}}{1 + z\bar{z}} s_0 \qquad\qquad \nabla_{\frac{\partial}{\partial \bar{z}}} s_0 = 0.$$

2. Using the nowhere vanishing section s_0 to trivialize $\iota^*\gamma$, show that $\nabla = \nabla^{\text{taut}} + A$, where
$$A = \frac{\bar{z}}{1 + z\bar{z}} dz$$

3. Show that $R^\nabla = \frac{1}{(1+z\bar{z})^2} d\bar{z} \wedge dz$.

4. Show that $\int_{\mathbb{C}} R^\nabla = 2\pi i$. Hint: use polar coordinates.

We remark that $d\bar{z} \wedge dz = 2i dx \wedge dy$, which implies by (3) that the curvature 2-form R^∇ is *purely imaginary*. This is in fact true for the connection 2-form for any Hermitian connection on a complex line bundle as we will discuss now. More generally, the curvature $\mathbb{R}^\nabla \in \Omega^2(M; \mathrm{End}(E))$ of a Hermitian connection ∇ is a section of a real sub bundle $\mathrm{Skew}(E)$ of the complex endomorphism bundle $\mathrm{End}(E)$ which is defined as follows.

Definition 2.60. Let V, W be complex vector spaces equipped with Hermitian inner products. The *adjoint* of a linear map $T\colon V \to W$ is the linear map

$$T^*\colon W \longrightarrow V \qquad \text{characterized by} \qquad \langle Tv, w\rangle = \langle v, T^*w\rangle \quad \text{for all } v \in V,\ w \in W.$$

An endomorphism $T \in \mathrm{End}(V)$ is *skew-adjoint* if $T^* = -T$. We will use the notation $\mathrm{Skew}(V) \subset \mathrm{End}(V)$ for the space of skew-adjoint endomorphisms. It should be emphasized that $\mathrm{Skew}(V)$ is only a *real* subspace of the complex vector space $\mathrm{End}(V)$, since $(iT)^* = -iT^*$, which follows directly from the defining property of the adjoint.

If E is a Hermitian vector bundle, let $\mathrm{Skew}(E) \subset \mathrm{End}(E)$ be the real sub bundle of the complex endomorphism vector bundle whose fiber over $p \in M$ consists of the skew-adjoint endomorphisms of E_p.

Homework 2.61. Let ∇ be a Hermitian connection on a Hermitian vector bundle $E \to M$.

1. Show that the curvature 2-form $R^\nabla \in \Omega^2(M; \mathrm{End}(E))$ takes values in the sub bundle $\mathrm{Skew}(E) \subset \mathrm{End}(E)$.

2. Show that $\nabla + A$ for $A \in \Omega^1(M; \mathrm{End}(E))$ is again a metric connection if and only if A takes values in the sub-bundle $\mathrm{Skew}(E) \subset \mathrm{End}(E)$.

While this homework problem shows that Hermitian connections on a Hermitian vector bundle E are closely related to the bundle $\mathrm{Skew}(E)$ of skew-symmetric endomorphisms, it doesn't really provide a conceptual reason for this relationship. The next few paragraphs are an attempt to furnish a conceptual link by giving a geometric way of thinking about a connection. First we will discuss the situation for real vector bundles, and then the completely analogous case of complex vector bundles.

It can be shown that a connection ∇ on a real vector bundle $E \to M$ determines for each smooth path $\gamma\colon [a,b] \to M$ a preferred isomorphism

$$||_\gamma\colon E_{\gamma(a)} \longrightarrow E_{\gamma(b)}$$

from the fiber $E_{\gamma(a)}$ over the starting point of γ to the fiber $E_{\gamma(b)}$ over the end point. This isomorphism is referred to as *parallel translation* along the path γ, and the notation is meant to remind the reader of "parallel". If the connection ∇ is metric, the associated parallel translation $||(\gamma)$ is an *isometry*, that is,

$$\langle ||_\gamma(v), ||_\gamma(w)\rangle = \langle v, w\rangle \qquad \text{for all } v, w \in E_{\gamma(a)}.$$

In particular, if γ is a loop, then $||_\gamma$ is an element of the *orthogonal group* $O(V)$ of the vector space $V = E_{\gamma(a)}$, which consists of all isometries from $V \to V$.

The parallel translation $||_\gamma$ is constructed from the connection ∇ by *integrating* (solving) an ODE. Conversely, we can use parallel translation along paths to define the derivative of a section s at a point p in the direction of a tangent vector $v \in T_pM$ as

$$\nabla_v s := \lim_{h\to 0} \frac{||_h^0 s(\gamma(h)) - s(\gamma(0))}{h}.$$

Here $\gamma\colon \mathbb{R} \to M$ is a path with $\gamma(0) = p$ and $\dot\gamma(0) = v$. With the path γ understood, the symbol $||_a^b\colon E_{\gamma(a)} \to E_{\gamma(b)}$ for $a < b$ is short-hand for the parallel translation along the path γ restricted to $[a,b]$, and $||_b^a\colon E_{\gamma(b)} \to E_{\gamma(b)}$ is the inverse of $||_a^b$.

We recall that if s is a vector valued function on M, the differential $ds(v)$ can be defined by

$$ds(v) := \lim_{h\to 0} \frac{s(\gamma(h)) - s(\gamma(0))}{h}.$$

We note that in the more general situation that s is a section of a vector bundle, this definition doesn't make sense: the difference $s(\gamma(h)) - s(\gamma(0))$ is meaningless since the vectors $s(\gamma(h))$ and $s(\gamma(0))$ typically belong to different vector spaces. However, if we have a parallel translation map $||_h^0$ we can use it to move $s(\gamma(h))$ back to the fiber E_p, and then we can subtract $s(\gamma(0))$.

This shows that passing from parallel translations along paths to the corresponding connection involves *differentiating*. Hopefully this discussion promotes the idea that if parallel translation involves the Lie group $O(V)$, then the corresponding covariant derivative involves its Lie algebra $\mathfrak{o}(V)$.

The situation for complex vector bundles is completely analogous: if parallel translations involves the *unitary group* $U(V)$ of isometries of a Hermitian vector space V, then the corresponding connection involves its Lie algebra $\mathfrak{u}(V)$.

Homework 2.62. Let V be a Hermitian inner product space. The unitary group $U(V)$ consisting of all isometries $T\colon V \to V$ is a Lie subgroup of $GL(V)$, the group of all invertible complex linear maps $T\colon V \to V$. In particular, its Lie algebra $\mathfrak{u}(V)$, the tangent space $T_eU(V)$ at the identity element $e \in U(V)$, is a subspace of $T_eGL(V) = \mathrm{End}(V)$. Show that $\mathfrak{u}(V)$ is equal to the space $\mathrm{Skew}(V)$ of skew-adjoint endomorphisms of V.

Now let us consider Hermitian line bundles, that is, complex vector bundles $L \to M$ of rank 1 equipped with a Hermitian metric. If ∇ is a Hermitian connection on L, then by part (1) of Homework 2.61 its curvature 2-form R^∇ is a 2-form with values in the vector bundle $\mathrm{Skew}(L)$ of skew-adjoint endomorphisms of L. This is a real sub bundle of the 1-dimensional complex vector bundle $\mathrm{End}(L)$ of endomorphisms of L. The identity section of $\mathrm{End}(L)$ determines an isomorphism $\underline{\mathbb{C}} \cong \mathrm{End}(L)$. This restricts to an isomorphism of real line bundles

$$\underline{i\mathbb{R}} \cong \mathrm{Skew}(L). \tag{2.63}$$

In particular, the curvature R^∇ of a Hermitian connection ∇ on a complex line bundle L can be interpreted as an element of $\Omega^2(M; i\mathbb{R}) \subset \Omega^2(M;\mathbb{C})$.

The next proposition is the analog of Proposition 2.51 for complex line bundles.

Proposition 2.64. *Let $L \to M$ be a complex line bundle with connection ∇.*

1. *The curvature 2-form $R = R^\nabla \in \Omega^2(M;\mathbb{C})$ is closed, that is, $dR = 0 \in \Omega^3(M;\mathbb{C})$.*

2. *Let $\nabla' = \nabla + A$ be the connection on L obtained from ∇ by adding a 1-form $A \in \Omega^1(M;\mathbb{C})$. Then $R^{\nabla'} = R^\nabla + dA \in \Omega^2(M;\mathbb{C})$.*

3. *The de Rham cohomology class $[R^\nabla] \in H^2_{dR}(M;\mathbb{C})$ is independent of the connection ∇.*

4. *If ∇ is a metric connection, then $R^\nabla \in \Omega^2(M; i\mathbb{R}) \subset \Omega^2(M;\mathbb{C})$.*

5. *A closed 2-form $\omega \in \Omega^2(M;\mathbb{C})$ is the curvature 2-form R^∇ of some connection ∇ on a complex line bundle L if and only if the de Rham cohomology class $[\frac{\omega}{2\pi i}] \in H^2_{dR}(M)$ is integral in the sense of the following definition.*

Definition 2.65. A de Rham cohomology class $\alpha \in H^k_{dR}(M;\mathbb{C})$ is *integral* if it belongs to the image of the composition

$$H^k(M;\mathbb{Z}) \longrightarrow H^k(M;\mathbb{C}) \cong H^k_{dR}(M;\mathbb{C}). \tag{2.66}$$

Here the first map is the homomorphism of singular cohomology groups induced by the standard inclusion $\mathbb{Z} \hookrightarrow \mathbb{C}$, and the second map is the de Rham isomorphism (2.33).

The proofs of parts (1), (2) and (3) of the proposition are completely analogous to the case for real line bundles in Proposition 2.51. Part (4) follows from part (1) of Exercise 2.61 and the isomorphism 2.63. The proof of part (5) is much more involved than the corresponding statement for real line bundles (part (5) of Proposition 2.51). It is based on the classification of complex line bundles up to isomorphisms. This classification as well as the proof of part (5) are the subject of the next section.

Finally we now return to our goal of constructing the classical field theory describing electromagnetism. For a hermitian line bundle L over a Lorentz manifold M we define the action functional

$$\mathcal{M}_L := \{\text{hermitian connections on } L\} \xrightarrow{\mathcal{S}_L} \mathbb{R} \tag{2.67}$$

by

$$\mathcal{S}_L(\nabla) := \frac{1}{2}(F,F) \quad \text{for } F = \frac{R^\nabla}{-2\pi i} \in \Omega^2(M). \tag{2.68}$$

The factor $2\pi i$ in the denominator is motivated by part (5) of Proposition 2.64. The additional minus sign is a convenient convention, since the integral closed 2-form $\frac{R^\nabla}{-2\pi i}$ has a name: it is called the first *Chern form* of the pair (L,∇) (see more on the first Chern form/class, including a discussion of minus sign, in the next subsection).

Mimicking the proof of Proposition 2.47 it is straightforward to show that a Hermitian connection ∇ on some Hermitian line bundle L is a critical point of $\mathcal{S}_L$ if and only if the integral closed 2-form $F = \frac{R^\nabla}{-2\pi i}$ satisfies the Maxwell equations.

We observe that the action functional $\mathcal{S}_L$ is a generalization of our old functional

$$\mathcal{S}\colon \Omega^1(M) \to \mathbb{R} \qquad A \mapsto \frac{1}{2}(dA, dA)$$

in the sense that if L is the trivial complex line bundle $\underline{\mathbb{C}}$, then the diagram

$$\begin{tikzcd}
\Omega^1(M) \arrow[d, "\cong"'] \arrow[dr, "\mathcal{S}"] & \\
\mathcal{M}_L \arrow[r, "\mathcal{S}_L"'] & \mathbb{R}
\end{tikzcd}$$

is commutative, where the vertical bijection sends a 1-form $A \in \Omega^1(M)$ to the Hermitian connection $\nabla^{\mathrm{taut}} + \frac{1}{-2\pi i}A$ on $L = \underline{\mathbb{C}}$. The commutativity of the diagram follows from part (2) of Proposition 2.64(2) for complex line bundles; the fact that the vertical map is a bijection is a consequence of homework problem 2.61(2).

2.2.8 Digression: Classification of vector bundles

The goal of this digression is discuss some classification results for vector bundles, in particular complex line bundles, and to use these results to prove part (5) of Proposition 2.64.

For $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ the topological space

$$G_k(\mathbb{F}^{n+k}) = \{k\text{-dimensional subspaces of } \mathbb{F}^{n+k}\}$$

is called the *Grassmannian.* We note that for $k = 1$ this is the projective space $\mathbb{FP}^n$ (the real projective space $\mathbb{RP}^n$ for $\mathbb{F} = \mathbb{R}$, the complex projective space $\mathbb{CP}^n$ for $\mathbb{F} = \mathbb{C}$). Generalizing the tautological line bundle over projective spaces, we have the *tautological vector bundle*

$$\gamma^k := \{(V, v) \mid V \in G_k(\mathbb{F}^{n+k}),\ v \in V\} \subset G_k(\mathbb{F}^{n+k}) \times \mathbb{F}^{n+k}$$

over the Grassmanian $G_k(\mathbb{F}^{n+k})$. This is a real (for $\mathbb{F} = \mathbb{R}$) or complex (for $\mathbb{F} = \mathbb{C}$) vector bundle of rank k.

If $f\colon X \to G_k(\mathbb{F}^{n+k})$ is a continuous map from a topological space X to the Grassmanian, the pull-back bundle $f^*\gamma^k$ is vector bundle of rank k over X. An important basic fact is that homotopic maps $f, g\colon X \to G_k(\mathbb{F}^{n+k})$ lead to isomorphic vector bundles $f^*\gamma^k$, $g^*\gamma^k$. In other words, we have a well-defined map

$$[X, G_k(\mathbb{F}^{n+k})] \longrightarrow \mathsf{Vect}^k_{\mathbb{F}}(X) \qquad \text{given by} \qquad [f] \mapsto [f^*\gamma^k]. \tag{2.69}$$

Here $[X, G_k(\mathbb{F}^{n+k})]$ is the set of homotopy classes of maps from X to the Grassmannian and given a map $f\colon X \to G_k(\mathbb{F}^{n+k})$, we write $[f] \in [X, G_k(\mathbb{F}^{n+k})]$ for its homotopy class. Moreover, $\mathsf{Vect}^k_{\mathbb{F}}(X)$ is the set of isomorphism classes of real (resp. complex) vector bundles of rank k over X.

We remark that if X is a smooth manifold, we should distinguish between vector bundles and *smooth vector bundles.* The latter involves more structure than the former, and there exist vector bundle isomorphisms between smooth vector bundles which aren't smooth. Fortunately, the forgetful map that sends a smooth vector bundle to its underlying vector bundle induces a bijection between isomorphisms classes of smooth vector bundles and the isomorphisms classes of vector bundles.

Theorem 2.70. *For $n = \infty$ the map* (2.69) *is a bijection.*

This important classification result shows that classifying vector bundles up to isomorphism is really a problem in homotopy theory. A proof can be found for example in section

5 of [MS]. The Grassmannians $G_k(\mathbb{R}^\infty)$ resp. $G_k(\mathbb{C}^\infty)$ are often called the *classifying spaces* for real (resp. complex) vector bundles of rank k, a terminology motivated by the theorem above. In general the problem of calculating the set $[X, G_k(\mathbb{F}^\infty)]$ can be pretty daunting, even in simple cases, like $X = S^n$.

Fortunately the case $k = 1$ we are interested in, is simple. First we observe that the space $G_1(\mathbb{F}^{n+1})$ of lines in $\mathbb{F}^{n+1}$ is well-known as the *projective space* $\mathbb{FP}^n$. In particular, the classifying space $G_1(\mathbb{R}^\infty)$ for real line bundles is the real projective space $\mathbb{RP}^\infty$, and the classifying space $G_1(\mathbb{C}^\infty)$ for complex line bundles is the complex projective space $\mathbb{CP}^\infty$. We also recall that the cohomology ring

$$H^*(\mathbb{CP}^\infty;\mathbb{Z}) := \bigoplus_{q=0}^{\infty} H^q(\mathbb{CP}^\infty;\mathbb{Z})$$

is the power series ring $\mathbb{Z}[x]$, where $x \in H^2(\mathbb{CP}^\infty;\mathbb{Z}) \cong \mathbb{Z}$ is a generator. It is customary to make a choice of this generator by requiring

$$\langle \iota^* x, [\mathbb{CP}^1]\rangle = 1.$$

Here $[\mathbb{CP}^1] \in H_2(\mathbb{CP}^1;\mathbb{Z})$ is the *fundamental homology class* of the 2-dimensional complex projective space $\mathbb{CP}^1$ (equipped with the orientation provided by its complex structure), and $\iota^* x \in H^2(\mathbb{CP}^1;\mathbb{Z})$ is the image of x under the map of cohomology groups

$$\iota^* \colon H_2(\mathbb{CP}^\infty;\mathbb{Z}) \longrightarrow H_2(\mathbb{CP}^1;\mathbb{Z})$$

induced by the inclusion map $\iota\colon \mathbb{CP}^1 \hookrightarrow \mathbb{CP}^\infty$. The pairing

$$\langle -, -\rangle \colon H^q(X;\mathbb{Z}) \times H_q(X;\mathbb{Z}) \longrightarrow \mathbb{Z}$$

is the *Kronecker product*, which evaluates a cohomology class on a homology class of the same dimension to produce an element of the coefficient ring (which is $\mathbb{Z}$ in the case at hand).

There is a geometric interpretation of the Kronecker product $\langle \alpha, [M]\rangle$ of a cohomology class $a \in H^n(M;\mathbb{Z})$ and the fundamental class $[M] \in H_n(M;\mathbb{Z})$ of an oriented closed n-manifold M:

$$\langle a, [M]\rangle = \int_M \alpha, \tag{2.71}$$

where $\alpha \in \Omega^n(M)$ is a closed n-form representing de Rham cohomology class of the image of $a \in H^n(M;\mathbb{Z})$ under the homomorphism (2.66).

Proposition 2.72. *The map*

$$[X, \mathbb{CP}^\infty] \longrightarrow H^2(X;\mathbb{Z}) \qquad \textit{given by} \qquad [f\colon X \to \mathbb{CP}^\infty] \mapsto -f^* x$$

is a bijection (we'll comment on the minus sign in Remark 2.74).

Combining this bijection with the bijection from Theorem 2.70 we obtain a bijection

$$\mathsf{Vect}^1_{\mathbb{C}}(X) \xleftarrow{\cong} [X, \mathbb{CP}^\infty] \xrightarrow{\cong} H^2(X;\mathbb{Z})$$

Definition 2.73. The *first Chern class* of a complex line bundle $L \to X$ is the cohomology class denoted $c_1(L) \in H^2(X;\mathbb{Z})$ which corresponds to L under the bijection above. In particular, for the tautological bundle $\gamma \to \mathbb{CP}^\infty$ we have $c_1(\gamma) = -x \in H^2(\mathbb{CP}^\infty;\mathbb{Z})$.

Remark 2.74. The reader might wonder about the minus sign in the definition of the first Chern class coming from the minus sign in the construction of the bijection of Proposition 2.72. This normalization is taylor-made such that for a compact complex curve Σ we have the equation

$$\langle c_1(T\Sigma), [\Sigma]\rangle = \chi(\Sigma) = 2 - 2g. \tag{2.75}$$

Here $\chi(\Sigma)$ is the Euler characteristic of Σ and g is its genus. In particular for $\Sigma = \mathbb{CP}^1$ we want $\langle c_1(T\mathbb{CP}^1), [\mathbb{CP}^1]\rangle = 2$. It can be shown that the complex line bundle $T\mathbb{CP}^1$ is isomorphic to $\bar\gamma^{\otimes 2}$, where $\bar\gamma \to \mathbb{CP}^1$ is the complex conjugate. This implies

$$c_1(T\mathbb{CP}^1) = c_1(\bar\gamma^{\otimes 2}) = 2c_1(\bar\gamma) = -2c_1(\gamma) = -2(-x) = 2x \in H^2(\mathbb{CP}^1;\mathbb{Z}).$$

Since $\langle x, [\mathbb{CP}^1]\rangle = 1$, this implies $\langle c_1(T\mathbb{CP}^1), [\mathbb{CP}^1]\rangle = 2$ as desired.

Proof of part (5) of Proposition 2.64. Still to come. □

2.3 Gravity (Einstein's theory of general relativity)

According to Newton, there is an attractive gravitational force between any two bodies which is proportional to the mass of each of the bodies and inverse proportional to the square of their distance. While that is a pretty good approximation of how planets move, what's wrong with it from a theoretical point of view is the "instantaneous" effect one body has on the gravitational force produced by this body acting on far away objects. This does not conform to the idea that information cannot travel faster than light, where "information" is what links a cause to its effect.

In Einstein's view, gravity affects the movement of bodies in a more indirect way. Namely matter causes a "gravitational field" which in turn influences the movement of other matter. A commonly used analogy is that of throwing stones into a pond which causes ripples to spread which in turn causes corks to bob up and down. So what is this "gravitational field" from a mathematical point of view? Whatever it is, it is supposed to influence the movement of objects, which are described mathematically as paths in space-time. Generally a "force" in physics is thought of as causing an "acceleration" of particles influenced by that force.

As we discussed in section 2.1.1, we need a (pseudo) metric on a manifold to talk about "acceleration" of paths in that manifold. So Einstein postulated that the gravitational field is a *Lorentz metric* on space-time (which is the datum we needed to write down the Maxwell equations for the electromagnetic field in section 2.2.4). He further postulated that objects move along geodesics in space time; so the gravitational field influences the movement of matter, since the geodesics in the space-time manifold depend on the Lorentz metric.

The question remains how matter influences the gravitational field, the metric on space time. A popular 2-dimensional analogy is to think of space time as made of thin rubber, like a balloon. Then masses, for example our sun, sit like stones on that surface, causing it to bend, influencing in this way the geometry (= metric) of space-time. The precise mechanism by which matter influences the gravitational field according to Einstein is given by the equation

$$\mathrm{Ric}_g - \frac{1}{2} s_g g + \lambda g = T \tag{2.76}$$

which is the *Einstein field equation*, see p. 97 in [Be]. Here

- g is a Lorentz metric on the 4-dimensional space-time manifold M. In physics lingo g is the gravitational field.
- T is the *stress-energy tensor* which encodes the mass distribution in space time. Mathematically, it is a section of the symmetric power S^2T^*M of the cotangent bundle. In particular, at any point $p \in M$ it gives a symmetric bilinear form

  $$T\colon T_pM \times T_pM \longrightarrow \mathbb{R}$$

  on the tangent space T_pM. For an observer traveling on a path $\gamma\colon \mathbb{R} \to M$ through space time with $\gamma(0) = p$ and $\dot\gamma(0) = v \in T_pM$ the quantity $T(v, v)$ is the energy density she observes at p. In particular, T vanishes in vacuum regions of space-time.
- Ric_g is the *Ricci curvature* of the Levi-Civita connection ∇ determined by g. It is a section of S^2T^*M obtained from the curvature tensor $R = R^\nabla$ by

  $$\mathrm{Ric}_g(x, y) := \mathrm{trace}(z \mapsto R(z, x)y) \qquad \text{for } x, y, z \in T_pM.$$
- $s_g \in C^\infty(M)$ is the *scalar curvature* given by $s_g(p) := tr_g((\mathrm{Ric}_g)_p)$. Here $\mathrm{tr}_g(h)$ for any symmetric bilinear form $h\colon T_pM \times T_pM \to \mathbb{R}$ is defined by

  $$\mathrm{tr}_g(h) := \mathrm{trace}(\tilde h\colon T_pM \to T_pM),$$

  where $\widetilde{h}$ is the unique linear map which is self-adjoint w.r.t. $\langle -, - \rangle_g$ and has the property

  $$h(v, w) = \langle \widetilde{h}(v), w \rangle_g \qquad \text{for all } v, w \in T_pM. \tag{2.77}$$

- $\lambda \in \mathbb{R}$ is called the *cosmological constant*.

We note that S^2T^*M is a vector bundle of rank 10, and hence using a local trivialization of S^2T^*M we can think of the stress energy tensor T as well as the gravitational field g as ten smooth functions. The ten components of the Ricci tensor Ric_g and the scalar curvature function s_g are obtained from the metric g by a differential operator of order 2. In other words, the Einstein field equation(2.76) can be regarded as a second order system of partial differential equations for the gravitational field g.

In vacuum the stress-energy tensor T vanishes, and so solving the Einstein equation (2.76) in the vacuum can be recast in as the question of finding a Lorentz metric g on M such that

$$\mathrm{Ric}_g = cg \tag{2.78}$$

for some constant c. From a mathematical point of view this equation makes sense for manifolds of any dimension and for pseudo metrics of any signature. Pseudo metrics satisfying this equation are called *Einstein metrics*.

2.3.1 Digression: Geometric interpretation of scalar and Ricci curvature

Scalar curvature and Ricci curvature have a geometric interpretation as a "volume distortions" which we want to explain in this section. Let M be a manifold of dimension n equipped with a Riemannian metric g. The value $s(p) \in \mathbb{R}$ of the scalar curvature at a point $p \in M$ measures how the volume of small balls around p compare with the volume of balls of the same radius in $\mathbb{R}^n$. More precisely:

- For $r > 0$ let $\mathrm{vol}(B_r(p, M))$ be the volume of $B_r(p, M)$, the ball of radius r around the point p in M, which consists of all points x which can be connected by a geodesic of length $\leq r$ with the point p. The volume of a subset $U \subset M$ is given by the integral $\int_U \mathrm{vol}$, where $\mathrm{vol} \in \Omega^n(M)$ is the volume form vol determined by the Riemannian metric.

- Let $\mathrm{vol}(B_r(0, \mathbb{R}^n))$ be the volume of the ball of radius r around the origin in $\mathbb{R}^n$.

The scalar curvature compares these volumes for small r. More precisely, $s(p)$ is up to a factor the coefficient of r^2 in the Taylor expansion of the volume ratio (see section 0.60 on p. 15 in [Be]):

$$\frac{\mathrm{vol}(B_r(p, M))}{\mathrm{vol}(B_r(0, \mathbb{R}^n))} = 1 - \frac{s(p)}{6(n+2)} + \dots \tag{2.79}$$

In particular, $s(p) > 0$ is equivalent to the statement that for sufficiently small r the volume of the ball $B_r(p, M)$ is *smaller* than the volume of $B_r(0, \mathbb{R}^n)$. For example, for the standard Riemannian metric on S^2 the area of a small ball around the north pole is smaller that the area of Euclidean balls of the same radius which shows that the scalar curvature is positive.

For a “mountain pass” surface in $\mathbb{R}^3$ the volume of the ball of radius r around the mountain pass p is *larger* then the area of the Euclidean ball of the same radius, and hence $s(p) < 0$.

While scalar curvature can be expressed in terms of ration of volume of balls, a more subtle volume comparison leads to Ricci curvature. We note that $B_r(p, M)$ can be viewed as the image of the ball $B_r(0, T_pM)$ in the tangent space T_pM under the exponential map

$$\exp_p \colon T_pM \longrightarrow M$$

which sends a tangent vector $v \in T_pM$ to $\gamma(1)$, where $\gamma\colon \mathbb{R} \to M$ is the geodesic in M determined by $\gamma(0) = p$, $\dot\gamma(0) = v$. In particular, if the exponential map is volume preserving, then the volume ratio (2.79) is equal to 1 for all r, and hence $s(p) = 0$. So the scalar curvature provides some measure for “volume distortion” of the exponential map $\exp_p$. What exactly does it mean to say that $\exp_p$ is volume preserving? The Riemannian metric determines the volume form $\mathrm{vol} \in \Omega^n(M)$ as well as the volume form $\mathrm{vol}^{T_pM} \in \Omega^n(T_pM)$ on the tangent space T_pM. The latter is the unique translation invariant n-form on T_pM whose value at $0 \in T_pM$ is determined by $\mathrm{vol}^{T_pM}(e_1, \dots, e_n) = 1$ for an oriented orthonormal basis $\{e_1, \dots, e_n\}$ of T_pM (which we identify with the tangent space of $0 \in T_pM$). The map $\exp_p$ is *volume preserving* if $\exp_p^* \mathrm{vol} = \mathrm{vol}^{T_pM}$. In general we have

$$\exp_p^* \mathrm{vol} = f\, \mathrm{vol}^{T_pM} \qquad \text{for some } f \in C^\infty(T_pM).$$

This function f is a measure for the extend that the exponential map distorts volumes. We concentrate on the volume distortion near the origin by using polar coordinates (x, r), $r \in [0, \infty)$, $x \in T_pM$ with $||x||^2 = 1$, and expanding $f(x, r)$ as a Taylor series in r. The result is

$$f(x, r) = 1 - \frac{1}{3}\,\mathrm{Ric}(x, x) r^2 + \dots .$$

A reference is again Besse’s book on Einstein manifolds [Be], where this discussion can be found in section 0.61 on pages 15 & 16. Regrettably, there is a typo in the crucial equation at the top of page 16, which should read

$$\exp_p^* \mathrm{vol} = \theta(x, r) d_px \otimes dr,$$

where d_px is the volume form on the unit sphere in T_pM. Since $\mathrm{vol}^{T_pM} = r^{n-1} d_px \otimes dr$, their function θ is related to our function f by $\theta(x, r) = r^{n-1} f(x, r)$.

2.3.2 Einstein field equations as Euler Lagrange equations

As in previous sections, we now want to address the question whether the partial differential equation (2.78) can be interpreted as the Euler-Lagrange equation of a suitable action functional. The answer is positive and goes back to David Hilbert who proved this in November of 1915, the same month that Einstein first wrote down his field equations.

Let M be a closed oriented manifold of dimension n, and let $\mathcal{M}$ be the space of pseudo metrics on M. Then the map

$$\mathcal{S}\colon \mathcal{M} \longrightarrow \mathbb{R} \qquad \text{defined by} \qquad \mathcal{S}(g) := \int_M s_g \operatorname{vol}_g$$

is called the *total scalar curvature* or *Einstein-Hilbert* functional. Here $s_g \in C^\infty(M)$ is the scalar curvature functional of the metric g, and $\operatorname{vol}_g \in \Omega^n(M)$ its volume form. We note that for $n > 2$ the functional $\mathcal{S}$ does not have any critical points, since if we rescale the metric g by multiplying it by a constant $\lambda > 0$, then $\operatorname{vol}_{\lambda g} = \lambda^{n/2} \operatorname{vol}_g$ and $s_{\lambda g} = \lambda^{-1} s_g$, and so

$$\mathcal{S}(g) = \int_M s_{\lambda g} \operatorname{vol}_{\lambda g} = \int_M \lambda^{-1} s_g \lambda^{n/2} \operatorname{vol}_g = \lambda^{\frac{n-2}{2}} \mathcal{S}(g).$$

This problem can be avoided by restricting the total scalar curvature functional of the subspace $\mathcal{M}_1 \subset \mathcal{M}$ consisting of those pseudo metrics g for which $\int_M \operatorname{vol}_g = 1$.

Theorem 2.80. *A pseudo metric $g \in \mathcal{M}_1$ on a closed oriented manifold M of dimension $n > 2$ is an Einstein metric if and only if it is a critical point of the total scalar functional $\mathcal{S}\colon \mathcal{M}_1 \to \mathbb{R}$.*

The crucial step in the proof of this result is the calculation of the derivative

$$d\mathcal{S}_g \colon T_g\mathcal{M} \longrightarrow \mathbb{R}$$

of $\mathcal{S}$ at a point $g \in \mathcal{M}$. To identify the tangent space $T_g\mathcal{M}$, we note that a pseudo metric g is a section of S^2T^*M such for each point $p \in M$ the symmetric bilinear form

$$g_p \colon T_pM \times T_pM \to \mathbb{R}$$

is non-degenerate. This shows that $\mathcal{M}$ is an open subset of the vector space $\Gamma(S^2T^*M)$. In particular, for any $g \in \mathcal{M}$ we can identify the tangent space $T_g\mathcal{M}$ with $\Gamma(S^2T^*M)$.

Before giving the formula for $d\mathcal{S}_g$, we need to introduce some notation. The pseudo metric g gives a scalar product, that is, a non-degenerate symmetric bilinear form $\langle -, - \rangle_g$ on each tangent space T_pM. Via the isomorphism $g\colon T_pM \to T_p^*M$ it gives a scalar product on the cotangent space T_p^*M as well. This in turn induces a scalar product on any vector bundle build from TM or T^*M by linear algebra operations, for example on the symmetric power $S^kT_p^*M$. Abusing language, we will denote all these bilinear forms by $\langle -, - \rangle_g$. The pseudo metric on the bundle $S^kT_p^*M$ given by the above construction induces a scalar product $(-,-)_g$ on the section space $\Gamma(S^kT^*M)$ in the usual way, defined by

$$(h_1, h_2)_g := \int_M \langle h_1, h_2 \rangle_g \operatorname{vol}_g .$$

Let us describe this explicitly for $k = 2$: given elements $h_1, h_2 \in S^2T_p^*M$ we define the scalar product

$$\langle h_1, h_2\rangle_g := \operatorname{tr}(\widetilde{h}_1\widetilde{h}_2),$$

where $\tilde{h}_i \colon T_pM \to T_pM$ is the self-adjoint endomorphism corresponding to the symmetric bilinear form $h_i \colon T_pM \times T_pM \to \mathbb{R}$ via equation (2.77). If h_1, h_2 are sections of S^2T^*M, then applying the above construction in each fiber gives a function $\langle h_1, h_2\rangle_g \in C^\infty(M)$.

Proposition 2.81. *For $g \in \mathcal{M}$ and $h \in T_g\mathcal{M} = \Gamma(S^2T^*M)$, the differential of the total scalar curvature functional $\mathcal{S} \colon \mathcal{M} \to \mathbb{R}$ is given by the formula*

$$d\mathcal{S}_g(h) = (\frac{1}{2}s_g g - \operatorname{Ric}_g, h)_g$$

The total scalar functional is build from the scalar curvature function s_g and the volume form vol_g. These can be viewed as maps

$$s \colon \mathcal{M} \longrightarrow C^\infty(M) \qquad \operatorname{vol} \colon \mathcal{M} \longrightarrow \Omega^n(M),$$

and we will need their derivatives.

Lemma 2.82. *For $g \in \mathcal{M}$ and $h \in T_g\mathcal{M} = \Gamma(S^2T^*M)$ we have*

1. $d\operatorname{vol}_g(h) = \frac{1}{2}\operatorname{tr}_g(h)\operatorname{vol}_g$.
2. $ds_g(h) = \Delta_g \operatorname{tr}_g(h) + \delta_g\delta_g(h) - \langle \operatorname{Ric}_g, h\rangle_g$.

Here $\delta_g \colon \Gamma(S^{k+1}T^*M) \to \Gamma(S^kT^*M)$ is a first order differential operator which is defined as the adjoint (with respect to the scalar products $(-,-)_g$ on $\Gamma(S^iT^*M)$ for $i = k, k+1$) to the first order differential operator δ_g^* given by the composition

$$\delta_g^* \colon \Gamma(S^kT^*M) \xrightarrow{\nabla} \Gamma(T^*M \otimes S^kT^*M) \xrightarrow{\text{projection}} \Gamma(S^{k+1}T^*M).$$

Here ∇ is the connection on the symmetric power S^kT^*M induced by the Levi-Civita connection on TM. We note that for $k = 0$, the linear map $\delta_g^* \colon C^\infty(M) \to \Omega^1(M)$ is just the de Rham differential and hence $\delta_g = d^* \colon \Omega^1(M) \to \Omega^0(M)$.

Proof. We will prove the first part of the lemma. For the proof of the second part, which is Theorem 1.174(e) in [Be], we refer the reader to that book.

Let $v_1, \dots, v_n$ be an oriented basis of T_pM. Then for any pseudo metric g the volume form vol_g is determined by

$$\operatorname{vol}_g(v_1, \dots, v_n) = \sqrt{|\det(g(v_i, v_j))_{1\le i,j\le n}|}$$

For sufficiently small $|t|$ the sum $g+th$ is again a pseudo metric on TM with

$$(g+th)(v,w) = g(v,w) + th(v,w) = g(v,w) + tg(\tilde{h}v,w) = g((\mathrm{id}+t\tilde{h})v,w).$$

To calculate the volume form for this pseudo-metric it will be convenient to use an oriented basis $v_1,\dots,v_n$ of eigenvectors for $\tilde{h}\colon T_pM \to T_pM$ with eigenvalues $\lambda_i \in \mathbb{R}$.

$$\begin{aligned}\mathrm{vol}_{g+th}(v_1,\dots,v_n) &= \sqrt{|\det((g+th)(v_i,v_j))_{1\le i,j\le n}|} = \sqrt{|\det(g((1+t\lambda_i)v_i,v_j))_{1\le i,j\le n}|}\\ &= \sqrt{\prod_i (1+t\lambda_i)}\sqrt{|\det(g(v_i,v_j))_{1\le i,j\le n}|} = \sqrt{\prod_i (1+t\lambda_i)}\,\mathrm{vol}_g(v_1,\dots,v_n)\end{aligned}$$

We note that $1+t\lambda_i$ is positive for sufficiently small t, and hence no absolute value is needed for this factor.

Then

$$\begin{aligned}(d\,\mathrm{vol}_g(h))(v_1,\dots,v_n) &= \frac{d}{dt}_{|t=0}\mathrm{vol}_{g+th}(v_1,\dots,v_n) = \frac{1}{2}(\sum_{i=1}^n \lambda_i)\,\mathrm{vol}_g(v_1,\dots,v_n)\\ &= \frac{1}{2}\operatorname{tr}(\tilde{h})\,\mathrm{vol}_g(v_1,\dots,v_n) = \frac{1}{2}\operatorname{tr}_g(h)\,\mathrm{vol}_g(v_1,\dots,v_n)\end{aligned}$$

□

Proof of Proposition 2.81.

$$\begin{aligned}d\mathcal{S}_g(h) &= \int_M d(s\,\mathrm{vol})_g(h) = \int_M (ds_g(h)\,\mathrm{vol}_g + s_g d\,\mathrm{vol}_g(h))\\ &= \int_M \left(\Delta_g \operatorname{tr}_g(h) + \delta_g\delta_g(h) - \langle \mathrm{Ric}_g, h\rangle_g + s_g\frac{1}{2}\langle g,h\rangle_g\right)\mathrm{vol}_g\\ &= (\Delta_g\operatorname{tr}_g(h),1)_g + (\delta_g\delta_g(h),1)_g + (\frac{1}{2}s_g g - \mathrm{Ric}_g, h)_g\end{aligned}$$

Here $1 \in C^\infty(M)$ is the constant function with value 1. We note that the first term vanishes since

$$(\Delta_g\operatorname{tr}_g(h),1)_g = (d^*d\operatorname{tr}_g(h),1)_g = (d\operatorname{tr}_g(h),d(1))_g = 0.$$

For the second term we note that $h \in \Gamma(S^2T^*M)$ and hence

$$\delta_g(h) \in \Gamma(S^1T^*M) = \Gamma(T^*M) = \Omega^1(M).$$

Moreover we recall that $\delta_g^*\colon \Omega^0(M) \to \Omega^1(M)$ is the de Rham differential, and hence

$$(\delta_g\delta_g(h),1)_g = (\delta_g(h),\delta_g^*(1))_g = (\delta_g(h),d(1))_g = 0.$$

This proves the proposition. □

Proof of Theorem 2.80. First we determine the condition on $h \in T_g\mathcal{M} = \Gamma(S^2T^*M)$ to belong to the codimension one subspace $T_g\mathcal{M}_1 \subset T_g\mathcal{M}$ by calculating the derivative of the map

$$\mathrm{Vol}\colon \mathcal{M} \longrightarrow \mathbb{R} \qquad g \mapsto \int_M \mathrm{vol}_g$$

which sends a (pseudo) metric g to the volume of M as measured in g.

$$d\,\mathrm{Vol}_g(h) = \int_M d\,\mathrm{vol}_g(h) = \int_M d\,\mathrm{vol}_g(h) = \int_M \frac{1}{2}\,\mathrm{tr}_g(h)\,\mathrm{vol}_g = \frac{1}{2}\int_M \langle g,h\rangle_g\,\mathrm{vol}_g = \frac{1}{2}(g,h)_g$$

This shows $T_g\mathcal{M}_1 = \{h \in \Gamma(S^2T^*M) \mid (g,h)_g = 0\}$.

Let $g \in \mathcal{M}_1$ be a pseudo metric which is a critical point for $\mathcal{S}\colon \mathcal{M}_1 \to \mathbb{R}$. Then Proposition 2.81 implies that

$$(\frac{1}{2}s_g g - \mathrm{Ric}_g, h)_g = 0 \qquad \text{for all } h \in T_g\mathcal{M}_1 = \{h \in \Gamma(S^2T^*M) \mid (g,h)_g = 0\}.$$

Our strategy is to add the term kg for some constant k such that

$$(kg + \frac{1}{2}s_g g - \mathrm{Ric}_g, h)_g = 0 \qquad \text{for } \textit{all } h \in \Gamma(S^2T^*M). \tag{2.83}$$

We note that for any k this holds for $h \in T_g\mathcal{M}_1$. Moreover, $\Gamma(S^2T^*M)$ is spanned by the codimension one subspace $T_g\mathcal{M}_1$ and $g \in \Gamma(S^2T^*M)$, since

$$(g,g)_g = \int_M \langle g,g\rangle_g\,\mathrm{vol}_g = \int_M \mathrm{tr}(\tilde{g}\tilde{g})\,\mathrm{vol}_g = \int_M \mathrm{tr}(\mathrm{id}_{T_pM})\,\mathrm{vol}_g = n\int_M \mathrm{vol}_g = n.$$

Hence we can achieve condition (2.83) by solving the equation $(kg + \frac{1}{2}s_g g - \mathrm{Ric}_g, g)_g = 0$ for k. Since $(-,-)_g$ is a scalar product on the section space $\Gamma(S^2T^*M)$, the condition implies

$$kg + \frac{1}{2}s_g g - \mathrm{Ric}_g = 0 \in \Gamma(S^2T^*M). \tag{2.84}$$

Applying tr_g we obtain the following equation in $C^\infty(M)$:

$$\begin{aligned} 0 &= k\,\mathrm{tr}_g(g) + \frac{1}{2}s_g\,\mathrm{tr}_g(g) - \mathrm{tr}_g(\mathrm{Ric}_g) = (k + \frac{1}{2}s_g)\,\mathrm{tr}(\tilde{g}) - s_g \\ &= (k + \frac{1}{2}s_g)n - s_g = kn + \frac{n-2}{2}s_g \end{aligned}$$

This shows that for $n \neq 2$ the scalar curvature function s_g is *constant*. Hence equation (2.84) implies that g is an Einstein metric. $\square$

2.4 Yang-Mills Theory

Yang-Mills Theory is a classical field theory which generalizes electromagnetism discussed in section 2.2. We recall that a field in electromagnetism is a metric connection on a Hermitian line bundle $L \to M$ over a 4-dimensional manifold M. The unit sphere bundle $S(L) \to M$ can be interpreted as a principal S^1-bundle (see Definition 2.85) and the metric connection on L can be interpreted as a principal connection on $S(L)$ (see Definition 2.91). This point of view leads to Yang-Mills theory by simply replacing S^1 by a more general Lie group. Before defining classical Yang-Mills theory (see Definition **??**), we need to define principal bundles, connections on principal bundles and their curvature.

2.4.1 Digression: Principal G-bundles

Definition 2.85. Let G be a topological group, that is, G is a topological space, and the multiplication map $G \times G \to G$ and the inverse map $G \to G$, $g \mapsto g^{-1}$ are continuous. A *principal G-bundle* over a topological space M is a locally trivial fiber bundle $\pi \colon P \to M$ with a right G-action $P \times G \to P$ on the total space P with the following properties.

- The projection map $\pi \colon P \to M$ is G-equivariant for the trivial G-action on M. Equivalently, the action preserves the fibers of π.
- The action is free and transitive on each fiber.

Example 2.86. Let $E \to M$ be a real vector bundle of rank n. Then the *frame bundle of E* is the fiber bundle $\pi \colon \mathrm{GL}(E) \to M$, where

$$\mathrm{GL}(E) := \{(x, f) \mid x \in X,\ f \colon \mathbb{R}^n \xrightarrow{\cong} E_x\}$$

and π sends a pair $(x, f) \in \mathrm{GL}(E)$ to $x \in M$. The linear isomorphism f between $\mathbb{R}^n$ and the fiber $E_x = \pi^{-1}(x)$ is called a *frame*. Equivalently, by looking at the images $f(e_i)$ of the standard basis $\{e_i\}$ of $\mathbb{R}^n$, a frame for E_x can be thought of as a basis of E_x. We note that the general linear group $\mathrm{GL}(\mathbb{R}^n)$ of linear automorphisms of $\mathbb{R}^n$ acts on the frame bundle $\mathrm{GL}(E)$ by pre-composition:

$$\mathrm{GL}(E) \times \mathrm{GL}(\mathbb{R}^n) \longrightarrow \mathrm{GL}(E) \qquad ((x, f), g) \mapsto (x, f \circ g)$$

It is easy to check that this action preserves the fibers, and acts freely and transitively on them. In other words, the frame bundle $\mathrm{GL}(E) \to M$ is a principal $\mathrm{GL}(\mathbb{R}^n)$-bundle.

There are many variations of frame bundles:

1. If E is equipped with a metric, we can require the frames $f \colon \mathbb{R}^n \to E_x$ to be isometries, which leads to the *orthonormal frame bundle* $O(E) \to M$. This is a principal bundle for the orthogonal group $O(n)$ of isometries of $\mathbb{R}^n$.

2. If E is oriented, we can require the frames $f\colon \mathbb{R}^n \to E_x$ to be orientation preserving, which leads to the *oriented frame bundle* $\mathrm{GL}^+(E) \to M$. This is a principal bundle for the group $\mathrm{GL}^+(\mathbb{R}^n)$ of orientation preserving automorphisms of $\mathbb{R}^n$.

3. If E is equipped with a metric and an orientation, we can require the frames to be isometries as well as orientation preserving, which leads to the *oriented orthonormal frame bundle* $SO(E) \to M$, which is a principal bundle for the special orthogonal group $SO(n)$.

4. If E is a complex vector bundle of rank n, its frame bundle $\mathrm{GL}(E) \to M$ is a principal $\mathrm{GL}(\mathbb{C}^n)$-bundle, where $\mathrm{GL}(\mathbb{C}^n)$ is the complex linear group of complex linear automorphisms of $\mathbb{C}^n$.

5. If E is a Hermitian vector bundle of rank n, we can require the frames to be complex linear isometries which leads to the *unitary frame bundle* $U(E)$, which is a principal bundle for the unitary group $U(n)$ of complex linear isometries of $\mathbb{C}^n$.

The frame bundle construction described above produces principal bundles from vector bundles. There is another important construction going from principal bundles to vector bundles.

Definition 2.87. Let $\pi\colon P \to M$ be a principal G-bundle and let $\rho\colon G \to \mathrm{GL}(V)$ be a representation. We write $P \times_{G,\rho} V$ (or simply $P \times_G V$ if the G-action on V is understood) for the quotient of the product $P \times V$ obtained by identifying (pg, v) with $(p, \rho(g)v)$ for $p \in P$, $g \in G$, $v \in V$. Alternatively, we can think of $P \times_G V$ as the quotient of $P \times V$ by the free left G-action given by sending (p, v) to $(pg^{-1}, \rho(g)v)$ (right multiplication by g^{-1} is a *left*-action on P). We will use the notation $[p, v]$ for the point in $P \times_G V$ represented by (p, v).

We note that the fibers of the projection map

$$q\colon P \times_G V \to M \qquad [p, v] \mapsto \pi(p) \tag{2.88}$$

are vector spaces isomorphic to V, since if we pick a point p_0 in the fiber $P_x = \pi^{-1}(x)$, then the map $V \to (P \times_G V)_x$, $v \mapsto [p_0, v]$ is a homeomorphism. In fact, this is a vector bundle, called the *associated vector bundle* (the local trivializations of $P \to M$ can be used to manufacture local trivializations for (2.88)).

Example 2.89. Let $E \to M$ be a vector bundle of rank n. Then the associated vector bundle $\mathrm{GL}(E) \times_{\mathrm{GL}(\mathbb{R}^n)} \mathbb{R}^n$ build from the frame bundle $\mathrm{GL}(E)$, a principal $\mathrm{GL}(\mathbb{R}^n)$-bundle, and the basic representation of $\mathrm{GL}(\mathbb{R}^n)$ on $\mathbb{R}^n$ is isomorphic to E. The vector bundle isomorphism

$$\mathrm{GL}(E) \times_{\mathrm{GL}(\mathbb{R}^n)} \mathbb{R}^n \xrightarrow{\cong} E \qquad \text{is given by} \qquad [(x, f), v] \mapsto (x, f(v)).$$

Homework 2.90. 1. Show that the category $\mathsf{Vect}^n_{\mathbb{R}}(M)$ of real vector bundles and vector bundle isomorphisms over M is equivalent to the category $\mathsf{Bun}^G(M)$ of principal G-bundles over M for $G = \mathrm{GL}(\mathbb{R}^n)$.

2. Show similarly that each of the variations of the frame bundle construction discussed above leads to an equivalence between $\mathsf{Bun}^G(M)$ for $G = O(n)$, $\mathrm{GL}^+(\mathbb{R}^n)$, $SO(n)$, $\mathrm{GL}(\mathbb{C}^n)$, or $U(n)$ and suitable categories of vector bundles with additional structure on M.

This shows that principal G-bundles can be regarded as generalizations of vector bundles. Our next task is to define connections on principal G-bundles in such a way that connections on frame bundles correspond to connections on the underlying vector bundle.

2.4.2 Digression: Connections on principal G-bundles

As we have seen, connections are about differentiation, which means that if we want to talk about connections on a principal G-bundle $P \to M$, we need to assume differentiability conditions for the objects G, P, and M. For simplicity we will assume that they are *smooth*, that is, that derivatives of all orders exist. More precisely, we will assume that

- M is a smooth manifold;
- G is a Lie group, that is, G is a smooth manifold and the multiplication map $G \times G \to G$ and the inversion map $G \to G$ are smooth;
- $P \to M$ is a *smooth* principal G-bundle (which means that $P \to M$ is a smooth bundle, and the action $P \times G \to P$ is smooth).

Definition 2.91. A *connection* on a principal G-bundle $P \to M$ is a sub bundle $\mathcal{H}$ of the tangent bundle TP with the following properties

- It is complementary to the vertical sub bundle $\mathcal{V} \subset TP$; that is, for every point $p \in P$, the tangent space T_pP is the direct sum $\mathcal{V}_p \oplus \mathcal{H}_p$ of the vertical and horizontal subspaces.
- It is equivariant; that is, for every $g \in G$ we have $(R_g)_*(\mathcal{H}_p) = \mathcal{H}_{pg}$, where $(R_g)_*\colon T_pP \to T_{pg}P$ is the differential of the map $R_g\colon P \to P$ given by right multiplication by g: $p \mapsto pg$.

The above definition is geometric and conceptual. For calculational purposes, it is convenient to describe the horizontal subspace $\mathcal{H}_p$ as the kernel of the projection map $T_PP = \mathcal{V}_p \oplus \mathcal{H}_p \to \mathcal{V}_p$. Identifying the vertical subspace $\mathcal{V}_p$ with the Lie algebra $\mathfrak{g}$ as discussed above, we can think of this projection map as map of vector bundles

$$\omega\colon TP \longrightarrow P \times \mathfrak{g} =: \underline{\mathfrak{g}},$$

or as a section of the bundle $\mathrm{Hom}(TP, \underline{\mathfrak{g}}) \cong T^*P \otimes \underline{\mathfrak{g}}$, that is, as a 1-form $\omega \in \Omega^1(P; \mathfrak{g})$ with values in the Lie algebra $\mathfrak{g}$.

Definition 2.92. A *connection 1-form* on a principal G-bundle $P \to M$ is a differential form $\omega \in \Omega^1(P; \mathfrak{g})$ with the following properties

- $\omega(X_\xi) = \xi$ for all $\xi \in \mathfrak{g}$. Here we write $X_\xi \in \mathcal{V}_p$ for the vertical tangent vector corresponding to $\xi \in \mathfrak{g}$ via the isomorphism $\mathfrak{g} \cong \mathcal{V}_p$.
- ω is G-equivariant; this can be written down in a number of equivalent ways, for example, we could say that the vector bundle map
$$\omega \colon TP \longrightarrow \mathfrak{g}$$
is equivariant, where $g \in G$ acts on $\mathfrak{g}$ via the *adjoint action*
$$\mathrm{ad}_g \colon \mathfrak{g} \longrightarrow \mathfrak{g},$$
which is the differential at $e \in TG$ of the adjoint action (= conjugation action) of $g \in G$ on G given by
$$\mathrm{Ad}_g \colon G \longrightarrow G, \qquad h \mapsto ghg^{-1}.$$

Next, we'll make a big table comparing connections on vector bundles with those on principal bundles.

1. $\{$connections on $E\} \leftrightarrow \Omega^1(M; \mathrm{End}(E))$ corresponds to $\{$connections on $P\} \leftrightarrow \Omega^1(M; \mathrm{ad}(P))$, $\mathrm{ad}(P) := P \times_{G,\mathrm{ad}} \mathfrak{g}$. This is given by $\omega \mapsto \omega - \omega_0 \in \Omega^1(P; \mathfrak{g})_{\mathrm{basic}} \cong \Omega^1(M; \mathrm{ad}(P))$.
2. Parallel translation along a piecewise smooth curve $\gamma \colon \mathbb{R} \to M$ gives
$$||_a^b(\gamma) \colon E_{\gamma(a)} \to E_{\gamma(b)} \qquad \text{given by} \qquad v \mapsto \tilde{\gamma}(b)$$
where $\tilde{\gamma} \colon \mathbb{R} \to E$ is the unique *horizontal lift* of γ with $\tilde{\gamma}(a) = v$. This means:
    - $\tilde{\gamma}$ is a *lift* of γ, i.e., $\pi \circ \tilde{\gamma} = \gamma$;
    - the lift $\tilde{\gamma}$, viewed a section of the pull-back $\gamma^*E \to \mathbb{R}$ is *horizontal* in the sense that $\nabla_{\frac{\partial}{\partial t}} \tilde{\gamma} = 0$.

    Show example of parallel translation in TS^2 via the Levi-Civita connection. The analog for principal bundles is a map $||_a^b(\gamma) \colon P_{\gamma(a)} \to P_{\gamma(b)}$ defined analogously. Talk about holonomy in both cases: if γ is a loop starting and ending at $x \in M$, then $\mathrm{Hol}(\gamma) \in GL(E_x)$ for vector bundle E, $\mathrm{Hol}(\gamma) \in \mathrm{Diff}(P_x)^G$
3. Curvature $R^\nabla(v, w) \in \mathrm{End}(E_x)$ for $v, w \in T_xM$. Analog for principal bundles. Formula $\Omega = d\omega + \frac{1}{2}[\omega, \omega]$.

2.5 Chern-Simons Theory

Chern-Simons Theory is a classical field theory whose space of fields $\mathcal{M}$ is the space of connections on a principal G-bundle $P \to M$ over a compact manifold M of dimension 3. So the space of fields is very similar to Yang-Mills Theory whose space of fields as described in section **??** is the space of connections on a principal bundle over a manifold of dimension 4. However, the Chern-Simons action functional is more subtle and to describe it we first need two digressions, one on classifying spaces for principal G-bundles, the other one on characteristic classes.

2.5.1 Digression: classifying spaces for principal G-bundles

We recall from section 2.2.8 that isomorphism classes of n-dimensional vector bundles over a topological space X are in bijective correspondence with homotopy classes of maps $X \to G_n(\mathbb{F}^\infty)$ to the Grassmann manifold of n-dimensional subspaces of $\mathbb{F}^\infty$, where $\mathbb{F} = \mathbb{R}$ for real vector bundles and $\mathbb{F} = \mathbb{C}$ for complex vector bundles. Moreover, this correspondence is given by sending a map $f\colon X \to G_n(\mathbb{F}^\infty)$ to the pullback vector bundle $f^*\gamma^n \to X$, where$\gamma \to G_n(\mathbb{F}^\infty)$ is the tautological vector bundle.

The following theorem is the analog of this result for principal G-bundles.

Theorem 2.93. *For any topological group G there is a bijection*

$$[X, BG] \longleftrightarrow \{\textit{principal } G\textit{-bundles over } X\}/\textit{isomophism}$$

for any topological space X. Here $[X, BG]$ is the set of homotopy classes of maps from X to a topological space BG associated to G called the classifying space *of G. The bijection associates to a map $f\colon X \to BG$ the principal bundle $f^*EG \to X$ obtained by pullback via f from a principal G-bundle $EG \to BG$ called the* universal principal G-bundle*. Any map $f\colon X \to BG$ corresponding to a principal bundle $P \to X$ via this bijection is called a* classifying map *for P.*

Remark 2.94. The classifying space construction is *functorial*, meaning that if $f\colon G \to H$ is a homomorphism of topological groups, then there is an associated map of classifying spaces $Bf\colon BG \to BH$ which is compatible with compositions in the obvious way. However, it is useful to know that the classifying space BG is characterized up to homotopy equivalence by the property that $EG \to BG$ is a principal G-bundle with contractible total space EG. In other words, if $E'G \to B'G$ is a principal G-bundle with contractible total space $E'G$, then $B'G$ is homotopy equivalent to BG. Moreover, the pullback of $E'G \to B'G$ via this homotopy equivalence is isomorphic to the principal G-bundle $EG \to BG$. In particular, the above result holds with $EG \to BG$ replaced by $E'G \to B'G$, and so abusing language every principal G-bundles $EG \to BG$ with contractible total space EG will be called *universal principal G-bundle.*

Here are some examples of universal principal G-bundles.

1. For $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ the infinite dimensional sphere $S(\mathbb{F}^\infty)$ is contractible and the group of unit length scalars $S(\mathbb{R}) = \{\pm 1\}$ resp. $S(\mathbb{C}) = S^1$ act freely on $S(\mathbb{F}^\infty)$. This implies that
$$E\mathbb{Z}/2 := S(\mathbb{R}^\infty) \to B\mathbb{Z}/2 := S(\mathbb{R}^\infty)/\mathbb{Z}/2 = \mathbb{RP}^\infty$$
and
$$E\mathbb{Z}/2 := S(\mathbb{C}^\infty) \to BS^1 := S(\mathbb{C}^\infty)/S^1 = \mathbb{CP}^\infty$$
are universal principal bundles.

2. Generalizing the previous example, the *Stiefel manifold* $V_n(\mathbb{F}^\infty)$, the space of all isometries $f\colon \mathbb{F}^n \to \mathbb{F}^\infty$, is contractible; we note that $V_1(\mathbb{F}^\infty)$ can be identified with $S(\mathbb{F}^\infty)$. The group $\mathrm{Isom}(\mathbb{F}^n)$ of isometries of $\mathbb{F}^n$, which is the orthogonal group $O(n)$ for $\mathbb{F} = \mathbb{R}$ and the unitary group $U(n)$ for $\mathbb{F} = \mathbb{C}$, acts freely on $V_n(\mathbb{F}^\infty)$ by pre-composition. We note that mapping an isometry $f\colon \mathbb{F}^n \to \mathbb{F}^\infty$ to its image gives a homeomorphism between the quotient space of this action and the Grassmann manifold $G_n(\mathbb{F}^\infty)$ of n-dimensional subspaces of $\mathbb{F}^\infty$. Hence
$$EO(n) := V_n(\mathbb{R}^\infty) \to BO(n) := V_n(\mathbb{R}^\infty)/O(n) = G_n(\mathbb{R}^\infty)$$
and
$$EU(n) := V_n(\mathbb{C}^\infty) \to BU(n) := V_n(\mathbb{C}^\infty)/U(n) = G_n(\mathbb{C}^\infty)$$
are universal principal bundles.

3. If $EG \to BG$ and $EH \to BH$ are universal principal bundles for G and H, respectively, then $EG \times EH \to BG \times BH$ is a universal principal bundle for the product group $G \times H$.

2.5.2 Digression: characteristic classes

Definition 2.95. A *characteristic class* for principal G-bundles is a rule that assigns to any principal G-bundle $P \to X$ (over an unspecified base space X) a cohomology class $c(P) \in H^*(X)$, which depends only on the isomorphism class of P. It is required to be compatible with pull-backs in the sense that for any map $f\colon X \to Y$ and a principal G-bundle $P \to Y$ we have
$$c(f^*P) = f^*c(P) \in H^*(X).$$
Here $H^*(X)$ is the cohomology of X with any coefficients; in fact, we can replace cohomology by generalized cohomology (i.e., drop the dimension axiom).

Characteristic classes provide us with invariants to show that some principal bundles are non-trivial: if $c(P) \in H^i(X)$ is a non-zero characteristic class in degree $i \geq 0$, then P is non-trivial. The argument is this: if P were trivial, it were isomorphic to the pull-back of the trivial principal bundle over the one-point-space pt via the projection map $p\colon X \to \mathrm{pt}$. This implies that $c(P)$ is in the image of the induced map $p^*\colon H^i(\mathrm{pt}) \to H^i(X)$ which is trivial, thus giving the desired contradiction.

Evaluating a characteristic class on the universal bundle $EG \to BG$ gives a cohomology class $c = c(EG) \in H^*(BG)$ of the classifying space. Theorem 2.93 implies that this provides a bijection between characteristic classes for principal G-bundles and the cohomology of the classifying space BG. This leads to the question how to calculate the cohomology groups of BG. If G is a discrete group, the cohomology of BG is known as *group cohomology*, which can be defined in purely algebraically.

Our main interest will be in compact connected Lie groups. We begin by recalling some standard Lie theory terminology. If G is a compact connected Lie group, a maximal abelian Lie subgroup $T \subset G$ is isomorphic to a product $S^1 \times \cdots \times S^1$ of S^1's, which is why such a subgroup is referred to as *maximal torus*. Any two maximal tori are conjugate; in particular, they have the same dimension which is called the *rank* of the Lie group G. Given a maximal torus $T \subset G$, let $NT \subset G$ be its normalizer consisting of all $g \in G$ such that $gTg^{-1} = T$. The group NT contains T as a normal subgroup; the quotient group $W := NT/T$ is called the *Weyl group*. The conjugation action of NT on T induces an action of W on T.

Example 2.96. A maximal torus for the unitary group $U(n)$ of linear isometries $f\colon \mathbb{C}^n \to \mathbb{C}^n$ is given by the subgroup T consisting of those f's with $f(e_i) = z_i e_i$, where $\{e_i\}$ is the standard basis of $\mathbb{C}^n$, and $z_i \in S^1$. The normalizer NT is generated by T and the permutation subgroup of $U(n)$ consisting of f's with $f(e_i) = e_{\sigma(i)}$, where $\sigma \in S_n$ is a permutation of n letters. Hence the Weyl group W can be identified with the symmetric group S_n which acts on $T = S^1 \times \cdots \times S^1$ by permuting the factors.

The inclusion $i\colon T \to G$ of a maximal torus induces a map of classifying spaces $Bi\colon BT \to BG$ which in turn induces a map on cohomology

$$Bi^*\colon H^*(BG) \longrightarrow H^*(BT).$$

Lemma 2.97. *The image of Bi^* is contained in the fixed point set $H^*(BT)^W$ of the induced action of the Weyl group W on the cohomology of BT.*

Proof. An element $g \in NT$ acts on G by conjugation leaving the subgroup $T \subset G$ invariant. In other words, we have a commutative diagram of Lie group homomorphisms

$$\begin{tikzcd} T \arrow[r, "i"] \arrow[d, "c_g"'] & G \arrow[d, "c_g"] \\ T \arrow[r, "i"] & G \end{tikzcd},$$

where the vertical maps are conjugation by g. This induces a corresponding diagram of classifying spaces, which in turn induces the following commutative diagram of maps between cohomology groups:

$$\begin{tikzcd} H^*(BT) & H^*(BG) \arrow[l, "Bi^*"'] \\ H^*(BT) \arrow[u, "Bc_g^*"] & H^*(BG) \arrow[l, "Bi^*"] \arrow[u, "Bc_g^*"'] \end{tikzcd}$$

Our assumption that G is connected means that the element $g \in NT$ can be connected to the identity element $e \in G$ by a path. This means that the maps

$$Bc_g\colon BG \to BG \qquad \text{and} \qquad Bc_e = B\,\mathrm{id}_G = \mathrm{id}_{BG}$$

are homotopic, and hence $Bc_g^* = Bc_e^* = \mathrm{id}_{H^*(BG)}$, which implies that for $\alpha \in H^*(BG)$ we have

$$Bc_g^* Bi^*(\alpha) = Bi^* Bc_g^*(\alpha) = Bi^*\alpha$$

as claimed.

We remark that the above argument *does not* show that Bc_g^* is the identity of $H^*(BT)$, since in general $g \in NT$ cannot be connected to the identity element e by a path in NT, but only conjugation by elements in NT gives an automorphism of T. □

Theorem 2.98. *Let G be a compact connected Lie group with maximal torus T and Weyl group W and let*

$$Bi^*\colon H^*(BG) \longrightarrow H^*(BT)^W$$

be the homomorphism induced by the inclusion map $i\colon T \to G$.

1. *This map is an isomorphism for cohomology with coefficients in $\mathbb{Q}$ or $\mathbb{R}$.*

2. *It is an isomorphism on integral cohomology for $G = U(n)$.*

Let us use the second statement to calculate the cohomology ring $H^*(BU(n);\mathbb{Z})$. The Weyl group W is the symmetric group S_n which acts on the maximal torus $T = S^1 \times \cdots \times S^1$ by permuting the factors. It follows that

$$BT = B(S^1 \times \cdots \times S^1) = BS^1 \times \cdots \times BS^1 = \mathbb{CP}^\infty \times \cdots \times \mathbb{CP}^\infty,$$

and S_n acts on BT by permuting the factors. We recall that the cohomology ring of the complex projective space $\mathbb{CP}^\infty$ is the polynomial ring $\mathbb{Z}[x]$, where x is one of the two generators of $H^2(\mathbb{CP};\mathbb{Z}) \cong \mathbb{Z}$. From our discussion in section 2.2.8 we know that we can choose

$$x = c_1(\gamma) \in H^2(\mathbb{CP}^\infty;\mathbb{Z})$$

to be the first Chern class of the universal complex line bundle $\gamma \to \mathbb{CP}^\infty$. The Künneth Theorem then implies the ring isomorphism

$$H^*(BT;\mathbb{Z}) \cong \mathbb{Z}[x_1,\ldots,x_n],$$

where $x_i = c_1(p_i^*\gamma)$ is the first Chern class of the pullback of the universal line bundle via the projection map

$$p_i\colon \mathbb{CP}^\infty \times \cdots \times \mathbb{CP}^\infty \longrightarrow \mathbb{CP}^\infty$$

onto the i-th factor. The second part of the theorem then yields the first of following the ring isomorphisms

$$H^*(BU(n);\mathbb{Z}) \cong H^*(BT;\mathbb{Z})^W \cong \mathbb{Z}[x_1,\ldots,x_n]^{S_n} \cong \mathbb{Z}[\sigma_1,\ldots,\sigma_n].$$

Here σ_i is the i-th elementary symmetric polynomial in $x_1,\ldots,x_n$, and the last isomorphism is the Fundamental Theorem of symmetric polynomials. Since σ_i is a polynomial of degree i in the x_i's, and each x_i is a 2-dimensional cohomology class, the elementary symmetric polynomial σ_i provides a cohomology class $c_i \in H^{2i}(BU(n);\mathbb{Z}) \cong H^{2i}(BT;\mathbb{Z})^W$ known as the *i-th Chern class* of the universal principal $U(n)$-bundle. As discussed in the beginning of this section, pulling this class back via the classifying map of a principal $U(n)$-bundle $P \to X$ leads to the i-th Chern class $c_i(P) \in H^{2i}(X;\mathbb{Z})$. This in turn leads to the construction of the i-th Chern class $c_i(E) \in H^{2i}(X;\mathbb{Z})$ of a complex vector bundle E of rank n by choosing a Hermitian metric on E, and defining $c_i(E) \in H^{2i}(X;\mathbb{Z})$ to be the i-th Chern class of the Hermitian frame bundle $P := U(E) \to X$.

2.5.3 Digression: the Chern-Weil construction

In this section we discuss the extremely useful consequences of part (1) of Theorem 2.98, which gives an isomorphism

$$H^*(BG;\mathbb{R}) \cong H^*(BT;\mathbb{R})^W$$

for every compact connected Lie group G. We've seen in the last section that the cohomology ring $H^*(BT;\mathbb{Z})$ is isomorphic to the polynomial ring $\mathbb{Z}[x_1,\ldots,x_n]$. This isomorphism can be made canonical, that is, without requiring us to pick an isomorphism between the maximal torus $T \subset G$ and the product $S^1 \times \cdots \times S^1$. Namely, one can construct a canonical isomorphism between the cohomology ring $H^*(BT;\mathbb{R})$ and the space

$$\mathsf{Sym}^*(\mathfrak{t}^\vee) := \text{symmetric algebra generated by } \mathfrak{t}^\vee = \{\text{polynomial functions } Q\colon \mathfrak{t} \to \mathbb{R}\}.$$

Here $\mathfrak{t}$ is the Lie algebra of the maximal torus, and $\mathfrak{t}^\vee$ its dual. The conjugation action of the Weyl group W on the maximal torus T induces an action on the algebra $\mathsf{Sym}^*(\mathfrak{t}^\vee)$. Part (1) of Theorem 2.98 then gives a ring isomorphism

$$H^*(BG;\mathbb{R}) \cong \mathsf{Sym}^*(\mathfrak{t}^\vee)^W = \{W\text{-invariant polynomial functions } Q\colon \mathfrak{t} \to \mathbb{R}\}.$$

We observe that any G-invariant polynomial $Q\colon \mathfrak{g} \to \mathbb{R}$ on the Lie algebra $\mathfrak{g}$ restricts to a W-invariant polynomial $Q_{|\mathfrak{t}}\colon \mathfrak{t} \to \mathbb{R}$ on the Lie algebra of the maximal torus. The restriction map

$$\mathsf{Sym}^*(\mathfrak{g}^\vee)^G \longrightarrow \mathsf{Sym}^*(\mathfrak{t}^\vee)^W$$

is in fact an algebra isomorphism, since the G-orbits in $\mathfrak{g}$ are in bijective correspondence with W-orbits in $\mathfrak{t}$; this bijection is given by sending a G-orbit to its intersection with $\mathfrak{t} \subset \mathfrak{g}$.

We conclude that a G-invariant polynomial

$$Q\colon \mathfrak{g} \longrightarrow \mathbb{R}$$

of degree d corresponds to a cohomology class in $H^{2d}(BG;\mathbb{R})$, and hence, as discussed in the last section, leads to a characteristic class $c_Q(P) \in H^{2d}(M;\mathbb{R})$ for principal G-bundles $P \to M$. This prompt the following question.

Question. Is there a direct construction which produces a de Rham cohomology class $c_Q(P) \in H^{2d}_{dR}(M)$ associated to a G-invariant polynomial Q of degree d?

The positive answer to this question is provided by the *Chern-Weil construction* which proceeds as follows. Let ω a connection on the principal G-bundle $P \to M$, and let $\Omega \in \Omega^2(M;\mathrm{ad}(P)) = \Gamma(\Lambda^2 T^*M \otimes \mathrm{ad}(P))$ be its curvature. We recall that the adjoint vector bundle $\mathrm{ad}(P)$ is the vector bundle

$$\mathrm{ad}(P) = P \times_G \mathfrak{g},$$

where G acts on its Lie algebra $\mathfrak{g}$ by the adjoint representation. The G-invariant polynomial $Q \in \mathsf{Sym}^d(\mathfrak{g}^\vee) \cong (\mathsf{Sym}^d \mathfrak{g})^\vee$ can be viewed as a G-equivariant linear map $Q\colon \mathsf{Sym}^d \mathfrak{g} \to \mathbb{R}$. Hence it induces a vector bundle map between the associated vector bundles

$$Q\colon \mathsf{Sym}^d(\mathrm{ad}(P)) \cong P \times_G \mathsf{Sym}^d \mathfrak{g} \longrightarrow P \times_G \mathbb{R} = \underline{\mathbb{R}}. \tag{2.99}$$

Abusing language, we again write Q for this map of vector bundles. Noting that

$$\underbrace{\Omega \wedge \cdots \wedge \Omega}_{d} \in \Omega^{2d}(M;\mathrm{ad}(P)^{\otimes d})$$

is invariant under permutations of the factors of $\mathrm{ad}(P)^{\otimes d}$, we can view $\Omega \wedge \cdots \wedge \Omega$ as an element of $\Omega^{2d}(M;\mathsf{Sym}^d(\mathrm{ad}(P)))$. Its image under the vector bundle map (2.99) is the *Chern-Weil form* $\mathrm{CW}_Q(P,\omega) \in \Omega^{2d}(M)$. According to the basic lemmas of Chern-Weil theory, for whose proofs we refer to [MS, Appendix C], this form enjoys the following properties:

- $\mathrm{CW}_Q(P,\omega)$ is a *closed* differential form.
- Its de Rham cohomology class $\mathrm{CW}_Q(P) := [\mathrm{CW}_Q(P,\omega)] \in H^{2d}_{dR}(M)$ is *independent of the connection* ω.

Example 2.100. For simple Lie groups G, let us compute the space of G-invariant polynomials $Q\colon \mathfrak{g} \to \mathbb{R}$ of degree one and two. We recall that a Lie group G is called *simple* if the adjoint representation of G on its Lie algebra $\mathfrak{g}$ is irreducible and non-trivial. For example, the group $U(n)$ of unitary complex $n \times n$-matrices is not simple, since its Lie algebra $\mathfrak{u}(n)$ consists of all skew-Hermitian matrices which can be decomposed as the direct sum of the trace-free skew-Hermitian matrices and purely imaginary multiples of the identity matrix (this is equivariant with the respect to the adjoint action of $U(n)$ on its Lie algebra which is by conjugation). The space of trace-free skew-Hermitian matrices can be interpreted as the Lie algebra of the special unitary group $SU(n)$, which is simple.

Let G be a compact connected simple Lie group.

- There are no non-trivial G-equivariant polynomials Q of degree $d = 1$, since the kernel of a such a map $\mathfrak{g} \to \mathbb{R}$ would be a G-equivariant proper subspace of $\mathfrak{g}$, contradicting our assumption that $\mathfrak{g}$ is irreducible.

- The space of G-equivariant polynomials Q of degree $d = 2$ is 1-dimensional. To see that this space is non-zero, pick an inner product on $\mathfrak{g}$ (a positive definite symmetric bilinear form on $\mathfrak{g}$), and construct from it a G-equivariant inner product by averaging over the compact group G using a Haar measure. Thinking of this G-invariant symmetric bilinear form as a G-invariant polynomial function $Q\colon \mathfrak{g} \to \mathbb{R}$ of degree 2 shows that this space is non-trivial.

  To show that the space is 1-dimensional, suppose that $\langle\ ,\ \rangle$ and $\{\ ,\ \}$ are two non-trivial G-invariant symmetric bilinear forms on $\mathfrak{g}$. These forms are necessarily non-degenerate since their annihilators would otherwise form G-invariant proper subspace of $\mathfrak{g}$ contradicting the irreducibility of $\mathfrak{g}$. This implies that there is an endomorphism $A\colon \mathfrak{g} \to \mathfrak{g}$, symmetric w.r.t. $\langle\ ,\ \rangle$ such that

$$\{v, w\} = \langle Av, w\rangle \qquad \text{for all } v, w \in \mathfrak{g}.$$

  The G-equivariance of $\langle\ ,\ \rangle$ and $\{\ ,\ \}$ implies G-equivariance of A. Hence by Schur's Lemma, A is a multiple of the identity which implies that $\{\ ,\ \}$ belongs to the space spanned by $\langle\ ,\ \rangle$.

In conjunction with part (1) of Theorem 2.98, this shows that for a compact connected simple Lie group G the real cohomology groups of the classifying space BG in degree ≤ 4 are given by

$$H^q(BG; \mathbb{R}) = \begin{cases} \mathbb{R} & q = 0, 4 \\ 0 & q = 1, 2, 3 \end{cases} \tag{2.101}$$

Moreover, the image of the coefficient homomorphism $H^4(BG; \mathbb{Z}) \to H^4(BG; \mathbb{R})$ is isomorphic to $\mathbb{Z}$, and hence it has *two* generators. Multiplication by -1 exchanges these two generators. Is there a good way to pick out one of these two? Yes, via the bijection provided by part (1) of Theorem 2.98 both of them correspond to non-zero G-invariant polynomial maps $\mathfrak{g} \to \mathbb{R}$ of degree 2, which by the discussion above are the same as G-invariant non-degenerate symmetric bilinear forms on $\mathfrak{g}$. Hence one of them is *positive definite*, while the other, as the negative of the first one, is negative definite.

2.5.4 Digression: Chern-Simons form and Chern-Simons invariant

The Chern-Simons form is a *relative* version of the Chern-Weil form. It depends on *two* connections on a principal G-bundle and a G-invariant polynomial on the Lie algebra $\mathfrak{g}$. It is defined as follows.

Definition 2.102. Let $Q\colon \mathfrak{g} \to \mathbb{R}$ be a G-invariant polynomial of degree d. Let ω_0, ω_1 be connection forms on a principal G-bundle $\pi\colon P \to M$. Abusing language, we continue to write ω_0, ω_1 for the connections on the pullback bundle $P \times [0,1] = p^*P \to M \times [0,1]$. Then $\omega := t\omega 1 + (1-t)\omega_0$ is a connection 1-form on $P \times [0,1]$, where $t \in C^\infty(P \times [0,1])$ is the projection onto the second coordinate. Integrating the Chern-Weil form $\mathrm{CW}_Q(P,\omega) \in \Omega^{2d}(M \times [0,1])$ over $[0,1]$ results in a form

$$\mathrm{CS}_Q(P,\omega_1,\omega_0) := \int_0^1 \mathrm{CW}_Q(P,\omega) \in \Omega^{2d-1}(M)$$

called the *Chern-Simons form*. Unlike the Chern-Weil form, this form is not in general closed, but rather its construction implies that

$$d\,\mathrm{CS}_Q(P,\omega_1,\omega_0) = \mathrm{CW}_Q(P,\omega_1) - \mathrm{CW}_Q(P,\omega_0).$$

The Chern-Simons form can be used to produce an invariant for connections on principal G-bundles over closed 3-manifolds. For simplicity we assume that G is a simple Lie group, in which case every principal G over a 3-manifold is trivializable. So we will only deal with connections on the trivial G-bundle, but we will comment on the more general case below.

Definition 2.103. Let G be a compact simple Lie group, and let ω be a connection on the trivial principal G-bundle $P = M \times G \to M$ over a closed 3-manifold M. Let $Q\colon \mathfrak{g} \to \mathbb{R}$ be the unique G-invariant positive definite quadratic form whose associated characteristic class $c_Q(EG) \in H^4(BG;\mathbb{R}) \cong \mathbb{R}$ is a generator of the image of $H^4(BG;\mathbb{Z}) \to H^4(BG;\mathbb{R})$.

Then the real number

$$\mathrm{CS}(\omega) := \int_M \mathrm{CS}_Q(P,\omega,\omega^{\mathrm{taut}})$$

is the *Chern-Simons invariant* of the connection ω; here ω^{taut} is the tautological connection on the trivial bundle P.

Remark 2.104. In the definition above we are suppressing an important aspect of the Chern-Simons invariant, namely its transformation properties under gauge transformations. These are G-equivariant diffeomorphisms $f\colon P \to P$ which commute with the projection map $P \to M$. At first sight, it might seem that $\mathrm{CS}_Q(\omega) = \mathrm{CS}_Q(f^*\omega)$ due to the functorial nature of our constructions. This would be true, if the pullback $f^*\omega^{\mathrm{taut}}$ of the tautological connection ω^{taut} were equal to ω^{taut}, but rather $f^*\omega^{\mathrm{taut}}$ is the tautological connection *with the respect to a new trivialization of P provided by f*. It turns out that

$$\int_M \mathrm{CS}_Q(P,\omega,f^*\omega^{\mathrm{taut}}) = \int_M \mathrm{CS}_Q(P,\omega,\omega^{\mathrm{taut}}) + \int_M \mathrm{CS}_Q(P,\omega^{\mathrm{taut}},f^*\omega^{\mathrm{taut}}),$$

and that the second summand on the right hand side in an integer thanks to our normalization of the G-invariant quadratic form Q. The upshot is that if $P \to M$ is a trivializable principal G-bundle with connection ω its Chern-Simons invariant $\mathrm{CS}_Q(P,\omega) \in \mathbb{R}$ depends on the choice of a trivialization, but the difference of Chern-Simons invariants corresponding to two choices of trivializations is always an integer, thus leading to a well-defined Chern-Simons invariant in $\mathbb{R}/\mathbb{Z}$. In particular, this invariant doesn't change under gauge transformations.

2.5.5 The classical Chern-Simons field theory

Definition 2.105. Let G be a compact connected simple Lie group, and let Q be the unique G-invariant inner product on its Lie algebra $\mathfrak{g}$ determined by the convention of Definition 2.103. Let M be a closed 3-manifolds. The space of fields of the *classical Chern-Simons field theory* determined by the data G, M consists of the connections on the trivial G-bundle $P = M \times G \to M$. Its action functional is the map

$$\mathcal{M} := \{\text{connections on } P = M \times G \to M\} \longrightarrow \mathbb{R} \qquad \text{given by} \qquad \omega \mapsto \mathrm{CS}_Q(\omega).$$

Homework 2.106. Show that a connection ω is a critical point of the Chern-Simons functional if and only if it is flat (that is, its curvature vanishes).

2.6 Non-linear σ-models

2.7 Summarizing classical field theories

After presenting examples of classical field theories, we are now ready to refine our preliminary Definition 2.1 of classical field theories.

Definition 2.107. A *classical field theory* is given by the following data:

- A finite dimensional oriented manifold M of dimension d called *space-time*.
- A finite dimensional fiber bundle $\mathcal{F} \to M$. Its space of sections $\Gamma(M, \mathcal{F})$ is the *space of fields*.
- A map $L\colon \Gamma(M, \mathcal{F}) \longrightarrow \Gamma(M, \Lambda^d T^*M) = \Omega^d(M)$ of spaces of sections called the *Lagrangian*. This map is required to be given by a differential operator, which means that for a section $\phi\colon M \to \mathcal{F}$ the value of $L(\phi) \in \Omega^d(M)$ at $x \in M$ depends only on ϕ and its derivatives *at the point* x. In particular, L is *local* in the sense that $L(\phi)(x) \in \Lambda^d T_x^* M$ depends only on the restriction of the field ϕ to a neighborhood of x. The Lagrangian L determines the map

$$\mathcal{S}\colon \Gamma(M, \mathcal{F}) \longrightarrow \mathbb{R} \qquad \text{by} \qquad \mathcal{S}(\phi) := \int_M L(\phi)$$

called the *action functional*.

Remark 2.108. In the above definition we assume that that space time manifold M is oriented so that integrating $L(\phi) \in \Omega^d(M)$ over M makes sense. A better, more general procedure is to define the *density line bundle* $\mathrm{Dens}_M \to M$ and to note that the change of variables formula for integration over open subsets in $\mathbb{R}^d$ implies that sections of Dens_M with compact support can be integrated over M. Then the range of the Lagrangian L should be modified by replacing $\Gamma(M, \Delta^d T^*M) = \Omega^d(M)$ by $\Gamma(M, \mathrm{Dens}_M)$. We these modifications, the assumption that M is oriented is superfluous.

Next we summarize the classical field theories we have discussed earlier in the following table. In each case the construction of the fiber bundle $\mathcal{F} \to M$ or the construction of the Lagrangian might require some additional structure on M which we indicate in one of the columns.

classical field theories	d	structure on space time manifold M	space of fields	Lagrangian
classical mechanics	1	Riemann metric g	$C^\infty(M,X) = \Gamma(M, M\times X)$	$\|\|d\phi\|\|^2 \operatorname{vol}_g$
non-linear σ-model	any	Riemann metric g	$C^\infty(M,X) = \Gamma(M, M\times X)$	$\|\|d\phi\|\|^2 \operatorname{vol}_g$
electro magnetism	4	Lorentz metric g & Hermitian line bundle L	$\left\{\begin{matrix}\text{Hermitian}\\ \text{connections } A \text{ on } L\end{matrix}\right\} \cong \Omega^1(M)$	$\|\|R\|\|^2 \operatorname{vol}_M$, $R = \frac{dA}{-2\pi i}$
Yang-Mills Theory	4	pseudo metric g & principal G-bundle $P\to M$	$\left\{\begin{matrix}\text{connections}\\ \omega \text{ on } P\end{matrix}\right\} \cong \Omega^1(M;\operatorname{ad}(P))$	$\|\|\Omega\|\|^2 \operatorname{vol}_g$, $\Omega = d\omega + \frac{1}{2}[\omega\wedge\omega]$
gravitation	any	none	$\left\{\begin{matrix}\text{pseudo metrics}\\ g \text{ on } M\end{matrix}\right\} \subset \Gamma(S^2T^*M)$	$s(g)\operatorname{vol}_g$
Chern-Simons theory	3	none	$\left\{\begin{matrix}\text{connections } \omega\\ \text{on } P = M\times G\end{matrix}\right\} = \Omega^1(M;\mathfrak{g})$	$\mathrm{CS}_Q(\omega,\omega^{\mathrm{taut}})$, $Q\colon \mathfrak{g}^{\otimes 2}\to\mathbb{C}$

3 Functorial field theories

Now we leave classical field theories to talk about "quantum field theories" and thus enter a much murkier realm. There are mathematical definitions of quantum field theories, in fact, quite a few, but finding the "correct" definition is very much an important 21st Century challenge for mathematicians.

Ideally, one would like to have a recipe called "quantization" which produces structures physicists refer to as quantum field theories that

3.1 From classical mechanics to quantum mechanics: the Feynman-Kac formula

In this section we want to look a simple physical system, namely a point particle moving in a manifold X equipped with a Riemannian metric g. We want to contrast the *classical* description with the *quantum mechanical* description of this system.

	classical mechanics	quantum mechanics
physical state of the system	a point $(x, v) \in TX$	a unit vector $\psi \in L^2(X, \mathbb{C})$
time evolution of the system	geodesic flow	$\psi_t = e^{-i\Delta t/\hbar}\psi_0$

Classically, the state of the system at a fixed time is determined by the position $x \in M$ and the velocity $v \in T_xX$ of the particle. The Riemannian metric on X determines an $\mathbb{R}$-action on TX known as *geodesic flow*. If $(x_0, v_0) \in TX$ is the physical state at time 0, the physical state $(x_t, v_t) \in TX$ is obtained by applying $t \in \mathbb{R}$ to (x_0, v_0). More explicitly, the geodesic flow is the $\mathbb{R}$-action on TX, which via the bijection

$$\{\text{geodesics } \phi\colon \mathbb{R} \to X\} \longrightarrow TX \qquad \phi \mapsto (\phi(0), \dot{\phi}(0))$$

corresponds to evident action of $t \in \mathbb{R}$ on geodesics by $(t\phi)(s) := \phi(t + s)$.

Quantum mechanically, the physical state of the system is characterized by the square integrable function $\psi\colon X \to \mathbb{C}$ physicists call a *wave function*. It is notoriously difficult to get an intuitive feeling for what ψ means. By contrast, it is easy to understand the meaning of the n-form $|\psi|^2 \operatorname{vol}_g$. We note that the condition that ψ is a unit vector in $L^2(X, \mathbb{C})$ implies that $|\psi|^2 \operatorname{vol}_g$ integrated over X yields 1. In other words, it can be interpreted as a probability measure on X. Evaluating this measure on an open subset $U \subset X$ is interpreted as the probability of finding the particle in U. The time evolution of the system is determined by the unitary operator $e^{-i\Delta t/\hbar}$: applying this operator to the wave function ψ_0 giving the physical state of the system at time 0 yields the wave function ψ_t giving the physical state at time t. Here Δ is the (positive definite) Laplace operator on X, and $\hbar$ is a constant that we will set equal to 1 from now on.

This short discussion hopefully conveys the feeling that the classical and the quantum mechanical description of the same physical system are very different from each other, and hence the process of "quantization" is subtle construction. It is in particular mysterious how the operator $e^{-i\Delta t}$, which describes time evolution quantum mechanically can be related to the *classical field theory* whose space of fields is the space of all paths $\phi\colon \mathbb{R} \to X$ and whose action functional is the energy functional.

Fortunately, there is a close relationship between the energy functional on paths and the operator $e^{-\Delta t}$. Formally, it is related to the unitary operator $e^{-i\Delta t}$ by replacing t by it, i.e., by passing to "imaginary time" which goes by the name of *Wick rotation*. The idea is that knowing the operator $e^{-\Delta t}$ should provide a lot of information about the operator $e^{-i\Delta t}$ by holomorphic continuation.

The operator $e^{-\Delta t}$ is known as *heat operator*, since if a function $f\colon X \to \mathbb{R}$ describes the heat distribution at time $t = 0$, then $f_t = e^{-\Delta t}f$ describes the heat distribution at time t. The heat operator is described in terms of paths and their energy by the following formula, known as the *Feynman-Kac formula*:

$$(e^{-\Delta t}f)(x) = \int_{\{\phi\colon [0,t]\to X|\phi(t)=x\}} f(\phi(0))\frac{e^{-\mathcal{S}(\phi)}\mathcal{D}\phi}{Z}. \tag{3.1}$$

Here

- $\mathcal{S}(\phi) = \int_0^t ||\dot{\phi}(s)||^2 ds$ is the energy of the path ϕ;
- $\mathcal{D}$ is the "volume form" on the infinite-dimensional Riemannian path manifold $W_x := \{\phi\colon [0,t] \to X \mid \phi(t) = x\}$;
- Z is a normalization constant which makes $\frac{e^{-\mathcal{S}(\phi)}\mathcal{D}\phi}{Z}$ a *probability measure* on the based path space W_x.

At this point we need to admit that this is only a *heuristic* discussion: while a Riemannian metric g on a finite dimensional oriented manifold X determines a volume form vol_g which can be interpreted as a measure on X, this is *not true on* ∞*-dimensional Riemannian manifolds* (for example, the standard Riemannian metric on $\mathbb{R}^n$ leads to the Lebesgue measure on $\mathbb{R}^n$, but there is no Lebesgue measure on $\mathbb{R}^\infty$!). Fortunately, the heuristic expression

$$\frac{e^{-\mathcal{S}(\phi)}\mathcal{D}\phi}{Z} \tag{3.2}$$

for a probability measure on W_x can be made rigorous as a probability measure known as *Wiener measure*. A beautiful discussion of this can be found in the paper *Finite-dimensional approximations to Wiener measure and path integral formulas on manifolds* by Andersson and Driver [AD]. They approximate the based path space W_x by spaces W_x^P of piecewise geodesics $\phi\colon [0,t] \to X$, for which jumps in the first derivative occur at a finite subset $P \subset [0,t]$. These are *finite dimensional* Riemannian manifolds for which the expression (3.2) indeed defines a probability measure on W_x^P. Their main result is that these probability measure converge to a probability measure on W_x (which needs to be much bigger than one might have suspected initially: it consists of all *continuous* paths; the subspace of differentiable paths has measure zero for the Wiener measure!).

Physical interpretation of the Feynman-Kac formula (3.1)**.** As mentioned above, if we think of the function $f(y)$ as describing the temperature at $y \in X$ at time 0, the left-hand side of the Feynman-Kac formula is the temperature at time t at the point x. This is affected by the temperature $f(y)$ at time 0 for any point $y \in X$. Physically, heat is propagated via chains of molecules bumping into each other. In particular, the temperature at time 0 at y influences the temperature at time t at x via chains of molecules forming paths connecting y and x. That leads to the idea that the temperature at time t at x should be the *weighted average* of the temperature $f(\phi(0))$ at time 0 at the starting point $\phi(0)$ over all paths $\phi\colon [0,t] \to X$ with endpoint x. Of course, shorter chains of bumping molecules are more likely than longer once; this makes the factor $e^{-\mathcal{S}(\phi)}$ plausible: paths of high energy lead to a small factor $e^{-\mathcal{S}(\phi)}$, and hence those paths contribute less to the weighted average provided by the right hand side of the Feynman-Kac formula.

3.2 Heuristic quantization of the non-linear σ-model

We recall from section 2.6 that the d-dimensional non-linear σ-model with target manifold X is a generalization of classical mechanics of a point particle moving in X. In fact, for $d = 1$ those two agree. In this section we attempt to "quantize" the d-dimensional non-linear σ-model by generalizing quantum mechanics to the situation at hand. There are two ingredients to quantum mechanics:

- The vector space $L^2(X)$;
- The operator $L^2(X) \to L^2(X)$ associated to the interval $[0,t]$ via the Feynman-Kac formula.

What are the analogous vector spaces resp. operators for $d > 1$?

Vector spaces The idea is to replace the one-point space pt for $d = 1$ by an arbitrary closed Riemannian manifold Y of dimension $d-1$, and to consider the space $\mathrm{map}(Y,X)$ of continuous maps $\phi\colon Y \to X$. For $Y = \mathrm{pt}$, this mapping space can be identified with X, but for example for $Y = S^1$, we obtain the loop space of X. Generalizing what we did in quantum mechanics, we take a suitable *space of functions* $E(Y) := \mathcal{F}(\mathrm{map}(Y,X))$ on the mapping space.

Operators The idea is to think of the interval $[0,t]$ as a 1-dimensional bordism from pt to pt equipped with a Riemannian metric. The analogous d-dimensional object is a d-dimensional bordism M from Y_0 to Y_1 equipped with a Riemannian metric. Imitating the Feyman-Kac formula (3.1) we try to produce an operator (= linear map)

$$E(M)\colon E(Y_0) \longrightarrow E(Y_1)$$

by defining $(E(M)f)(\phi_1) \in E(Y_1) = \mathcal{F}(\mathrm{map}(Y_1,X))$ for $f \in E(Y_0) = \mathcal{F}(\mathrm{map}(Y_0,X))$, $\phi_1 \in \mathrm{map}(Y_1,X)$ by

$$(E(M)f)(\phi_1) = \int_{\{\phi\colon M\to X | \phi_{|Y_1}=\phi_1\}} f(\phi_{|Y_0}) \frac{e^{-\mathcal{S}(\phi)}\mathcal{D}\phi}{Z} \tag{3.3}$$

Comments as to why this doesn't work for $d > 1$.

At this point, as a mathematician faced with the difficulties of constructing the vector spaces and operators needed for the quantization of the classical non-linear σ-model, you can

- give up, or
- axiomatize the expected/desired properties of the vector spaces and operators you strive to construct.

The second option has the advantage that somebody more clever than you or with an original new point of view or a creative new construction might still succeed where you failed. Also, in the end there might be more than one fruitful way to approach this construction, and in any of these scenarios an axiomatization is useful.

Expected properties:

1. Concerning the topological vector space $E(Y)$ associated to a closed $(d-1)$-manifold Y, let us assume that Y is the disjoint union $Y_1 \amalg Y_2$ of two closed $(d-1)$-manifolds Y_1, Y_2. Then the space $\mathrm{map}(Y_1 \amalg Y_2, X)$ of maps to X can be identified with the Cartesian product $\mathrm{map}(Y_1, X) \times \mathrm{map}(Y_2, X)$. We expect to have an isomorphism of topological vector spaces
$$\mathcal{F}(\mathrm{map}(Y_1, X) \times \mathrm{map}(Y_2, X)) \cong \mathcal{F}(\mathrm{map}(Y_1, X)) \otimes \mathcal{F}(\mathrm{map}(Y_2, X)) \tag{3.4}$$
for a suitable version of function spaces $\mathcal{F}(\ )$ on these mapping spaces and an appropriate notion of "tensor product" of topological vector spaces. For example:
   (a) for measure spaces Ω, Ω' the Hilbert space $L^2(\Omega \times \Omega')$ of square-integrable functions on the Cartesian product $\Omega \times \Omega'$ is isomorphic to the Hilbert space tensor product $L^2(\Omega) \otimes L^2(\Omega')$;
   (b) for manifolds M, M' the topological vector space $C^\infty(M \times M')$ of smooth functions on the product manifold $M \times M'$ equipped with the standard Frechet topology on the space of smooth functions is isomorphic to the *projective tensor product* $C^\infty(M) \otimes C^\infty(M')$, a suitable completion of the algebraic tensor product.

2. Let M^1 be a bordism from Y_0 to Y_1, let M^2 be a bordism from Y_1 to Y_2, and let $M = M^2 \cup_{Y_1} M^1$ be the bordism from Y_0 to Y_2 obtained by gluing the bordisms M^1 and M^2 along their common boundary Y_1. We claim that if the measure $e^{-\mathcal{S}(\Phi)}\mathcal{D}\Phi$ made sense, and we consequently could define the operators $E(M^1)$, $E(M^2)$, and $E(M)$, then
$$E(M^2 \cup_{Y_1} M^1) = E(M^2) \circ E(M^1) \colon E(Y_0) \to E(Y_2).$$
Here is an outline of the argument. For $f \in E(Y_0)$, and $\phi_2 \in \mathrm{map}(Y_2, X)$ we calculate:
$$\begin{aligned} &E(M^2)(E(M^1)f_0)(\phi_2) \\ &= \int_{\{\Phi^2\colon M^2 \to X \mid \Phi^2_{|Y_2} = \phi_2\}} e^{-\mathcal{S}(\Phi^2)}\mathcal{D}\Phi^2 (E(M^1)f_0)(\Phi_{|Y_1}) \\ &= \int_{\{\Phi^2\colon M^2 \to X \mid \Phi^2_{|Y_2} = \phi_2\}} e^{-\mathcal{S}(\Phi^2)}\mathcal{D}\Phi^2 \int_{\{\Phi^1\colon M^1 \to X \mid \Phi^1_{|Y_1} = \Phi^2_{|Y_1}\}} e^{-\mathcal{S}(\Phi^1)}\mathcal{D}\Phi^1 f_0(\Phi^1_{|Y_0}) \end{aligned}$$

The space of pairs (Φ^1, Φ^2) of maps $\Phi^i \colon M^i \to X$ with $\Phi^1_{|Y_1} = \Phi^2_{|Y_1}$ can be identified with the space of maps $\Phi\colon M \to X$. In particular, the restriction map

$$\{\Phi\colon M \to X \mid \Phi_{Y_2} = \phi_2\} \longrightarrow \{\Phi_2\colon M_2 \to X \mid (\Phi_2)_{|Y_2} = \phi_2\} \qquad \Phi \mapsto \Phi_{|M_2}$$

is a fiber bundle, whose fiber over a fixed $\Phi_2\colon M_2 \to X$ can be identified with the space

$$\{\Phi_1\colon M_1 \to X \mid (\Phi_1)_{|Y_1} = (\Phi_2)_{|Y_1}\}.$$

In particular, the above double integral can be interpreted as an integral over the fibers of this vector bundle followed by an integral over its base space. For the heuristic "volume forms" $\mathcal{D}\Phi^1$ on the fiber, and $\mathcal{D}\Phi^2$ on the base, we expect that their product is the "volume form" corresponding on Riemannian metric on the total space. Morever, since the energy of a map $\Phi\colon M \to X$ is the sum $\mathcal{S}(\Phi^1) + \mathcal{S}(\Phi^2)$ of the energy of its restrictions $\Phi^i := \Phi_{|M^i}$, the "measure" $e^{-\mathcal{S}(\Phi^2)}\mathcal{D}\Phi^2 e^{-\mathcal{S}(\Phi^1)}\mathcal{D}\Phi^1$ should be equal to $e^{-\mathcal{S}(\Phi)}\mathcal{D}\Phi$. Hence the double integral above is equal to

$$\int_{\{\Phi\colon M\to X \mid \Phi_{|Y_2}=\phi_2\}} e^{\mathcal{S}(\Phi)}\mathcal{D}\Phi f_0(\Phi_{|Y_0} = (E(M)f_0)(\phi_2),$$

thus "proving" the claim.

3.3 Definition of a topological quantum field theory

This section represents our first stab at the definition of a "quantum field theory". Generally, a quantum field theory should be thought of as the mathematical structure obtained by "quantizing" a classical field theory. Based on the heuristic discussion of quantization of the non-linear σ-model in the last section will be our guide, we expect that the "quantization of the d-dimensional non-linear σ-model with target manifold X" yields the following data:

1. For each closed $(d-1)$-manifold Y a topological vector space $E(Y)$.

2. For each bordism M from Y_0 to Y_1 a continuous linear map $E(M)\colon E(Y_0) \to E(Y_1)$.

3. For closed $(d-1)$-manifolds Y_0, Y_1 an isomorphism $E(Y_0 \amalg Y_1) \cong E(Y_0) \otimes E(Y_1)$.

Moreover, the operators associated to bordisms should have the "composition property" that if M^1 is a bordism from Y_0 to Y_1 and M^2 is a bordism from Y_1 to Y_2, then

$$E(M^2 \cup_{Y_1} M^1) = E(M^2) \circ E(M^1). \tag{3.5}$$

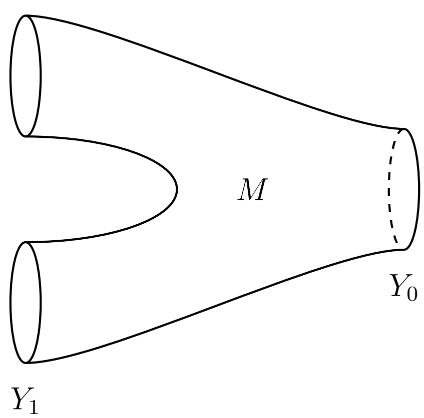
We recognize that the data (1)&(2) plus equation (3.5) can be interpreted as a *functor* from a from a 'bordism category' whose objects are closed $(d-1)$-manifolds, and whose morphism are d-dimensional bordisms to the category TV of topological vector spaces and continuous linear maps.

Remark 3.6. A word of warning: often at category is named after its *objects* rather than its morphisms. For example, we talk about the 'category of vector spaces' rather than the 'category of linear maps'. This is fine as long as the morphisms can be guessed from the description of the objects. Unfortunately, this is not the case with more involved categories like the bordism category. Being told that the objects of this category are closed $(d-1)$-manifolds, the natural guess would be that the morphisms between them are diffeomorphisms or smooth maps rather than bordisms. For this reason some categories, like the bordism category are named after their *morphisms*.

Definition 3.7. The *d-dimensional bordism category*, denoted $d\text{-}\mathsf{Bord}$, is the category defined as follows.

Objects: An object Y of $d\text{-}\mathsf{Bord}$ is a closed manifold of dimension $d-1$.

Morphisms: A morphism from Y_0 to Y_1 is a bordism from Y_0 to Y_1; that is, a d-manifold M equipped with a diffeomorphism $Y_1 \amalg Y_0 \cong \partial M$ between the disjoint union of Y_1 and Y_0 and the boundary of M. We write $i_j\colon Y_j \to \partial M \subset M$ for the restriction of the diffeomorphism to $Y_j \subset Y_1 \amalg Y_0$. Here is a picture of such a bordism for $d = 2$:



More precisely, a morphism from Y_0 to Y_1 is an equivalence class of bordisms from Y_0 to Y_1, where we declare two such bordisms M, M' as equivalent if they are *diffeomorphic relative boundary*; that is, if there is a diffeomorphism $f\colon M \to M'$ making the diagram

$$\begin{tikzcd} & M \arrow[dd, "f"] & \\ Y_1 \arrow[ur, "i_1"] \arrow[dr, "i'_1"'] & & Y_0 \arrow[ul, "i_0"'] \arrow[dl, "i'_0"] \\ & M' & \end{tikzcd}$$

commutative.

Composition: If M^1 is a bordism from Y_0 to Y_1 and M^2 is a bordism from Y_1 to Y_2, the composition $M^2 \circ M^1$ of the two morphisms represented by M^2 and M^1 is given by the bordism $M^2 \cup_{Y_1} M^1$ obtained from the disjoint union of M^2 and M^1 by identifying the part of the boundary of M^2 corresponding to points of Y_1 with the corresponding part of the boundary of M^1.

Identities: If Y is a closed $(d-1)$-manifold, then the product $Y \times [a, b]$ with some interval $[a, b]$ is a bordism from Y to itself. If M is any bordism from Y to some other closed $(d-1)$-manifold Y', then using the tubular neighborhood theorem, it is easy to show that $M \cup_Y Y \times [a, b]$ is diffeomorphic to M relative boundary. This plus the analogous statement for bordisms with target Y shows that $Y \times [a, b]$ is in fact the identity morphism of Y.

Remark 3.8. There are a number of reasons why the morphisms of the bordism category d-Bord should be defined as *equivalence classes* of bordisms.

- The equivalence classes of bordisms from Y_0 to Y_1 form a *set*, but the bordisms from Y_0 to Y_1 only give a *class*.
- If M^i is a bordism from Y_{i-1} to Y_i for $i = 1, 2, 3$, then the bordisms

  $$(M^3 \cup_{Y_2} M^2) \cup_{Y_1} M^1 \qquad \text{and} \qquad M^3 \cup_{Y_2} (M^2 \cup_{Y_1} M^1)$$

  are canonically homeomorphic, but *not identical*, since the first is a quotient space of the disjoint union $(M^3 \amalg M^2) \amalg M^1$, while the second is a quotient of $M^3 \amalg (M^2 \amalg M^1)$, and the operation disjoint union is not strictly associative, it's only associative up to canonical homeomorphism.
- It's even worse if we take smooth structures seriously and start asking about how to equip the topological manifold $M^2 \cup_{Y_1} M^1$ with a *smooth structure*. This can be done, but it requires the choice of a collar neighborhood of Y_1 in M^1 and M^2. Different choices typically lead to *different smooth structures* on the topological manifold $M^2 \cup_{Y_1} M^1$ (i.e., the transition maps between charts constructed using different choices of collars typically aren't smooth). Fortunately, these smooth manifolds are always *diffeomorphic* via a homeomorphism from $M^2 \cup_{Y_1} M^1$ to itself which can be chosen to be the identity outside a small neighborhood of $Y_1 \subset M^2 \cup_{Y_1} M^1$.
- If M is a bordism from Y_0 to Y_1, and we consider the product $Y_0 \times [a, b]$ as a bordism from Y_0 to itself, then

  $$M \cup_{Y_0} (Y_0 \times [a, b]) \qquad \text{is diffeomorphic to, but not equal to} \qquad M \tag{3.9}$$

  relative the boundary $Y_1 \amalg Y_0$.

Summarizing what we have done so far, we can say that the hoped for data (1)&(2) of a quantization plus equation (3.5) can be expressed concisely by saying that we have a functor

$$E\colon d\text{-}\mathsf{Bord} \longrightarrow \mathsf{TV}$$

from the d-dimensional bordism category to the category TV of topological vector spaces and continuous linear maps.

How can the third datum, the isomorphism

$$E(Y_0 \amalg Y_1) \cong E(Y_0) \otimes E(Y_1)$$

be phrased in categorical terms? The disjoint union is an operation allowing us to produce from two objects Y_0, Y_1 of the bordism category d-Bord a new object $Y_0 \amalg Y_1$. Similarly, the tensor product on the right hand side produces from two topological vector spaces a new topological vector space by tensoring them (for the definition of this tensor product see Definition **??**). We can also take the disjoint union of bordisms, giving us a way to product a new morphism in d-Bord from two given morphisms; similarly, we can tensor two continuous linear maps $T_1\colon V_1 \to W_1$ and $T_1\colon V_1 \to W_1$ to get a continuous linear map $T_1 \otimes T_2\colon V_1 \otimes V_2 \to W_1 \otimes W_2$. Such a structure on a category C is called a *monoidal structure*.

Before giving the precise definition we would like to comment. It is straightforward to implement a way to combine two objects (resp. morphisms) of a category C to a new one: we simply postulate a functor

$$\otimes\colon \mathsf{C} \times \mathsf{C} \longrightarrow \mathsf{C}.$$

The objects of the product category $\mathsf{C} \times \mathsf{C}$ are pairs (V, W) of objects of C; its morphisms are pairs of morphisms $(f\colon V \to V', g\colon W \to W')$. The image of the object (V, W) under the functor $\otimes$ is written $V \otimes W$; the image of the morphism (f, g) is denoted $f \otimes g\colon V \otimes W \to V' \otimes W'$, and hence applying the functor $\otimes$ to objects (resp. morphisms) gives us the desired way to produce a new object (resp. morphism) from two given ones.

The tricky part comes when it comes to requiring properties for the functor $\otimes$; for example, we want an associativity condition. It is tempting to require

$$(U \otimes V) \otimes W = U \otimes (V \otimes W)$$

for any objects $U, V, W \in \mathsf{C}$. However, this is not strictly speaking true in the cases we are interested in; for example, for sets U, V, W there is an obvious bijection between $(U \amalg V) \amalg W$ and $U \amalg (V \amalg W)$, but these two sets are not the *same set*. Let's be more detailed to make this point clear. Let us define the disjoint union $V \amalg W$ of two sets to be the subset of $\{1, 2\} \times \{V \cup W\}$ consisting of the elements $(1, v)$ for $v \in V$ and $(2, w)$ for $w \in W$. Then an element of $v \in V$ viewed as an element of $(U \otimes V) \otimes W$ is $(1, (2, v))$, whereas as an element of $U \amalg (V \amalg W)$ it is $(2, (1, v))$. Does this seem picky? Yes, it is; it is in fact the

> **Nitpicking Principle of Category Theory:** Never identify objects in a category with each other; rather, specify an isomorphism between them.

The drawback of this strategy is pretty clear – you carry a heavy bag of isomorphisms with you at all times. What is the advantage of doing this? The answer is that keeping track of the various isomorphisms can uncover interesting information that otherwise would be lost. As a simple example let us consider the *fundamental groupoid* ΠX of a topological space X. This is a groupoid (= category whose morphisms are all invertible) whose objects are the points of X; the morphisms from a point $x \in X$ to a point $y \in X$ consist of paths from x to y; more precisely, a morphism is an equivalence class of such paths, where two paths are identified if there is a homotopy between them that leaves the endpoints fixed (the inverse morphism is given by running this path backwards). The first question about two points $x, y \in X$ is whether they are isomorphic in ΠX, which is equivalent to them being in the same path connected component of X. If they are, finer questions can be asked: imagine that there are two different constructions that lead to isomorphisms $f, g\colon x \xrightarrow{\cong} y$ in ΠX. Then we can ask whether the composition $f^{-1} \circ g\colon x \to x$ is the identity morphism of ΠX; or, formulated geometrically, whether the loop formed by concatenating these paths represents a non-trivial element in the fundamental group of X.

Some reader might object that we violated the Nitpicking Principle of Category Theory in the example directly following it by identifying paths from x to y which are homotopic; rather we should keep track of the homotopy used! This is a very valid point; we recovered the fundamental group of X in addition to the set $\pi_0 X$ of path connected components of X by keeping the paths, but we threw away the information above the second homotopy group $\pi_2(X)$ by looking just at homotopy classes of paths relative endpoints. Keeping the homotopies (up to homotopy between these homotopies) we would recover $\pi_2(X)$. However, we would have to leave the world of categories to accommodate homotopies between paths; these homotopies can be regarded as 2-morphisms in a 2-category. Continuing along this route it is possible to capture $\pi_i(X)$ for $0 \le i \le n$ in n-category associated to X (called the *fundamental n-groupoid of X*). Higher categories have become quite popular in topology, in particular after Lurie's proof of the Baez-Dolan Cobordism Hypothesis [Lu].

Definition 3.10. A *monoidal structure* on a category C consists of the following data

- a functor
$$\otimes\colon \mathsf{C} \times \mathsf{C} \longrightarrow \mathsf{C}.$$
- An object $I \in \mathsf{C}$ called the *unit object*.
- Isomorphisms $\alpha_{U,V,W}\colon (U \otimes V) \otimes W \cong U \otimes (V \otimes W)$ called *associators* for objects $U, V, W \in \mathsf{C}$ which are natural in U, V, W.
- Natural isomorphisms $\lambda_V\colon I \otimes V \cong V$ and $\rho_V \cong V \otimes I \cong V$ called *left unit isomorphism* and *right unit isomorphism*, respectively.

Saying that the family of isomorphisms $\alpha_{U,V,W}$ is *natural* in U means that for any morphism $f\colon U \to U'$ the diagram

$$\begin{tikzcd}
(U \otimes V) \otimes W \arrow[r, "\alpha_{U,V,W}"] \arrow[d, "(f\otimes 1)\otimes 1"'] & U \otimes (V \otimes W) \arrow[d, "f\otimes(1\otimes 1)"] \\
(U' \otimes V) \otimes W \arrow[r, "\alpha_{U',V,W}"] & U' \otimes (V \otimes W)
\end{tikzcd}$$

is commutative. The naturality of $\alpha_{U,V,W}$ in V, W, and the naturality of λ_V, ρ_V in V is defined analogously. There are *coherence conditions* for the associators and the unit isomorphisms:

- For all $U, V, W, X \in \mathsf{C}$, the pentagon diagram

$$\begin{tikzcd}
((U \otimes V) \otimes W) \otimes X \arrow[r, "\alpha_{U,V,W}\otimes 1"] \arrow[d, "\alpha_{U\otimes V,W,X}"'] & (U \otimes (V \otimes W)) \otimes X \arrow[r, "\alpha_{U,V\otimes W,Z}"] & U \otimes ((V \otimes W) \otimes X) \arrow[d, "1\otimes\alpha_{V,W,X}"] \\
(U \otimes V) \otimes (W \otimes X) \arrow[rr, "\alpha_{U,V,W\otimes X}"] & & U \otimes (V \otimes (W \otimes X))
\end{tikzcd}$$

  commutes;

- For all $V, W \in \mathsf{C}$ the triangle diagram

$$\begin{tikzcd}
(V \otimes I) \otimes W \arrow[rr, "\alpha_{V,I,W}"] \arrow[dr, "\rho_V\otimes 1"'] & & V \otimes (I \otimes W) \arrow[dl, "1\otimes\lambda_W"] \\
& V \otimes W &
\end{tikzcd}$$

  commutes.

Good references for these these definitions are the book *Categories for the working mathematician* by Mac Lane [McL] or the paper *Braided tensor categories* by Joyal and Street [JS]. We note that the commutativity of the pentagon and triangle diagram in the definition imply that *every diagram* that can be build from the associators, and the left- and right unit isomorphisms is commutative. For example, the diagram

$$\begin{tikzcd}
I \otimes I \arrow[r, bend left, "\lambda_I"] \arrow[r, bend right, "\rho_I"'] & I
\end{tikzcd} \tag{3.11}$$

is commutative. This statement is known as *Mac Lane's Coherence Theorem*. It is interesting to note that the observation that the commutativity of the pentagon and triangle diagram imply the commutativity of (3.11) is due to Kelly; a proof can also be found in the very nice paper by Joyal and Street [JS, Prop. 1.1]. In Mac Lane's book [McL, Ch. XI.1] he includes diagram (3.11) as a requirement in the definition of a monoidal category.

The monoidal structure given by the tensor product of vector spaces, or by the disjoint union of sets or topological spaces has an additional property, namely the existence of a natural family of isomorphisms $\beta_{V,W} \colon V \otimes W \cong W \otimes V$.

Definition 3.12. A *symmetric monoidal category* C is a monoidal category $(\mathsf{C}, \otimes, I)$ equipped with a natural family of isomorphisms

$$\beta_{V,W} \colon V \otimes W \cong W \otimes V.$$

These are required to satisfy the following coherence properties:

- For all V, W the composition

$$V \otimes W \xrightarrow{\beta_{V,W}} W \otimes V \xrightarrow{\beta_{W,V}} V \otimes W \qquad \text{is the identity;} \tag{3.13}$$

- For all $U, V, W \in \mathsf{C}$ the following hexagon is commutative:

$$\begin{tikzcd}
& (V \otimes U) \otimes W \arrow[r, "\alpha_{V,U,W}"] & V \otimes (U \otimes W) \arrow[dr, "1 \otimes \beta_{U,W}"] & \\
(U \otimes V) \otimes W \arrow[ur, "\beta_{U,V} \otimes 1"] \arrow[dr, "\alpha_{U,V,W}"'] & & & V \otimes (W \otimes U) \\
& U \otimes (V \otimes W) \arrow[r, "\beta_{U,V \otimes W}"'] & W \otimes (U \otimes V) \arrow[ur, "\alpha_{V,W,U}"'] &
\end{tikzcd} \tag{3.14}$$

While the definition of a symmetric monoidal category is quite sophisticated, we are all very familiar with many examples of symmetric monoidal categories:

- the category of sets equipped with monoidal structure given by disjoint union or Cartesian product;
- the category of groups with monoidal structure given by the free product or the Cartesian product;
- the category of vector spaces with monoidal structure given by the direct sum or tensor product.

Remark 3.15. There are very interesting examples of monoidal categories equipped with a natural family of isomorphisms $\beta_{V,W}\colon V \otimes W \cong W \otimes V$ which *do not* satisfy the symmetry property (3.13), but which make the hexagon diagram (3.14), as well as its mirror diagram commutative. Such categories are called *braided monoidal categories* [JS, Def. 2.2] (the "mirror diagram" I'm referring to is diagram (B2) in [JS]; the hexagon diagram (3.14) is diagram (B1)). Examples include the category of *positive energy representations of loop groups*, also known as *modules over Kac-Moody algebras*, equipped with a monoidal structure which goes under the name *fusion product*. A more basic example is the *braid category*, whose morphisms are braids [JS, Example 2.1] (another category named after its morphisms!). The terminology *braided monoidal category*, as well as the convention of calling the isomorphisms $\beta_{V,W}$ *braiding isomorphisms* is motivated by this example.

Example 3.16.

1. The bordism category d-Bord;
2. The category TV

Definition 3.17. Symmetric monoidal functor

Definition 3.18. (Atiyah, Segal, Kontsevich) A *d-dimensional topological (quantum) field theory*, or d-TFT, is a symmetric monoidal functor

$$E\colon d\text{-}\mathsf{Bord} \longrightarrow \mathsf{TV}.$$

from the symmetric monoidal category d-Bord to the symmetric monoidal category TV of topological vector spaces.

We observe that a closed d-manifold M can be interpreted as a bordism from $\emptyset$ to itself. So if E is a d-dimensional topological field theory, we can apply the functor E to the morphism $M\colon \emptyset \to \emptyset$ to obtain a morphism $E(M)\colon E(\emptyset) \to E(\emptyset)$ in TV. Since $\emptyset$ is the monoidal unit of d-Bord, the object $E(\emptyset) \in \mathsf{TV}$ is isomorphic to the monoidal unit $\mathbb{C}$ of TV. In particular, the vector space $E(\emptyset)$ has dimension 1, and hence the endomorphism space $\mathsf{TV}(E(\emptyset), E(\emptyset))$ can be identified with $\mathbb{C}$. It follows that

$$E(M) \in \mathsf{TV}(E(\emptyset), E(\emptyset)) = \mathbb{C}$$

is a *numerical invariant* of the closed manifold M. We note that if M is diffeomorphic to M', then $E(M) = E(M')$, since M and M' represent the *same* morphism in d-Bord.

Homework 3.19. Let E be a topological field theory of dimension d and let M, M' be closed d-manifolds. Show that $E(M \amalg M') = E(M) \cdot E(M')$.

Homework 3.20. Let E be a topological field theory of dimension d.

1. Show that the vector space $E(Y)$ associated to a closed $(d-1)$-manifold Y comes equipped with a canonical bilinear form

$$\langle\ ,\ \rangle\colon E(Y) \times E(Y) \longrightarrow \mathbb{C}.$$

   Hint: Consider the d-manifold $Y \times [0,1]$.

2. Let M_1 and M_2 be d-manifolds with boundary Y, considered as bordisms from $\emptyset$ to Y, and let $E(M_i) \in \mathrm{Hom}(E(\emptyset), E(Y)) = \mathrm{Hom}(\mathbb{C}, E(Y)) = E(Y)$ be the algebraic data that E associates to M_i. Show that the numerical invariant $E(M_2 \cup_Y M_1) \in \mathbb{C}$ can be calculated via the "gluing formula"

$$E(M_2 \cup_Y M_1) = \langle E(M_2), E(M_1)\rangle.$$

These statements show that topological field theories of dimension d can be viewed as refinements of numerical invariants for closed d-manifolds which are exponential and satisfy a gluing formula. Here is an example of a d-TFT for d even based on the Euler characteristic.

Example 3.21. (Euler characteristic topological field theory) For a compact d-manifold M (with or without boundary) the homology groups $H_i(M;\mathbb{Z}/2)$ are finite dimensional, and vanish for $i > d$. This makes it possible to define its *Euler characteristic*

$$\chi(M) := \sum_{i=0}^{d} (-1)^i \dim H_i(M;\mathbb{Z}/2) \in \mathbb{Z}.$$

If M is closed, then $\dim H_i(M;\mathbb{Z}/2) = \dim H^{d-i}(M;\mathbb{Z}/2) = \dim H_{d-i}(M;\mathbb{Z}/2)$ by Poincaré duality and the Universal Coefficient Theorem. In particular, its Euler characteristic vanishes for d odd (the advantage of defining the Euler characteristic in terms of the $\mathbb{Z}/2$-homology groups of M is that Poincaré duality for (co)homology with $\mathbb{Z}/2$-coefficients doesn't require the assumption that M is orientable). The additivity of homology groups implies that the Euler characteristic enjoys the additivity property

$$\chi(M_1 \amalg M_2) = \chi(M_1) + \chi(M_2).$$

More generally, if M_1 and M_2 are manifolds of dimension $d \equiv 0 \mod 2$ intersect in a compact manifold Y, then, like for the cardinality of finite sets, we have

$$\chi(M_2 \cup_Y M_1) = \chi(M_2) + \chi(M_1) - \chi(Y).$$

In particular, if $M_2 \cup_Y M_1$ is the manifold obtained by gluing M_2 and M_1 along a closed codimension zero submanifold Y of their boundary, then $\dim Y$ is odd and hence $\chi(Y) = 0$.

The discussion above shows that the Euler characteristic is an *additive* invariant. To obtain an *exponential* numerical invariant $E(M)$ for closed manifolds M of dimension $d \equiv 0 \mod 2$, we simply pick a non-zero complex number $\lambda \in \mathbb{C}$ and define

$$E(M) := \lambda^{\chi(M)} \in \mathbb{C}.$$

To extend this to a topological field theory, we define the vector space $E(Y)$ to be $\mathbb{C}$ for every closed $(d-1)$-manifold Y, and for any bordism M from Y_1 to Y_2 we define

$$E(M) \in \operatorname{Hom}(E(Y_1), E(Y_2)) = \operatorname{Hom}(\mathbb{C}, \mathbb{C}) = \mathbb{C} \qquad \text{by} \qquad E(M) := \lambda^{\chi(M)}.$$

The properties of the Euler characteristic discussed above show that E is a topological field theory.

We end this section by mentioning the following important property of topological field theories.

Theorem 3.22. *Let E be a topological field theory of dimension d, and let Y be a closed manifold of dimension $d-1$. Then the associated vector space $E(Y)$ is finite dimensional.*

Since we have no additional information about either the field theory E nor the closed manifold Y, the proof of this result is necessarily of an abstract argument. It is based on a *categorical characterization* of the finite dimensionality of vector spaces. This characterization and the proof of the theorem above is provided in the following digression.

3.3.1 Digression: dualizable objects in symmetric monoidal categories

Before defining the abstract notion of dualizable objects in symmetric monoidal categories, let us look at an example of a symmetric monoidal category we are all quite familiar with, namely the category Vect_k of vector spaces over a field k and linear maps. We want to emphasize that the objects of Vect_k are *not* required to be finite dimensional. Let $V \in \mathsf{Vect}_k$ be a vector space over k, let $V^\vee := \operatorname{Hom}(V, k)$ be the dual vector space and let

$$\operatorname{ev}\colon V^\vee \otimes V \longrightarrow k$$

be the *evaluation map* defined by $\operatorname{ev}(f \otimes v) := f(v)$. If V is finite-dimensional, there is also a *coevaluation map*

$$\operatorname{coev}\colon k \longrightarrow V \otimes V^\vee,$$

which is characterized by $\operatorname{coev}(1) = \sum_i e_i \otimes e^i$, where $1 \in k$ is the unit of k, $\{e_i\}$ is a basis of V, and e^i is the dual basis of $V^\vee$. We want to emphasize that the element $\sum_i e_i \otimes e^i \in V \otimes V^\vee$ is *independent of the choice of the basis* $\{e_i\}$, but the condition $\dim V < \infty$ is required since only *finite* sums of elementary tensors $v \otimes f \in V \otimes V^\vee$ belong to $V \otimes V^\vee$.

Homework 3.23. Let V be a finite dimensional vector space over k. Show that the following compositions built from the evaluation and coevaluation map are the identities on V resp. $V^\vee$.

$$V \cong k \otimes V \xrightarrow{\text{coev} \otimes 1} V \otimes V^\vee \otimes V \xrightarrow{1 \otimes \text{ev}} V \otimes k \cong V$$

$$V^\vee \cong V^\vee \otimes k \xrightarrow{1 \otimes \text{coev}} V^\vee \otimes V \otimes V^\vee \xrightarrow{\text{ev} \otimes 1} k \otimes V^\vee \cong V^\vee$$

This example motivates the following definition.

Definition 3.24. Let $(\mathsf{C}, \otimes, I)$ be a symmetric monoidal category. An object $X \in \mathsf{C}$ is *dualizable* if there exists an object $X^\vee \in \mathsf{C}$ and morphisms

$$\text{ev}\colon X^\vee \otimes X \longrightarrow I \qquad \text{coev}\colon I \longrightarrow X \otimes X^\vee$$

called *evaluation* and *co-evaluation* maps, respectively, such that the compositions

$$X \cong I \otimes X \xrightarrow{\text{coev} \otimes 1} X \otimes X^\vee \otimes X \xrightarrow{1 \otimes \text{ev}} X \otimes I \cong X$$

$$X^\vee \cong X^\vee \otimes I \xrightarrow{1 \otimes \text{coev}} X^\vee \otimes X \otimes X^\vee \xrightarrow{\text{ev} \otimes 1} I \otimes X^\vee \cong X^\vee$$

are the identities on X and $X^\vee$, respectively.

The discussion above and Exercise 3.23 show that a finite dimensional vector space V is a dualizable object in the symmetric monoidal category Vect_k of vector spaces over a field k (with monoidal structure provided by the tensor product and monoidal unit given by k, considered as a vector space over itself). In fact, the converse holds as well, and this fact provides us with the desired categorical characterization of finite dimensional vector spaces.

Proposition 3.25. *A vector space V over a field k is finite dimensional if and only if it is a dualizable object in the symmetric monoidal category Vect_k of vector spaces over k.*

Homework 3.26. Prove this fact. Hint: Assuming that V is dualizable with categorical dual $V^\vee$ and (co)evaluation maps ev, coev, use the factorization of the identity of V to argue that V is generated by a finitely many vectors.

We want to emphasize that we should distinguish between the *vector space dual* of $V \in \mathsf{Vect}_k$ and its *categorical dual* in the sense of the definition above. The vector space dual is a particular vector space we can always construct from V without a finiteness assumption for $\dim V$, while the categorical dual exists only for $\dim V < \infty$ and it is unique only *up to isomorphism*.

Example 3.27. The classical notion of Spanier-Whitehead duality in stable homotopy theory fits in this abstract context. This is a notion of duality in the category of spectra, equipped with the monoidal structure given by the smash product $X \wedge Y$ of spectra; the monoidal unit is provided by the sphere spectrum S. We refer to the stable homotopy literature for the definition of spectra and their smash product, since we won't use these notions beyond this example, which is just included to illustrate the usefulness of this general notion of dualizability.

The dual $X^\vee$ of a dualizable spectrum X is called the *Spanier-Whitehead dual* of X. We remark that there is a natural isomorphism $H^i(X^\vee) \cong H_{-i}(X)$, and this holds more generally for any generalized cohomology theory and the associated homology theory. We remark that the minus sign in the subscript is no typo: unlike the (co)homology groups of topological spaces, the (co)homology groups of spectra can be non-zero in negative degrees. According to a classical result the suspension spectrum $\Sigma^\infty Y_+$ of a finite CW complex Y is a dualizable spectrum.

Lemma 3.28. *Every object in the symmetric monoidal category d-Bord is dualizable.*

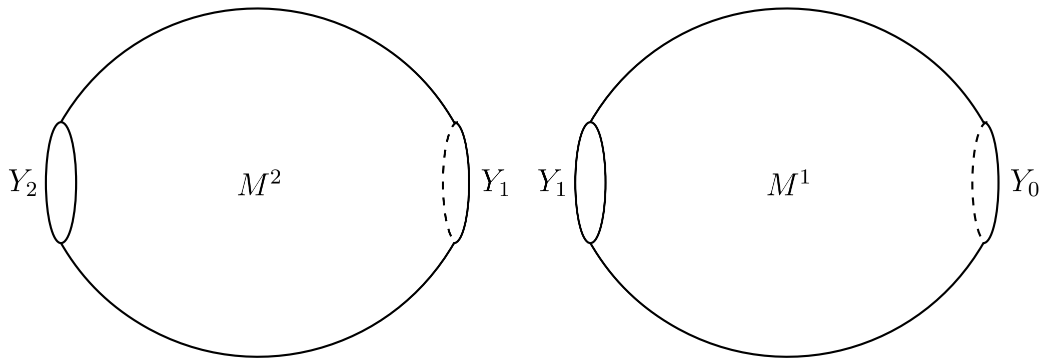
Outline of proof of Theorem 3.22. Let Y be an object of the bordism category d-Bord. Since Y is dualizable, so is its image $E(Y)$ under the symmetric monoidal functor $E\colon d\text{-}\mathsf{Bord} \to \mathsf{TV}$. Analogously to dualizability in the symmetric monoidal category Vect_k, a topological vector space $V \in \mathsf{TV}$ is dualizable if and only if it is finite dimensional. $\square$

3.4 Riemannian field theories

The goal of this section is to modify Definition **??** of d-dimensional topological field theory to include quantum mechanics for $d = 1$, and more generally, the quantizations of classical d-dimensional non-linear σ-models, discussed heuristically in section **??**, hopefully should provide examples.

From our discussion in the previous section, it is clear how to attempt to modify our definition: the classical non-linear σ-model is not a *topological* field theory in the sense that the definition of the Lagrangian of this classical field theory requires a *Riemannian metric* on the space-time manifold M. Since these space-time manifolds become the *bordisms* in the quantum field theory according to our heuristic discussion in section **??**, we should modify the definition of the bordism category d-Bord by the bordisms (that form its morphisms) as well as the closed $(d-1)$-manifolds (that are its objects) to be equipped with Riemannian metrics.

Problem. Consider the following bordisms M^1 and M^2 equipped with the Riemannian metric they inherit as submanifolds of $\mathbb{R}^3$ equipped with the standard metric.



The three boundary circles Y_0, Y_1 and Y_2 are all isometric. We consider M^1 as a bordism from Y_0 to Y_1 and M^2 as a bordism from Y_1 to Y_2. Composing these two Riemannian bordisms by gluing them along the two boundary circles labeled Y_1 does *not result in a smooth Riemannian manifold* as is evident from the picture.

A possible strategy to avoid this problem is to insist that the Riemannian metric on a bordism M is the *product metric* near the boundary of M. In other words, that there is a neighborhood of ∂M which is isometric to the product metric on $\partial M \times [0, \epsilon)$ for some ϵ. It is not hard to see that this requirement allows us to define the composition of such Riemannian bordisms by gluing them. However, this assumption is *extremely restrictive*. We recall that from one point of view a topological field theory is as a refinement of a numerical invariant for closed d-manifolds that allows the calculations of this invariant by cutting the closed manifold up into bordisms and composing the linear maps corresponding to these bordisms. Similarly, the Riemannian field theories we are aiming to define in this section can be thought of as refinements of numerical invariants for closed Riemannian manifolds. We could attempt to calculate the complex number a 2-dimensional Riemannian field theory associates to 2-sphere equipped with its standard metric by decomposing S^2 as the composition of say two discs, one viewed as a bordism from the empty set to S^1, the other one as a bordism from S^1 to the empty set. The problem is that there is no such decomposition if we insist on product metrics near the boundary: if $Y \subset S^2$ is the circle along which we cut the sphere into two pieces, then the condition that the metric is a product metric near Y forces Y to be a closed geodesic and hence Y is a great circle on S^2. However, no neighborhood of a great circle is isometric to a product, and hence the round sphere cannot be cut into bordisms which have product metrics near their boundary. Unfortunately, this is the *typical* situation for closed Riemannian manifolds.

The next idea to try is to compose Riemannian bordisms of dimension d not by gluing along their boundaries, but by "overlapping" neighborhoods of their boundary. This means that the objects of our d-dimensional Riemannian bordism category are not just closed manifolds of dimension $d-1$, but rather Riemannian d-manifolds with boundary, thought of as a *neighborhood* of the closed $(d-1)$-manifold we originally had in mind.

4 Factorization algebras

Kevin Costello and Owen Gwilliam have used the notion of 'factorization algebra' to describe observables in both, classical as well as quantum field theories, see [Gw], [CG], and this section is based on their work. We begin by defining what a *factorization algebra* is; this is done in two steps: first we define a structure called *prefactorization algebra*; a *factorization algebra* is a prefactorization algebra satisfying a *locality property*, also called *descent property*. There is a formal analogy between (pre)sheaves and (pre)factorization algebras, and so we begin by a digression on (pre)sheaves.

4.1 Strict factorization algebras

4.1.1 Digression: sheaves

Before giving the formal definition of sheaves, let us look at an example first, and then distill the formal properties we observe in the example to define what a sheaf is.

Motivating example: the sheaf of continuous functions on a topological space. To a topological space M we can associate the following data:

- To each open subset $U \subset M$, the complex vector space $C^0(U)$ of continuous functions $f\colon U \to \mathbb{C}$ (this is in fact an algebra, but we choose to ignore that additional structure, since we want to talk about sheaves of *vector spaces*). The inclusion of subsets $U \subset V$ induces a restriction map
$$r_V^U\colon C^0(V) \longrightarrow C^0(U).$$
These restriction maps are linear maps which are compatible with composition in the sense that for $U \subset V \subset W$ we have $r_U^V \circ r_V^W = r_U^W$. It is easy to formalize these data as a contravariant functor from the category of open subsets and inclusion maps to the category $\mathsf{Vect}_{\mathbb{C}}$ of complex vector spaces and linear maps. We will resist this temptation here, since the analogous statement for factorization algebras would require the more elaborate language of multicategories.

These data have the following properties.

1. If $\{U_j\}_{j\in J}$ are disjoint open subsets of M, and $U = \bigcup_{j\in J} U_j$ is their union, then the product
$$C^0(U) \xrightarrow{\cong} \prod_{j\in J} C^0(U_j) \tag{4.1}$$

of the restriction maps $r^U_{U_j}\colon C^0(U) \to C^0(U_j)$ is an isomorphism. We recall that the *direct product* $\prod_{j\in J} V_j$ of a family of vector spaces V_j indexed by $j \in J$ consists of collections of vectors $(v_j)_{j\in J}$ with $v_j \in V_j$. There are no restrictions on these collections, unlike the situation for the direct sum $\bigoplus_{j\in J} V_j$, whose elements are also collections $(v_j)_{j\in J}$ with $v_j \in V_j$, but here *only finitely many v_j's are allowed to be non-zero.*

We want to point out an extreme case of the isomorphism (4.1): if U is the empty set, the vector space $C^0(U)$ of continuous functions on U is the trivial vector space $\{0\}$. Concerning the right hand side, an open cover of U is provided by the collection of open subsets of M indexed by the empty set J. The product of vector spaces indexed by the empty set is by definition the trivial vector space.

2. The previous point shows how to express the vector space of continuous function on the union $U = \bigcup_{j\in J} U_j$ of disjoint open sets U_j as the direct product $\prod_{j\in J} C^0(U_j)$. More generally, we could ask whether if $\{U_j\}_{j\in J}$ is an open cover of U, we can express $C^0(U)$ in terms of the vector spaces $C^0(U_k)$ for $k \in J$ and $C^0(U_i \cap U_j)$ for $i, j \in J$. For any $i, j \in J$ we have a commutative diagram of inclusion maps

$$\begin{tikzcd} U & U_i \arrow[l] \\ U_j \arrow[u] & U_i \cap U_j \arrow[l] \arrow[u] \end{tikzcd}$$

and the corresponding commutative diagram of the restriction maps

$$\begin{tikzcd} C^0(U) \arrow[r, "r^U_{U_i}"] \arrow[d, "r^U_{U_j}"'] & C^0(U_i) \arrow[d, "r^{U_i}_{U_i\cap U_j}"] \\ C^0(U_j) \arrow[r, "r^{U_j}_{U_i\cap U_j}"] & C^0(U_i \cap U_j) \end{tikzcd}$$

Taking products of these maps, we obtain three maps

$$\begin{tikzcd} C^0(U) \arrow[r, "r"] & \prod_{k\in J} C^0(U_k) \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & \prod_{i,j\in J} C^0(U_i \cap U_j). \end{tikzcd}$$

More precisely, r is the product of the restriction maps $r^U_{U_k}$, and f (resp. g) is the product of the maps $r^{U_i}_{U_i\cap U_j}$ (resp. $r^{U_j}_{U_i\cap U_j}$). The commutativity of the diagram above implies that $f \circ r = g \circ r$, and hence the map r induces a linear map

$$C^0(U) \xrightarrow{r} \mathrm{eq}\left(\begin{tikzcd} \prod_{k\in J} C^0(U_k) \arrow[r, shift left, "f"] \arrow[r, shift right, "g"'] & \prod_{i,j\in J} C^0(U_i \cap U_j) \end{tikzcd} \right) \tag{4.2}$$

to the *equalizer* of f and g, which is defined as follows:

$$\mathrm{eq}\left(A \overset{f}{\underset{g}{\rightrightarrows}} B \right) := \{a \in A \mid f(a) = g(a)\} = \ker(f-g). \tag{4.3}$$

We claim that the map r of equation (4.2) is an isomorphism. It is injective since if $f \in C^0(U)$ restricts to the zero function on all U_i, then f is the zero function. It is surjective, since if $(f_k)_{k\in J} \in \prod_{k\in J} C^0(U_k)$ belongs to the equalizer, then the functions f_k agree on the intersections, and hence they fit together to give a well-defined continuous function on U.

We note that the isomorphism (4.1) is a special case of the isomorphism (4.2). In fact, if the open subsets U_j of a cover of U are mutually disjoint, then $U_i \cap U_j = \emptyset$ for all $i, j \in J$, and hence $C^0(U_i \cap U_j) = \{0\}$. If follows that the equalizer on the right hand side of (4.2) is just the product $\prod_{k\in J} C^0(U_k)$.

Collecting the formal properties we observe in this example leads to the notion of a sheaf on a topological space, defined as follows.

Definition 4.4. A *presheaf* $\mathcal{S}$ *on a topological space* M *with values in complex vector spaces* assigns to each open subset $U \subset M$ a complex vector space $\mathcal{S}(U)$, and to each inclusion $U \subset V$ a linear map $r^U_V \colon \mathcal{S}(U) \to \mathcal{S}(V)$. The maps r^U_V are required to be compatible with composition. A presheaf $\mathcal{S}$ is a *sheaf* if for every open cover $\{U_i\}_{i\in I}$ the map

$$\mathcal{S}(U) \xrightarrow{r} \mathrm{eq}\left(\prod_{k\in I} \mathcal{S}(U_k) \overset{f}{\underset{g}{\rightrightarrows}} \prod_{i,j\in I} \mathcal{S}(U_i \cap U_j) \right) \tag{4.5}$$

is in an isomorphism (here r, f, g are constructed as products of restriction maps as in the motivating example discussed above).

We think of the sheaf condition as a *locality property* for a presheaf $\mathcal{S}$, since it allows us to reconstruct the vector space $\mathcal{S}(U)$ associated to an open set in terms of the vector spaces $\mathcal{S}(U_k)$ and $\mathcal{S}(U_i \cap U_j)$ associated to the smaller subsets U_k and $U_i \cap U_j$.

Example 4.6. Let M be a closed complex manifold. Then the *sheaf of holomorphic functions on* M, denoted $\mathcal{O}_M$, assigns to each open subset $U \subset M$ the vector space $\mathcal{O}_M(U)$ of holomorphic functions $U \to \mathbb{C}$.

We note that the only holomorphic functions on the closed manifold M are the *constant functions*. In particular, no information about the complex manifold can be obtained from studying holomorphic functions on all of M. This is very unlike the case of continuous functions, where the topological space can be reconstructed from the algebra of continuous functions on it. However, a lot of information about the complex manifold M can be uncovered by studying holomorphic functions on open subsets of M, that is, by analyzing the sheaf $\mathcal{O}_M$.

Sheaves with values in a category C. In our motivating example for a sheaf $\mathcal{S}$ we associated to an open subset U the vector space $\mathcal{S}(U) = C^0(U)$ of continuous maps $U \to \mathbb{C}$. Many variants can be obtained by replacing the target $\mathbb{C}$; for example:

- if Y is a set, we can take $\mathcal{S}(U)$ to be the set of all maps $U \to Y$;
- if Y is a topological space, we can take $\mathcal{S}(U)$ to be the topological space of all maps $U \to Y$ equipped with the compact open topology;
- if V is a topological vector space, we can take $\mathcal{S}(U)$ to be the topological vector space of all maps $U \to Y$ equipped with the compact open topology;
- if A is an algebra over $\mathbb{R}$ or $\mathbb{C}$, we can take $\mathcal{S}(U)$ to be the algebra of continuous maps $U \to A$.

In all of these examples, we observe formal properties of $U \mapsto \mathcal{S}(U)$ completely analogous to the sheaf properties of $U \mapsto C^0(U)$, except that now $\mathcal{S}(U)$ is not a vector space, but rather an object of some category C. This forces us to contemplate what we mean by the *product* $\prod_{j\in J} C_j$ of objects C_j and the *equalizer* $\mathrm{eq}\left(A \overset{f}{\underset{g}{\rightrightarrows}} B \right)$ of morphisms $f, g\colon A \to B$ in the category C. Fortunately, these are categorical notions which can be characterized by universal properties in any category.

Definition 4.7. Let C_j for $j \in J$ be a collection of objects of a category C. An object $C \in \mathsf{C}$ is a *product* of the C_j's if there are morphisms $p_j\colon C \to C_j$ such that for every object $B \in \mathsf{C}$ the map

$$\mathsf{C}(B, C) \longrightarrow \prod_{j\in J} \mathsf{C}(B, C_j) \qquad f \mapsto (p_j \circ f)_{j\in J}$$

is a bijection. In other words, it's easy to understand morphisms with target C: they are just collections of morphisms with target C_j for $j \in J$. The product is denoted $\prod_{j\in J} C_j$.

As usual with universal properties, it characterizes the object C up to isomorphism, and an object with this universal property might or might not exist. In the category of vector spaces the product $\prod_{j\in J} V_j$ does exist; it can be constructed explicitly by saying it consists of all collections of vectors $(v_j)_{j\in J}$ with $v_j \in V_j$. In other words, the direct products of vector spaces in (4.1) and (4.2) can now be interpreted as categorical products in Vect.

Definition 4.8. Let $f, g\colon A \to B$ be morphisms in a category C. An *equalizer* for $A \overset{f}{\underset{g}{\rightrightarrows}} B$ is an object $C \in \mathsf{C}$ and a morphism $r\colon C \to A$ such that $f \circ r = g \circ r$ and that (C, r) is universal with the property in the sense that if $r'\colon C' \to A$ is another map with $f \circ r' = g \circ r'$, then the map

$$\mathcal{S}(C', C) \longrightarrow \{r'\colon C' \to A \mid f \circ r' = g \circ r'\} \qquad f \mapsto r \circ f$$

is a bijection. The notation $\mathrm{eq}\left(A \underset{g}{\overset{f}{\rightrightarrows}} B \right)$ is used for the equalizer of f, g.

It is straightforward to check that the vector space defined in (4.3) is in fact an equalizer, thus showing that equalizers exist in the category of vector spaces.

Definition 4.9. Let C be a category with products. A *presheaf $\mathcal{S}$ on a topological space M with values in* C assigns to each open subset $U \subset M$ an object $\mathcal{S}(U) \in \mathsf{C}$, and to each inclusion $U \subset V$ a morphism $r_V^U \colon \mathcal{S}(U) \to \mathcal{S}(V)$. The maps r_V^U are required to be compatible with composition. A presheaf $\mathcal{S}$ is a *sheaf* if for every open cover $\{U_i\}_{i \in I}$ the map

$$\mathcal{S}(U) \xrightarrow{r} \mathrm{eq}\left(\prod_{k \in I} \mathcal{S}(U_k) \underset{g}{\overset{f}{\rightrightarrows}} \prod_{i,j \in I} \mathcal{S}(U_i \cap U_j) \right) \tag{4.10}$$

is in an isomorphism (here r, f, g are constructed as products of restriction maps).

In the definition above we need to assume that the category C has products in order to make sense of the maps $f, g \colon \prod_{k \in I} \mathcal{S}(U_k) \to \prod_{i,j \in I} \mathcal{S}(U_i \cap U_j)$. It should be emphasized that we *do not* need to assume that equalizers exit in C, since the sheaf condition (4.10) can be interpreted as requiring that the equalizer of f, g exists and that the map r from $\mathcal{S}(U)$ to the equalizer is an isomorphism. Equivalently, we can require that the map $r \colon \mathcal{S}(U) \to \prod_{k \in I} \mathcal{S}(U_k)$ satisfies the universal property characterizing equalizers.

4.1.2 Definition of strict factorization algebras

The goal of this section is to define strict factorization algebras. Before giving the formal definition we will discuss a motivating example, namely the factorization algebra of classical observables. This is a strict factorization algebra on the space-time of a classical field theory. We will describe the factorization algebra data associated to a general classical field theory, as well as the specific example of a point particle moving in a Riemannian manifold X.

Motivating example: the factorization algebra of classical observables.

classical field theory	point particle moving in a Riemannian manifold X
space-time M	$M = \mathbb{R}$
space of fields $\mathcal{M}$	$\mathcal{M} = \{\phi \colon \mathbb{R} \to X \mid \phi \text{ is smooth}\}$
$\mathrm{EL}(U) := \left\{ \begin{array}{l} \text{manifold of solutions to the} \\ \text{Euler-Lagrange equation} \\ \text{on an open subset } U \subset M \end{array} \right\}$	$\mathrm{EL}(U) = \{\phi \colon U \to X \mid \phi \text{ is geodesic}\}$ $\mathrm{EL}((a,b)) \cong TX$ $\phi \mapsto (\phi(c), \dot{\phi}(c))$ for $c \in (a,b)$
$\mathrm{Obs}^c(U) := \left\{ \begin{array}{l} \text{classical observables} \\ \text{in space-time region } U \end{array} \right\}$ $:= C^\infty(EL(U))$	$\mathrm{Obs}^c((a,b)) = C^\infty(\mathrm{EL}((a,b))) = C^\infty(TX)$

As we did for the case of continuous functions on open subsets U of a topological space M, we now discuss the data and formal properties provided by the vector spaces of classical observables. First the data:

- To each open subset $U \subset M$ we associate the complex vector space $\mathrm{Obs}^c(U)$ of classical observables in U (again we simply ignore the algebra structure since we strive for the definition of a factorization algebra with values in $\mathsf{Vect}_{\mathbb{C}}$). The inclusion of subsets $U \subset V$ induces a restriction map

$$r_V^U \colon \mathrm{EL}(V) \longrightarrow \mathrm{EL}(U),$$

which in turn induces a linear map

$$m_V^U = (r_V^U)^* \colon \mathrm{Obs}^c(U) = C^\infty(\mathrm{EL}(U)) \longrightarrow \mathrm{Obs}^c(V) = C^\infty(\mathrm{EL}(V)).$$

It is clear that the maps m_V^U are compatible with composition; in other words, this construction gives a functor from the category of open subsets of M to $\mathsf{Vect}_{\mathbb{C}}$. We note that the case of a pre-sheaf on M the functor $U \mapsto \mathrm{Obs}^c(U)$ is *covariant*, and so it might be referred to as a (pre) cosheaf.

- More generally, if $U_1, \dots, U_k$ are disjoint open subsets of an open subset $V \subset M$, then there is a linear map

$$m_V^{U_1,\dots,U_k} \colon \mathrm{Obs}^c(U_1) \otimes \cdots \otimes \mathrm{Obs}^c(U_k) \longrightarrow \mathrm{Obs}^c(V). \tag{4.11}$$

It is given by the composition

$$\mathrm{Obs}^c(U_1) \otimes \cdots \otimes \mathrm{Obs}^c(U_k) \xrightarrow{m_V^{U_1} \otimes \cdots \otimes m_V^{U_k}} \mathrm{Obs}^c(V) \otimes \cdots \otimes \mathrm{Obs}^c(V) \longrightarrow \mathrm{Obs}^c(V) \tag{4.12}$$

where the unlabeled second map is given by multiplication in the algebra $\mathrm{Obs}^c(V) = C^\infty(\mathrm{EL}(V))$.

Remark 4.13. The attentive reader might wonder about two points here: we did not use the assumption that the U_i's are disjoint, and why would we list the maps $m_V^{U_1,\dots,U_k}$ as additional data if in fact they are determined by the maps $m_V^{U_i}$? Addressing the second point first, the vector space $\mathrm{Obs}^c(V) = C^\infty(\mathrm{EL}(V))$ is an *algebra*, a structure that the vector space $\mathcal{F}(V)$ a factorization algebra associates to an open set $V \subset M$ typically *does not have*. Even if $\mathcal{F}(V)$ happens to be an algebra, the maps $m_V^{U_1,\dots,U_k}$ might not be given by composition (4.12) as Example **??** shows. We will also see that the construction of the map (4.11) in that example requires that the sets U_i are disjoint.

From a categorical point of view, the big difference between sheaves and factorization algebras is that for a pre-sheaf $\mathcal{S}$ on M and open subsets $U_1, \ldots, U_k$ of $U \subset M$ the restriction maps $r^U_{U_i} \colon \mathcal{S}(U) \to \mathcal{S}(U_i)$ uniquely determine a map

$$\mathcal{S}(U) \longrightarrow \mathcal{S}(U_1) \oplus \cdots \oplus \mathcal{S}(U_k)$$

since the direct sum of finitely many vector spaces is their *categorical product*. For a pre-factorization algebra $\mathcal{F}$ the maps $m^{U_i}_U \colon \mathcal{F}(U_i) \to \mathcal{F}(U)$ do not determine a map

$$\mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) \longrightarrow \mathcal{F}(U),$$

since the tensor product is *not the categorical coproduct* of vector spaces (the tensor product is the categorical coproduct of *algebras* which is what we made use of to construct the multiplication map $m^{U_1,\ldots,U_k}_V$ in our motivating example).

Next we assume that U is a union of disjoint open subsets U_i or, more generally, that $\{U_i\}_{i\in I}$ is an open cover for U, and we discuss whether the vector space $\mathcal{F}(U) := \mathrm{Obs}^c(U) = C^\infty(\mathrm{EL}(U))$ can be expressed in terms of data associated to the U_i's. This is the analog of our discussion of the vector space $C^0(U)$ of continuous functions on U which lead to equations (4.1) and (4.2).

Finite disjoint union. If $U_1, \ldots, U_k$ are disjoint open subsets with union U, then a solution of the Euler Lagrange equation on U is just a collection of solutions on each U_i and hence we obtain an isomorphism

$$\mathrm{EL}(U) \xrightarrow{\cong} \mathrm{EL}(U_1) \times \cdots \times \mathrm{EL}(U_k).$$

Then the induced map of functions

$$m^{U_1,\ldots,U_k}_U \colon C^\infty(\mathrm{EL}(U_1)) \otimes \cdots \otimes C^\infty(\mathrm{EL}(U_k)) \longrightarrow C^\infty(\mathrm{EL}(U)) \tag{4.14}$$

is an isomorphism, provided the tensor product is interpreted not as the algebraic tensor product, but rather as the projective tensor product of the topological vector spaces $C^\infty(\mathrm{EL}(U_i))$ equipped with the usual Frechet topology. We observe that for $U = \emptyset$, the space $\mathrm{EL}(U)$ of solutions of the Euler Lagrange equation consist of one point, and hence $\mathrm{Obs}^c(\emptyset) = C^\infty(\mathrm{EL}(\emptyset))$ can be identified with $\mathbb{C}$.

Open cover. One might expect that the assignment $\mathcal{F}(U) := \mathrm{Obs}^c(U)$ for an open subset $U \subset M$ can be calculated in terms of $\mathcal{F}(U_i)$ and $\mathcal{F}(U_i \cap U_j)$ for an open cover $\{U_i\}_{i\in I}$ of U. More precisely, one might expect that $U \mapsto \mathcal{F}(U)$ has the following *cosheaf property*, the analog of the sheaf property (4.10) taking into account the covariant nature of the assignment $U \mapsto \mathcal{F}(U)$: the map

$$\mathcal{F}(U) \xleftarrow{m} \mathrm{coeq}\left(\bigoplus_{k\in I} \mathcal{F}(U_k) \overset{f}{\underset{g}{\leftleftarrows}} \bigoplus_{i,j\in I} \mathcal{F}(U_i \cap U_j) \right) \tag{4.15}$$

is an isomorphism. Here f (resp. g) is the linear map induced by the inclusion of $U_i \cap U_j$ in U_i (resp. U_j), and the *coequalizer* of two linear maps $f, g\colon A \to B$ is the cokernel of $f - g$, that is, the quotient of B obtained by modding out by the image of $f - g$. However, the map m, induced by the inclusions $U_k \hookrightarrow U$ is *typically not an isomorphism.* In fact, it conflicts directly with the disjoint union property as we will see now.

For this argument and others, it is useful to observe that the above coequalizer can be simplified, since the maps f and g agree on the $i = j$ summand, and hence that summand does not contribute to the coequalizer. Similarly, the image of the summand $\mathcal{F}(U_i \cap U_j)$ under the map $f - g$ is the same as the image of the subspace $\mathcal{F}(U_j \cap U_i)$. Hence we don't change the coequalizer if we pick a total order on I and replace the domain of f, g by the direct sum of $i, j \in I$ with $i < j$.

Now suppose U is the union of disjoint open subsets U_1, U_2. Then $\{U_1, U_2\}$ is an open cover of U and the corresponding coequalizer is given by

$$\begin{aligned}
&\operatorname{coeq}\left(\mathcal{F}(U_1) \oplus \mathcal{F}(U_2) \underset{g}{\overset{f}{\leftleftarrows}} \bigoplus_{i,j=1}^{2} \mathcal{F}(U_i \cap U_j) \right) \\
= \ &\operatorname{coeq}\left(\mathcal{F}(U_1) \oplus \mathcal{F}(U_2) \underset{g}{\overset{f}{\leftleftarrows}} \bigoplus_{i<j} \mathcal{F}(U_i \cap U_j) \right) \\
= \ &\operatorname{coeq}\left(\mathcal{F}(U_1) \oplus \mathcal{F}(U_2) \underset{g}{\overset{f}{\leftleftarrows}} \mathcal{F}(\emptyset) = \mathbb{C} \right) \\
= \ &\mathcal{F}(U_1) \oplus \mathcal{F}(U_2)/(\Omega_1, \Omega_2)
\end{aligned}$$

where $\Omega_i \in \mathcal{F}(U_i)$ is the *vacuum vector* given by the image of $1 \in \mathbb{C}$ under the map $\mathbb{C} = \mathcal{F}(\emptyset) \to \mathcal{F}(U_i)$ induced by the inclusion map $\emptyset \to U$.

We see that the expectation that the map (4.15) is an isomorphism for any open cover $\{U_i\}$ of U is incompatible with the product property $\mathcal{F}(U) \cong \mathcal{F}(U_1) \otimes \mathcal{F}(U_2)$ if U is the union of two disjoint open subsets U_1, U_2 (see (**??**)). Rather mysteriously, the map (4.15) *is* an isomorphism for a very restrictive class of open covers which called *Weiss covers*, a fact that we won't prove here. These open covers were used by the topologist Michael Weiss, who was a post-doc at Notre Dame, in his 'calculus of embeddings'.

Definition 4.16. A collection of open subsets $U_i \subset U$ of a topological space U is a *Weiss cover* of U if for any finite subset $S \subset U$ there is some U_i that contains S.

Example 4.17. Here are some examples of Weiss covers.

1. Let M be a Riemannian manifold and $\epsilon > 0$. The collection of open subsets of M consisting of disjoint unions of balls of radius $< \epsilon$ is a Weiss cover (we note that the collection of open balls of radius $< \epsilon$ is not a Weiss cover).

2. The collection of punctured intervals $U_i := [0,2] \setminus \{\frac{1}{i}\}$ is a Weiss cover of the closed interval $[0,2]$. We will use this Weiss cover for calculations. We note that a Weiss cover of $[0,2]$ necessarily consists of infinitely many open subsets.

Abstracting the data and the properties of our motivating example, we now define prefactorization algebras and strict factorization algebras. In the example above, we associate to each open subset U a vector space. This is the main example we are interested in, but more generally, we will define (pre) factorization algebras with values in a monoidal category $(\mathsf{C}, \otimes)$.

Definition 4.18. Let M be a topological space. A *prefactorization algebra* $\mathcal{F}$ on M with values in a monoidal category $(\mathsf{C}, \otimes)$ assigns

- to each open subset $U \subset M$ a vector space $\mathcal{F}(U)$, and
- to a collection $U_1, \ldots, U_k$ of disjoint open subsets of an open subset $V \subset M$ a linear map
$$m_U^{U_1,\ldots,U_k} \colon \mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) \longrightarrow \mathcal{F}(V).$$

These maps are required to be compatible with composition in the sense that if $U_{i1}, \ldots, U_{in_i}$ are disjoint open subsets of V_i, and $V_1, \ldots, V_k$ are disjoint open subsets of W, then the diagram

$$\begin{tikzcd}
\bigotimes_{i=1}^{k} \bigotimes_{j=1}^{n_i} \mathcal{F}(U_{ij}) \arrow[rr, "\bigotimes_{i=1}^{k} m_{V_i}^{U_{i1},\ldots,U_{in_i}}"] \arrow[dr, "m_W^{U_{11},\ldots,U_{kn_k}}"'] & & \bigotimes_{i=1}^{k} \mathcal{F}(V_i) \arrow[dl, "m_W^{V_1,\ldots,V_k}"] \\
& \mathcal{F}(W) &
\end{tikzcd}$$

is commutative.

A *strict factorization algebra* is a prefactorization algebra which satisfies the following two properties:

Factorization axiom. For any finite collection of disjoint open subsets $U_1, \ldots, U_k \subset M$ with union $U = U_1 \cup \cdots \cup U_k$ the map

$$m_U^{U_1,\ldots,U_k} \colon \mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) \longrightarrow \mathcal{F}(U)$$

is an isomorphism.

Locality axiom. For every open subset $U \subset M$ and every Weiss cover $\{U_i\}_{i\in I}$ of U the map

$$\mathcal{F}(U) \xleftarrow{m} \operatorname{coeq}\left(\bigoplus_{k\in I} \mathcal{F}(U_k) \underset{g}{\overset{f}{\leftleftarrows}} \bigoplus_{i,j\in I} \mathcal{F}(U_i \cap U_j) \right) \tag{4.19}$$

is an isomorphism. Here we assume that the category $\mathcal{C}$ has coproducts that we denote by the symbol $\bigoplus$, since the coproduct in the category category of vector spaces is given by the direct sum. We not require that all coequalizers in $\mathcal{C}$ exits; rather, the locality axiom is the requirement that $\mathcal{F}(U)$ and the morphism from $\bigoplus_{k\in I}\mathcal{F}(U_k)$ to $\mathcal{F}(U)$ induced by the inclusions $U_k \to U$ satisfy the universal property that characterizes the coequalizer.

4.1.3 Examples of strict factorization algebras

In this section we will give some examples of strict factorization algebras. In general we won't provide a proof of the locality axiom; the factorization algebra

Example 4.20. (A factorization algebra on $\mathbb{R}$ associated to a unital algebra A.) For an open subset $U \subset \mathbb{R}$ which is the union of disjoint open intervals $U_1, \dots, U_k$, say $U_i = (a_i, b_i)$ with $a_1 < b_1 \le a_2 < b_2 \le \dots a_k < b_k$ we define

$$\mathcal{F}(U) := \mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) := \underbrace{A \otimes \cdots \otimes A}_{k}.$$

If $V = (a, b)$ is an open interval containing all intervals U_i, we define the multiplication map

$$\begin{aligned} m_V^{U_1,\dots,U_k} \colon \mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) = A \otimes \cdots \otimes A &\longrightarrow \mathcal{F}(V) = A \\ a_1 \otimes \cdots \otimes a_k &\mapsto a_1 \cdot \cdots \cdot a_k. \end{aligned}$$

A general open subset of $\mathbb{R}$ is a union of open intervals, possibly infinitely many. So the above example is incomplete, since we haven't defined the vector space associated to a union of infinitely many disjoint open intervals. This issue is addressed in general by the following exercise.

Homework 4.21. Let $\mathcal{F}$ be a strict factorization algebra on a topological space M, and let U_i, $i \in I$ be a collection of disjoint open subsets of M. Show that there is an isomorphism

$$\mathcal{F}(\bigcup_{i\in I} U_i) \cong \bigotimes_{i\in I} \mathcal{F}(U_i).$$

Here the tensor product is defined as the colimit over the tensor products $\bigotimes_{i\in J}\mathcal{F}(U_i)$ where $J \subset I$ ranges over all the finite subsets of I; for $J \subset J'$, the map

$$\bigotimes_{i\in J} \mathcal{F}(U_i) \longrightarrow \bigotimes_{i\in J'} \mathcal{F}(U_i)$$

sends $\bigotimes_{i\in J} v_i$ to $\bigotimes_{i\in J'} v_i$, where $v_i = \Omega_i \in \mathcal{F}(U_i)$ is the vacuum vector for $i \notin J$ (see Definition **??** for the definition of colimits). Hint: show that a Weiss cover for $U = \bigcup_{i\in I} U_i$ is given by the collection of subsets $U_J := \bigcup_{i\in J} U_i$ of U where J runs through the finite subsets of I. Show that the coequalizer on the right hand side of (4.19) for this Weiss cover is isomorphic to the colimit over the finite subsets $J \subset I$ described above.

Example 4.22. (A factorization algebra from an algebra A and a pair of modules.)

Example 4.23. (Quantum mechanics of a point particle in X.) This is an example of a factorization algebra on $\mathbb{R}$ with values in the monoidal category of topological vector spaces equipped with the projective tensor product. We define

$$\mathcal{F}((a,b)) := \mathrm{End}(V) \qquad \text{for } V := C^\infty(X)$$

for any open interval $(a,b) \subset \mathbb{R}$, $a, b \in [-\infty, \infty]$, $a < b$. For $a < b < c$ the linear map

$$\mathcal{F}((a,b)) \otimes \mathcal{F}((b,c)) \longrightarrow \mathcal{F}((a,c)) \qquad \text{is given by } (S,T) \mapsto S \circ T.$$

The vacuum vector $\Omega_{(a,b)} \in \mathcal{F}((a,b))$, the image of $1 \in \mathbb{C} = \mathcal{F}(\emptyset)$ under the map induced by the inclusion $\emptyset \hookrightarrow (a,b)$, is given by

$$\Omega_{(a,b)} := e^{-(b-a)\Delta_X} \in \mathrm{End}(V) = \mathcal{F}((a,b)),$$

where Δ_X is the Laplace operator. We note that the vector space $\mathcal{F}((a,b))$ depend only on the manifold X, while the vacuum vector $\Omega_{(a,b)}$ is determined by the Laplace operator Δ_X which in turn depends on the Riemannian metric on X.

If U is a union of finitely many intervals, the factorization axiom requires us to define $\mathcal{F}(U)$ as the tensor product $\mathrm{End}(V) \otimes \cdots \otimes \mathrm{End}(V)$ of a corresponding number of copies of $\mathrm{End}(V)$.

Homework 4.24. Let $U_1, \dots, U_k$ be a disjoint open intervals, say $U_i = (a_i, b_i)$ with $a_1 < b_1 \leq a_2 < b_2 \leq \dots a_k < b_k$, which are contained in an open interval $U = (a,b)$. Using the compatibility of the multiplication maps with composition, show that the multiplication map

$$m_U^{U_1,\dots,U_k} \colon \mathcal{F}(U_1) \otimes \cdots \otimes \mathcal{F}(U_k) = \mathrm{End}(V) \otimes \cdots \otimes \mathrm{End}(V) \longrightarrow \mathcal{F}(U) = \mathrm{End}(V)$$

is given by

$$(T_1, \dots, T_k) \mapsto e^{-(a_1-a)\Delta_X} T_1 e^{-(a_2-b_1)\Delta_X} T_2 e^{-(a_3-b_2)\Delta_X} \dots e^{-(a_k-b_{k-1})\Delta_X} T_k e^{-(b-b_k)\Delta_X}.$$

1. quantum mechanics of point particle in X;
2. Let A be a unital algebra. factorization algebra on $\mathbb{R}$ associated to unital algebra A;
3. factorization algebra on $[0,2]$ associated to unital algebra and pointed modules M_A, ${}_AN$;

4.2 Factorization algebras

The goal of this section is to define factorization algebras with values in the category of cochain complexes Ch of complex vector spaces. The tensor product of cochain complexes gives Ch the structure of a symmetric monoidal category, and so specializing Definition **??** to this target category we have already defined *strict factorization algebras with values in* Ch. As the terminology suggests, a *factorization algebra with values in* Ch is a "looser" version of this definition, obtained in the following way:

- In the factorization and locality axiom, replace the requirement of being an *isomorphism* by the weaker requirement of being a *quasi-isomorphism* or *weak equivalence* of cochain complexes (which is a chain map inducing an isomorphism on cohomology).
- In the locality axiom, replace the diagram

$$\bigoplus_{k\in I}\mathcal{F}(U_k) \underset{g}{\overset{f}{\leftleftarrows}} \bigoplus_{i,j\in I}\mathcal{F}(U_i\cap U_j)$$

  by a larger diagram which includes terms obtained by applying $\mathcal{F}$ to the intersection of arbitrary many U_i's.
- replace the coequalizer of the small diagram by the *homotopy* colimit of the larger diagram (the coequalizer of $A \underset{g}{\overset{f}{\leftleftarrows}} B$ is the same as its colimit).

Remark 4.25. From a philosophical point of view, what is the difference between the category of cochain complexes, and the category of vector spaces? Well, two cochain complexes can be weakly equivalent without being isomorphic, whereas there is no good notion of 'weak equivalence' for vector spaces. Another example of a category with 'weak equivalences' is the category of topological spaces. A continuous map $f\colon X\to Y$ between topological spaces is a *weak equivalence* if it induces an isomorphism on all homotopy groups (with respect to all base points).

There are many other categories C with a distinguished class of morphisms $\mathcal{W}$ called *weak equivalences*, which is closed under composition and satisfies the 2-out-of-6 condition: if f,g,h are composable morphisms (i.e., the composition fgh makes sense), and the two morphisms fg and gh are weak equivalences, then also f, g, h and fgh are weak equivalences (so all six morphisms that can be build from f, g and h are weak equivalences).

This extra structure on a category C makes it possible to do "homotopy theory" in C, although often more structure, say a model category structure in the sense of Quillen, is available and makes some constructions easier. If C is a symmetric monoidal category with a compatible class of weak equivalences, factorization algebras with values in C can be defined. Due to our interests, we restrict ourselves to factorization algebras with values in cochain complexes.

4.2.1 Digression: Homotopy colimits

4.2.2 Definition of factorization algebras

4.2.3 A factorization algebra build from an algebra

Remark 4.26. From a categorical point of view the direct product of vector spaces is their categorical product and their direct sum is their categorical coproduct. Let us recall the definition of the (co)product of a collection $\{C_j\}_{j\in J}$ of objects of a category C. An object $C \in \mathsf{C}$ is called a *product* of the C_j's if there exist morphisms $p_j\colon C \to C_j$ such that for every object $D \in \mathsf{C}$ the map

$$\mathsf{C}(D, C) \longrightarrow \prod_{j\in J} \mathsf{C}(D, C_j) \qquad f \mapsto (p_j \circ f)_{j\in J}$$

is a bijection. In other words, it's easy to understand morphisms with target C: they are just collections of morphisms with target C_j for $j \in J$. As usual with universal properties, it characterizes the object C up to isomorphism, and an object with this universal property might or might not exist. The product is denoted $\prod_{j\in J} C_j$.

Dually, an object $C \in \mathsf{C}$ is a *coproduct* of the C_j's if there are morphisms $i_j\colon C_j \to C$ with the property that for every object $D \in \mathsf{C}$ the map

$$\mathsf{C}(C, D) \longrightarrow \prod_{j\in J} \mathsf{C}(C_j, D) \qquad f \mapsto (f \circ i_j)_{j\in J}$$

is a bijection. The coproduct is denoted $\coprod_{j\in J} C_j$ or $\bigoplus_{j\in J} C_j$.

4.3 Free field theories

Traditionally, a classical field theory is called *free* if

- its space of fields $\mathcal{E}$ is the vector space of sections of a vector bundle $E \to M$ over the space-time manifold E, and
- the Lagrangian $L(\phi)$ is a *quadratic* function of the field $\phi \in \Gamma(M, E)$.

Following the book by Costello-Gwilliam [CG], here we want to consider free field theories of *Batalin-Vilikovisky type*. Here the vector space of fields $\mathcal{E}$ is a dg-vector vector space; as a graded vector space, $\mathcal{E}$ is the space of sections of a $\mathbb{Z}$-graded vector bundle $E = \bigoplus_{i\in\mathbb{Z}} E_i \to M$. The differential of $\mathcal{E}$ is Q:

4.3.1 Digression: Symplectic manifolds

4.3.2 The density line bundle

Let M be a manifold of dimension d and $\omega \in \Omega^d(M)$ a top-dimensional differential form on M with compact support. Then ω can be integrated over M *provided M comes equipped with an orientation*; passing to the opposite orientation changes the integral $\int_M \omega$ by a minus sign.

It is desirable to have quantities that can be integrated over M, regardless of whether M is orientable. Moreover, even if M happens to be orientable, we would like the result of this integration to be independent of the choice of the orientation. For example:

- We would like to compute the volume of a non-orientable compact Riemannian manifold M by integrating some local quantity determined by the Riemannian metric over M.
- We want to define classical field theories on non-orientable space-time manifolds M. For these, the Lagrangian $L(\phi)$ of a field ϕ needs to be a local quantity that can be integrated over M in order to define the action functional $\mathcal{S}$ by $\mathcal{S}(\phi) := \int_M L(\phi)$.

Here is the definition of quantities that can be integrated over manifolds.

Definition 4.27. A *density* on a real vector space V of dimension d is a function

$$f\colon \Lambda^d V \to \mathbb{R} \qquad \text{with the property} \qquad f(r\lambda) = |r| f(\lambda) \text{ for all } r \in \mathbb{R},\ \lambda \in \Lambda^d V.$$

We denote the one-dimensional vector space of densities on V by Dens_V.

Here are two examples of densities:

(1) An element $\omega \in \Lambda^d(V^\vee)$ determines a density

$$|\omega|\colon \Lambda^d(V) \to \mathbb{R} \qquad \text{given by} \qquad v_1 \wedge \cdots \wedge v_d \mapsto |\omega(v_1, \ldots, v_d)|.$$

(2) If V is an inner product space, the map $\Lambda^d(V) \to \mathbb{R}$ which sends $v_1 \wedge \cdots \wedge v_d$ to the volume of the parallelepiped spanned by $v_1, \ldots, v_d$ is a density. Since the vector space $\Lambda^d(V)$ is 1-dimensional, this map is in fact determined by $e_1 \wedge \cdots \wedge e_d \mapsto 1$, where $\{e_1, \ldots, e_d\}$ is an orthonormal basis of V.

We note that an orientation on V determines a volume form $\mathrm{vol}_V \in \Lambda^d(V^\vee)$. While vol_V depends on the choice of orientation, the density $|\mathrm{vol}_V|$ built from vol_V as in example (1) does not, and in fact agrees with the density described in the paragraph above.

Definition 4.28. The *density line bundle* on a manifold M is the real line bundle Dens_M whose fiber over a point $x \in M$ is the 1-dimensional space Dens_{T_xM} of densities on the tangent space T_xM. The sections of Dens_M are called *densities*. Example (2) above shows that a Riemannian metric on M determines an associated nowhere vanishing section of Dens_M we denote by $|\,\mathrm{vol}_M\,|$ (somewhat abusing notation, since vol_M might not exist due to a lack of orientability of M).

The real line bundles $\Lambda^d T^*M$ and Dens_M on a d-manifold M are closely related:

1. Any real line bundle $L \to M$ determines a double covering of M, whose fiber at $x \in M$ consists of the two connected components of $L_x \setminus \{0\}$. The double covering associated to $\Lambda^d T^*M$ can be identified with the orientation cover of M. In particular, the line bundle $\Lambda^d T^*M$ is trivializable if and only if M is orientable.

   By contrast, the double covering associated to Dens_M is always trivial, since the two connected components of Dens_{T_xM} consists of maps $f\colon \Lambda^d(T_xM) \to \mathbb{R}$ with $f(\lambda) > 0$ (resp. $f(\lambda) < 0$) for all $0 \neq \lambda \in \Lambda^d(T_xM)$. It follows that the density line bundle is always trivializable.

2. A Riemannian metric on M determines a section $|\,\mathrm{vol}_M\,|$ of the density line bundle Dens_M by applying construction (2) above in each fiber. Similarly, a Riemannian metric and an orientation on M determine a volume form $\mathrm{vol}_M \in \Omega^d(M)$.

3. If $f\colon M \to N$ is a smooth map between d-manifolds, the differential of f at $x \in M$ induces linear maps

   $$f_x^*\colon \Lambda^d(T^*_{f(x)}N) \to \Lambda^d(T^*_xM) \qquad \text{and} \qquad f_x^*\colon \mathrm{Dens}_{T_{f(x)}N} \to \mathrm{Dens}_{T_xM}\,.$$

   In particular, if M, N are open subsets of $\mathbb{R}^d$, the standard Riemannian metric and orientation provide non-zero vectors for these 1-dimensional vector spaces, allowing us to identify these linear maps with real numbers. In the case of d-forms, this number is the determinant of the Jacobian of f at x, while for the densities, it is the absolute value of that determinant.

Lemma 4.29. *densities can be integrated.*

4.3.3 Definition of free BV-field theories

Definition 4.30. A *free Batalin-Vilkovisky field theory* consists of the following data:

- A smooth manifold M, the *space-time manifold*.

- For any open subspace $U \subset M$ a topological differential graded vector space $\mathcal{E}(U)$, the space of *fields* on U. In other words, $\mathcal{E}(U)$ is a collection of topological vector spaces $\mathcal{E}_i(U)$ for $i \in \mathbb{Z}$, and a sequence of continuous linear maps

$$\dots \xrightarrow{Q} \mathcal{E}_{i-1}(U) \xrightarrow{Q} \mathcal{E}_i(U) \xrightarrow{Q} \mathcal{E}_{i+1}(U) \xrightarrow{Q} \dots . \tag{4.31}$$

We recall that for classical field theories we required that the space of fields is a space of sections of a bundle over space-times, which expresses the idea that the fields should be "local in space time". For the dg-vector space $\mathcal{E}(U)$ this locality is implemented by the following conditions:

  – There are vector bundles $E_i \to M$ such that the topological vector space $\mathcal{E}_i(U)$ is the space $\Gamma(U, E_i)$ of smooth sections of E_i restricted to U, equipped with the usual Frechet topology. Equivalently we can say that there is a $\mathbb{Z}$-graded vector bundle $E \to M$ such that the $\mathbb{Z}$-graded topological vector space $\mathcal{E}(U)$ is the space of sections $\Gamma(U, E)$.
  – The operator $Q \colon \Gamma(U, E_i) \longrightarrow \Gamma(U, E_{i+1})$ is a differential operator.
  – The complex of differential operators (4.31) is *elliptic*. The precise meaning of this condition will not be relevant for us, and we refer the reader to [] for the definition. What matters for us is that ellipticity implies that the cohomology groups of the cochain complex (4.31) are *finite dimensional* if U is compact and without boundary (note that $\mathcal{E}_i(U)$ is infinite dimensional unless E_i has rank 0 or $\dim M = 0$). Moreover, ellipticity is a condition that is easily verified in concrete examples.

- A continuous bilinear pairing

$$\langle\ ,\ \rangle \colon \mathcal{E}_c(U) \times \mathcal{E}_c(U) \longrightarrow \mathbb{C}$$

on the space $\mathcal{E}_c(U)$ of sections with compact support.

This pairing is required to have the following properties:

degree -1: the pairing has degree -1, which means that for homogeneous elements $\phi, \psi \in \mathcal{E}$ of degree $|\phi|$, $|\psi|$ the degree of $\langle \phi, \psi\rangle$ is $|\phi| + |\psi| - 1$. In particular, $\langle \phi, \psi\rangle$ is trivial unless $|\phi| + |\psi| = 0$).

graded anti-symmetry: $\langle \phi, \psi\rangle = -(-1)^{|\phi||\psi|}\langle \psi, \phi\rangle$.

locality: it is induced by a 'local pairing', a vector bundle homomorphism

$$\langle\ ,\ \rangle_{\mathrm{loc}} \colon E \otimes E \longrightarrow \mathrm{Dens}_M$$

of degree -1 in the sense that for $\phi, \psi \in \mathcal{E}_c(U) = \Gamma_c(U, E)$, the pairing $\langle \phi, \psi \rangle \in \mathbb{C}$ is given by

$$\langle \phi, \psi \rangle = \int_U \langle \phi, \psi \rangle_{\mathrm{loc}}.$$

non-degenerate: the local pairing $\langle\ ,\ \rangle_{\mathrm{loc}} \colon E \otimes E \to \mathrm{Dens}_M$ induces a vector bundle isomorphism

$$E \xrightarrow{\cong} E^\vee \otimes \mathrm{Dens}_M = \mathrm{Hom}(E, \mathrm{Dens}_M) \qquad \text{via} \qquad v \mapsto (w \mapsto \langle v, w \rangle_{\mathrm{loc}})$$

Remark 4.32. Looking at the above definition of a free BV-field theory, the attentive reader might miss an essential ingredient of a classical field theory, namely the Lagrangian L and the action functional $\mathcal{S}$. In terms of the data featured in the above definition, for a field $\phi \in \mathcal{E}(U)$ they are given by

$$L(\phi) := \langle \phi, Q\phi \rangle_{\mathrm{loc}} \qquad \mathcal{S}(\phi) = \int_U L(\phi) = \int_U \langle \phi, Q\phi \rangle_{\mathrm{loc}} = \langle \phi, Q\phi \rangle.$$

We remark that both, the Lagrangian and the action functional, are quadratic functions of ϕ, as they should be in a *free* field theory. We also observe that L and S have degree 0, an important feature, that helps to explain why the pairings are required to have degree -1.

Example 4.33. (Free scalar field theory) Let M be a Riemannian manifold.

4.3.4 The factorization algebra of classical observables of a free BV-theory

4.3.5 The factorization algebra of quantum observables of a free BV-theory

Integration by parts: It is well-known that the Laplace operator is self-adjoint on closed manifolds, which is proved by integration by parts. For manifolds with boundary, there is in general a boundary term given by an integral over the boundary. For the proof of Theorem **??** we will need to understand the boundary term in the simple situation of the Laplace operator on an open interval $(a, b) \subset \mathbb{R}$. Then $\Delta F = -F''(x)$, and for functions $F, G \in C^\infty(a, b)$ we have

$$\begin{aligned}
\langle \Delta F, G \rangle &= -\int_a^b F''(x)G(x)dx = \int_a^b F'(x)G'(x)dx - [F'(x)G(x)]_a^b \\
&= -\int_a^b F(x)G''(x)dx + [F(x)G'(x) - F'(x)G(x)]_a^b \\
&= \langle F, \Delta G \rangle + [F(x)G'(x) - F'(x)G(x)]_a^b
\end{aligned}$$

We conclude that
$$\langle \Delta F, G\rangle - \langle F, \Delta G\rangle = \langle F, G\rangle_\partial,$$
where the boundary term $\langle F, G\rangle_\partial$ is defined by
$$\langle F, G\rangle_\partial := \left[F(x)G'(x) - F'(x)G(x)\right]_a^b.$$

Next we will use the boundary term to define a pairing for the cohomology of the complex $\mathcal{E}_c(a,b)[1]$. Written out, it has the following form

$$\begin{array}{ccc} \mathcal{E}_c(a,b)[1]: & C_c^\infty(a,b) \xrightarrow{\Delta} & C_c^\infty(a,b) \\ & & \\ \text{degree} & -1 & 0 \end{array}$$

We recall that the Laplace operator Δ is injective on functions with compact support, which implies that the cohomology groups $H^i(\mathcal{E}_c(a,b)[1])$ are trivial except for $i \neq 0$. We define a pairing
$$\alpha\colon H^0(\mathcal{E}_c(a,b)[1]) \otimes H^0(\mathcal{E}_c(a,b)[1]) \longrightarrow \mathbb{C}$$
by setting
$$\alpha(f,g) := \langle F, G\rangle_\partial \qquad \text{for } F, G \in \mathcal{E}(a,b)[1]_{-1} \text{ with } [\Delta F] = f,\ [\Delta G] = g.$$

We remark that if we drop the compact support condition and look at the complex $\mathcal{E}(a,b)[1]$, the Laplace operator is *onto*. In particular, for any $\widetilde{f}, \widetilde{g} \in C_c^\infty(a,b)$ representing the cohomology classes f, g, we can find $F, G \in C^\infty(a,b)$ whose image under Δ is $\widetilde{f}$ resp. $\widetilde{g}$ (note that unlike $\widetilde{f}, \widetilde{g}$ the functions F, G are not required to have compact support).

Homework 4.34. Show that for the elements $p, q \in H^0(\mathcal{E}_c(a,b))$ we have $\alpha(p,q) = 1$, $\alpha(p,p) = 0$, $\alpha(q,q) = 0$.

Proposition 4.35. *Let* $f, g \in\in H^0(\mathcal{E}_c(a,b)[1]) \subset H^0(\mathsf{Sym}^*(\mathcal{E}_c(a,b)[1]), \widehat{Q})$. *Then*
$$f \star g - g \star f = \hbar\, \alpha(f,g).$$

Proof. □

5 Solutions to selected homework problems

Solution to problem 2.52. For vector fields X, Y and a section s we calculate

$$\begin{aligned}\nabla'_X \nabla'_Y s &= (\nabla_X + A(X))(\nabla_Y + A(Y))s \\ &= \nabla_X \nabla_Y s + \nabla_X (A(Y)s) + A(X)\nabla_Y s + A(X)A(Y)s \\ &= \nabla_X \nabla_Y s + (XA(Y))s + A(Y)\nabla_X s + A(X)\nabla_Y s + A(X)A(Y)s\end{aligned}$$

Here the last equation uses the derivation property for ∇_X. To calculate

$$\mathbb{R}^{\nabla'}(X,Y)s = \nabla'_X \nabla'_Y s - \nabla'_Y \nabla'_X s - \nabla'_{[X,Y]} s,$$

we write out the three terms on the right hand side:

$$\begin{aligned}\nabla'_X \nabla'_Y s &= \nabla_X \nabla_Y s + (XA(Y))s + A(Y)\nabla_X s + A(X)\nabla_Y s + A(X)A(Y)s \\ \nabla'_Y \nabla'_X s &= \nabla_Y \nabla_X s + (YA(X))s + A(X)\nabla_Y s + A(Y)\nabla_X s + A(Y)A(X)s \\ \nabla'_{[X,Y]} s &= \nabla_{[X,Y]} s + A([X,Y])s\end{aligned}$$

Collecting the terms, we see that

$$\begin{aligned}R^{\nabla'}(X,Y)s &= R^{\nabla}(X,Y)s + (XA(Y))s - (YA(X))s - A([X,Y])s \\ &= R^{\nabla}(X,Y)s + dA(X,Y)s\end{aligned}$$

by the formula for the exterior differential of a one-form. □

Solution to Problem 2.59. The complex line bundle $L \to \mathbb{CP}^1$ is a sub bundle of the trivial bundle $\mathbb{C}^2$, and hence its pullback $\iota^* L \to \mathbb{C}$ is a sub bundle of the trivial bundle $\mathbb{C} \times \mathbb{C}^2$. More explicitly, the fiber $(\iota^* L)_z$ for $z \in C$ is the 1-dimensional subspace of $\mathbb{C}^2$ spanned by $\iota(z) = (1, z)$. In particular, we obtain a nowhere vanishing section

$$s_0 \colon \mathbb{C} \longrightarrow \iota^* L \subset \mathbb{C} \times \mathbb{C}^2 \qquad \text{by defining} \qquad s_0(z) := (1, z).$$

Strictly speaking, a point in the total space of the trivial bundle $\mathbb{C} \times \mathbb{C}^2 \to \mathbb{C}$ over $\mathbb{C}$ is a triple $(z; (w_0, w_1))$ with $z \in \mathbb{C}$, $(w_0, w_1) \in \mathbb{C}^2$. However, the first component isn't really relevant, since this is a trivial vector bundle, and so we suppress it in our notation.

For calculating the connection of $\iota^* L$ as a sub bundle of $\underline{\mathbb{C}}^2$, it will be convenient to calculate in terms of the complex coordinate functions $z = x + iy, \bar{z} = x - iy \in C^\infty(\mathbb{C}, \mathbb{C})$ and the corresponding vector fields

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \qquad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right).$$

These are *complex* vector fields, that is, sections of the complexification of the tangent bundle, but as discussed during the digression on connections on complex vector bundles 2.2.7, a connection on a complex vector bundle allows us to differentiate a section in the direction of the complex vector field.

Thinking of s_0 as a section of $\underline{\mathbb{C}}^2$, we can differentiate it and obtain

$$\left(\frac{\partial}{\partial z}s_0\right)(z) = \frac{\partial}{\partial z}(1,z) = (0,1) \qquad \left(\frac{\partial}{\partial \bar z}s_0\right)(z) = \frac{\partial}{\partial \bar z}(1,z) = (0,0)$$

This implies

$$\nabla_{\frac{\partial}{\partial z}} s_0 = \mathrm{proj}^L(\frac{\partial}{\partial z}s_0) = \mathrm{proj}^L(0,1) = \frac{\langle (0,1),(1,z)\rangle}{\langle (1,z),(1,z)\rangle}(1,z) = \frac{\bar z}{1+z\bar z}(1,z) = \frac{\bar z}{1+z\bar z}s_0 \quad (5.1)$$

and $\nabla_{\frac{\partial}{\partial \bar z}} s_0 = 0$. Here $\mathrm{proj}^L\colon \underline{\mathbb{C}}^2 \to L$ is the orthogonal projection to the subbundle $L \subset \underline{\mathbb{C}}^2$.

We can use the nowhere vanishing section s_0 to identify $\iota^* L$ with the trivial complex line bundle. This allows us to write the connection ∇ in the form

$$\nabla = \nabla^{\mathrm{taut}} + A \qquad \text{for } A \in \Omega^1(\mathbb{C};\mathbb{C}).$$

Here ∇^{taut} is the tautological connection on the trivial line bundle $\mathbb{C}$ which we consider as a connection on L via the bundle isomorphism $\underline{\mathbb{C}} \cong L$ determined by s_0. Since the constant section 1 of $\underline{\mathbb{C}}$ corresponds to s_0, we have $\nabla^{\mathrm{taut}}_X s_0 = 0$ for every vector field X on $\mathbb{C}$. It follows that

$$\nabla_{\frac{\partial}{\partial z}} s_0 = \nabla^{\mathrm{taut}}_{\frac{\partial}{\partial z}} s_0 + A(\frac{\partial}{\partial z})s_0 = A(\frac{\partial}{\partial z})s_0$$
$$\nabla_{\frac{\partial}{\partial \bar z}} s_0 = \nabla^{\mathrm{taut}}_{\frac{\partial}{\partial \bar z}} s_0 + A(\frac{\partial}{\partial z})s_0 = A(\frac{\partial}{\partial \bar z})s_0$$

Comparison with equation (5.1) then implies $A(\frac{\partial}{\partial z}) = \frac{\bar z}{1+z\bar z}$ and $A(\frac{\partial}{\partial \bar z}) = 0$, and we conclude

$$A = \frac{\bar z}{1+\bar z z}dz.$$

Then the curvature 2-form R is given by

$$R = dA = \frac{\partial}{\partial \bar z}\left(\frac{\bar z}{1+z\bar z}\right) d\bar z \wedge dz = \frac{1(1+z\bar z) - \bar z z}{(1+z\bar z)^2} d\bar z \wedge dz = \frac{1}{(1+z\bar z)^2} d\bar z \wedge dz.$$

To integrate this form over $\mathbb{C}$, we will use polar coordinates r, θ. These are related to the complex coordinates z, $\bar z$ via

$$z = x+iy = r\cos\theta + ir\sin\theta = re^{\theta}$$
$$\bar z = x-iy = r\cos\theta - ir\sin\theta = re^{-\theta}$$

It follows that

$$dz = \frac{\partial}{\partial r}(re^{i\theta})dr + \frac{\partial}{\partial \theta}(re^{i\theta})d\theta = e^{i\theta}dr + ire^{i\theta}d\theta$$
$$d\bar{z} = \frac{\partial}{\partial r}(re^{-i\theta})dr + \frac{\partial}{\partial \theta}(re^{-i\theta})d\theta = e^{-i\theta}dr - ire^{-i\theta}d\theta$$

It follows that

$$d\bar{z} \wedge dz = (e^{-i\theta}dr - ire^{-i\theta}d\theta) \wedge (e^{i\theta}dr + ire^{i\theta}d\theta) = 2irdr \wedge d\theta.$$

Finally, we obtain

$$\int_{\mathbb{C}} R = \int_{\mathbb{C}} \frac{1}{(1+z\bar{z})^2} d\bar{z} \wedge dz = 2i \int_{\mathbb{C}} \frac{r}{(1+r^2)^2} dr \wedge d\theta = 4\pi i \int_0^\infty \frac{r}{(1+r^2)^2} dr$$

Substituting $u = 1 + r^2$, $du = 2r$, this integral is equal to

$$2\pi i \int_1^\infty \frac{1}{u^2} du = 2\pi i [-u^{-1}]_1^\infty = 2\pi i.$$

□

References

[AD] Andersson, Lars; Driver, Bruce K. *Finite-dimensional approximations to Wiener measure and path integral formulas on manifolds.* J. Funct. Anal. 165 (1999), no. 2, 430–498.

[At] Atiyah, Michael *Topological quantum field theories.* Inst. Hautes tudes Sci. Publ. Math. No. 68 (1988), 175–186 (1989)

[BD] Beilinson, Alexander; Drinfeld, Vladimir, *Chiral algebras.* American Mathematical Society Colloquium Publications, 51. American Mathematical Society, Providence, RI, 2004. vi+375 pp.

[Be] Besse, Arthur L. *Einstein manifolds.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 10. Springer-Verlag, Berlin, 1987. xii+510 pp.

[Bo] R. Borcherds, *Vertex algebras, Kac-Moody algebras, and the Monster*, Proc. Natl. Acad. Sci. USA. vol. 83 (1986), 3068–3071

[CG] K. Costello and O. Gwilliam *Factorization algebras in quantum field theory*, draft monograph, available at `http://www.math.northwestern.edu/~costello/factorization.pdf`

[Gw] O. Gwilliam, *Factorization algebras and free field theories*, Northwestern thesis (2012), available at `http://math.berkeley.edu/~gwilliam/thesis.pdf`.

[JS] A. Joyal and R. Street, *Braided tensor categories*, Adv. Math. 102 (1993), no. 1, 20–78.

[Lu] Lurie, Jacob *On the classification of topological field theories*, Current developments in mathematics, 2008, 129?280, Int. Press, Somerville, MA, 2009.

[McL] S. Mac Lane, *Categories for the working mathematician*, Second edition. Graduate Texts in Mathematics, 5. Springer-Verlag, New York, 1998. xii+314 pp

[MS] Milnor, John W.; Stasheff, James D. *Characteristic classes*. Annals of Mathematics Studies, No. 76. Princeton University Press, Princeton, N. J.; University of Tokyo Press, Tokyo, 1974. vii+331 pp.

[Se2] G. Segal, *The definition of conformal field theory*, Topology, geometry and quantum field theory, 423 – 577, London Math. Soc. Lecture Note Ser., 308, Cambridge Univ. Press, Cambridge, 2004.

[ST] S. Stolz and P. Teichner, *Supersymmetric field theories and generalized cohomology*, Proc. Symp. Pure Math. Vol. 83, available at `http://arxiv.org/pdf/1108.0189.pdf`.