Chapter 2
Physics

2.1 Classical and Quantum Physics

2.1.1 Introduction

In this section, we will describe some important principles at a heuristic level. We hope this will be useful as a guide to some of the sequel which is more formal, but whenever the meaning of this section appears unclear, the reader should proceed to the more formal treatment below. There are many textbooks available on the mathematical aspects of quantum mechanics, for instance [53].

Classically, a particle is represented as or described by a point in some state space M. It moves in time along some trajectory $x(t)$ that is a solution of a system of second-order ODEs (a dot denoting a derivative with respect to time t),

$$\ddot{x}(t) = f(x, \dot{x}) \tag{2.1.1}$$

that is derived from an action principle. This principle consists in minimizing the Lagrangian action

$$S(x) := \int F(x(t), \dot{x}(t))\, dt, \tag{2.1.2}$$

the integral w.r.t. time over some Lagrangian that is a function of x and its first temporal derivative.[1] As will be discussed in more detail in Sect. 2.3.1 below, a minimizing $x(t)$ satisfies the corresponding Euler–Lagrange equations

$$\frac{d}{dt}\frac{dF}{d\dot{x}} - \frac{dF}{dx} = 0. \tag{2.1.3}$$

Here, the space M is d-dimensional, and in local coordinates $x = (x^1, \ldots, x^d)$. Alternatively, one may utilize the $2d$-dimensional phase space N with coordinates $(x^1, \ldots, x^d, x^{d+1} = \dot{x}^1, \ldots, x^{2d} = \dot{x}^d)$.

$\frac{dF}{d\dot{x}}$ stands for the covector of partial derivatives $(\frac{\partial F}{\partial \dot{x}^1}, \ldots, \frac{\partial F}{\partial \dot{x}^d})$. When one introduces this covector as a new variable, that is, puts

$$p := \frac{dF}{d\dot{x}}, \quad \text{i.e.,} \quad p_j := \frac{\partial F}{\partial \dot{x}^j} \tag{2.1.4}$$

[1] We consider here only the autonomous case; in the non-autonomous case, the density may also explicitly depend on t, $F(t, x(t), \dot{x}(t))$, and not only implicitly through its dependence on $x(t)$ and $\dot{x}(t)$.

J. Jost, *Geometry and Physics*,
DOI 10.1007/978-3-642-00541-1_2, © Springer-Verlag Berlin Heidelberg 2009

one arrives at the Hamiltonian formulation. This involves the Hamiltonian

$$H(p,x) := \dot{x}\frac{dF}{d\dot{x}} - F \tag{2.1.5}$$

where $\dot{x}\frac{dF}{d\dot{x}} = \sum_j \dot{x}^j \frac{\partial F}{\partial \dot{x}^j} = \sum_j \dot{x}^j p_j$. A solution is then obtained from the Hamilton equations

$$\dot{p} = -\frac{dH}{dx}, \qquad \dot{x} = \frac{dH}{dp}. \tag{2.1.6}$$

The Hamiltonian formalism singles out time and is therefore not relativistically invariant. Consequently, in our treatment of QFT, we shall mainly employ the Lagrangian formalism.

The standard example is

$$F = \frac{m}{2}|\dot{x}|^2 - V(x), \tag{2.1.7}$$

where m is the mass of the particle and V the potential. The Euler–Lagrange equations (2.1.3) are then

$$m\ddot{x} = -\frac{dV}{dx} \tag{2.1.8}$$

(in components: $m\ddot{x}^i = -\frac{\partial V}{\partial x^i}$). The Hamiltonian is then

$$H = \frac{m}{2}|\dot{x}|^2 + V(x) = \frac{p^2}{2m} + V(x), \tag{2.1.9}$$

and (2.1.6) becomes

$$\dot{p} = -\frac{dV}{dx}, \qquad \dot{x} = \frac{p}{m}. \tag{2.1.10}$$

For a solution $(x(t), p(t))$ of (2.1.6), we can then also compute the time evolution of any function $A(x, p)$ via

$$\frac{dA}{dt} = \frac{\partial A}{\partial x^i}\dot{x}^i + \frac{\partial A}{\partial p^i}\dot{p}^i = \frac{\partial A}{\partial x^i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial x^i} =: \{A, H\}, \tag{2.1.11}$$

where the last expression is called the Poisson bracket. It satisfies all the properties of a Lie bracket, as well as the canonical relations (Heisenberg commutation relations)

$$\{x^i, x^j\} = 0 = \{p_i, p_j\} \quad \text{and} \quad \{x^i, p_j\} = \delta^i_j. \tag{2.1.12}$$

Equation (2.1.11) is obviously a generalization of (2.1.6) (in the sense that $\dot{x}^i = \{x^i, H\}$, $\dot{p}_j = \{p_j, H\}$), and it also tells us that conserved quantities, that is, time-independent quantities, are precisely those whose Poisson bracket with the Hamiltonian H vanishes.

Quantum mechanics was discovered by Heisenberg and developed with Born and Jordan as the description of quantum theory through the correspondence with classical mechanics via matrix algebra. We now describe this, employing more modern terminology, of course. Quantum mechanically, in place of a point x in M, we have a probability distribution $|\phi(x)|^2$ derived from a function $\phi : M \to \mathbb{C}$ with

$$\|\phi\|_{L^2}^2 \left(= \int_M |\phi|^2(x) dvol(x) \right) = 1. \tag{2.1.13}$$

$|\phi(x)|^2$ can thus be interpreted as the probability density for finding the particle under consideration at the point x. Here, for the L^2-norm, we need a volume form $dvol$ on M; that volume form could come from a Riemannian metric. The classical case is recovered as the limit where this probability distribution becomes concentrated at a single point, that is, a delta function(al). In quantum mechanics, the observables are self-adjoint operators on the Hilbert space $\mathcal{H} := L^2(M, \mathbb{C})$. As self-adjoint operators, they have a purely real spectrum. The eigenvalues corresponding to eigenstates of such an operator then represent sharp observations. These operators, however, are typically unbounded which leads to certain mathematical difficulties, as will be described in more detail in Sect. 2.1.3 below.

In the formalism of canonical quantization, the momentum p_j becomes the operator $\frac{\hbar}{i}\frac{\partial}{\partial x^j}$. The total energy, the Hamilton function above, thus also becomes an operator, the Hamiltonian operator H, and the state ϕ evolves in time t according to the Schrödinger equation

$$i\hbar \frac{\partial \phi(x,t)}{\partial t} = H\phi(x,t). \tag{2.1.14}$$

For the Lagrangian (2.1.7), the Schrödinger equation (2.1.14) becomes

$$i\hbar \frac{\partial \phi(x,t)}{\partial t} = -\frac{\hbar^2}{2m}\Delta\phi(x,t) + V(x)\phi(x,t). \tag{2.1.15}$$

The ansatz $\phi(x,t) = \phi(x)\exp(-\frac{i}{\hbar}Et)$ of separated variables leads to

$$-\frac{\hbar^2}{2m}\Delta\phi(x) + V(x)\phi(x) = E\phi(x), \tag{2.1.16}$$

the time-independent Schrödinger equation.

We can arrive at (2.1.15) from the ansatz of representing $\phi(x,t)$ as a wave:

$$\phi(x,t) = \frac{1}{2\pi^{3/2}} \exp \frac{i}{\hbar}(p_\nu x^\nu - Et) =: \langle x,t|p,E\rangle, \tag{2.1.17}$$

where we have already introduced Dirac's notation to be explained below. Then

$$\frac{\partial}{\partial x^\nu}\phi(x,t) = \frac{i}{\hbar}p_j\phi(x,t), \tag{2.1.18}$$

$$\frac{\partial}{\partial t}\phi(x,t) = -\frac{i}{\hbar}E\phi(x,t). \tag{2.1.19}$$

On the basis of this computation, we then put

$$\frac{\partial}{\partial x^\nu} = \frac{i}{\hbar}p_j, \tag{2.1.20}$$

$$\frac{\partial}{\partial t} = -\frac{i}{\hbar}E. \tag{2.1.21}$$

For the Hamiltonian (2.1.9), we are then naturally led to (2.1.15), and with the ansatz $\phi(x,t) = \phi(x)\exp(-\frac{i}{\hbar}Et)$ at (2.1.16).

Remark We use here the so-called Schrödinger picture where the states ϕ are evolving in time. In the complementary Heisenberg picture, instead the observables, represented as self-adjoint operators A, evolve according to

$$i\hbar\frac{dA}{dt} = [A, H], \tag{2.1.22}$$

in analogy to (2.1.11), see (2.1.92), (2.1.93).

In the quantum mechanical view, the field $\phi : M \to \mathbb{C}$ is obtained from the quantization of a point particle. There is, however, another interpretation of ϕ that turns out to be more fruitful for our purposes. Namely, we can view ϕ also as a classical field on M. It then need no longer satisfy the normalization $\|\phi\|_{L^2} = 1$. Also, it need no longer take its values in $\mathbb{C}$ only, but it can also assume values in the fibers of some vector or principal bundle or some manifold. As a classical situation, it can then be quantized again, and one then speaks of a second or field quantization. The analog of the Schrödinger equation is then a PDE on some function space, that is, a PDE with infinitely many variables.

There is an important generalization of this picture: When the particle possesses some internal symmetry, described by some Lie group G, the space $\mathbb{C}$ gets replaced by a (Hermitian) vector space that carries a (unitary) representation of G. Thus, a particle is described by some $\psi \in L^2(M, V)$, again of norm 1, so that $\|\psi\|^2$ (where $\|.\|$ is the Hermitian norm) can again be interpreted as a probability density. The vector space V enters here in order to distinguish different states that are not G-invariant, as G leaves the space V invariant, but not the individual elements of V. This is needed because not all physical forces will be G-invariant. An example is the electron with its spin. Since there are only two possible values of the spin, here the vector space is finite, $\mathbb{Z}_2$, and the corresponding Hilbert space is finite-dimensional, $\mathbb{C}^2$. Quantum electrodynamics (QED) then couples the Maxwell equation with the Dirac equation for the electron spin on a relativistic space time. The standard model of elementary particle physics interprets the observed multitude of particles through symmetry breaking from some encompassing Lie group $\bar{G}$ that contains all the symmetry groups of the individual particles. Of course, we shall explain this in more detail below. A quick and useful introduction to the topics of this section can be found in [90].

2.1.2 Gaussian Integrals and Formal Computations

Before proceeding with quantum physics, we introduce a basic formal tool, Gaussian integrals, that serve as a heuristic transmission line from finite-dimensional exponential integrals to infinite-dimensional functional integrals.

We start with the bosonic case. Let A be a symmetric $n \times n$-matrix with eigenvalues

$$\lambda_i > 0, \tag{2.1.23}$$

and let b be a vector.

The Gaussian integral ($x = (x^1, \dots, x^n)$) is

$$I(A) := \int \exp\left(-\frac{1}{2} x^t A x\right) dx^1 \cdots dx^n = \left(\frac{(2\pi)^n}{\det A}\right)^{\frac{1}{2}} \tag{2.1.24}$$

with $\det A = \prod_{i=1}^n \lambda_i$, as follows easily by diagonalizing A. A formal extension of this formula to infinite dimensions is often based on expressing the determinant of A in terms of a zeta function; we define the zeta function of the operator A as

$$\zeta_A(s) := \sum_{k=1}^n \frac{1}{\lambda_k^s}, \quad \text{for } s \in \mathbb{C}. \tag{2.1.25}$$

Since $\lambda_k^{-s} = e^{-s \log s}$, we obtain for the derivative of the zeta function

$$\zeta_A'(s) = -\sum_{k=1}^n \frac{\log \lambda_k}{\lambda_k^s}. \tag{2.1.26}$$

Therefore, we can express the determinant of A in terms of the derivative of the zeta function at 0:

$$\det A = \prod_{i=1}^n \lambda_i = e^{-\zeta_A'(0)}. \tag{2.1.27}$$

The general Gaussian integral

$$I(A, b) := \int dx^1 \cdots dx^n \exp\left(-\frac{1}{2} x^t A x + b^t x\right) \tag{2.1.28}$$

is reduced to this case by putting

$$x := A^{-1} b + y$$

(note that $x_0 := A^{-1}b$ minimizes the quadratic form $\frac{1}{2}x^t Ax - b^t x$). Namely, we obtain

$$\begin{aligned} I(A,b) &= \exp\left(\frac{1}{2}b^t A^{-1}b\right)\int \exp\left(-\frac{1}{2}y^t A\, y\right)dy^1\cdots dy^n \\ &= \exp\left(\frac{1}{2}b^t A^{-1}b\right)\left(\frac{(2\pi)^n}{\det A}\right)^{\frac{1}{2}} \\ &= (2\pi)^{n/2}\exp\left(\frac{1}{2}b^t A^{-1}b\right)e^{\frac{1}{2}\zeta_A'(0)}, \end{aligned} \tag{2.1.29}$$

using (2.1.27) for the last line.

In many cases, the vector b has an auxiliary or dummy role. Namely, we wish to compute moments

$$\begin{aligned} \langle x^{i_1}\cdots x^{i_m}\rangle &:= \frac{\int x^{i_1}\cdots x^{i_m}\exp(-\frac{1}{2}x^t Ax)\,dx^1\cdots dx^n}{\int \exp(-\frac{1}{2}x^t Ax)\,dx^1\cdots dx^n} \\ &= \frac{1}{I(A)}\frac{\partial}{\partial b^{i_1}}\cdots\frac{\partial}{\partial b^{i_m}}I(A,b)|_{b=0}. \end{aligned} \tag{2.1.30}$$

In particular, the second-order moment or propagator is

$$\langle x^i x^j\rangle = \left(A^{-1}\right)_{ij}. \tag{2.1.31}$$

When m is odd, the moment (2.1.30) vanishes because the (quadratic) exponential is even at $b = 0$.

For even m, we have Wick's theorem

$$\langle x^{i_1}\cdots x^{i_m}\rangle = \sum_{\substack{\text{all possible}\\ \text{pairings of}\\ (i_1,\ldots,i_m)}} \left(A^{-1}\right)_{i_{p_1}i_{p_2}}\cdots\left(A^{-1}\right)_{i_{p_{m-1}}i_{p_m}} \tag{2.1.32}$$

as follows directly from (2.1.29) and (2.1.30).

As a preparation for the functional integrals to follow, we now wish to consider x^i as an operator on the finite-dimensional Hilbert space $E^n = \mathbb{R}^n$ with its Euclidean product. We then have the matrix elements

$$\begin{aligned} &\langle e_i|x^{i_1}\cdots x^{i_m}|e_j\rangle \\ &\quad = \frac{\int x^{i_1}\cdots x^{i_m}\delta(x^i-1)\delta(x^j-1)\exp(-\frac{1}{2}x^t Ax)\,dx^1\cdots dx^n}{\int \exp(-\frac{1}{2}x^t Ax)\,dx^1\cdots dx^n}. \end{aligned}$$

Instead of the δ-functions, one can then also make arbitrary insertions into the functional integral, that is, functions of x.

In the Grassmann case, we start with a Grassmann algebra generated by $\eta^1, \dots, \eta^n, \bar{\eta}^1, \dots, \bar{\eta}^n$

$$\begin{aligned} J_0(A) &:= \int d\eta^1\, d\bar{\eta}^1 \cdots d\eta^n\, d\bar{\eta}^n \exp\bigl(-\bar{\eta}^t A \eta\bigr) \\ &= \int d\eta^1\, d\bar{\eta}^1 \cdots d\eta^n\, d\bar{\eta}^n \prod_{i=1}^{n} \prod_{j=1}^{n} \bigl(1 - \bar{\eta}^i A_{ij} \eta^j\bigr) \\ &= \sum_{\text{permutations } p} \operatorname{sign}(p) A_{1p(1)} \cdots A_{np(n)} \\ &= \det A. \end{aligned} \tag{2.1.33}$$

We next compute, for a Grassmann algebra generated by $\vartheta^1, \dots, \vartheta^{2n}$,

$$J(A) = \int d\vartheta^1 \cdots d\vartheta^{2n} \exp\left(-\frac{1}{2} \vartheta^t A \vartheta\right). \tag{2.1.34}$$

We may assume that A is antisymmetric, as the symmetric terms cancel because the ϑ's anticommute:

$$\begin{aligned} J(A) &= \int d\vartheta^1 \cdots d\vartheta^{2n} \prod_{i<j} \bigl(1 - \vartheta^i A_{ij} \vartheta^j\bigr) \\ &= \sum_{\substack{\text{permutations } p \\ \text{with } p(2i-1)<p(2i), \\ p(2i-1)<p(2i+1) \\ \text{for } i=1,\dots,n, \text{ or } n-1, \text{ resp.}}} \operatorname{sign}(p) A_{p(1)p(2)} A_{p(3)p(4)} \cdots A_{p(n-1)p(n)} \\ &=: \operatorname{Pf}(A) \quad \text{(Pfaffian)}. \end{aligned} \tag{2.1.35}$$

We have

$$J^2(A) = \int d\vartheta^1 \cdots d\vartheta^{2n}\, d\vartheta'^1 \cdots \vartheta'^{2n} \exp\left(-\frac{1}{2}\bigl(\vartheta^t A \vartheta + \vartheta'^t A \vartheta'\bigr)\right). \tag{2.1.36}$$

The coordinate transformation

$$\begin{aligned} \eta^k &:= \frac{1}{\sqrt{2}}\bigl(\vartheta^k + i\vartheta'^k\bigr), \\ \bar{\eta}^k &:= \frac{1}{\sqrt{2}}\bigl(\vartheta^k - i\vartheta'^k\bigr), \end{aligned}$$

has the Jacobian $(-1)^n$ and satisfies $\vartheta^i \vartheta^j + \vartheta'^i \vartheta'^j = \bar{\eta}^i \eta^j - \bar{\eta}^j \eta^i$.

Using the antisymmetry of A, we obtain

$$J^2(A) = \int d\eta^1 \, d\bar{\eta}^1 \cdots d\eta^n \, d\bar{\eta}^n \exp\bigl(-\bar{\eta}^t A \eta\bigr)$$
$$= \det A \tag{2.1.37}$$

from (2.1.33). From (2.1.34), (2.1.35), (2.1.37) we see

$$\mathrm{Pf}^2(A) = \det A, \tag{2.1.38}$$

that is,

$$J(A) = (\det A)^{\frac{1}{2}}. \tag{2.1.39}$$

As in the ordinary case, we also have

$$J(A,b) = \int d\vartheta^1 \cdots d\vartheta^{2n} \exp\left(-\frac{1}{2}\vartheta^t A \vartheta + b^t \vartheta\right)$$
$$= J(A) \exp\left(\frac{1}{2} b^t A^{-1} b\right) \tag{2.1.40}$$

and likewise

$$\langle \vartheta^i \vartheta^j \rangle = \frac{1}{J(A)} \frac{\partial}{\partial b_j} \frac{\partial}{\partial b_i} J(A,b)_{|_{b=0}} = (A^{-1})_{ij}.$$

Another formal tool that is useful in this context are formal computations with Dirac functions. In the physics literature, linear functionals on space of functions are systematically expressed by their integral kernels. Thus, the evaluation of a function φ at a point y

$$\varphi \mapsto \varphi(y) \tag{2.1.41}$$

is written in terms of the Dirac δ-functional

$$\varphi(y) = \int dz \varphi(z) \delta(z-y) = \delta_y(\varphi). \tag{2.1.42}$$

If we change the variable $z = f(w)$, this becomes

$$\varphi(y) = \int dw \left| \det \frac{\partial f}{\partial w} \right| \varphi(f(w)) \delta(f(w) - y). \tag{2.1.43}$$

This formula is useful for calculating $\varphi(f(w))$ at $f(w) = y$, without having to solve the latter equation explicitly.

In analogy to (2.1.42), we also have

$$\frac{\partial}{\partial y^j} \varphi(y) = \int dz \frac{\partial}{\partial y^j} \varphi(z) \delta(z-y) = -\int dz \varphi(z) \frac{\partial}{\partial y^j} \delta(z-y) \tag{2.1.44}$$

if we formally integrate by parts. Thus, we can define the derivative $\frac{\partial}{\partial y^j}\delta(z-y)$ of the delta function $\delta(z-y)$ as the functional

$$\varphi \mapsto -\frac{\partial}{\partial y^j}\varphi(y). \tag{2.1.45}$$

We now consider a functional Φ

$$\varphi \mapsto \Phi(\varphi)$$

defined on some Banach or Fréchet space $\mathcal{B}$. The Gateaux derivative $\delta\Phi$ in the direction $\delta\varphi$ is defined as

$$\delta\Phi(\varphi)(\delta\varphi) := \frac{\delta\Phi(\varphi)}{\delta\varphi} := \lim_{t\to 0}\frac{1}{t}(\Phi(\varphi + t\delta\varphi) - \Phi(\varphi)), \tag{2.1.46}$$

provided this limit exists. Here, $\delta\varphi$ is, of course, also assumed to be in $\mathcal{B}$. When the limit in (2.1.46) exists uniformly for all variations $\delta\varphi$ in some neighborhood of 0, we speak of a Fréchet derivative. Some formal examples:

$$\delta\varphi(f)^n(\delta\varphi) = n\delta\varphi(f)\varphi(f)^{n-1}, \tag{2.1.47}$$

$$\delta e^{\varphi(f)}(\delta\varphi) = \delta\varphi(f)e^{\varphi(f)}. \tag{2.1.48}$$

We are usually interested in Lagrangian functionals,

$$L(u) := \int F(\xi, u(\xi), du(\xi))\, d\xi. \tag{2.1.49}$$

L is usually defined on some Sobolev space of functions. We then have

$$\delta L(u)(\delta u) = \frac{\delta L(u)}{\delta u} = \frac{d}{ds}\int F(\xi, u(\xi) + s\delta u(\xi), d(u(\xi) + s\delta u(\xi)))\, d\xi_{|s=0}. \tag{2.1.50}$$

The question arises as to which variations δu one may take here. One class of variations is given by the test functions, that is, the functions from the space $\mathcal{D} := C_0^\infty$ of infinitely often differentiable functions with compact support. That space is not a Banach space, but only a limit of Fréchet spaces with topology generated by the seminorms $|f|_{k,K} := \sup_{x\in K}|D^k f(x)|$ for nonnegative integers k and compact sets K. Its dual space $\mathcal{D}'$, that is, the space of continuous functionals on $\mathcal{D}$, is the space of distributions. The best-known distribution is of course the Dirac distribution already displayed in (2.1.42) above. Conceiving the Dirac functional as an element of $\mathcal{D}'$, that is, as an operation on smooth functions (in fact, continuous functions are good enough here), is the Schwartz point of view. A different point of view, which does not need topologies with unpleasant properties (for example, the implicit function theorem is very cumbersome in Fréchet spaces) and is more useful in nonlinear analysis, is the one of Friedrichs, which considers the Dirac function as a limit of smooth integral kernels with compact support that in the limit shrinks to

a point. One then, in effect, never needs to carry out any formal manipulation with the Dirac function(al) itself, but only ones with such smooth integral kernels. The Dirac point of view, finally, simply performs formal operations with the Dirac function. Thus, the different approaches consist in justifying, avoiding, or performing the Dirac function computations. The Dirac approach is prominent in the physics literature, and we shall also follow that here, because we are assured that these operations can be made mathematically rigorous by either of the other two approaches.

Having said that, we then also take functional derivatives in the direction of Dirac functions. That means considering

$$\frac{\delta\Phi}{\delta\varphi(z)} := \lim_{t\to 0}\frac{1}{t}\left(\Phi(\varphi(y)+t\delta(y-z))-\Phi(\varphi(y))\right). \tag{2.1.51}$$

For (2.1.49), (2.1.50), we then have the formal relation

$$\delta L(\delta u) = \int d\xi\,\delta u(\xi)\frac{\delta L}{\delta u(\xi)}, \tag{2.1.52}$$

with the consistency relation

$$\frac{\delta L}{\delta u(z)} = \int d\xi\,\delta(z-\xi)\frac{\delta L}{\delta u(\xi)}. \tag{2.1.53}$$

Thus, $\frac{\delta L}{\delta u(z)}$ measures the response of L to a change in u supported at z.

We also have

$$\frac{\delta}{\delta\varphi(z)}\varphi(x) = \delta(x-z) \tag{2.1.54}$$

and

$$\frac{\delta}{\delta\varphi(z)}\int dx\varphi^n(x) = n\varphi^{n-1}(z). \tag{2.1.55}$$

Looking at (2.1.42), (2.1.51), (2.1.54), we see that the operations with the Dirac δ-function are simply formal extensions of the ones with the Kronecker symbol in the finite-dimensional case.

In the Grassmann case, the Dirac δ-function is

$$\delta(\vartheta) = \vartheta = \int d\eta\exp(\eta\vartheta) \tag{2.1.56}$$

satisfying

$$\int d\vartheta\,\delta(\vartheta)f(\vartheta) = f(0), \quad \text{for } f(\vartheta) = f(0)+a\vartheta. \tag{2.1.57}$$

For more details on the formal calculus, see [114].

2.1.3 Operators and Functional Integrals

In this section, we want to amplify the discussion of Sect. 2.1.1 and introduce path integrals. We want to investigate the time evolution of a quantized particle. This is described by a complex-valued wave function $\phi(x,t)$ whose squared norm $|\phi(x,t)|^2$ represents the probability density for finding the particle at time t at the position $x \in M$. Here, $\phi(x,t)$ is assumed to be an L^2-function of x so that the total probability can be normalized:

$$\int_M |\phi(x,t)|^2 dvol(x) = 1 \tag{2.1.58}$$

for all t. For a measurable subset B of M, the probability for finding the particle in B at time t is then given by

$$\int_B |\phi(x,t)|^2 dvol(x). \tag{2.1.59}$$

More abstractly, a pure state $|\psi\rangle$ of a quantum mechanical system is a one-dimensional subspace, which we then represent by a unit vector ψ, in some Hilbert space $\mathcal{H}$. The scalar product is written as $\langle\phi|\psi\rangle$; here, by duality, we may also consider $\langle\phi|$ as an element of the dual space $\mathcal{H}^*$. For a pure state ψ, we let P_ψ be the projection onto the one-dimensional subspace defined by ψ. As a projection, P_ψ is idempotent, that is, $P_\psi^2 = P_\psi$. Then

$$|\langle\phi,\psi\rangle|^2 = \langle P_\phi\psi,\psi\rangle = \operatorname{tr} P_\phi P_\psi \tag{2.1.60}$$

is the probability of finding the system in the state ϕ when knowing that it is in the state ψ. Let us assume that for some map T on the states of $\mathcal{H}$, we have

$$|\langle T\phi, T\psi\rangle|^2 = |\langle\phi,\psi\rangle|^2 \tag{2.1.61}$$

for all ϕ,ψ, that is, the probabilities are unchanged by applying T to all states. By a theorem of Wigner, T can then be represented by a unitary or antiunitary operator U_T of $\mathcal{H}$, that is $T\psi = U_T\psi$ for all ψ.[2]

The observables are self-adjoint (Hermitian) operators A on $\mathcal{H}$, typically unbounded. Being self-adjoint, their spectrum is real. The state $|\psi\rangle$ then also defines an observable, the projection P_ψ. The expectation value of the observable A in the state $|\psi\rangle$ is given by

$$\langle\psi, A\psi\rangle = \operatorname{tr} A P_\psi \tag{2.1.62}$$

[2] In particular, connected groups of automorphisms G of $\mathcal{H}$ are represented by unitary transformations of $\mathcal{H}$—with the following note of caution: U_T is determined by T only up to multiplication by a factor of norm 1. Therefore, in general, we only obtain a projective representation of G, that is, we only obtain the group law $U_{gh} = c(g,h)U_gU_h$ for some scalar factor $c(g,h)$ of absolute value 1. It is, however, possible, to obtain an honest unitary representation by enlarging the group G.

(assuming that ψ is contained in the domain of definition of A). This includes (2.1.60) as a special case. When ψ is an eigenstate of A, that is,

$$A\psi = \lambda\psi \tag{2.1.63}$$

for some (real) eigenvalue λ, then this λ is the expectation value of A in the state ψ. The variance of the probability distribution for the observations of the values of A for a system in state ψ is then

$$\langle\psi, A^2\psi\rangle - \langle\psi, A\psi\rangle^2. \tag{2.1.64}$$

This variance vanishes iff (2.1.63) holds, that is, iff ψ is an eigenstate of A. That means that an observable A takes precise values precisely on its eigenstates.

Let G be a group, like $SO(3)$ or $SU(2)$, acting on $\mathcal{H}$ by unitary transformations. An observable A is called a scalar operator when it commutes with the action of G. Then, if ψ is an eigenstate of A with eigenvalue λ, for all $g \in G$,

$$Ag\psi = gA\psi = g\lambda\psi = \lambda g\psi. \tag{2.1.65}$$

Thus, the space of eigenstates with eigenvalue λ is invariant under the action of G. As such an invariant subspace, it could be reducible or irreducible. In the latter case, the degeneracy of the eigenvalue λ equals the dimension of the corresponding irreducible representation of G. These dimensions are known by representation theory, see [45, 75]. If one then perturbs the operator A to an operator A' that is no longer invariant under the action of G, the multiplicity of the eigenvalue λ will decrease. This is important for understanding many experimental results. The operator A might be the Hamiltonian H_0 of a system invariant under some group G, say of spatial rotations. Its eigenvalues are the energy levels, and because of the invariance, they are degenerate. H_0 then is perturbed to $H = H_0 + H_1$ by some external magnetic field in some direction which then destroys rotational invariance. Then the energy levels, the eigenvalues of the new Hamiltonian H split up into several values. Often H_1 is small compared to H_0, and one can then approximate these energy levels by a perturbative expansion of H around H_0.

We return to the general theory. When the spectrum of A is discrete, and $|a\rangle$ runs through a complete set of orthonormal eigenvectors of A, we have the relation

$$\sum_{|a\rangle} |a\rangle\langle a| = \mathrm{id}, \tag{2.1.66}$$

the identity operator on $\mathcal{H}$. Applying this to $|\psi\rangle \in \mathcal{H}$ yields

$$\sum_{|a\rangle} |a\rangle\langle a|\psi\rangle = |\psi\rangle, \tag{2.1.67}$$

which is simply the expansion of $|\psi\rangle$ in terms of a Hilbert space basis. When the eigenvalue of A corresponding to the eigenstate $|a\rangle$ is denoted by a, that is, $A|a\rangle =$

$a|a\rangle$, we have the relationships

$$A|\psi\rangle = \sum_{|a\rangle} A|a\rangle\langle a|\psi\rangle = \sum_{|a\rangle} a|a\rangle\langle a|\psi\rangle \quad \text{and} \tag{2.1.68}$$

$$\langle\phi|A|\psi\rangle = \sum_{|a\rangle}\langle\phi|A|a\rangle\langle a|\psi\rangle = \sum_{|a\rangle} a\langle\phi|a\rangle\langle a|\psi\rangle. \tag{2.1.69}$$

When the spectrum of the operator A is continuous, the sums in the preceding relations are replaced by integrals; for instance

$$\langle\phi|A|\psi\rangle = \int_a da\, a\langle\phi|a\rangle\langle a|\psi\rangle. \tag{2.1.70}$$

This is rigorously investigated in von Neumann's spectral theory of (unbounded) self-adjoint operators on Hilbert spaces. Let us briefly describe this, referring e.g. to [100, 110] for details (the knowledgeable reader may of course skip this).

A family of projections $E(a)$ $(-\infty < a < \infty)$ in a Hilbert space $\mathcal{H}$ is called a resolution of the identity iff for all $a, b \in \mathbb{R}$

1.

$$E(a)E(b) = E(\min(a, b)), \tag{2.1.71}$$

2.

$$E(-\infty) = 0, \qquad E(\infty) = Id \tag{2.1.72}$$

(here, $E(-\infty)|\phi\rangle := \lim_{a\downarrow-\infty} E(a)|\phi\rangle$, $E(\infty)|\phi\rangle := \lim_{a\uparrow\infty} E(a)|\phi\rangle$ for $|\phi\rangle \in \mathcal{H}$),

3.

$$E(a+0) = E(a) \tag{2.1.73}$$

$(E(a+0)|\phi\rangle := \lim_{b\downarrow 0} E(b)|\phi\rangle)$.

For a continuous function $f : \mathbb{R} \to \mathbb{C}$, one can then define

$$\int_{A_1}^{A_2} f(a)\, dE(a)|\phi\rangle := \lim_{\max|a_{k+1}-a_k|\to 0} \sum_k f(\alpha_k)(E(a_{k+1}) - E(a_k))|\phi\rangle \tag{2.1.74}$$

for $A_1 = a_1 < a_2 \cdots < a_n = A_2$ and $\alpha_k \in (a_k, a_{k+1}]$ (a limit of Riemann sums), and

$$\int f(a)\, dE(a)|\phi\rangle := \lim_{A_1\downarrow-\infty, A_2\uparrow\infty} \int_{A_1}^{A_2} f(a)\, dE(a)|\phi\rangle \tag{2.1.75}$$

whenever that limit exists. This is the case precisely if

$$\int |f(a)|^2\, d\|E(a)|\phi\rangle\|^2 < \infty \tag{2.1.76}$$

where $\|.\|$ is the norm in the Hilbert space $\mathcal{H}$. For such $|\phi\rangle$,

$$|\psi\rangle \to \int f(a)\, d\langle\psi|E(a)|\phi\rangle \tag{2.1.77}$$

defines a bounded linear functional on $\mathcal{H}$. In other words, we have a self-adjoint operator G with

$$\langle\psi|G|\phi\rangle = \int f(a)\, d\langle\psi|E(a)|\phi\rangle \tag{2.1.78}$$

defined for those $|\phi\rangle$ and all $|\psi\rangle \in \mathcal{H}$. The central result is that every self-adjoint operator A on the Hilbert space $\mathcal{H}$ admits a unique spectral resolution, that is, can be uniquely written as

$$A = \int a\, dE(a), \tag{2.1.79}$$

in the sense that

$$\langle\psi|A|\phi\rangle = \int a\, d\langle\psi|E(a)|\phi\rangle \tag{2.1.80}$$

for all $|\phi\rangle$ in the domain of definition of A and all $|\psi\rangle \in \mathcal{H}$. This is the meaning of (2.1.70).

On this basis, one can define the functional calculus for self-adjoint operators and put

$$f(A) := \int f(a)\, dE(a) \tag{2.1.81}$$

for a function $f : \mathbb{R} \to \mathbb{C}$ when $(E(a))$ is the spectral resolution of A. $f(A)$ is then also a self-adjoint operator. When f is an exponential function, for example, this leads to the same result as defining e^A directly through the power series of the exponential function.

The correspondence between classical and quantum mechanics consists in the requirement that the quantum mechanical operators $\hat{x}$, $\hat{p}$ corresponding to position x and momentum p satisfy the operator analogs of (2.1.12), that is,

$$[\hat{x}^i, \hat{x}^j] = 0 = [\hat{p}_i, \hat{p}_j] \quad \text{and} \quad [\hat{x}^j, \hat{p}_k] = i\hbar\delta^j_k, \tag{2.1.82}$$

with the commutator of operators,

$$[A, B] = AB - BA \tag{2.1.83}$$

in place of the Poisson bracket. The factor i in (2.1.82) comes from the fact that the commutator of two Hermitian operators is skew Hermitian.

$|x\rangle$ then denotes the state where the particle is localized at the point $x \in M$, i.e., the probability to find it at x is 1, and 0 elsewhere. When M is the real line $\mathbb{R}$, that is, one-dimensional, this is an eigenstate of the position operator $\hat{x}$, that is,

$$\hat{x}|x\rangle = |x\rangle x \tag{2.1.84}$$

corresponding to the eigenvalue x. In $\mathbb{R}^d$, the components x^i, $i = 1, \ldots, d$, are the eigenvalues of the corresponding operators $\hat{x}^i$, and

$$\hat{x}^i|x\rangle = |x\rangle x^i. \tag{2.1.85}$$

One should note that these operators are unbounded, and in our above L^2-space, the states $|x\rangle$ are represented by Dirac functionals $\delta(x)$, that is, they are *not* L^2-functions. Functional analysis provides concepts for making this entirely rigorous. In fact, according to spectral theory as presented above, we consider the Hilbert space $L^2(\mathbb{R})$ and write the operator as

$$A|\phi\rangle(x) = x|\phi\rangle(x), \tag{2.1.86}$$

which then admits the spectral resolution

$$A = \int a\, dE(a) \tag{2.1.87}$$

with

$$E(a)|\phi\rangle(x) = \begin{cases} |\phi\rangle(x) & \text{for } x \le a, \\ 0 & \text{for } x > a. \end{cases} \tag{2.1.88}$$

Returning to (2.1.84), we see that the spectrum of the position operator $\hat{x}$ on $\mathbb{R}$ consists of the entire real line.

The established notation usually leaves out the hats, that is, writes x both for the position at a point in M and the corresponding operator on $\mathcal{H}$ that has been called $\hat{x}$ in (2.1.84). We shall also do that from this point on.

With these conventions, the Schrödinger equation (2.1.14) becomes

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle. \tag{2.1.89}$$

H, the Hamiltonian, here is a self-adjoint (Hermitian) operator, that is, an observable, in fact the most basic one of the whole theory.

The solution of (2.1.89) can be expressed by functional calculus as

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}tH}|\psi(0)\rangle. \tag{2.1.90}$$

Here the exponential of $-H$ is defined through the usual power series of the exponential function, or better, via (2.1.81). Since H is Hermitian, the operators $e^{-\frac{i}{\hbar}tH}$ are unitary. Thus, the state $|\psi\rangle$ evolves by unitary transformations.

By taking $\frac{\partial}{\partial t}$ of (2.1.90), we see that formally it satisfies the Schrödinger equation (2.1.89), indeed. From (2.1.90), we also infer the semigroup property

$$|\psi(t)\rangle = e^{-\frac{i}{\hbar}(t-\tau)H}|\psi(\tau)\rangle = e^{-\frac{i}{\hbar}(t-\tau)H}|e^{-\frac{i}{\hbar}\tau H}|\psi(0)\rangle. \tag{2.1.91}$$

Thus, the solution at time t is obtained from the solution at time τ by applying the solution operator for the remaining time $t-\tau$ $(0 < \tau < t)$. We express the relation between (2.1.110) and (2.1.109) also by saying that $-\frac{i}{\hbar}H$ is the infinitesimal generator of the semigroup $e^{-\frac{i}{\hbar}tH}$. For an account of the mathematical theory of semigroups for partial differential equations, we refer to [63].

In this Schrödinger picture, the states evolve in time, whereas the observables don't. In the Heisenberg picture, this relation is reversed. The states are time-independent, whereas the operators representing the observables change in time, according to

$$i\hbar\frac{dA}{dt} = [A, H] \tag{2.1.92}$$

whose solution is

$$A(t) = e^{\frac{i}{\hbar}Ht}A(0)e^{-\frac{i}{\hbar}Ht}. \tag{2.1.93}$$

The Schrödinger and the Heisenberg picture are equivalent insofar as they yield the same probability density for the outcome of observations. This is expressed by the relation

$$\langle\phi(t)|A(0)|\psi(t)\rangle = \langle\phi(0)|e^{\frac{i}{\hbar}Ht}A(0)e^{-\frac{i}{\hbar}Ht}|\psi(0)\rangle = \langle\phi(0)|A(t)|\psi(0)\rangle \tag{2.1.94}$$

where we have used (2.1.90) and (2.1.93).

From (2.1.92), we also see that A is conserved precisely if it commutes with H. In that case, the quantity in (2.1.94) is constant in time, equaling

$$\langle\phi(0)|A(0)|\psi(0)\rangle. \tag{2.1.95}$$

Experimental interactions are formally described by the S-matrix. It is assumed that a state is prepared to have a definite particle content α for $t \to -\infty$ (that is, before the interaction takes place); this is the in state ψ_α^+. The interactions take place at finite time, and one then measures the out state ψ_β^- with particle content β for $t \to \infty$ (that is, after the interaction has taken place). Then, the probability amplitude for the transition is

$$S_{\beta\alpha} = \langle\psi_\beta^-, \psi_\alpha^+\rangle. \tag{2.1.96}$$

The $S_{\beta\alpha}$ are the components of the S-matrix. It is assumed here that these values are computed for complete sets of orthonormal in and out states, so that the S-matrix has to be unitary.

We consider once more the Hamiltonian (2.1.9). x is the position operator, and a particle that is located at a point $x \in M$ (note that we use the same symbol for the position and the position operator) is then in the eigenstate $|x\rangle$ of the position operator (note that this eigenstate in general will not be contained in the Hilbert space $L^2(M)$, but is instead given by a delta functional δ_x). If the particle is in the state $|x\rangle$ at time 0, then by the solution of the Schrödinger equation (2.1.90), at time t it will be in the state

$$e^{-\frac{i}{\hbar}tH}|x\rangle. \tag{2.1.97}$$

More generally, the probability amplitude $\langle x'', t''|x', t'\rangle$ that a particle starting at x' at time t' will be at x'' at the time $t'' > t'$ is given by

$$\langle x'', t''|x', t'\rangle = \langle x''|e^{-\frac{i}{\hbar}H(t''-t')}|x'\rangle, \tag{2.1.98}$$

the projection of the state $e^{-\frac{i}{\hbar}H(t''-t')}|x'\rangle$ obtained from the solution of the Schrödinger equation onto the state $|x''\rangle$.

By formal functional calculus, this is expressed as a Feynman functional integral

$$\langle x''|e^{-\frac{i}{\hbar}H(t''-t')}|x'\rangle = \int Dx \exp\left(\frac{i}{\hbar}L(x)\right) \tag{2.1.99}$$

$$= \int Dx \exp\left(\frac{i}{\hbar}\int_{\tau=t'}^{t''}\left(\frac{m}{2}|\dot{x}(\tau)|^2 - V(x(\tau))\right)d\tau\right) \tag{2.1.100}$$

for our standard example (2.1.7). Here, Dx symbolizes a formal measure on the space of all paths starting at time t' at x' and ending at time t'' at x'', according to the interpretation given in many texts. One should point out here, however, that this measure by itself is not well defined. What one can hope to attach a mathematical meaning to is only the entire integrand $Dx\exp(iL(x))$ in (2.1.99) as a functional measure on the path space. This is in analogy with the Wiener measure where one considers the probability density $p(x'',t''|x',t')$ for a particle starting at time t' at x' and ending up at time t'' at x'' under the influence of the potential $V(x)$, that is, governed by the Lagrangian (2.1.7),

$$F = \frac{m}{2}|\dot{x}|^2 - V(x). \tag{2.1.101}$$

The probability density evolves according to the heat equation

$$\frac{\partial\phi(x,t)}{\partial t} = m\Delta\phi(x,t) - V(x)\phi(x,t). \tag{2.1.102}$$

For comparison, we recall the Schrödinger equation (2.1.14) for the Lagrangian (2.1.7),

$$i\hbar\frac{\partial\phi(x,\tau)}{\partial\tau} = -\frac{\hbar^2}{2m}\Delta\phi(x,\tau) + V(x)\phi(x,\tau). \tag{2.1.103}$$

Let us assume $\hbar = m = 1$ to make the comparison a little simpler. Then, in fact, setting $\tau = -it$ transforms (2.1.103) into (2.1.102). Thus, the Schrödinger equation is the heat equation for imaginary time. With this change of time (called analytic continuation or Wick rotation in the physics literature), the corresponding functional integrals are also transformed into each other. This is useful at the formal level, but perhaps not as much so for the more detailed mathematical analysis.

Returning to (2.1.102), Wiener then showed that, under appropriate conditions on V, the solution can be represented as a path integral

$$p(x'',t''|x',t') = \int Dx \exp(-L(x))$$

$$= \int Dx \exp\left(-\int_{\tau=t'}^{t''}\left(\frac{m}{2}|\dot{x}(\tau)|^2 - V(x(\tau))\right)d\tau\right). \tag{2.1.104}$$

For the discussion to follow, it will be convenient to rewrite (2.1.104) slightly as

$$p(x'',t''|x',t') = \int [Dx]_{x',t'}^{x'',t''} \exp(-L(x)) \tag{2.1.105}$$

to indicate the initial and terminal points of the paths over which we integrate. One then has the property

$$p(x'',t''|x',t') = \int_x p(x'',t''|x,t)\,p(x,t|x',t')dx \tag{2.1.106}$$

for every $t' < t < t''$, which simply expresses the fact that every path leading from x' at time t' to x'' at time t'' has to pass through some x at the intermediate time t. Thus, we may cut the path at time t and integrate over all possible cutting points x. In terms of functional integrals, this becomes

$$p(x'',t''|x',t') = \int_x dx \int [Dx]_{x',t'}^{x,t} \exp(-L(x)) \int [Dx]_{x,t}^{x'',t''} \exp(-L(x)). \tag{2.1.107}$$

The difference between (2.1.102) and (2.1.104) is the i versus the -1 in the exponent in the integral. In the Wiener case, the minus sign leads to a rapid dampening of the influence of those paths with large values of the Lagrangian action, and to a concentration of the functional measure near the minimum of the action. In the Feynman case, in contrast, paths with large values of the action cause rapid fluctuations in the integral, making the analysis substantially harder, see [2]. We do not enter the details here. For more on this, see [49].

In analogy to (2.1.106), we have the cutting relation

$$\begin{aligned}
&\langle x''|e^{-\frac{i}{\hbar}H(t''-t')}|x'\rangle \\
&\quad = \int [Dx]_{x',t'}^{x'',t''} \exp\left(\frac{i}{\hbar}L(x)\right) \\
&\quad = \int_x dx \int [Dx]_{x',t'}^{x,t} \exp\left(\frac{i}{\hbar}L(x)\right) \int [Dx]_{x,t}^{x'',t''} \exp\left(\frac{i}{\hbar}L(x)\right)
\end{aligned} \tag{2.1.108}$$

for $t' < t < t''$.

Written more abstractly, this is the analog of (2.1.106),

$$\langle x'',t''|x',t'\rangle = \int_x dx \langle x'',t''|x,t\rangle\langle x,t|x',t'\rangle. \tag{2.1.109}$$

This fits together well with the operator formalism as in (2.1.70). In particular, we can now insert a position operator (writing $x(t)$ in place of $\hat{x}(t)$ as announced above) and compute

$$\begin{aligned}
\langle x'',t''|x(t)|x',t'\rangle &= \int [Dx]_{x',t'}^{x'',t''} x(t) \exp\left(\frac{i}{\hbar}L(x)\right) \\
&= \int_x dx \langle x'',t''|x,t\rangle x \langle x,t|x',t'\rangle.
\end{aligned} \tag{2.1.110}$$

Here, the $x(t)$ in the integral is a number,[3] not an operator. That number in the integral then translates into the operator $x(t)$ in the inner product on the l.h.s. In physics, these inner products are viewed as the matrix elements of an infinite-dimensional matrix.

This process of cutting the path in the integral can be iterated, and we can insert two intermediate positions $x(t_1), x(t_2)$ (or more, but the principle emerges for two already), to get

$$\int [Dx]_{x',t'}^{x'',t''} x(t_1)x(t_2) \exp\left(\frac{i}{\hbar}L(x)\right) = \langle x'', t''|x(t_2)x(t_1)|x', t'\rangle. \tag{2.1.111}$$

Here, in the integral, the temporal order of t_1 and t_2 is irrelevant because in the integral, $x(t_1), x(t_2)$ are real numbers.[4] In the r.h.s. of (2.1.111), however, they are operators, and since operators in general do not commute, the order does matter here. Since the paths are traversed in increasing time, we need to put them into the temporal order, that is, always apply the operator corresponding to the smaller time first. This is called temporal ordering. Formally, one can define the temporally ordered operator

$$T[x(t_2)x(t_1)] := \begin{cases} x(t_2)x(t_1) & \text{if } t_1 < t_2, \\ x(t_1)x(t_2) & \text{if } t_1 > t_2 \end{cases} \tag{2.1.112}$$

and write the r.h.s. of (2.1.111) as

$$\langle x'', t''|T[x(t_2)x(t_1)]|x', t'\rangle. \tag{2.1.113}$$

We may also write

$$T[x(t_2)x(t_1)] = \theta(t_2 - t_1)x(t_2)x(t_1) + \theta(t_1 - t_2)x(t_1)x(t_2) \tag{2.1.114}$$

where

$$\theta(s) := \begin{cases} 1 & \text{for } s \geq 0, \\ 0 & \text{for } s < 0 \end{cases} \tag{2.1.115}$$

is the Heaviside function. Considered as a functional, its derivative is the Dirac functional,

$$\frac{d}{ds}\theta(s) = \delta(s). \tag{2.1.116}$$

[3] More precisely, when the path x takes its values in Euclidean space $\mathbb{R}^d$, $x(t)$ is a vector with d components. The operations in (2.1.110) and subsequent formulae are to be understood for each component. In particular, when we later on, in (2.1.111) and subsequently, insert expressions like $x(t_1)\cdots x(t_m)$, this is understood as the vector $(x^1(t_1)\cdots x^1(t_m), \ldots, x^d(t_1)\cdots x^d(t_m))$ obtained by componentwise multiplication.

[4] See the preceding footnote.

We now make some general observations about functional integrals of the form

$$\int [Dx]_{x',t'}^{x'',t''} \exp\left(\frac{i}{\hbar} L(x)\right). \tag{2.1.117}$$

In place of the position operator $x(t)$, we can also insert other operators $f(x)$ in (2.1.110). The formula becomes

$$\langle x'', t''|f(x)|x', t'\rangle = \int [Dx]_{x',t'}^{x'',t''} f(x) \exp\left(\frac{i}{\hbar} L(x)\right). \tag{2.1.118}$$

Again, on the l.h.s., $f(x)$ stands for an operator, on the r.h.s. for a number.

The analogy between ordinary (Gaussian) integrals and functional integrals says that the finitely many ordinary degrees of freedom, the coordinate values of the integration variable, are replaced by the infinitely many function values $x(t)$. Therefore, integration by parts should yield that

$$0 = \int [Dx]_{x',t'}^{x'',t''} \frac{\delta}{\delta x(t)} \exp\left(\frac{i}{\hbar} L(x)\right), \tag{2.1.119}$$

that is, the integral of a total derivative vanishes. This yields

$$0 = \frac{i}{\hbar} \int [Dx]_{x',t'}^{x'',t''} \exp\left(\frac{i}{\hbar} L(x)\right) \frac{\delta}{\delta x(t)} L(x). \tag{2.1.120}$$

Recalling (2.1.118), this is written as

$$\langle x'', t''| \frac{\delta L(x)}{\delta x(t)} |x', t'\rangle = 0. \tag{2.1.121}$$

Now, $\frac{\delta L(x)}{\delta x(t)}$ represents the Euler–Lagrange operator (see also (2.3.8) below), and

$$\frac{\delta L(x)}{\delta x(t)} = 0 \tag{2.1.122}$$

is the classical equation of motion. Comparing (2.1.122) and (2.1.121), we see that the classical equation of motion is translated into an operator equation in the quantum mechanical picture. In this interpretation, x', t' and x'', t'' represent arbitrary initial and final conditions for our paths $x(t)$.

Returning to our integration by parts, (2.1.119) generalizes to

$$\begin{aligned} &\frac{i}{\hbar} \int [Dx]_{x',t'}^{x'',t''} \exp\left(\frac{i}{\hbar} L(x)\right) \frac{\delta L(x)}{\delta x(t)} f(x) \\ &\quad = \int [Dx]_{x',t'}^{x'',t''} \frac{\delta}{\delta x(t)} \left(\exp\left(\frac{i}{\hbar} L(x)\right)\right) f(x) \\ &\quad = -\int [Dx]_{x',t'}^{x'',t''} \exp\left(\frac{i}{\hbar} L(x)\right) \frac{\delta}{\delta x(t)} f(x). \end{aligned} \tag{2.1.123}$$

When none of the fields (see below) contained in $f(x)$ is evaluated at time t, the r.h.s. vanishes. This, of course, confirms the interpretation of (2.1.122) as a quantum mechanical operator equation.

Naturally, we are now curious what happens when some of the fields are present at time t. So, we insert $x(t_0)$ for some $t' < t_0 < t''$. This yields

$$\begin{aligned}
&\frac{i}{\hbar}\int [Dx]_{x',t'}^{x'',t''}\, x(t_0)\exp\left(\frac{i}{\hbar}L(x)\right)\frac{\delta}{\delta x(t)}L(x)\\
&\quad=\int [Dx]_{x',t'}^{x'',t''}\frac{\delta}{\delta x(t)}\left(\exp\left(\frac{i}{\hbar}L(x)\right)\right)x(t_0)\\
&\quad=-\int [Dx]_{x',t'}^{x'',t''}\exp\left(\frac{i}{\hbar}L(x)\right)\frac{\delta}{\delta x(t)}x(t_0)\\
&\quad=-\int [Dx]_{x',t'}^{x'',t''}\exp\left(\frac{i}{\hbar}L(x)\right)\delta(t-t_0). \qquad (2.1.124)
\end{aligned}$$

As an operator equation, this is interpreted as

$$\frac{i}{\hbar}x(t_0)\frac{\delta}{\delta x(t)}L(x)=\delta(t-t_0). \qquad (2.1.125)$$

As remarked above, $\frac{\delta L(x)}{\delta x(t)}$ represents the Euler–Lagrange operator (2.3.8). We consider here the example $\frac{d^2}{dt^2}$, and the classical Euler–Lagrange equation becomes the operator equation

$$\frac{d^2}{dt^2}x(t)=0. \qquad (2.1.126)$$

From (2.1.114), (2.1.116), (2.1.126), (2.1.45), we obtain

$$\frac{d^2}{dt^2}T[x(t_2)x(t_1)]=\delta(t_2-t_1)\left[\frac{d}{dt}x(t_1),x(t_1)\right]=-i\hbar\delta(t_2-t_1) \qquad (2.1.127)$$

using the Heisenberg commutation relation (2.1.12) for the last equation. This, of course, coincides with (2.1.125). In Sect. 2.5.1 below, this equation will lead us to the normal ordering scheme for operators.

A recent reference on path integrals is [112].

2.1.4 Quasiclassical Limits

In this section, we briefly discuss some analytical aspects of the relationship between classical and quantum mechanics.

In classical mechanics, stable equilibria are characterized by the principle of locally minimal potential energy, whereas dynamical processes are described by the

principle of stationary motion. Both are variational principles. We consider a physical system with d degrees of freedom $x^1, \dots, x^d$. We want to determine the motion of the system by expressing the x^i as functions of the time t. The mechanical properties of the system are described by the kinetic and the potential energy. The kinetic energy is typically of the form

$$T = \sum_{i,j=1}^{d} A_{ij}(x^1, \dots, x^d, t)\dot{x}^i \dot{x}^j. \tag{2.1.128}$$

Thus, T is a function of the velocities $\dot{x}^1, \dots, \dot{x}^d$, the coordinates $x^1, \dots, x^d$, and time t; often, T does not depend explicitly on t, and one may then investigate equilibria. Here, T is a quadratic form in the generalized velocities $\dot{x}^1, \dots, \dot{x}^d$.

The potential energy is of the form

$$V = V(x^1, \dots, x^d, t), \tag{2.1.129}$$

that is, it does not depend on the velocities.

In order not to have to worry about the justification of taking various derivatives, we assume that V and T are of class C^2.

Hamilton's principle now postulates that motion between two points in time t_0 and t_1 occurs in such a way that the Lagrangian action

$$L(x) := \int_{t_0}^{t_1} (T - V)dt \tag{2.1.130}$$

is stationary in the class of all functions $x(t) = (x^1(t), \dots, x^d(t))$ with fixed initial and final states $x(t_0)$ and $x(t_1)$ respectively. The Lagrangian action is the integral over the Lagrangian

$$F(t, x, \dot{x}) := T - V, \tag{2.1.131}$$

the difference between kinetic and potential energy.

Thus, one does not necessarily look for a minimum under all motions which carry the system from an initial state to a final state, but only for a stationary value of the integral. For such a stationary motion, the Euler–Lagrange equations hold:

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{x}^i} - \frac{\partial}{\partial x^i}(T - V) = 0 \quad \text{for } i = 1, \dots, d. \tag{2.1.132}$$

If V and T do not explicitly depend on the time t, then equilibrium states are constant in time, that is, $\dot{x}^i = 0$ for $i = 1, \dots, d$, and hence $T = 0$, therefore by (2.1.132)

$$\frac{\partial V}{\partial x^i} = 0 \quad \text{for } i = 1, \dots, d. \tag{2.1.133}$$

Thus, in a state of equilibrium, V must have a critical point, and in order for this equilibrium to be stable, V must even have a minimum there. That minimum, however, need not be unique. For example, when the state space is simply the real line

$\mathbb{R}$, and

$$V(x) = (x^2 - a^2)^2 \tag{2.1.134}$$

for some $a \in \mathbb{R}$, the classical equilibrium states are $x = \pm a$.

Quantum mechanically, we have an L^2-function $\phi : \mathbb{R} \to \mathbb{C}$ (but, for simplicity, we shall consider real-valued functions ϕ in the present section) with the normalization

$$\|\phi\|_{L^2} = 1 \tag{2.1.135}$$

and the asymptotic behavior

$$\lim_{x \to \pm\infty} \phi(x) = 0. \tag{2.1.136}$$

The potential energy $V(x)$ is now replaced by the energy

$$\int_{\mathbb{R}} \left(\frac{\hbar^2}{2} \left| \frac{d\phi(x)}{dx} \right|^2 + V(x)|\phi(x)|^2 \right) dx. \tag{2.1.137}$$

The corresponding Euler–Lagrange equation is

$$\left(-\frac{\hbar^2}{2} \frac{d^2}{dx^2} + V(x) \right) \phi = 0. \tag{2.1.138}$$

Since there is no kinetic term in (2.1.137), quantum mechanics tries to find the eigenfunctions of the operator in (2.1.138), the Hamiltonian. That is, we look for solutions of

$$\left(-\frac{\hbar^2}{2} \frac{d^2}{dx^2} + V(x) \right) \phi_i = E_i \phi_i. \tag{2.1.139}$$

The eigenvalues E_i are the energy levels. A solution ϕ_0 for the smallest possible energy E_0 corresponds to the vacuum. E_0 is positive since the potential V is positive, see (2.1.134). The solution ϕ_0 (normalized by $\|\phi_0\|_{L^2} = 1$ according to (2.1.135)) is symmetric, that is, $\phi_0(x) = \phi_0(-x)$, with maxima at $\pm a$, a local minimum at 0, and asymptotic decay $\lim_{x \to \pm\infty} \phi(x) = 0$ required by (2.1.136). The eigenfunctions for different eigenvalues are L^2-orthogonal. In particular, the eigenfunction ϕ_1 for the second smallest energy level E_1 satisfies $\int \phi_0 \phi_1 dx = 0$. It is antisymmetric, that is $\phi_1(x) = -\phi_1(-x)$ and thus changes sign at $x = 0$. In the quasiclassical limit, that is, for $\hbar \to 0$, we have

$$E_0, E_1 \sim a\hbar \tag{2.1.140}$$

and

$$E_1 - E_0 \sim c_0 \exp\left(-\frac{c_1}{\hbar} \right) \tag{2.1.141}$$

for constants a, c_0, c_1. Thus, the difference between these energy levels goes to 0 exponentially. Therefore, in that quasiclassical limit, the energy levels of ϕ_0 and ϕ_1

become indistinguishable. Thus, for $\hbar \to 0$, the limits of $\phi_\pm := \frac{1}{\sqrt{2}}(\phi_0 \pm \phi_1)$ also become minima, with

$$\lim_{\hbar \to 0} |\phi_\pm|^2 = \delta(x \mp a). \tag{2.1.142}$$

ϕ_+ and ϕ_- break the symmetry between a and $-a$. Classically, any linear combination of $\delta(x-a)$ and $\delta(x+a)$ is a possible minimum. The quantum mechanical vacuum, however, is symmetric.

It is also instructive to consider a nonlinear problem. We take

$$\int \left(\frac{\hbar^2}{2} \left| \frac{d\phi(x)}{dx} \right|^2 + W(\phi(x)) \right) dx. \tag{2.1.143}$$

This time, ϕ need not be real-valued, but could assume values in some other space, like a Riemannian manifold N. We first consider the real-valued case. The domain, denoted by M, however, is allowed to be of higher dimension. We suppose again that the potential W has two minima (W is then called a two-well potential). Now, in the quasiclassical limit, we do not obtain a concentration at two points, the two minima, in the domain, but rather the concentration at two values of ϕ. This time, in contrast to the linear case, the symmetry can also be broken for $\hbar > 0$. Since this is not a linear problem, we no longer have the concept of eigenfunctions available. For $\hbar \to 0$, the solution becomes piecewise constant, the values being the two minima of W, of course. When one imposes suitable constraints, by a result of Modica [82], the set of discontinuity of the limit for $\hbar \to 0$ of the solutions for $\hbar > 0$ is a hypersurface of constant mean curvature in the domain. Of course, this is meaningful only for a higher-dimensional domain.

When the domain is one-dimensional, that is, the real line $\mathbb{R}$, but the target is of higher dimension, we may have quantum mechanical tunneling solutions, i.e.,

$$\lim_{x \to \pm\infty} \phi(x) = a_\pm \tag{2.1.144}$$

between the vacua, that is, minima of W, denoted by a_+, a_-. These tunneling solutions are gradient flow lines of W when the target is a Riemannian manifold.

There exist some generalizations of this problem that lead to analytical constructions of great interest:

1. Let L be a real line bundle over $M \backslash S$ where S is a submanifold of codimension 2 in the domain M, with prescribed holonomy for L around S. A section of L then has a zero set of codimension 1 in M with boundary S. Considering the quasiclassical limit of those zero sets for minimizers of the above functional for sections of L yields a minimal hypersurface in M with boundary S, according to [44].
2. Let L now be a complex line bundle over M. The above functional then leads to a vortex equation of the type studied by Taubes [99], Bethuel et al. [13], and Ding et al. [28–30].

2.2 Lagrangians

2.2.1 Lagrangian Densities for Scalars, Spinors and Vectors

A type of particle is represented by a vector bundle E over some Lorentz manifold M. The particle transforms according to some representation of the Lorentz group or its double cover, the spin group.[5] Thus, it transforms as a tensor or as a spinor. The states of collections of such particles are represented by sections ψ of E, so-called fields.[6]

We are considering here the semiclassical situation, i.e., before field quantization, and so ψ has to satisfy the Euler–Lagrange equations of some action functional that is invariant under the representation of the Lorentz or spin group according to which the particle transforms. In addition, there are internal symmetries that affect only the values of the fields, but not of the coordinates, and leave the action invariant. In fact, the symmetries and certain general considerations often suffice to allow us to construct the appropriate Lagrangian for the action as we shall see.

Notation: $\partial_\mu = \frac{\partial}{\partial x^\mu}$, $\partial^\mu = g^{\mu\nu}\partial_\nu$.

For the moment, we can think of $g_{\mu\nu}$ as a (Lorentz) metric on $\mathbb{R}^{1,3}$, and the indices μ, ν then run from 0 to 3.

We consider the action functional (Lagrangian)[7]

$$S(\phi) = \int \left\{ \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 \right\} \sqrt{-g(x)}d(x) \tag{2.2.1}$$

for a free scalar field, with the Lagrangian density

$$F(\phi, D\phi) = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2, \tag{2.2.2}$$

on $\mathbb{R}^{1,3}$, or, more generally, on some Riemannian or Lorentzian manifold. (Note that in (2.2.1), we use $\sqrt{-g(x)}d(x)$ for the volume form, since we are assuming a Lorentzian metric. Subsequently, when we switch to the Riemannian case, the minus sign has to be deleted.)

The corresponding Euler–Lagrange equation is the Klein–Gordon equation

$$\Box\phi + m^2\phi = 0$$

where $\Box$ is the Minkowski Laplacian (1.1.106).

---

[5] In fact, according to Wigner's principle as explained in Sect. 1.3.4, we should consider a particle as an irreducible unitary representation not only of the Lorentz or spin group, but of the Poincaré group or the double covering $Sl(2, \mathbb{C}) \ltimes \mathbb{R}^{1,3}$. While this is fundamental for determining the types of possible elementary particles from the theory of group representations, in this section, we shall be mainly concerned with internal symmetries that arise from invariance w.r.t. to the action of some compact group.

[6] A section represents a state containing possibly several particles of a given type, since in quantum field theory, particle numbers need not be preserved.

[7] In the mathematical literature, an action functional is often called a Lagrange functional.

Remark

1. As a classical action functional for a field, we consider the action for a particle $q(t) = (q^1(t), \ldots, q^m(t))$ with m degrees of freedom

$$\int \sum_{j=1}^{m} \left\{ \frac{1}{2}(\dot{q}^j(t)\dot{q}_j(t) - m^2 q^j(t) q_j(t)) \right\} dt. \tag{2.2.3}$$

   When we compare this with our quantum field theoretic setting, we see that the index j corresponds to the spatial variable (x^1, x^2, x^3) above. In this sense, $\phi(t, x)$ is a particle with infinitely many degrees of freedom, one degree of freedom for each point of M.
2. In the physics literature, the field ϕ in (2.2.1), (2.2.2) is usually taken as complex valued instead of real valued, that is, one considers

$$S_1(\phi) = \int \left\{ \frac{1}{2}\partial_\mu \phi \partial^\mu \bar{\phi} - \frac{1}{2}m^2 |\phi|^2 \right\} \sqrt{-g(x)} d(x), \tag{2.2.4}$$

   in line with the basic formalism of quantum mechanics, see (2.1.13). Our reason for starting with a real valued ϕ here is, besides its simplicity, that this is better suited for subsequent generalizations to nonlinear models where the field will take its values in a Riemannian manifold.

We now turn to Lagrangians for spinors.
For two left-handed spinors (see (1.3.49)) ϕ, χ,

$$\phi\chi := \varepsilon_{\alpha\beta}\phi^\alpha \chi^\beta$$

transforms as a scalar under the spinor representation, see (1.3.56).
Similarly

$$\phi^\alpha \sigma^\mu_{\alpha\dot{\alpha}} \bar{\chi}^{\dot{\alpha}}$$

transforms as a vector, for $\mu = 0, 1, 2, 3$, see (1.3.57). We may then write a Lagrangian for a left-handed spinor ϕ as

$$\begin{aligned} F &= \mathrm{Re}(i\phi\sigma^\mu \partial_\mu \bar{\phi} + 2m\phi\phi) \\ &= \frac{i}{2}(\phi\sigma^\mu \partial_\mu \bar{\phi} - \partial_\mu \phi \sigma^\mu \bar{\phi}) + m(\phi\phi + \overline{\phi\phi}) \end{aligned} \tag{2.2.5}$$

(here, $\bar{\phi}$ is the complex conjugate of ϕ—subsequently, we shall employ a somewhat different convention when we consider full spinors).

The equation of motion for

$$S(\phi) = \int F(\phi) \tag{2.2.6}$$

is

$$i\partial_\mu \phi \sigma^\mu - m\bar{\phi} = 0, \tag{2.2.7}$$

or equivalently

$$i\sigma^\mu \partial_\mu \bar{\phi} + m\phi = 0. \tag{2.2.8}$$

In quantum field theory (QFT), charged particles correspond to complex-valued fields, and the Lagrangian has to remain invariant under multiplication of the fields by $e^{i\lambda}(\lambda \in \mathbb{R})$ since the phase is not observable. Instead of imposing the normalization $\|\phi\| = 1$, we can then consider states as corresponding to lines in a Hilbert space.

The preceding Lagrangian satisfies this invariance property only for $m = 0$. Since it does not have this property in general, it corresponds to a neutral fermion. In the standard model to be discussed below, these neutral fermions are the neutrinos.

In order to obtain a Lagrangian for charged fermions, we need full spinors

$$\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}.$$

Then in the Weyl representation,

$$\bar{\psi}\gamma^\mu \psi$$

transforms as a vector, see (1.3.61).

The Dirac–Lagrangian is then

$$F = i\bar{\psi}\gamma^\mu \partial_\mu \psi - m\bar{\psi}\psi = i\langle \bar{\psi}, \not{D}\psi \rangle - m\bar{\psi}\psi \tag{2.2.9}$$

(recalling the Dirac operator $\not{D}$ defined in (1.3.22)). The mass term mixes the left and the right spinor, since $\bar{\psi}\psi = \bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L$. This time, we do have invariance under multiplication of ψ by $e^{i\lambda}$ for constant real λ also in the general case $m \neq 0$.

The corresponding Dirac equation is

$$i\gamma^\mu \partial_\mu \psi - m\psi = 0. \tag{2.2.10}$$

Perhaps the factor i in (2.2.9) in front of the Dirac operator $\not{D} = \gamma^\mu \partial_\mu$ needs some explanation. The reason is that, upon integration, the corresponding term is purely imaginary, and the factor i then makes it real. It is instructive to consider an example, and since we shall mainly investigate the Riemannian in place of the Lorentzian setting in the sequel, we shall also use a Riemannian example. As in our treatment of the supersymmetric sigma model below (2.4.3), we consider the two-dimensional case and use the following representation of the Clifford algebra $Cl(2, 0)$:

$$e_1 \to \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad e_2 \to \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, \tag{2.2.11}$$

which is different from the one described in Sect. 1.3.2 (but of course equivalent to it). A spinor ω is thus identified with an element $(\omega_1 = \alpha_1 + i\beta_1, \omega_2 = \alpha_2 + i\beta_2)$ of

$\mathbb{C}^2$. In local coordinates x, y, then

$$\gamma^\mu \partial_\mu \omega = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \frac{\partial}{\partial x} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \frac{\partial}{\partial y} \begin{pmatrix} \omega_1 \\ \omega_2 \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{\partial \alpha_1}{\partial x} + i \frac{\partial \beta_1}{\partial x} + \frac{\partial \alpha_2}{\partial y} + i \frac{\partial \beta_2}{\partial y} \\ \frac{\partial \alpha_2}{\partial x} + i \frac{\partial \beta_2}{\partial x} - \frac{\partial \alpha_1}{\partial y} - i \frac{\partial \beta_1}{\partial y} \end{pmatrix} \tag{2.2.12}$$

and

$$\begin{aligned} \bar{\omega} \gamma^\mu \partial_\mu \omega = & -\alpha_1 \frac{\partial \alpha_1}{\partial x} - \alpha_1 \frac{\partial \alpha_2}{\partial y} - \beta_1 \frac{\partial \beta_1}{\partial x} - \beta_1 \frac{\partial \beta_2}{\partial y} + \alpha_2 \frac{\partial \alpha_2}{\partial x} - \alpha_2 \frac{\partial \alpha_1}{\partial y} \\ & + \beta_2 \frac{\partial \beta_2}{\partial x} - \beta_2 \frac{\partial \beta_1}{\partial y} + i \Big(\beta_1 \frac{\partial \alpha_1}{\partial x} - \alpha_1 \frac{\partial \alpha_2}{\partial y} - \alpha_1 \frac{\partial \beta_1}{\partial x} + \beta_1 \frac{\partial \alpha_2}{\partial y} - \beta_2 \frac{\partial \alpha_2}{\partial x} \\ & + \beta_2 \frac{\partial \alpha_1}{\partial y} + \alpha_2 \frac{\partial \beta_2}{\partial x} - \alpha_2 \frac{\partial \beta_1}{\partial y} \Big). \end{aligned} \tag{2.2.13}$$

Upon integration, the real part vanishes by integration by parts, and only the imaginary part remains. This comes about because the coefficients of ω commute. Were they to anticommute, only the real part would remain. We are making this observation here because in our subsequent treatment of supersymmetry, we shall use spinor fields with anticommuting coefficients.

We have now seen action functionals for scalars and spinors, where these names describe the transformation behavior under Lorentz transformations, i.e., coordinate changes. An electromagnetic field, however, is described by a potential that transforms as a vector or covector. We consider

$$A = A_\mu(x) dx^\mu.$$

A is called a vector particle, because A^μ transforms as a vector. Mathematically, A is a connection, see (1.2.12), on a vector bundle with fiber $\mathbb{C}$ and the Abelian structure group $U(1) = SO(2)$. We also recall the transformation behavior (1.2.32).

The field strength is described by the tensor

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu,$$

(2 times) the[8] curvature of the connection A, see (1.2.23) (note that the brackets $[A_\mu, A_\nu]$ vanish here, because the structure group is Abelian), and the corresponding Lagrangian is the Maxwell density (the Abelian case of the Yang–Mills density)

$$-\frac{1}{4} F_{\mu\nu} F^{\mu\nu} = \frac{1}{2} (\partial_\mu A_\nu \partial^\mu A^\nu - \partial_\mu A_\nu \partial^\nu A^\mu). \tag{2.2.14}$$

[8] Note the different conventions between the present section and Sect. 1.2.2.

An important property of this Lagrangian is the gauge invariance, namely its invariance under replacing A_μ by

$$A_\mu + \partial_\mu \xi$$

where ξ is a scalar function. This is the present, Abelian, version of (1.2.13), (1.2.32). Of course, the field strength $F_{\mu\nu}$ is already invariant under such a gauge transformation (see (1.2.25), (1.2.33) for the general result).

The equations of motion for

$$S(A) = -\frac{1}{4}\int F_{\mu\nu}F^{\mu\nu} \tag{2.2.15}$$

are

$$\partial_\mu F^{\mu\nu} = 0. \tag{2.2.16}$$

If we add a "mass term"

$$m^2 A_\mu A^\mu,$$

then the gauge invariance no longer holds.

As described, the mathematical interpretation of A is that of a covariant derivative for sections of a line bundle, see Sect. 1.2.2. Thus, for a scalar field ϕ taking values in this bundle, we put

$$(d_A\phi)_\mu = \partial_\mu\phi + A_\mu\phi,$$

and we may consider the interaction Lagrangian

$$\frac{1}{2}(\partial_\mu\phi + A_\mu\phi)(\partial^\mu\phi^* + (A^\mu\phi)^*) = \frac{1}{2}\|d_A\phi\|^2. \tag{2.2.17}$$

Here, we assume that the line bundle is Hermitian, and for simplicity, we write the metric as $\|\phi\|^2 = \phi\phi^*$; of course, in general this only holds in suitable coordinates; we also assume that A is unitary w.r.t. this metric—we shall return to this point in a moment.

The replacement of

$$\partial_\mu\phi \quad \text{with } \partial_\mu\phi + A_\mu\phi \tag{2.2.18}$$

is for the following reason. The Lagrangian

$$\frac{1}{2}\partial_\mu\phi\partial^\mu\phi^* - \frac{1}{2}m^2\phi\phi^*$$

is invariant under $U(1)$, i.e., under replacements

$$\phi \mapsto e^{i\vartheta}\phi \quad \text{with } \vartheta \in \mathbb{R}.$$

It thus has a global internal symmetry. It is not invariant, however, under general local symmetries, i.e.,

$$\phi \mapsto e^{i\vartheta(x)}\phi$$

if $\vartheta(x)$ is a nontrivial function of x. However, if, according to (1.2.14), we also replace[9]

$$A_\mu \mapsto A_\mu - i\,\partial_\mu \vartheta,$$

then the above interaction Lagrangian (2.2.17) remains invariant. Thus, we have a gauge invariant Lagrangian. The procedure (2.2.18) of replacing an ordinary by a covariant derivative is called minimal coupling.

In fact, we have a free parameter here: We consider the exterior derivative d as the trivial connection ("vacuum") on the trivial bundle $M \times \mathbb{R}$. We can then view the affine space of connections A as the vector space $\Omega^1(M)$ of 1-forms. We can therefore multiply A by some factor q and choose the covariant derivative

$$D_A := \partial + qA \tag{2.2.19}$$

and gauge transform A to

$$A - \frac{i}{q}\partial\vartheta. \tag{2.2.20}$$

q here is interpreted as the charge of the electromagnetic field. It is the Noether charge associated to the $U(1)$ gauge symmetry.

The full Lagrangian for a complex scalar field ϕ interacting with an electromagnetic field A is

$$\frac{1}{2}\|d_A\phi\|^2 - \frac{1}{2}m^2\|\phi\|^2 + \frac{1}{4q^2}\|F\|^2. \tag{2.2.21}$$

The same discussion applies to spinor fields ψ, and we may form the interaction Lagrangian

$$i\bar\psi\gamma^\mu(\partial_\mu + A_\mu)\psi - m\bar\psi\psi + \frac{1}{4q^2}\|F\|^2. \tag{2.2.22}$$

Let us see the details once more: Replacing $\psi(x)$ by $e^{i\vartheta(x)}\psi(x)$ changes the spinor Lagrangian

$$i\bar\psi\gamma^\mu\partial_\mu\psi - m\bar\psi\psi \tag{2.2.23}$$

by $-\bar\psi\gamma^\mu\psi\,\partial_\mu\vartheta$, and this is again compensated when we replace ∂_μ by $\partial_\mu + qA_\mu$ and require that A transforms to $A - \frac{i}{q}\partial\vartheta$ as before. Thus,

$$i\bar\psi\gamma^\mu(\partial_\mu + A_\mu)\psi - m\bar\psi\psi \tag{2.2.24}$$

remains gauge invariant.

[9]Note that the convention here is different from the one in Sect. 1.2.3; here, elements of the Lie algebra $\mathfrak{u}(1)$ of $U(1)$, and similarly of other Lie algebras $\mathfrak{g}$, are written as $i\vartheta$, with a *real* ϑ. This will lead to various factors i and -1 when compared to Sect. 1.2.3. This is the standard convention employed in the physics literature.

As was realized by Yang and Mills,[10] this can be generalized to an arbitrary internal symmetry group G with Lie algebra $\mathfrak{g}$, and a field ϕ that takes its value in a vector bundle (or, similarly, in a spinor bundle—the physically more important case, see Sect. 2.2.3 below)[11] with structure group G. The mathematical formalism for this has been described in Sect. 1.2.3. In abstract physical terms, the gauge principle says that the symmetries should determine the forces. The particles conveying these forces are called gauge bosons.

To implement this, we simply consider $A = A_\mu dx^\mu$, a 1-form with values in $\mathfrak{g}$, and form the covariant derivative (1.2.12) of the field ϕ, a section of the vector bundle with structure group G on which A operates as a covariant derivative,

$$d_A\phi = d\phi + A\phi.$$

The replacements

$$\phi(x) \mapsto g(x)\phi(x), \quad \text{with } g(x) \in G \text{ for all } x,$$

i.e., g is an element of the group of gauge transformations, and (1.2.32), that is,

$$A \mapsto gAg^{-1} - (\partial g)g^{-1}$$

then leave

$$\|d_A\phi\|^2$$

invariant (assuming of course that the metric $\|\cdot\|$ is G-invariant).

The gauge field strength is now (two times) the curvature (1.2.22), (1.2.23)

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$$

where $[\cdot,\cdot]$ is the Lie algebra bracket of $\mathfrak{g}$.

As before, we may form the Lagrangian involving the Yang–Mills action (1.2.34) and coupling it with the action for the field ϕ

$$\frac{1}{2}\|d_A\phi\|^2 - \frac{1}{2}m^2\|\phi\|^2 - \frac{1}{4q^2}\operatorname{Tr} F_{\mu\nu}F^{\mu\nu}. \tag{2.2.25}$$

The same discussion applies to spinor fields ψ with values in a vector bundle on which G acts, that is, sections of $S \otimes E$ for some vector bundle E over M with a G-action, and we may form the interaction Lagrangian

$$i\bar\psi\gamma^\mu(\partial_\mu + A_\mu)\psi - m\bar\psi\psi - \frac{1}{4q^2}\operatorname{Tr} F_{\mu\nu}F^{\mu\nu}. \tag{2.2.26}$$

[10] Such ideas were first conceived by Hermann Weyl.

[11] Of course, a spinor bundle is a vector bundle, but in physics, it is important to distinguish between vector and spinor representations, that is, whether a representation of the spin group lifts to one of the orthogonal group or not.

If the representation of G on our vector bundle is not irreducible, but decomposes into subrepresentations indexed by j, we may use a more general Lagrangian for the gauge field strength because we can take a combination

$$-\sum_j \gamma_j \operatorname{Tr} F^{(j)}_{\mu\nu} F^{(j)\mu\nu}.$$

The γ_j are the so-called coupling constants. Mathematically, they parametrize the ad-invariant bilinear forms on the Lie algebra of G.

Important remark: It is undesirable to have too many constants whose values are not theoretically deduced, but can only be experimentally determined. As we have just seen, such a situation comes about if the representation of G under consideration is not irreducible. One possible solution of this problem would be to suppose that there is some larger group $\tilde{G} \supset G$ in the background with an irreducible representation that induces the (reducible) representation of G, and so determines all the constants except one. It may be possible that the symmetry group $\tilde{G}$ cannot be experimentally observed because of a symmetry-breaking mechanism that reduces $\tilde{G}$ to G. The Higgs mechanism, to be described below, is such a mechanism.

Let us recapitulate that A is a 1-form with values in $\mathfrak{g}$, and the symmetries therefore are two-fold: Under Lorentz or space–time symmetries, the 1-form part is transformed, whereas under local internal symmetries (i.e., those coming from G), the $\mathfrak{g}$ part is affected. Similarly, ϕ transforms as a scalar under space–time symmetries and by the action of G under local internal symmetries.

More generally, one may also introduce some nonlinearities into the ϕ and ψ equations by adding some polynomial terms to the Lagrangian. These polynomial terms, however, are constrained by the requirement of renormalizability. In dimension 4, the most general renormalizable Lagrangian is

$$-\frac{1}{4q^2}\operatorname{Tr} F_{\mu\nu}F^{\mu\nu} + \frac{1}{2}\|d_A\phi\|^2 - \frac{1}{2}m^2\|\phi\|^2 + i\bar{\psi}\gamma^\mu(\partial_\mu A_\mu)\psi$$
$$+ g_1\|\phi\|^4 + g_2\phi\bar{\psi}\psi + \text{ lower-order terms.} \tag{2.2.27}$$

Here, g_1, g_2, like q, are coupling constants. The term $\phi\bar{\psi}\psi$ is called a Yukawa term. A version of (2.2.27) is also the Lagrangian of the standard model to which we shall turn after a brief discussion of the scaling behavior of Lagrangians. The gauge group of the standard model is $SU(3) \times SU(2) \times U(1)$.

2.2.2 Scaling

An important criterion for field theories in physics and variational problems in mathematics is their scaling behavior. That means that one scales the independent variables

$$x \to \lambda^{-1}x =: y \tag{2.2.28}$$

where x is n-dimensional, $x \in \mathbb{R}^n$, and $\lambda > 0$, and computes the resulting scaling behavior of the integrals of the fields. The starting point is the scaling of the volume form

$$d^n y = \lambda^{-n} dx. \tag{2.2.29}$$

Also,

$$\frac{\partial}{\partial y} = \lambda \frac{\partial}{\partial x}. \tag{2.2.30}$$

Putting $\phi_\lambda(y) := \phi(\lambda y) = \phi(x)$, one obtains

$$\int_{\mathbb{R}^n} |d\phi_\lambda(y)|^p d^n y = \lambda^{p-n} \int_{\mathbb{R}^n} |d\phi(x)|^p d^n x \tag{2.2.31}$$

and

$$\int_{\mathbb{R}^n} |\phi_\lambda(y)|^q d^n y = \lambda^q \int_{\mathbb{R}^n} |\phi(x)|^q d^n x \tag{2.2.32}$$

for exponents $p, q > 0$. Therefore, the L^q-norm of ϕ_λ, $(\int_{\mathbb{R}^n} |\phi_\lambda(y)|^q d^n y)^{1/q}$ has a scaling behavior dominated by the L^p-norm of the derivative $d\phi_\lambda$, $(\int_{\mathbb{R}^n} |d\phi_\lambda(y)|^p d^n y)^{1/p}$, if

$$q \leq \frac{np}{n-p} \quad \text{for } p < n. \tag{2.2.33}$$

This is exploited in the Sobolev embedding theorem, see e.g. [63]. For instance, in the most important case $p = 2$, in dimension 2, any polynomial in ϕ is controlled by $\int |d\phi|^2$, in dimension 3, a polynomial of order ≤ 6, and in dimension 4, only those of order ≤ 4.

In the light of (2.2.29), (2.2.30), the scaling law (2.2.31) can also be interpreted in the way that $\int_{\mathbb{R}^n} |d\phi|^2$ remains invariant if the field ϕ is scaled as

$$\phi \to \lambda^{\frac{n-2}{2}} \phi. \tag{2.2.34}$$

In particular, $\int |d\phi|^2$ is scaling invariant in dimension 2; in fact, this integral is even conformally invariant in dimension 2, as we shall explore below. In other dimensions, it becomes invariant only after a rescaling of the field ϕ according to (2.2.34). In order to compensate for the different scaling laws for $\int_{\mathbb{R}^n} |d\phi|^2 d^n x$ and $\int_{\mathbb{R}^n} |\phi(x)|^q d^n x$, one may also introduce coupling constants that are then scaled appropriately. To see this, let us consider

$$\int_{\mathbb{R}^n} (|d\phi|^2 - m^2 |\phi(x)|^2 + g_1 |\phi(x)|^3 + g_2 |\phi(x)|^4) d^n x; \tag{2.2.35}$$

here, m is the mass of the field as before. In order that this integral be scaling invariant, because of (2.2.29), each such polynomial term has to scale with a factor λ^n. If now ϕ scales according to (2.2.34), this leads to

$$m \to \lambda m, \qquad g_1 \to \lambda^{\frac{6-n}{2}} g_1, \qquad g_2 \to \lambda^{\frac{4-n}{2}} g_2. \tag{2.2.36}$$

Of course, this just re-expresses our discussion of (2.2.32), (2.2.33). For instance, in dimension ≤ 4, the integral of the polynomial ϕ^4 is controlled by that of $|d\phi|^2$, and therefore, we can afford a nonnegative scaling exponent for the coupling constant g_2. In general, an interaction term is called perturbatively renormalizable when the coupling constant scales with exponent 0, and superrenormalizable when it scales with a positive exponent.

Similarly, when we consider a term

$$\int_{\mathbb{R}^n} \gamma |\phi|^q |d\phi|^2 d^n x, \tag{2.2.37}$$

we see that for $q > 0$, the coupling constant γ scales with exponent $\frac{(2-n)q}{2}$ which is negative for dimension > 2. Thus, such an interaction is renormalizable only for $n = 2$, but not for $n > 2$. This applies to the nonlinear sigma model discussed below (see (2.4.27), (2.4.28)),

$$\int_{\mathbb{R}^n} g_{ij}(\phi(x)) \frac{\partial \phi^i(x)}{\partial x^\alpha} \frac{\partial \phi^j(x)}{\partial x^\alpha} d^n x \tag{2.2.38}$$

where g_{ij} denotes the metric tensor of the target N. When we expand (in normal coordinates)

$$g_{ij}(\phi) = \delta_{ij} + g_{ij,kl}\phi^k \phi^l + \text{higher-order terms}, \tag{2.2.39}$$

we see that this model is not renormalizable for $n > 2$.

Next, if we have a Dirac term for a spinor field as in (2.2.9) in the Lagrangian, that is, if we have an action of the form

$$\int_{\mathbb{R}^n} i\langle \bar{\psi}, \not{D}\psi \rangle d^n x, \tag{2.2.40}$$

we need the scaling behavior

$$\psi \to \lambda^{\frac{n-1}{2}} \psi \tag{2.2.41}$$

to make it invariant. When we then have a mass term

$$m\bar{\psi}\psi, \tag{2.2.42}$$

we obtain once more the scaling law $m \to \lambda m$ as in (2.2.36).

We finally consider gauge fields as in (2.2.27),

$$D_A = \partial + qA, \tag{2.2.43}$$

with curvature $F = qdA + q^2 A \wedge A$, by (1.2.22). We can then expand the Lagrangian action (2.2.21) (in shorthand notation, as we are only interested in the growth orders),

$$\frac{1}{4q^2} \int_{\mathbb{R}^n} \|F\|^2 d^n x = \frac{1}{4} \int_{\mathbb{R}^n} ((dA)^2 + qA^2 dA + q^2 A^4) d^n x. \tag{2.2.44}$$

As above, from the first term we get the scaling law (2.2.34)

$$A \to \lambda^{\frac{n-2}{2}} A, \tag{2.2.45}$$

which then, taking always (2.2.29) into account, yields

$$q \to \lambda^{\frac{4-n}{2}} q. \tag{2.2.46}$$

Thus, a gauge field Lagrangian action is renormalizable for dimension 4, but not above. The reader will then check that the same applies to the full interaction Lagrangians (2.2.21) and (2.2.26). Thus, we see that the Lagrangian action defined by (2.2.27) is indeed perturbatively renormalizable in dimension 4.

In contrast to this, let us consider the Einstein–Hilbert functional (1.1.163) of general relativity,

$$\frac{1}{16\pi G} \int_{\mathbb{R}^n} R(g) d^n x \tag{2.2.47}$$

for the scalar curvature R of the metric g. Here, we have introduced the factor $\frac{1}{16\pi G}$ that we had neglected in the discussion of (1.1.163) above, G being Newton's gravitational constant. Also, we write $d^n x$ for the volume form because we expand around the flat metric,

$$g_{ij} = \delta_{ij} + \gamma_{ij}\sqrt{G}. \tag{2.2.48}$$

We then obtain, with a similar shorthand notation as above, for the Einstein–Hilbert action (2.2.47)

$$\frac{1}{16\pi} \int_{\mathbb{R}^n} ((dh)^2 + \sqrt{G}h(dh)^2 + \text{ higher-order terms}) \, d^n x, \tag{2.2.49}$$

whence the scaling behavior $h \to \lambda^{\frac{n-2}{2}} h$ as before, and then

$$G \to \lambda^{2-n} G. \tag{2.2.50}$$

Thus, the Lagrangian action of general relativity is renormalizable only in dimension 2, but not in dimension 4. This indicates that there should be difficulties in unifying gravity in dimension 4 with the other forces that are governed by a renormalizable Lagrangian of the form (2.2.27).

Here, we have only discussed perturbative renormalization (using [106]), but not nonperturbative renormalization, which is a more difficult issue. Some references for renormalization theory are [52, 113].

2.2.3 Elementary Particle Physics and the Standard Model

We now interrupt the process of setting up mathematical structures to discuss how this relates to elementary particles and in particular to their the contemporary theory

as incorporated in the so-called standard model and its extensions. Physicists will, of course, know all this and may skip this section.

There exist four known basic physical forces: the electromagnetic, weak and strong forces and gravity. The standard model includes the first three of them, but leaves out gravity. In fact, it is the fundamental challenge of high-energy theoretical physics to construct a unified theory of all known forces, including gravity.

In any case, in a relativistic theory of elementary particles without gravitational effects, the Lagrangian should be invariant under the action of the Poincaré group or the double cover $G := Sl(2,\mathbb{C}) \ltimes \mathbb{R}^{1,3}$, see Sect. 1.3.4. This was most clearly formulated by Wigner who identified an elementary particle with an irreducible unitary representation of G satisfying certain physical restrictions, like $m^2 \geq 0$, where m is the mass. This principle is still fundamental, with the modification that one needs to consider groups that are larger than G, in order to account for internal symmetries of the particles beyond the spin. The principal for identifying that group combines the mathematical theory of group representations with scattering experiments designed to break the symmetry. To take an example from quantum mechanics, the Hamiltonian of a particle in a rotationally symmetric potential, $H = \frac{p^2}{2m} + V(|x|)$ (see (2.1.9)), commutes with the angular momentum operator $x \times p = \frac{\hbar}{i}(x \times \nabla)$, and therefore its eigenvalues, the energy levels, are degenerate. When an external magnetic field is applied, this symmetry gets broken and the energy levels, that is, the eigenvalues of the Hamiltonian, become distinct (Stern–Gerlach experiment, Zeeman effect). A further splitting of the energy levels of an electron in the presence of an external field is caused by its spin.

When it became clear that the proton and the neutron were very similar, except for their electrical charge, Heisenberg suggested that there was a single underlying particle, the nucleon, with a so-called isotopic spin, for short isospin, symmetry that was broken in the presence of electromagnetic interactions. This should correspond to the $L = \frac{1}{2}$ representation of $SU(2)$, as described at the end of Sect. 1.3.4. The proton and the neutron should correspond to the eigenvalues $\frac{1}{2}$ and $-\frac{1}{2}$ of $h = -it_3$. The subsequently discovered pions π^+, π^0, π^- should likewise correspond to the representation for $L = 1$ with the eigenvalues $1, 0, -1$ of h. This was supported by pion–nucleon scattering experiments. In those scattering experiments, the total charge Q as well as the total baryon number B was conserved (proton and neuron have $B = 1$, the pions have $B = 0$). In order to also incorporate the decay properties of other particles, Gell-Mann and Nishijima introduced another quantum number S, called strangeness, to be also preserved. There is a fundamental relation between the preceding numbers

$$Q = h + \frac{1}{2}(B + S). \tag{2.2.51}$$

Gell-Mann and Ne'eman then interpreted this as a consequence of the embedding of the isospin and the "hypercharge" $B + S$ into the larger Lie group $SU(3)$ of "flavor" symmetry, and this led Gell-Mann to suggest the existence of "quarks", particles corresponding to the basic representation of $SU(3)$ on $\mathbb{C}^3$. Finally, "color", another internal $SU(3)$ degree of freedom, was proposed. The modern theory of the

strong interaction was then called quantum chromodynamics, after the Greek word for color.

We now turn to the unification of the fundamental forces. Electromagnetic and weak interactions (responsible for certain decay processes, like the beta decay of neutrons in nuclei) had been unified earlier, in 1967, in the so-called electroweak theory developed by Glashow, Salam and Weinberg, and in the early 1970s, the standard model combined this theory with quantum chromodynamics, the theory of strong interactions between quarks developed by Gell-Mann, Zweig and others. There are two types of particles in the theory: the fermions, which represent matter, and the bosons, which transmit forces between the fermions. Fermions have half-integer spin and satisfy the Pauli exclusion principle which states that two fermions cannot be in identical quantum states. This aspect will subsequently be incorporated into the formal framework by letting the fermions be odd Grassmann-valued, that is, anticommuting. Bosons have integer spins and do not have to satisfy the Pauli principle, and they will therefore be even Grassmann-valued, that is, commuting. The Lagrangian then has to couple the fermions and the bosons. There are four categories of bosons in the model: the photon that mediates electromagnetic interaction, the $W^{\pm}$ and Z boson for the weak force, eight types of gluons for the strong nuclear force, and finally the Higgs boson that induces spontaneous symmetry breaking of the gauge group for the electroweak interactions by a mechanism described in Sect. 2.2.4 below and that thereby provides masses to particles. While all the other bosons have been experimentally confirmed, with several of them predicted by the theory before their experimental observation, the Higgs boson has not yet been detected, but it may be detected soon with more powerful particle accelerators, because it should be seen at the energy scale where the unification between the electromagnetic and weak forces takes place (this is about 10^{12} electron volts, or about 10^{-16} times the Planck scale). Except for the Higgs boson and the W and Z bosons (which, in contrast to the photon, are massive, by the Higgs mechanism), the bosons are gauge particles, meaning that their contribution to the Lagrangian is invariant under gauge transformations from some internal symmetry group, as described in Sect. 2.2.1. The Lagrangians for bosons are of Yang–Mills type, as described in Sect. 1.2.3. The gauge group of electromagnetism is the Abelian group $U(1)$, as already explained. For the electroweak interaction, the gauge group is $SU(2) \times U(1)$. We should note, however, that the $SU(2)$ here is not the symmetry group of the weak interaction, which is not a gauge theory anyway. In fact, below the energy for the unification of the weak and the electromagnetic interactions, $SU(2)$ does not represent any symmetries. The gauge group $U(1)$ of electromagnetism is not the $U(1)$-factor in $SU(2) \times U(1)$, but rather a combination of the $U(1)$-subgroup of $SU(2)$ and the $U(1)$-factor in $SU(2) \times U(1)$ (Weinberg angle). For the strong interaction, the gauge group is $SU(3)$. The gauge group of the standard model is therefore the product $SU(3) \times SU(2) \times U(1)$. A class of extensions of the standard model, the so-called grand unified theories (GUTs), postulate that these groups are subgroups of a single large symmetry group, for example $SU(5)$. The grand unified theories have the advantage that they reduce the number of free parameters in the theory, see the remark at the end of Sect. 2.2.1. This symmetry, however, is only

present at very high energies, but is reduced to $SU(3) \times SU(2) \times U(1)$ at lower energies (including those achievable by current particle accelerators) by a process of spontaneous symmetry breaking, see Sect. 2.2.4 below. Most of these grand unified theories, including $SU(5)$, had to be given up because (in contrast to the standard model) they predicted proton decay at a rate not observed in nature.

Supersymmetry, to be discussed below, postulates an additional symmetry between bosonic and fermionic particles.

For the fermions, we have 12 different types ("flavors", each representing a particle and its antiparticle) in the standard model. That flavor is changed by the weak interaction, mediated by the heavy W and Z gauge bosons. The fermions come in two classes, leptons (including the electron and the electron neutrino) and quarks. Only the latter ones participate in strong interactions, by a property called "color", as already mentioned above. Since, in contrast to the electroweak forces, the strong force grows with the distance between quarks, they become confined in hadrons, colorless combinations. These can consist either of three quarks, like the protons and neutrons, and therefore be fermionic (baryons), or of a quark–antiquark pair, and then be bosonic (mesons), like the pions. In particular, these particles, protons, neutrons, pions and so on, are not elementary, but composite. The fermions are also classified into three generations. Each fermion in one generation has counterparts in the other generations that only differ in their masses, for example the electron, the muon and the tau lepton. Ordinary matter consists of fermions of the first generation, that is, the electron, the electron neutrino and the up and down quarks, as the ones in the other generations are substantially more massive and quickly decay into lower-generation ones.

While the standard model is well confirmed (with some revision to account for the experimentally observed neutrino masses that had not been predicted by the original model), renormalizable and generally accepted, it cannot yet be the ultimate answer because it does not include gravity and does not fare well at the cosmological level. Also, it is not entirely satisfactory because it contains too many free parameters that are not theoretically derived, but can only be experimentally determined. (As mentioned, the number of these free parameters is reduced in the grand unified theories.)

The concepts behind the standard model, however, are theoretically very appealing. When extended by the more recent ideas of superstring theory, they may well lead to a general theory of all known physical forces, at least according to the present opinion of many, if not most, theoretical physicists.[12]

In the present book, we are not concerned with the detailed physical aspects of the standard model, but only with the underlying mathematical concepts. Therefore, we shall essentially only treat a toy model, the so-called sigma model, that itself does not pretend to describe actual physics, but which on the one hand exhibits many of the conceptual issues in a particularly transparent manner, and on the other hand

[12] As opinions in theoretical physics can change rapidly, this statement may no longer be up to date when this book goes to print, and perhaps not even at the time of writing. There seems to be at least a tendency towards growing scepticism with regard to the prospects of superstring theory.

constitutes the starting point for string theory which, in contrast, aims at physically valid predictions.

2.2.4 The Higgs Mechanism

For a d-component scalar field $\phi = (\phi^1, \ldots, \phi^d)$, we may consider the Lagrangian

$$F = \frac{1}{2}\partial_\mu \phi^i \partial^\mu \phi_i - \frac{1}{2} a_i^j \phi^i \phi_j \quad \text{(w.l.o.g. } a_i^j = a_j^i\text{)}.$$

(Since the metric δ_{ij} on d-dimensional Euclidean space is flat, we can freely move indices up and down to conform to the usual summation conventions.)

By diagonalizing the quadratic form $a_i^j \phi^i \phi_j$ by an orthogonal transformation—which leaves $\partial_\mu \phi^i \partial^\mu \phi_i$ invariant—we can bring F into the form

$$F = \frac{1}{2}\partial_\mu \phi^i \partial^\mu \phi_i - \frac{1}{2}\sum_i \mu_i \phi^i \phi_i$$

(the μ_i are the eigenvalues of the symmetric matrix (a_{ij})).

If all $\mu_i \geq 0$, this Lagrangian describes d scalar particles of masses $m_i = \sqrt{\mu_i}$. Such an interpretation is no longer possible for negative μ_i.

More generally, for a multicomponent scalar field, we may consider the Lagrangian

$$F = \frac{1}{2}\partial_\mu \phi^i \partial^\mu \phi_i - V(\phi)$$

where the potential $V(\phi)$ incorporates self-interactions. Typically, V contains a quadratic term $a_{ij}\phi^i \phi^j$ and a higher-order term.

The classical vacuum corresponds to the minimum of V. The problem of symmetry breaking arises, namely that while V itself is invariant, the vacuum may not be invariant under the full symmetry group G of F. In that case, the vacuum consists of a whole G orbit, i.e., is degenerate.

Let us consider a simple example: ϕ is a real scalar field,

$$V(\phi) = \frac{1}{2}\mu\phi^2 + \frac{\lambda}{4}\phi^4.$$

In order to make V bounded from below, we assume $\lambda > 0$. If $\mu \geq 0$, then $\phi = 0$ is the vacuum, and the term $\frac{\lambda}{4}\phi^4$ may simply be treated as a higher-order perturbation for the Lagrangian

$$F_0 = \frac{1}{2}\partial_\mu \phi \partial^\mu \phi - \frac{1}{2}m^2\phi^2 \quad (m = \sqrt{\mu})$$

for a scalar particle of mass m. If $\mu < 0$, the situation changes, and this interpretation is no longer possible. The vacuum is now located at

$$\phi_0 = v = \pm\sqrt{\frac{-\mu}{\lambda}}.$$

In order to make a perturbation around the vacuum, one now has to consider the shifted field

$$\tilde{\phi} = \phi - v,$$

which breaks the symmetry between ϕ and $-\phi$. In terms of $\tilde{\phi}$, the Lagrangian becomes

$$F = \frac{1}{2}\partial_\mu\tilde{\phi}\partial^\mu\tilde{\phi} + \frac{1}{2}\mu\tilde{\phi}^2 - \lambda v\tilde{\phi}^3 - \frac{\lambda}{4}\tilde{\phi}^4 \quad (+\text{an irrelevant constant term}).$$

Since μ is negative, we can interpret $\tilde{\phi}$ as a scalar particle of mass $m = \sqrt{-\mu}$.

We now apply a similar analysis for a massive complex scalar particle coupled with a gauge field (i.e., a massless vector particle) and consider the Lagrangian (cf. (2.2.1), (2.2.14), (2.2.17))

$$F = \frac{1}{2}(\partial_\mu\phi + A_\mu\phi)(\partial^\mu\phi^* - A^\mu\phi^*) - \lambda(|\phi|^2 - \tau^2)^2 - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} \quad (\tau > 0). \tag{2.2.52}$$

(The minus sign in front of A^μ arises because A^μ is in $\mathfrak{u}(1) \cong i\mathbb{R}$, and we have to take the complex conjugate $(A^\mu\phi)^* = -A^\mu\phi^*$.)

For $\lambda > 0$, the vacuum now is at

$$|\phi| = \tau,$$

i.e.,

$$\phi = \tau e^{i\vartheta}.$$

Thus, the vacuum is a nontrivial $U(1)$ orbit, i.e., degenerate. We therefore impose an additional gauge condition

$$\operatorname{Im}\phi = 0, \qquad \operatorname{Re}\phi > 0$$

which uniquely locates the vacuum at

$$\phi = \tau.$$

Again, we want to expand around the vacuum and put

$$\tilde{\phi} = \phi - \tau.$$

The Lagrangian becomes (up to a constant term)

$$\begin{aligned} F &= \frac{1}{2}\partial_\mu\tilde{\phi}\partial^\mu\tilde{\phi} - \lambda\tilde{\phi}^2(\tilde{\phi}+2\tau)^2 + \frac{1}{2}A_\mu A^\mu(\tilde{\phi}+\tau)^2 - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} \\ &= \frac{1}{2}\partial_\mu\tilde{\phi}\partial^\mu\tilde{\phi} - 4\lambda\tau^2\tilde{\phi}^2 + \frac{1}{2}\tau^2 A_\mu A^\mu - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} + \text{higher-order terms.} \end{aligned}$$

Thus, up to these higher-order terms, F describes a scalar particle $\tilde{\phi}$ of mass $m = 2\tau\sqrt{2\lambda}$ and a vector particle A of mass τ. Because of our above gauge condition, the scalar particle now has only one real degree of freedom left; the other degree of freedom has been gauged into the vector particle that has acquired a mass.

Alternatively, we write

$$\phi = e^{i\frac{\eta}{\tau}}(\tau+\xi) = \tau + \xi + i\eta + \mathcal{O}(\xi^2+\eta^2)$$

and consider

$$\xi = e^{-i\frac{\eta}{\tau}}\phi - \tau,$$

$$A'_\mu = A_\mu - \frac{1}{\tau}\partial_\mu\eta.$$

Then F becomes

$$\begin{aligned} F = \frac{1}{2}\partial_\mu\xi\,\partial^\mu\xi - \frac{1}{2}\tau^2 A'_\mu A'^\mu - 4\lambda\tau^2\xi^2 - \frac{1}{4e^2}F_{\mu\nu}F^{\mu\nu} \\ + \text{higher-order terms in } \xi \text{ and } A', \end{aligned}$$

but η has disappeared, gauged into the vector particle A' that has acquired a mass.

Let us discuss the Higgs mechanism in more generality. Again, we start with a simple scalar field ϕ and a Lagrangian

$$F(\phi) = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi - V(\phi).$$

We assume that ϕ takes values in a vector bundle with structure group G, and that V is G-invariant.

Let v be a classical vacuum, i.e., a minimizer of the potential V. Then

$$\frac{dV}{d\phi} = 0 \quad \text{at } \phi = v.$$

We assume that the symmetry group G is broken in the sense that v is only invariant under a smaller group $H \subset G$, but not under all of G.

We choose generators $\vartheta^1, \ldots, \vartheta^N$ of the Lie algebra $\mathfrak{g}$ of G in such a manner that $\vartheta^1, \ldots, \vartheta^M$ generate the Lie algebra $\mathfrak{h}$ of H. Thus

$$\vartheta^i v = 0 \quad \text{for } i = 1, \ldots, M,$$

$$\vartheta^j v \neq 0 \quad \text{for } j = M+1, \ldots, N.$$

Since V is G-invariant, the derivative of V in the directions tangent to the G-orbit of ϕ vanishes, that is,

$$\frac{dV}{d\phi}(\vartheta^j\phi) = \frac{d}{dt}V(\exp(t\vartheta^j)\phi)_{|_{t=0}} = 0 \quad \text{for all } j.$$

Differentiating this relation w.r.t. ϕ yields

$$\frac{d^2V}{d\phi^2}\vartheta^j\phi + \frac{dV}{d\phi}\vartheta^j = 0 \quad \text{for all } j.$$

At $\phi = v$, $\frac{dV}{d\phi} = 0$, and hence

$$\frac{d^2V(v)}{d\phi^2}\vartheta^j v = 0.$$

Since for $j = M+1, \ldots, N$, $\vartheta^j v \neq 0$, $\vartheta^j v$ is an eigenvector of the Hessian $\frac{d^2V(v)}{d\phi^2}$ with zero eigenvalue. Since this Hessian gives the quadratic term in the expansion of our Lagrangian $F(\phi)$ at the vacuum $\phi = v$, and since the eigenvalues of this quadratic form are interpreted as squared masses, we interpret $\exp(\vartheta^j)v$ for $j = M+1, \ldots, N$ as a massless boson.

Thus, for each broken generator of the symmetry group, we have found a so-called massless Goldstone boson. To get the Higgs mechanism, we introduce a gauge field A, with values in $\mathfrak{g}$, that couples to our scalar field ϕ and consider the Lagrangian

$$F(\phi, A) = -\frac{1}{4}F^i_{\mu\nu}F_i^{\mu\nu} + \frac{1}{2}(\partial_\mu + A_\mu\phi)(\partial^\mu + A^\mu\phi) - V(\phi),$$

with a G-invariant potential V as before.

Again, we assume that the vacuum v is only H-invariant. We expand

$$\begin{aligned}\phi &= v + \sum_{i=1}^{M}\xi_i\vartheta^i + \sum_{j=M+1}^{N}\eta_j\vartheta^j v + O(\xi^2+\eta^2)\\ &= \left\{\exp\left(\sum_{j=M+1}^{N}\eta_j\vartheta^j\right)\right\}\left(v + \sum_{i=1}^{M}\xi_i\vartheta^i\right).\end{aligned}$$

With this expansion and with the notation $A_\mu = A_{i\mu}\vartheta^i$, we obtain the term

$$\operatorname{Tr}\vartheta^i v\vartheta_j v A_{i\mu}A^{j\mu}$$

in $F(\phi, A)$, which we interpret as the mass term for the vector fields. The gauge change

$$A_{j'\mu} = A_{j\mu} - \partial_\mu\eta_j,$$

together with the fact that $\vartheta^j v$ for $j = M+1, \ldots, N$ is a 0-eigenvector of the Hessian of V at v, makes the η^j disappear in the expansion of $F(\phi, A)$ up to second-order. Thus, we obtain $N - M$ massive vector fields that have absorbed the $N - M$ Goldstone bosons.

The idea of a gauge theoretic interpretation of spontaneous symmetry breaking was conceived independently by Englert and Brout and by Higgs in the early 1960s.

2.2.5 *Supersymmetric Point Particles*

We now take a step backwards and consider a one-dimensional domain. In geometric terms, the aim of this section is to derive a supersymmetric version of the action functional (1.1.119) for geodesics, discussed in Sect. 1.1.4. In physical terms, we want to introduce Lagrangians that exhibit a symmetry between bosonic and fermionic fields. The supersymmetric point particle is the simplest instance of this.

We start with the Euclidean case. We consider $(t, \theta) \in \mathbb{R}^{1|1}$ as coordinates (see Sect. 1.5.2), as well as scalar superfields

$$X^a(t, \theta) = \phi^a(t) + \psi^a(t)\theta, \quad a = 1, \ldots, d, \tag{2.2.53}$$

with ϕ^a even and ψ^a odd.[13] Our Lagrangian is

$$L_1 = \frac{1}{2}(\dot{\phi}^a \dot{\phi}_a + \psi^a \dot{\psi}_a) \tag{2.2.54}$$

and the action is

$$S_1 = \frac{1}{2}\int (\dot{\phi}^a \dot{\phi}_a + \psi^a \dot{\psi}_a) dt. \tag{2.2.55}$$

Here, t and θ are the independent variables, ϕ and ψ the dependent ones. ϕ is called a bosonic field, or boson for short, ψ a fermionic one or fermion.

Thus, both the arguments and the values of X are Grassmannian. Here, this makes X even. We also note that it is important that ψ be anticommuting in (2.2.55), as otherwise an integration by parts would imply that $\int \psi^a \dot{\psi}_a \, dt$ vanishes identically.

Remark

1. In the physics literature, one usually puts a factor i in front of the term $\psi^a \dot{\psi}_a$ in the Lagrangian F_1, and likewise in the other supersymmetric Lagrangians we shall treat here. Since the expression is Grassmann valued, this becomes a matter of convention, in contrast to the real-valued situation of (2.2.12). The convention with the i is compatible with the following convention usually adopted in the

[13] As ψ and θ are both odd, they anticommute. Otherwise, at this point, they have nothing to do with each other.

physics literature for defining a complex conjugation * on Grassmann variables. This conjugation should satisfy

$$(\tau_1 + \tau_2)^* = \tau_1^* + \tau_2^*, \qquad (\tau_1\tau_2)^* = \tau_2^*\tau_1^*. \tag{2.2.56}$$

One defines the complex conjugate of an ordinary complex number as the ordinary complex conjugate, and one assumes that the generators $\vartheta_1, \ldots, \vartheta_N$ of the Grassmann algebra are real, i.e.,

$$\vartheta_i^* = \vartheta_i. \tag{2.2.57}$$

Thus

$$(\vartheta_{\alpha_1}\vartheta_{\alpha_2}\cdots\vartheta_{\alpha_k})^* = \vartheta_{\alpha_k}\cdots\vartheta_{\alpha_2}\vartheta_{\alpha_1}. \tag{2.2.58}$$

The elements of the Grassmann algebra are called supernumbers. A supernumber τ is then called real if $\tau^* = \tau$, imaginary if $\tau^* = -\tau$. Thus, $\vartheta_{\alpha_1}\cdots\vartheta_{\alpha_k}$ is real if $\frac{1}{2}k(k-1)$ is even, imaginary if $\frac{1}{2}k(k-1)$ is odd. With this convention, the term $\psi^a\dot{\psi}^a$, being the product of two real odd quantities, is purely imaginary, and the factor i then serves to make it real. In any case, a factor i in the Lagrangian in front of the ψ term would then require also a compensating factor in the supersymmetry transformations (2.2.68) below.

2. We may, in fact, put any factor κ in front of the term $\psi^a\dot{\psi}^a$ in the Lagrangian F_1 (and likewise, we may put factors in front of other terms we shall add to our Lagrangians to make them supersymmetric). We then simply need to compensate for this in our variations (2.2.68) below, for example by inserting a factor $1/\kappa$ into the right-hand side of the variation for ψ. With such a factor κ, we can then perform expansions of the Lagrangian and other quantities in terms of κ which is a useful device often seen in physics texts.

The Euler–Lagrange equations for L_1 are

$$\ddot{\phi}^a = 0, \tag{2.2.59}$$

$$\dot{\psi}^a = 0, \tag{2.2.60}$$

and L_1 describes a free superpoint particle. a is a vector index, and the setting can be generalized to particles moving on a Riemannian manifold M with metric $g_{ab}(y)dy^a \otimes dy^b$. $\dot{\phi}(t)$ then transforms as a tangent vector to M, and one postulates that $\psi(t)$ likewise transforms as a tangent vector in $T_{\phi(t)}M$, i.e., under a change of coordinates $y = f(y')$ in M, one has

$$\psi^a = \frac{\partial y^a}{\partial y'^b}\psi'^b \quad (\psi \text{ is thus a vector with Grassmann coefficients}). \tag{2.2.61}$$

(Note that here we are transforming the coordinates in the image; anticipating the discussion below, this is perfectly compatible with ψ being a spinor field, because

this refers to the transformation behavior w.r.t. the independent variables.) Thus, ψ is an odd vector field along the map ϕ. In particular, the scalar product

$$\langle \dot{\phi}, \psi \rangle = g_{ab}(\phi)\dot{\phi}^a \psi^b \tag{2.2.62}$$

is invariant under coordinate transformations on M. We may use the Lagrangian

$$L_2 := \frac{1}{2} g_{ab}(\phi)\dot{\phi}^a \dot{\phi}^b + \frac{1}{2} g_{ab}(\phi)\psi^a \nabla_{\frac{d}{dt}} \psi^b, \tag{2.2.63}$$

with

$$\nabla_{\frac{d}{dt}} \psi^b = \frac{d}{dt}\psi^b + \dot{\phi}^a \Gamma^b_{ac}(\phi)\psi^c. \tag{2.2.64}$$

The Euler–Lagrange equations for

$$S_2 = \int L_2(\phi, \psi)dt \tag{2.2.65}$$

are

$$\nabla_{\frac{d}{dt}} \dot{\phi}^a - \frac{1}{2} R^a_{bcd}\dot{\phi}^b \psi^c \psi^d = 0, \tag{2.2.66}$$

$$\nabla_{\frac{d}{dt}} \psi^a = 0. \tag{2.2.67}$$

In contrast to (2.2.59), (2.2.60), these field equations couple ϕ and ψ. Equation (2.2.66) is the supersymmetric generalization of the geodesic equation $\nabla_{\frac{d}{dt}} \dot{x}^a = 0$, cf. (1.1.124), (1.1.125). When we consider ψ as an ordinary field, the equations (2.2.66), (2.2.67) describe a spinning particle in a gravitational field. Unless $\psi = 0$, the particle no longer moves along a geodesic, because of the presence of the second term in the first equation.

We return to the action S_1, and we perform the variations

$$\begin{aligned} \delta\phi^a &= -\varepsilon\psi^a, \\ \delta\psi^a &= \varepsilon\dot{\phi}^a \end{aligned} \tag{2.2.68}$$

with an odd parameter ε. The variation of the Lagrangian L_1 of S_1 is

$$\begin{aligned} \delta L_1 &= \dot{\phi}^a \delta\dot{\phi}_a + \frac{1}{2}(\delta\psi^a)\dot{\psi}_a + \frac{1}{2}\psi^a \delta\dot{\psi}_a \\ &= -\varepsilon\dot{\phi}^a\dot{\psi}_a + \frac{1}{2}\varepsilon\dot{\phi}^a\dot{\psi}_a + \frac{1}{2}\psi^a\varepsilon\ddot{\phi}_a \\ &= -\varepsilon\dot{\phi}^a\dot{\psi}_a + \frac{1}{2}\varepsilon\dot{\phi}^a\dot{\psi}_a + \frac{1}{2}\frac{d}{dt}(\psi^a\varepsilon\dot{\phi}_a) - \frac{1}{2}\dot{\psi}^a\varepsilon\dot{\phi}_a \\ &= -\varepsilon\frac{d}{dt}\left(\frac{1}{2}\psi^a\dot{\phi}_a\right) \quad \text{(using } \dot{\psi}^a\varepsilon = -\varepsilon\dot{\psi}^a\text{, as } \varepsilon \text{ and } \psi^a \text{ are odd).} \end{aligned} \tag{2.2.69}$$

Since δL_1 is a total derivative, it follows that S_1 is invariant under the variation (2.2.68). The point of the superspace formalism is now that this variation is induced by a variation of the independent variables t, θ. Namely, we consider the so-called supersymmetry generators

$$Q := \tau := \theta\partial_t + \partial_\theta, \tag{2.2.70}$$

$$P := \partial_t. \tag{2.2.71}$$

Here, Q and P are the notations usually employed in physics texts. The operator $Q = \tau$ should be compared with the operator $D := \partial_\theta - \theta\partial_t$ introduced in (1.5.34) in Sect. 1.5.2. Then

$$\varepsilon QX^a(t,\theta) = \varepsilon Q(\phi^a(t) + \psi^a(t)\theta) = \varepsilon\dot{\phi}^a\theta - \varepsilon\psi^a$$
$$\text{(since } \theta^2 = 0 \text{ and } \theta \text{ commutes with } \dot{\phi}^a\text{, but anticommutes with } \psi^a\text{)}, \tag{2.2.72}$$

which yields (2.2.68). We have

$$[Q, Q] \equiv 2Q^2 = 2(\theta\partial_t + \partial_\theta)(\theta\partial_t + \partial_\theta) = 2\partial_t = 2P$$
$$\text{(since, e.g., } \theta \text{ and } \partial_\theta \text{ anticommute)}. \tag{2.2.73}$$

Similarly

$$[D, D] = -2\partial_t = -2P, \tag{2.2.74}$$

$$[Q, P] = (\theta\partial_t + \partial_\theta)\partial_t - \partial_t(\theta\partial_t + \partial_\theta) = 0$$
$$\text{(since } \partial_\theta \text{ anticommutes with } \theta \text{ and commutes with } \partial_t\text{)}, \tag{2.2.75}$$

$$[D, P] = 0, \tag{2.2.76}$$

$$[P, P] = 0. \tag{2.2.77}$$

Equations (2.2.73), (2.2.75), (2.2.77) mean that Q and P generate a super Lie algebra, see (1.5.5), i.e., a mod 2 graded vector space S over $\mathbb{C}$, endowed with a superbracket $[\cdot,\cdot]$ that is bilinear, mod 2 graded additive and superanticommutative, i.e.,

$$[A, B] = [B, A] \quad \text{if } A, B \text{ both are of odd degree,}$$
$$[A, B] = -[B, A] \quad \text{otherwise, i.e., if } A \text{ or } B \text{ is even}$$

and that satisfies the super-Jacobi identity (1.5.6),

$$(-1)^{ac}[A,[B,C]] + (-1)^{ba}[B,[C,A]] + (-1)^{cb}[C,[A,B]] = 0,$$

where a, b, c are the degrees of A, B and C, resp. Returning to (2.2.69), we have

$$\delta L_1 = -\varepsilon\dot{I}$$

with

$$I := \frac{1}{2}\psi^a \dot{\phi}^a.$$

We also have

$$\delta I = \frac{1}{2}\varepsilon \dot{\phi}^a \dot{\phi}^a + \frac{1}{2}\varepsilon \psi^a \dot{\psi}^a = \varepsilon L_1.$$

These variations for L_1 and I are quite similar to the ones for ϕ^a and ψ^a, compare (2.2.68), except that the roles of bosons and fermions have been exchanged. In the physics literature, the representation of the supersymmetry algebra on the ϕ^a, ψ^a space is called a "bosonic multiplet", whereas the one on the L_1, I space is called a "fermionic multiplet".

We are now going to consider a functional on $\mathbb{R}^{1|2}$ with coordinates (t, θ^1, θ^2), and the supersymmetry generators

$$Q_\alpha = \theta^\alpha \partial_t + \partial_{\theta^\alpha}, \tag{2.2.78}$$

$$P = \partial_t. \tag{2.2.79}$$

They span a super Lie algebra with

$$\begin{aligned} [Q_\alpha, Q_\beta] &= 2\delta_{\alpha\beta} P, \\ [Q_\alpha, P] &= 0 = [P, P]. \end{aligned} \tag{2.2.80}$$

We try a superfield

$$X^a(t, \theta^\alpha) = \phi^a(t) + \psi^a_\alpha(t)\theta^\alpha. \tag{2.2.81}$$

We have

$$\varepsilon Q_1 X^a(t, \theta^1, \theta^2) = -\varepsilon \psi^a_1 + \varepsilon \dot{\phi}^a \theta^1 - \varepsilon \dot{\psi}^a_2 \theta^1 \theta^2, \tag{2.2.82}$$

and similarly for εQ_2. This is different from the previous situation as we now also get a $\theta^1\theta^2$ term that can neither be considered as a variation of ϕ^a nor as a variation of ψ^a. This problem stems from the fact that we now have only one bosonic field ϕ, but two fermionic fields ψ_1, ψ_2. Since supersymmetry mixes bosonic and fermionic fields, we need to introduce an additional bosonic field and consider the superfield

$$Y^a(t, \theta^\alpha) = \phi^a(t) + \psi^a_\alpha(t)\theta^\alpha + F^a(t)\theta^1\theta^2 \tag{2.2.83}$$

with an even F^a. We now get

$$\varepsilon Q_1 Y^a(t, \theta^\alpha) = -\varepsilon \psi^a_1 + \varepsilon \dot{\phi}^a \theta^1 + \varepsilon F^a \theta^2 - \varepsilon \dot{\psi}^a_2 \theta^1 \theta^2. \tag{2.2.84}$$

Thus, we get the variations

$$\begin{aligned}\delta\phi^a &= -\varepsilon\psi_1^a,\\ \delta\psi_1^a &= \varepsilon\dot{\phi}^a,\\ \delta\psi_2^a &= \varepsilon F^a,\\ \delta F^a &= -\varepsilon\dot{\psi}_2^a\end{aligned} \tag{2.2.85}$$

for Q_1 and analogous variations for Q_2.

We consider the action

$$S_3 = \int \frac{1}{2}(\dot{\phi}^a\dot{\phi}_a + \psi_\alpha^a\dot{\psi}_{a\alpha} + F^a\ F_a)dt. \tag{2.2.86}$$

The Euler–Lagrange equations for L_3 are

$$\begin{aligned}\ddot{\phi}^a &= 0,\\ \dot{\psi}_\alpha^a &= 0,\\ F^a &= 0.\end{aligned} \tag{2.2.87}$$

Thus, the equations for the F^a are trivial, and the F^a are auxiliary variables that do not evolve and can be eliminated. They are only needed to close the supersymmetry algebra. We also observe that on-shell, i.e., if the equations of motion (2.2.87) are satisfied, we have

$$\varepsilon Q_1 X^a(t,\theta^\alpha) = -\varepsilon\psi_1^a + \varepsilon\dot{\phi}^a\theta^1 \tag{2.2.88}$$

so that the supersymmetry algebra closes here without the F^a field. On-shell, the number of degrees of freedom of the ψ_α^a fields is reduced so that we no longer need the F^a field in order to restore the balance between bosonic and fermionic fields, and on-shell, F^a vanishes anyway.

We may also write things in a more invariant manner. Namely, with

$$X(t,\theta) = \phi(t) + \psi(t)\theta, \tag{2.2.89}$$

we have

$$DX = -\dot{\phi}\theta - \psi \tag{2.2.90}$$

and

$$D(DX) = -(\dot{\phi} + \dot{\psi}\theta), \tag{2.2.91}$$

and so we have

$$S_1 = \int \frac{1}{2}DXD(DX)dtd\theta. \tag{2.2.92}$$

Since the supersymmetry generator Q anticommutes with D, we now see directly, without any need for further computation, that L_1 remains invariant under supersymmetry transformations. Similarly, to represent S_3, we consider the operators

$$D_\alpha = -\theta^\alpha \partial_t + \partial_{\theta^\alpha}. \tag{2.2.93}$$

From (2.2.83), we obtain

$$S_3 = \frac{1}{4}\int \epsilon^{\alpha\beta} D_\alpha Y D_\beta Y dt d\theta^2 d\theta^1, \tag{2.2.94}$$

for the field Y, where the antisymmetric ϵ-tensor satisfies

$$\epsilon^{12} = -\epsilon^{21} = 1. \tag{2.2.95}$$

We observe that, in contrast to (2.2.92), (2.2.94) contains only two Ds, the reason being that here we have two odd variables, θ^1 and θ^2, that are integrated. The Euler–Lagrange equations (2.2.87) for L_3 then become

$$\epsilon^{\alpha\beta} D_\alpha D_\beta Y = 0. \tag{2.2.96}$$

We now wish to include self-interaction terms in the functional and consider a (smooth) potential function of the superfields Y^a,

$$W(Y^a) = W(\phi^a + \psi^a_\alpha \theta^\alpha + F^a \theta^1 \theta^2). \tag{2.2.97}$$

Thus, we have the expansion

$$W(Y) = w(\phi) + \frac{\partial w(\phi)}{\partial \phi^a}(\psi^a_\alpha \theta^\alpha + F^a \theta^1 \theta^2) + \frac{1}{2}\frac{\partial^2 w(\phi)}{\partial \phi^a \partial \phi^b}\psi^a_\alpha \theta^\alpha \psi^b_\beta \theta^\beta. \tag{2.2.98}$$

We introduce an interaction Lagrangian

$$L_{int} = -\int W(Y(t, \theta^1, \theta^2)) dt d\theta^2 d\theta^1 \tag{2.2.99}$$

$$= -\int \left(\frac{\partial w(\phi(t))}{\partial \phi^a(t)} F^a + \frac{\partial^2 w(\phi(t))}{\partial \phi^a(t) \partial \phi^b(t)} \psi^a_1 \psi^b_2 \right) dt. \tag{2.2.100}$$

The total Lagrangian is

$$L_3 + L_{int}. \tag{2.2.101}$$

The corresponding Euler–Lagrange equations include the following equation for F

$$F^a = \frac{\partial w(\phi)}{\partial \phi^a}. \tag{2.2.102}$$

Again, this is an algebraic equation and thus eliminates F^a, and we may write

$$L_3 + L_{int} = \int dt \left(\frac{1}{2}\dot{\phi}^a \dot{\phi}_a + \frac{1}{2}\psi^a_\alpha \dot{\psi}_{a\alpha} - \frac{1}{2}\frac{\partial w(\phi)}{\partial \phi^a}\frac{\partial w(\phi)}{\partial \phi_a} - \frac{\partial^2 w(\phi)}{\partial \phi^b \partial \phi^a}\psi^a_1 \psi^b_2 \right). \tag{2.2.103}$$

2.3 Variational Aspects

2.3.1 The Euler–Lagrange Equations

Here, we present a brief summary of the calculus of variations as needed for treating action functionals and their symmetries. For more details, we refer to [66]. We consider a Lagrangian

$$L = \int F(x, u(x), du(x))dx \tag{2.3.1}$$

and variations

$$u(x) \to u(x) + s\delta u(x); \tag{2.3.2}$$

here, s is a parameter, and $\delta u(x)$ is the variation of u at x.[14] This means the following:

$$\delta L(u)(\delta u) := \frac{\delta L(u)}{\delta u} := \frac{d}{ds}\int F(x, u(x) + s\delta u(x), d(u(x) + s\delta u(x)))dx_{|s=0}. \tag{2.3.3}$$

More generally, one may consider a C^2-family of diffeomorphisms $h_s(u)$ of the dependent variables, defined for s in some neighborhood of 0, with h_0 being the identity, and

$$\frac{d}{ds}h_s(u(x))_{|s=0} = \delta u(x), \tag{2.3.4}$$

and

$$\delta L(u)(\delta u) = \frac{d}{ds}\int F(x, h_s(u(x)), d(h_s(u(x))))dx_{|s=0}. \tag{2.3.5}$$

Since we consider only infinitesimal variations, (2.3.3) and (2.3.5) are the same, and we may use either formulation.

We now assume that

$$\delta L(u)(\delta u)\left(= \frac{\delta L(u)}{\delta u}\right) = 0 \tag{2.3.6}$$

for a variation δu. We compute

$$\delta L(u)(\delta u) = \int \left(F_u\,(x, u(x), du(x))\,\delta u(x) + F_{p^\alpha}\,(x, u(x), du(x))\,\frac{\partial}{\partial x^\alpha}\delta u(x)\right)dx$$

[14] The integration is supposed to take place on some domain Ω, but as that domain will play no essential role, we suppress it in our notation. In many situations, the variations δu are required to satisfy certain conditions at the boundary of Ω (because u itself is constrained there), but again, that will not be essential in the present context. Of course, it is important to realize that the integral in the definition of the Lagrangian is a definite one. Likewise, we suppress other constraints that u may have to satisfy, and that need to be preserved by its variations.

$$= \int \left(F_u(x, u(x), du(x)) - \frac{d}{dx^\alpha} F_{p^\alpha}(x, u(x), du(x)) \right) \delta u(x) dx \tag{2.3.7}$$

where p^α is a dummy variable for the place where $\frac{\partial u}{\partial x^\alpha}$ is inserted, and subscripts denote partial derivatives, e.g. $F_u := \frac{\partial F}{\partial u}$. (In fact, u might be vector valued, and in that case F_u stands for all the partial derivatives of F w.r.t. the components of u.) For the last line, we have integrated by parts, assuming that the variation δu is such that no boundary term occurs. We note the full derivative $\frac{d}{dx^\alpha}$ that indicates that we need to differentiate $F_{p^\alpha}(x, u(x), du(x))$ for all three occurrences of x.

Comparing (2.3.7) with (2.1.52), we obtain

$$\frac{\delta L(u)}{\delta u(x)} = F_u(x, u(x), du(x)) - \frac{d}{dx^\alpha} F_{p^\alpha}(x, u(x), du(x)). \tag{2.3.8}$$

Thus, $\frac{\delta L(u)}{\delta u(x)}$ represents the Euler–Lagrange operator.

We now assume that u is a stationary point, e.g., a minimizer of L, in the sense that (2.3.6) holds for all variations δu satisfying an appropriate boundary condition. We then obtain the **Euler–Lagrange equations**

$$\frac{\delta L(u)}{\delta u(x)} = F_u(x, u(x), du(x)) - \frac{d}{dx^\alpha} F_{p^\alpha}(x, u(x), du(x)) = 0. \tag{2.3.9}$$

When we wish to derive things in a geometrically invariant way, we should change the preceding formalism slightly. The reason is that the integration measure dx employed in (2.3.1) is not geometrically invariant. More natural is the volume form

$$\sqrt{\det g_{ij}}\, dx \text{ for a Riemannian metric } g_{ij}. \tag{2.3.10}$$

Thus, in place of (2.3.1), we should consider

$$L = \int G(x, u(x), du(x)) \sqrt{\det g_{ij}}\, dx. \tag{2.3.11}$$

We abbreviate

$$\sqrt{g} := \sqrt{\det g_{ij}}. \tag{2.3.12}$$

The Euler–Lagrange equations (2.3.9) then become

$$\frac{\delta L(u)}{\delta u(x)} = G_u(x, u(x), du(x)) - \frac{1}{\sqrt{g}} \frac{d}{dx^\alpha} (\sqrt{g} G_{p^\alpha}(x, u(x), du(x))) = 0. \tag{2.3.13}$$

2.3.2 *Symmetries and Invariances: Noether's Theorem*

We consider an action

$$L = \int F(x, u(x), du(x)) dx \tag{2.3.14}$$

that is infinitesimally invariant under some variation

$$u(x) \to u(x) + s\eta(x). \tag{2.3.15}$$

As just explained, the invariance means that

$$\delta L := \frac{d}{ds} \int F(x, u(x) + s\eta(x), d(u(x) + s\eta(x)))dx_{|s=0} = 0. \tag{2.3.16}$$

In contrast to the preceding, here we consider arbitrary fields u, but only particular variations η—above, we had considered arbitrary variations δu for a particular u.

Again, we may alternatively consider a C^2-family of diffeomorphisms $h_s(u)$ of the dependent variables, defined for s in some neighborhood of 0, with h_0 being the identity, and $\frac{d}{ds}h_s(u(x))_{|s=0} = \eta(x)$. We now assume

$$\int F(x, h_s(u(x)), dh_s(u(x)))dx = \int F(x, u(x), du(x))dx$$

for all s near 0 and all admissible u. The interpretation that a variation arises from a diffeomorphism of the dependent variables that leaves the action invariant is useful when one wants to analyze invariances in the context of global analysis.

As in (2.3.7), we obtain

$$0 = \int \left(F_u(x, u(x), du(x))\eta(x) + F_{p^\alpha}(x, u(x), du(x)) \frac{\partial}{\partial x^\alpha} \eta(x) \right) dx. \tag{2.3.17}$$

We now consider a more general variation

$$u(x) \to u(x) + s(x)\eta(x), \tag{2.3.18}$$

that is, where the variation parameter s may also depend on x. Since the variation δL vanishes for constant s, it must now be proportional to the derivative of s, that is,

$$\delta L = \int F_{p^\alpha}(x, u(x), du(x))\eta(x) \frac{\partial}{\partial x^\alpha} s(x)dx. \tag{2.3.19}$$

If we now assume in addition that u is stationary, that is, δL vanishes for all variations, (2.3.19) vanishes as well, and we conclude

$$\frac{d}{dx^\alpha}(F_{p^\alpha}(x, u(x), du(x))\eta(x)) = 0. \tag{2.3.20}$$

This is a special version of **Noether's theorem**. We define the Noether current

$$j^\alpha(x) := F_{p^\alpha}(x, u(x), du(x))\eta(x). \tag{2.3.21}$$

Noether's theorem thus says that j is conserved in the sense that

$$\operatorname{div} j = \frac{\partial}{\partial x^\alpha} j^\alpha(x) = 0. \tag{2.3.22}$$

As at the end of Sect. 2.3.1, when we consider a functional of the form (2.3.10), we obtain the geometric version of Noether's theorem,

$$\operatorname{div} j = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^\alpha} (\sqrt{g}\, j^\alpha(x)) = 0, \tag{2.3.23}$$

with $j^\alpha(x) := G_{p^\alpha}(x, u(x), du(x))\eta(x)$.

For the general version of Noether's theorem, we also allow for variations of the independent variable x. That means that we consider

$$x \to x' := x + s\delta x, \tag{2.3.24}$$

$$u(x) \to u'(x') := \psi(u(x)) := u(x) + s\delta\psi(x). \tag{2.3.25}$$

When we write

$$u'(x) = u(x) + s\eta(x) \tag{2.3.26}$$

we have

$$\eta = \delta\psi - \frac{du}{dx^\beta}\delta x^\beta. \tag{2.3.27}$$

Since now the integration measure dx also varies under (2.3.24), the Noether current becomes

$$j^\alpha := F_{p^\alpha}(x, u(x), du(x))\left(\delta\psi - \frac{du}{dx^\beta}\delta x^\beta\right) + F(x, u(x), du(x))\delta x^\alpha, \tag{2.3.28}$$

and again a conserved quantity,

$$\operatorname{div} j = \frac{\partial}{\partial x^\alpha} j^\alpha(x) = 0. \tag{2.3.29}$$

Here as well, in the Riemannian setting, we instead have

$$j^\alpha := G_{p^\alpha}(x, u(x), du(x))\left(\delta\psi - \frac{du}{dx^\beta}\delta x^\beta\right) + G(x, u(x), du(x))\delta x^\alpha \tag{2.3.30}$$

and (2.3.29) should be replaced by

$$\operatorname{div} j = \frac{1}{\sqrt{g}} \frac{\partial}{\partial x^\alpha} (\sqrt{g}\, j^\alpha(x)) = 0, \tag{2.3.31}$$

cf. (2.3.23).

Finally, in many situations, the Lagrangian is only invariant up to a divergence term, that is,

$$\delta L + \int \operatorname{div} A(u, \delta x, \delta u) = 0. \tag{2.3.32}$$

In this case, we obtain that (in abbreviated notation when compared to (2.3.28))

$$\operatorname{div}(F_p \delta u + (F - F_p du)\delta x + A(u, \delta x, \delta u)) = 0. \tag{2.3.33}$$

The standard example is the conservation of energy for time-invariant Lagrangians. We consider the action

$$S_0 = \int \left(\frac{1}{2} \dot{\phi}^a \dot{\phi}_a - V(\phi) \right) dt. \tag{2.3.34}$$

Since the integrand does not depend explicitly on t, it is invariant under a variation $t \to t + \delta t$, and from (2.3.28), we conclude that the negative of the Hamiltonian (= energy) is preserved, this being given by

$$\frac{1}{2} \dot{\phi}^a \dot{\phi}_a + V(\phi). \tag{2.3.35}$$

The same happens for our supersymmetric Lagrangian (2.2.55),

$$S_1 = \frac{1}{2} \int (\dot{\phi}^a \dot{\phi}_a + \psi^a \dot{\psi}_a) dt; \tag{2.3.36}$$

again, the Noether current is

$$-\frac{1}{2} \dot{\phi}^a \dot{\phi}_a, \tag{2.3.37}$$

that is, minus the Hamiltonian. We observe that the Noether current here contains only the bosonic field ϕ, not the fermionic one ψ. We may also consider supersymmetry invariance in this framework. When we perform the variations (2.2.68), we compute that the associated current is given by

$$-\epsilon \dot{\phi} \psi. \tag{2.3.38}$$

The superspace formalism represented the supersymmetry variation as a variation of the independent variables. Since S_1 in (2.2.92), however, also contains terms of the form $D^2 X$, the formalism needs to be slightly extended to carry over. The Lagrangian S_3 as written in (2.2.94) does not present this problem, and so, in that case, the conserved current can be computed from a variation of the independent variables without the need for an extension of the formalism, except that of course we now need to take the signs into account as always when performing supercomputations.

We now finally derive some implications of Noether's theorem in the Minkowski setting. According to (2.3.29), invariance implies a conserved current:

$$\partial_\alpha j^\alpha = 0. \tag{2.3.39}$$

From this, we obtain a conserved charge:

$$Q = \int d^{d-1} x j^0, \tag{2.3.40}$$

where j^0 is the time component of j and $d^{d-1}x$ denotes the integration over a space-like slice,

$$\begin{aligned}\frac{d}{dt}Q &= \int d^{d-1}x\ \partial_0 j^0 \\ &= -\int d^{d-1}x\ \partial_a\ j^a \quad (a \text{ running over spatial indices) by (2.3.24)}, \\ &= 0 \end{aligned} \tag{2.3.41}$$

by the divergence theorem when j vanishes sufficiently quickly at spatial infinity.

2.4 The Sigma Model

In this section, we discuss an action functional that is fundamental to conformal field theory and string theory, the sigma model and its nonlinear and supersymmetric versions. In the mathematical literature, the corresponding theory appears under the name of harmonic maps, and we refer to [65] for a detailed treatment with proofs and references; for the supersymmetric version, the harmonic map needs to be coupled with a Dirac field as treated in [16]. A monograph about this topic from physics is [73].

2.4.1 The Linear Sigma Model

We let M be a Riemannian manifold of dimension m, with metric tensor in local coordinates $(\gamma_{\alpha\beta})_{\alpha,\beta=1,\ldots,m}$.[15]

We recall the following notation:

$$\begin{aligned}(\gamma^{\alpha\beta})_{\alpha,\beta=1,\ldots,m} &= (\gamma_{\alpha\beta})^{-1}_{\alpha,\beta} \quad \text{(inverse metric tensor)}, \\ \gamma &= \det(\gamma_{\alpha\beta}), \\ \Gamma^{\alpha}_{\beta\eta} &= \frac{1}{2}\gamma^{\alpha\delta}(\gamma_{\beta\delta,\eta} + \gamma_{\eta\delta,\beta} - \gamma_{\beta\eta,\delta}) \quad \text{(Christoffel symbols)}.\end{aligned}$$

For a function $\phi : M \to \mathbb{R}$ of class C^1, we consider

$$\gamma^{\alpha\beta}(x)\frac{\partial\phi(x)}{\partial x^\alpha}\frac{\partial\phi(x)}{\partial x^\beta} \tag{2.4.1}$$

[15] The conventions here are different from the ones established in Sect. 1.1.1 where the metric of M was denoted by g_{ij}. The reason is that for the nonlinear sigma model, another manifold will come into play, the physical space(–time) whose metric will then be denoted by g_{ij}. M will be the world sheet instead.

in local coordinates $(x^1, \dots, x^m)$ on M. The quantity (2.4.1) is simply the square of the norm of the differential $d\phi = \frac{\partial \phi}{\partial x^\alpha} dx^\alpha$, which is a section of the cotangent bundle $T^\star M$, that is,

$$\gamma^{\alpha\beta}(x) \frac{\partial \phi(x)}{\partial x^\alpha} \frac{\partial \phi(x)}{\partial x^\beta} = \langle d\phi, d\phi \rangle_{T^\star M} = \|d\phi\|^2 \tag{2.4.2}$$

in hopefully self-explanatory notation. Therefore, it is clear that (2.4.1) is invariant under coordinate changes.

The Dirichlet integral of ϕ is then

$$S(\phi) = \frac{1}{2} \int_M \gamma^{\alpha\beta}(x) \frac{\partial \phi}{\partial x^\alpha} \frac{\partial \phi}{\partial x^\beta} \sqrt{\gamma} dx^1 \cdots dx^m = \frac{1}{2} \int_M \|d\phi\|^2 dvol_\gamma(M). \tag{2.4.3}$$

Minimizers are harmonic functions; they solve the Laplace–Beltrami equation (see (1.1.103)) (the Euler–Lagrange equation for $S(\phi)$)

$$\Delta_M \phi = \frac{1}{\sqrt{\gamma}} \frac{\partial}{\partial x^\alpha} \left(\sqrt{\gamma} \gamma^{\alpha\beta} \frac{\partial}{\partial x^\beta} \phi \right) = 0. \tag{2.4.4}$$

We now specialize this to the case where M is two-dimensional, that is, a surface equipped with some Riemannian metric. According to the conventions set up in Sect. 1.1.2, we can then let the indices α, β stand for $z, \bar{z}$, where $z = x^1 + ix^2$ is a complex coordinate; when we want to avoid indices, we shall also write $z = x + iy$, as in Sect. 1.1.2. Thus

$$\|d\phi\|^2 = \gamma^{zz} \frac{\partial \phi}{\partial z} \frac{\partial \phi}{\partial z} + 2\gamma^{z\bar{z}} \frac{\partial \phi}{\partial z} \frac{\partial \phi}{\partial \bar{z}} + \gamma^{\bar{z}\bar{z}} \frac{\partial \phi}{\partial \bar{z}} \frac{\partial \phi}{\partial \bar{z}} \tag{2.4.5}$$

and, recalling (1.1.76),

$$\begin{aligned} S(\phi) &= \frac{1}{2} \int_M \gamma^{\alpha\beta} \frac{\partial \phi}{\partial x^\alpha} \frac{\partial \phi}{\partial x^\beta} \sqrt{\gamma_{11}\gamma_{22} - \gamma_{12}^2}\, dx \wedge dy \\ &= \frac{1}{2} \int_M \left(\gamma^{zz} \frac{\partial \phi}{\partial z} \frac{\partial \phi}{\partial z} + 2\gamma^{z\bar{z}} \frac{\partial \phi}{\partial z} \frac{\partial \phi}{\partial \bar{z}} + \gamma^{\bar{z}\bar{z}} \frac{\partial \phi}{\partial \bar{z}} \frac{\partial \phi}{\partial \bar{z}} \right) \\ &\quad \times \sqrt{\gamma_{zz}\gamma_{\bar{z}\bar{z}} - \gamma_{z\bar{z}}^2}\, dz \wedge d\bar{z}. \end{aligned} \tag{2.4.6}$$

A fundamental point in the sequel will be to consider S not only as a function of the field ϕ, but also of the metric γ. We thus write

$$S(\phi, \gamma). \tag{2.4.7}$$

Naturally, we then also want to study the effect of variations $\delta\gamma$ of the metric on S; it is in fact more convenient to study variations of the inverse metric γ^{-1}. Observing that $\sqrt{\gamma_{11}\gamma_{22} - \gamma_{12}^2} = (\sqrt{\gamma^{11}\gamma^{22} - (\gamma^{12})^2})^{-1}$, we compute

$$\delta S(\phi, \gamma) = \int T_{\alpha\beta} \delta\gamma^{\alpha\beta} \sqrt{\gamma_{11}\gamma_{22} - \gamma_{12}^2}\, dx \wedge dy \tag{2.4.8}$$

with

$$T_{\alpha\beta} = \frac{\partial\phi}{\partial x^\alpha}\frac{\partial\phi}{\partial x^\beta} - \frac{1}{2}\gamma_{\alpha\beta}\gamma^{\epsilon\eta}\frac{\partial\phi}{\partial x^\epsilon}\frac{\partial\phi}{\partial x^\eta}. \tag{2.4.9}$$

We call $T_{\alpha\beta}$ the **energy–momentum tensor**[16] and observe that T is trace-free, that is,

$$T_\alpha^\alpha = 0. \tag{2.4.10}$$

Here, the dimension 2 is essential. The reason why T is trace free is the **conformal invariance** of S in dimension 2. This simply means that when we change the metric to $e^\sigma\gamma$ for some function $\sigma : M \to \mathbb{R}$, then S stays invariant:

$$S(\phi, e^\sigma\gamma) = S(\phi, \gamma). \tag{2.4.11}$$

Infinitesimally, the variation of γ^{-1} is $-\delta\sigma\ \gamma^{-1}$, and from (2.4.8), (2.4.11), we get

$$\begin{aligned} 0 &= \int T_{\alpha\beta}\ \delta\sigma\ \gamma^{\alpha\beta}\sqrt{\gamma_{11}\gamma_{22} - \gamma_{12}^2}\, dx \wedge dy \\ &= \int T_\alpha^\alpha\ \delta\sigma\sqrt{\gamma_{11}\gamma_{22} - \gamma_{12}^2}\, dx \wedge dy \end{aligned} \tag{2.4.12}$$

for all variations $\delta\sigma$, which implies (2.4.10).

We now consider this from a slightly different point of view. A Riemannian metric γ on a surface induces the structure of a Riemann surface Σ, as defined in Sect. 1.1.2, via the uniformization theorem (for a detailed treatment, we refer to [64]). As a consequence, we can find holomorphic coordinates $z = x + iy$ for which the metric is diagonal, that is,

$$\gamma_{12} = 0 \quad \text{and} \quad \gamma_{11} = \gamma_{22}, \tag{2.4.13}$$

or equivalently,

$$\gamma_{zz} = 0 = \gamma_{\bar{z}\bar{z}}. \tag{2.4.14}$$

We can then express the metric tensor by a single (nonvanishing) scalar function λ, that is, as

$$\lambda^2 dz d\bar{z}; \tag{2.4.15}$$

cf. (1.4.18). As explained in Sect. 1.4.2, a Riemann surface Σ can be considered as a conformal equivalence class of metrics of the form (2.4.15). When we choose

[16] Since we consider a Euclidean instead of a Minkowskian situation, we cannot distinguish between temporal and spatial directions, and thus also not between energy and momentum as quantities that are preserved because of temporal or spatial invariance, according to Noether's theorem. This explains the name energy–momentum tensor here. In general relativity, the energy–momentum tensor emerges because Lorentz invariance combines temporal and spatial invariance.

coordinates so that this holds, our functional S also simplifies:

$$S(\phi,\gamma)=\frac{1}{2}\int_M\left(\frac{\partial\phi}{\partial x}\frac{\partial\phi}{\partial x}+\frac{\partial\phi}{\partial y}\frac{\partial\phi}{\partial y}\right)dx\wedge dy=\int_M\frac{\partial\phi}{\partial z}\frac{\partial\phi}{\partial\bar{z}}\,i\,dz\wedge d\bar{z}. \tag{2.4.16}$$

Thus, the dependence on the metric disappears, except that the holomorphic coordinates $z=x+iy$ have been chosen so as to diagonalize the metric. In other words, S is a function of the equivalence class of metrics encoded by Σ, and we can write it as

$$S(\phi,\Sigma). \tag{2.4.17}$$

In these conformal coordinates, that is, where (2.4.13), (2.4.14) hold, the condition (2.4.10) becomes

$$T_{z\bar{z}}=0. \tag{2.4.18}$$

Since we take the field ϕ to be real-valued here, we also have $\frac{\partial\phi}{\partial\bar{z}}=\overline{\frac{\partial\phi}{\partial z}}$, and so $T_{\bar{z}\bar{z}}=\overline{T_{zz}}$. Therefore T is determined by its component T_{zz}. Taking its the transformation behavior into account as well, the energy–momentum tensor becomes a quadratic differential

$$T_{zz}dz^2=\left(\frac{\partial\phi}{\partial z}\right)^2dz^2 \tag{2.4.19}$$

from (2.4.9).

When the metric takes the form (2.4.15), the Laplace–Beltrami equation satisfied by critical points of S also simplifies:

$$\frac{4}{\lambda^2}\frac{\partial^2\phi}{\partial z\partial\bar{z}}=0, \tag{2.4.20}$$

which is equivalent to the simpler equation

$$4\frac{\partial^2\phi}{\partial z\partial\bar{z}}=0. \tag{2.4.21}$$

In this presentation, the dependence on the metric is no longer visible. This simply comes from the fact that we write the equation in local coordinates, and local coordinate neighborhoods are conformally equivalent to domains in the Euclidean complex line $\mathbb{C}$. When we return to the global aspects, as explained, we have a functional $S(\phi,\Sigma)$ that depends on the Riemann surface Σ. In Sect. 1.4.2, we have considered the moduli space of Riemann surfaces, and we can thus consider S as a functional on the moduli space M_p of Riemann surfaces of some given genus p. As explained in that section, that moduli space is not compact for $p>0$, and one should then consider an extension of S to a compactification of M_p. Such a compactification was constructed by pinching homotopically nontrivial closed curves (represented by closed geodesics w.r.t. a hyperbolic metric for $p>1$), thus creating

surfaces with singularities. Those singularities were then removed by compactifying the resulting surfaces by two points, one for each side of the closed geodesic. This now connects well with the behavior of the functional S, because its critical points, the solutions of (2.4.20), (2.4.21), are harmonic functions. And bounded harmonic functions can be smoothly extended across isolated singularities. That means that if we consider a sequence of degenerating Riemann surfaces Σ_n and controlled harmonic functions u_n (with some suitable norm bounded independently of n) on them, we can pass to the limit (of some subsequence) that then defines a harmonic function u on the Riemann surface Σ obtained by the described compactification of the limit of the Riemann surfaces. That harmonic function is then smooth on all of Σ, and in particular, it does not feel the presence or the position of the puncture, that is, of the points added for the compactification. In particular, the functional S then naturally extends not only to the Deligne–Mumford compactification, but also to the Baily–Satake compactification $\overline{\overline{M}}_p$ (see Sect. 1.4.3 of the moduli space M_p). For more details, see [62].

There is one point here that will become important below in Sect. 2.5. While the equation of motion, our Euler–Lagrange equation (2.4.20), is conformally invariant in the sense that the conformal factor $\frac{1}{\lambda^2}$ plays no role, the corresponding differential operator, the Laplace–Beltrami operator $\frac{4}{\lambda^2}\frac{\partial^2}{\partial z\partial\bar{z}}$, is not conformally invariant itself.

From (2.4.20), we see directly that the energy–momentum tensor as given by (2.4.19) is holomorphic at a solution of (2.4.20):

$$\frac{\partial T_{zz}}{\partial \bar{z}} = 0. \tag{2.4.22}$$

In conclusion, the energy–momentum tensor yields a holomorphic quadratic differential $T_{zz}dz^2 = (\frac{\partial\phi}{\partial z})^2 dz^2$ on our Riemann surface Σ.

There is a deeper reason why T is holomorphic. As we shall now explain, S is invariant under diffeomorphisms, and by Noether's theorem, this yields a conserved current, that is, a divergence-free quantity. That latter equation then turns out to be equivalent to (2.4.22). The reason is simply that (2.4.6), or equivalently (2.4.16), is invariant under coordinate changes. In mathematical terms, as explained in Sect. 1.1.1, this means that we compose the field ϕ with a diffeomorphism h of our surface and simultaneously pull the metric γ in (2.4.6) or the area form $dx \wedge dy$ in (2.4.16) back by that diffeomorphism. In other words, we have

$$S(\phi \circ h, h^\star\gamma) = S(\phi, \gamma). \tag{2.4.23}$$

In the formalism of physics, we move the points in the domain by an infinitesimal diffeomorphism, that is, a vector field, and consider the variation

$$x^\alpha + \epsilon\delta x^\alpha \quad \text{or, in complex coordinates, } z + \epsilon\delta z. \tag{2.4.24}$$

By (2.3.28), the conserved current is

$$j^z = -\left(\frac{\partial\phi}{\partial\bar{z}}\right)^2 \delta\bar{z}, \qquad j^{\bar{z}} = -\left(\frac{\partial\phi}{\partial z}\right)^2 \delta z, \tag{2.4.25}$$

and with $j_z = \gamma_{z\bar{z}} j^{\bar{z}}$ (cf. 1.1.2 and note that $\gamma_{zz} = 0$ by (2.4.14)), (2.3.31) becomes

$$0 = \frac{\partial}{\partial \bar{z}} j_z = \frac{\partial}{\partial \bar{z}} \left(\frac{\partial \phi}{\partial z} \right)^2 \delta z. \tag{2.4.26}$$

When we take holomorphic variations, $\frac{\partial}{\partial \bar{z}} \delta z = 0$, that is, respect the Riemann surface structure, this becomes (2.4.22), the holomorphicity of the energy–momentum tensor at a solution of the Euler–Lagrange equations, that is, (2.4.20).

We now wish to connect this discovery with 7 in Sect. 1.4.2. There, we had also found a holomorphic quadratic differential as a (co)tangent vector to the moduli space of Riemann surfaces. When we consider $S(\phi, \gamma)$ as a function of the metric γ, its derivative with respect to γ should be a tangent vector to the space of all metrics on our underlying surface. Here, we have been considering variations with respect to the inverse metric γ^{-1}, and thus, we obtain a cotangent instead of a tangent vector to the space of metrics. In 7 of Sect. 1.4.2, we have distinguished three types of variations of metrics, the ones through diffeomorphisms, the ones by conformal factors, and the residual ones that correspond to tangent directions of the Riemann moduli space. Now our functional $S(\phi, \gamma)$ is invariant under the first two types of variations: diffeomorphism invariance led to the holomorphicity (2.4.22), and conformal invariance made the energy–momentum tensor trace-free, (2.4.18). Therefore, it must correspond to a cotangent direction of the Riemann moduli space, and thus the agreement with the condition (1.4.20) is no coincidence.

2.4.2 The Nonlinear Sigma Model

In the nonlinear sigma model, the field ϕ takes its values in some Riemannian manifold N with metric g_{ij}, instead of in the real line $\mathbb{R}$. In the physics literature, one is usually interested in the case where N is the sphere S^n, that is, a homogeneous space for the Lie group $O(n+1)$ (one then speaks of the nonlinear $O(n+1)$ sigma model), or more generally, where N is the homogeneous space for some other compact Lie group. The case where N itself is a compact Lie group G leads to the Wess–Zumino–Witten model (see for instance [38, 73]). For the mathematical theory, however, one can consider an arbitrary Riemann manifold N, and this generality should make the structure more transparent. In fact, this will also be necessary for the applications to Morse theory presented below.

The action functional for the nonlinear sigma model is formally the same as (2.4.3),

$$S(\phi) = \frac{1}{2} \int_M \|d\phi\|^2 dvol(M), \tag{2.4.27}$$

where the norm of the differential is now given by

$$\|d\phi\|^2 = \gamma^{\alpha\beta}(x) g_{ij}(\phi(x)) \frac{\partial \phi^i(x)}{\partial x^\alpha} \frac{\partial \phi^j(x)}{\partial x^\beta}. \tag{2.4.28}$$

Expressed more abstractly, the differential of ϕ,

$$d\phi = \frac{\partial \phi^i}{\partial x^\alpha} dx^\alpha \otimes \frac{\partial}{\partial \phi^i}, \tag{2.4.29}$$

is now a section of the bundle $T^\star M \otimes \phi^{-1} TN$ where the latter bundle is the pullback of the tangent bundle of the target N by the map ϕ. Since the bundle thus depends on the field ϕ, the situation is intrinsically nonlinear. In particular, the Euler–Lagrange equations are also nonlinear:

$$\tau^i(\phi) := \frac{1}{\sqrt{\gamma}} \frac{\partial}{\partial x^\alpha} \left(\sqrt{\gamma} \gamma^{\alpha\beta} \frac{\partial}{\partial x^\beta} \phi^i \right) + \gamma^{\alpha\beta}(x) \Gamma^i_{jk}(\phi(x)) \frac{\partial}{\partial x^\alpha} \phi^j \frac{\partial}{\partial x^\beta} \phi^k = 0, \tag{2.4.30}$$

with the Christoffel symbols as in (1.1.60). (The expression $\tau(\phi)$ is called the tension field of ϕ.) Whereas this nonlinearity makes the analysis more subtle and much harder, see [65], most of the formal aspects remain unchanged when compared with the linear version of Sect. 2.4. Solutions of (2.4.30) are called harmonic maps in the mathematical literature.

We are again interested in the situation when the underlying domain M is a Riemann surface. As in the linear case, the action (2.4.27) is conformally invariant, and so we can consider it either as a function $S(\phi, \gamma)$ of the domain metric γ or as a function $S(\phi, \Sigma)$, with the Riemann surface Σ considered as the equivalence class of conformal metrics that γ belongs to. Thus, conformal invariance is preserved in the nonlinear case, and so is, obviously, diffeomorphism invariance. Therefore, we again obtain a holomorphic energy–momentum tensor as before,

$$T_{zz} dz^2 = \left\langle \frac{\partial \phi}{\partial z}, \frac{\partial \phi}{\partial z} \right\rangle_N dz^2 = g_{ij}(\phi) \frac{\partial \phi^i}{\partial z} \frac{\partial \phi^j}{\partial z} dz^2 \tag{2.4.31}$$

where we use the scalar product defined by the metric of N. When one does the computation right, it is the same as in the linear case and therefore need not be repeated here.

For example, from (2.4.8), we also see that S is critical for variations of the metric γ when the energy–momentum tensor vanishes. According to (2.4.19), this means

$$\left\langle \frac{\partial \phi}{\partial z}, \frac{\partial \phi}{\partial z} \right\rangle_N = 0 \tag{2.4.32}$$

(note that we are not taking a Hermitian product here, and so this quantity can well be 0 without $\frac{\partial \phi}{\partial z}$ being 0 itself—when that happens, we say that $\frac{\partial \phi}{\partial z}$ is isotropic), or in real coordinates x, y with $z = x + iy$, from (2.4.9)

$$\left\langle \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial x} \right\rangle_N = \left\langle \frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial y} \right\rangle_N, \qquad \left\langle \frac{\partial \phi}{\partial x}, \frac{\partial \phi}{\partial y} \right\rangle_N = 0. \tag{2.4.33}$$

When those relations hold, the map $\phi : \Sigma \to N$ is **conformal**. Since by a special case of the Riemann–Roch theorem, see Sect. 1.4.2, every holomorphic quadratic

differential on the sphere S^2 vanishes, we conclude that on S^2, the energy–momentum tensor associated with a harmonic map automatically vanishes, and therefore, any harmonic map $\phi : S^2 \to N$ into any Riemann manifold N is conformal. For Riemann surfaces of genus > 0, this is not true.

2.4.3 The Supersymmetric Sigma Model

We now extend the sigma model to include supersymmetry, proceeding as in Sect. 2.2.5. We work with the Clifford algebra $Cl(2,0)$, which admits a real representation as explained in Sect. 1.3.2. In fact, this representation is a dimensional reduction of that of $Cl(2,1)$, and so the two-dimensional formalism to be developed here is a dimensional reduction of a three-dimensional one. As explained in [23], a three-dimensional space (with Minkowski signature) is the basic setting for $N = 1$ supersymmetry, but for our purposes, conformal invariance is a crucial underlying feature of our variational problems, and therefore, we continue to focus on the two-dimensional case and consider (2|2) dimensions here (with Euclidean signature). We choose local even coordinates x^1, x^2 and odd ones θ^1, θ^2.

In order to conform to the conventions employed in [23], we use the following representation of $Cl(2,0)$:

$$e_1 \to \gamma^1 = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad e_2 \to \gamma^2 = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \tag{2.4.34}$$

which is different from the one described in Sect. 1.3.2 (but of course equivalent to it). We recall that the γ^μ satisfy

$$\{\gamma^\mu, \gamma^\nu\} = 2\delta^{\mu\nu}. \tag{2.4.35}$$

This is a real two-dimensional euclidean representation, and so we have real euclidean Majorana spinors satisfying

$$\bar\psi = \psi^\dagger = \psi^T = (\psi_1, \psi_2). \tag{2.4.36}$$

In particular, we could leave out the bars for complex conjugation (and we shall do so sometimes). Since these spinors are supposed to anticommute, we also have

$$\bar\psi\chi = -\bar\chi\psi, \qquad \bar\psi\gamma^\mu\chi = -\bar\chi\gamma^\mu\psi. \tag{2.4.37}$$

For a spinor field ψ, we then have the Dirac form

$$\begin{aligned} \bar\psi \not{D}\psi &= (\psi_1, \psi_2)\left(\gamma^1\begin{pmatrix}\frac{\partial\psi_1}{\partial x^1} \\ \frac{\partial\psi_2}{\partial x^1}\end{pmatrix} + \gamma^2\begin{pmatrix}\frac{\partial\psi_1}{\partial x^2} \\ \frac{\partial\psi_2}{\partial x^2}\end{pmatrix}\right) \\ &= -\psi_1\frac{\partial\psi_1}{\partial x^1} + \psi_2\frac{\partial\psi_2}{\partial x^1} - \psi_1\frac{\partial\psi_2}{\partial x^2} - \psi_2\frac{\partial\psi_1}{\partial x^2}. \end{aligned} \tag{2.4.38}$$

For later purposes, we observe that this can also be expressed in complex notation as

$$\bar{\psi}\not{D}\psi = -\left((\psi_1 + i\psi_2)\frac{\partial}{\partial z}(\psi_1 + i\psi_2) + (\psi_1 - i\psi_2)\frac{\partial}{\partial \bar{z}}(\psi_1 - i\psi_2)\right) \tag{2.4.39}$$

with $z = x^1 + ix^2$.

We define the vector fields

$$\begin{aligned} D_1 &:= \partial_{\theta^1} - \theta^1\partial_{x^1} - \theta^2\partial_{x^2}, \\ D_2 &:= \partial_{\theta^2} - \theta^1\partial_{x^2} + \theta^2\partial_{x^1}, \end{aligned} \tag{2.4.40}$$

$$\begin{aligned} Q_1 &:= \partial_{\theta^1} + \theta^1\partial_{x^1} + \theta^2\partial_{x^2}, \\ Q_2 &:= \partial_{\theta^2} + \theta^1\partial_{x^2} - \theta^2\partial_{x^1}. \end{aligned} \tag{2.4.41}$$

They satisfy

$$\begin{aligned} &[D_1, D_1] = -2\partial_{x^1}, \qquad [D_1, D_2] = -2\partial_{x^2}, \qquad [D_2, D_2] = 2\partial_{x^1}, \\ &[Q_1, Q_1] = 2\partial_{x^1}, \qquad [Q_1, Q_2] = 2\partial_{x^2}, \qquad [Q_2, Q_2] = -2\partial_{x^1}. \end{aligned} \tag{2.4.42}$$

We are now ready to introduce the supersymmetric sigma model. We consider a superfield Y with expansion

$$Y = \phi(x) + \psi_\alpha(x)\theta^\alpha + F(x)\theta^1\theta^2 \tag{2.4.43}$$

and the action

$$S_4 = \int \frac{1}{4}\epsilon^{\alpha\beta}\langle D_\alpha Y, D_\beta Y\rangle d^2x d\theta^2 d\theta^1 \tag{2.4.44}$$

where $d\theta$ indicates that a Berezin integral has to be taken; namely, we recall from Sect. 1.5.2 that for an expression $Z = z + z_\alpha\theta^\alpha + z_{12}\theta^1\theta^2$,

$$\int Z d^2\theta = \int Z d\theta^2 d\theta^1 = z_{12}, \tag{2.4.45}$$

that is, the θ-integration picks out the $\theta^1\theta^2$ term, see (1.5.27). Moreover, the antisymmetric ϵ-tensor is defined by

$$\epsilon^{12} = -\epsilon^{21} = 1. \tag{2.4.46}$$

S_4 is the Wess–Zumino action (in flat Euclidean space). After expanding and carrying out the Berezin integral, this becomes

$$S_4 = \frac{1}{2}\int d^2x(\partial^\mu\phi^a\partial_\mu\phi_a + \bar{\psi}^a\gamma^\mu\partial_\mu\psi_a + F^a\ F_a). \tag{2.4.47}$$

In complex notation, this looks as follows: We set

$$\psi_+ := \psi_1 - i\psi_2, \qquad \psi_- := \psi_1 + i\psi_2 \tag{2.4.48}$$

and

$$\theta_+ := \theta_1 + i\theta_2, \qquad \theta_- := \theta_1 - i\theta_2 \tag{2.4.49}$$

and define the operators

$$D_+ := \partial_{\theta_+} + \theta_+ \partial_z, \qquad D_- := \partial_{\theta_-} + \theta_- \partial_{\bar{z}}. \tag{2.4.50}$$

With this notation, (2.4.43) becomes

$$Y = \phi + \frac{1}{2}(\psi_+\theta_+ + \psi_-\theta_-) + \frac{i}{2}F\theta_+\theta_-. \tag{2.4.51}$$

We then have

$$\begin{aligned} S_4 &= \int \frac{1}{2} D_- Y D_+ Y d^2x d\theta_- d\theta_+ \\ &= \int \frac{1}{2}\left(4\partial_z\phi^a\partial_{\bar{z}}\phi_a - \psi_+\frac{\partial}{\partial\bar{z}}\psi_+ - \psi_-\frac{\partial}{\partial z}\psi_- + F^a F_a\right)d^2x. \end{aligned} \tag{2.4.52}$$

S_4 is invariant under the supersymmetry transformations

$$\delta\phi^a = \overline{\varepsilon}\psi^a, \tag{2.4.53}$$

$$\delta\psi^a = \gamma^\mu\varepsilon\partial_\mu\phi^a - \overline{\varepsilon}F^a, \tag{2.4.54}$$

$$\delta F^a = \varepsilon\gamma^\mu\partial_\mu\psi^a. \tag{2.4.55}$$

Indeed, the variation is

$$\begin{aligned} \delta S_4 = \frac{1}{2}\int dx^2 \big(&2\overline{\varepsilon}\partial^\mu\phi^a\partial_\mu\psi_a + \overline{\varepsilon}\gamma^\nu\partial_\nu\phi^a\gamma^\mu\partial_\mu\psi_a + \overline{\varepsilon}F^a\gamma^\mu\partial_\mu\psi_a \\ &+ \overline{\psi}^a\gamma^\mu\partial_\mu(\gamma^\nu\varepsilon\partial_\nu\phi_a + \varepsilon F_a) - 2\overline{\varepsilon}\gamma^\mu\partial_\mu\psi^a\ F_a + \partial_\mu(\overline{\varepsilon})\psi^a\partial^\mu\phi_a\big), \end{aligned} \tag{2.4.56}$$

which vanishes after integration by parts and using (2.4.35) and (2.4.37), when we assume that $\overline{\varepsilon}$ is constant. Locally, the latter can be assumed, and we do so for the moment, but later on, in Sect. 2.4.7, we shall return to the global issue.

The Euler–Lagrange equations for S_4 are

$$\epsilon^{\alpha\beta}D_\alpha D_\beta Y = 0, \tag{2.4.57}$$

or in components,

$$\Delta\phi = 0, \tag{2.4.58}$$

$$\gamma^\mu\partial_\mu\psi = 0, \tag{2.4.59}$$

$$F = 0. \tag{2.4.60}$$

In complex notation, these equations become

$$D_- D_+ Y = 0, \tag{2.4.61}$$

or in components

$$\partial_z \partial_{\bar{z}} \phi = 0, \tag{2.4.62}$$

$$\partial_{\bar{z}} \psi_+ = 0, \qquad \partial_z \psi_- = 0, \tag{2.4.63}$$

$$F = 0. \tag{2.4.64}$$

Again, F is a nonpropagating, auxiliary field that is only introduced to close the supersymmetry algebra off-shell. On-shell, (2.4.53) and (2.4.54) become

$$\delta \phi^a = \overline{\varepsilon} \psi^a, \tag{2.4.65}$$

$$\delta \psi^a = \gamma^\mu \varepsilon \partial_\mu \phi^a. \tag{2.4.66}$$

(Note that (2.4.59) implies that $\delta F^a = 0$ on-shell, i.e., the term that obstructs the closing of the algebra is proportional to one to the equations of motion.)

We now turn to the supersymmetric *nonlinear* sigma model. In fact, the formalism remains the same; we just need to expand its interpretation.

Thus, we consider a map

$$Y : M \to N \tag{2.4.67}$$

from a (2|2)-dimensional supermanifold to some Riemannian manifold N. We expand Y as before:

$$Y = \phi(x) + \psi_\alpha(x)\theta^\alpha + F(x)\theta^1\theta^2. \tag{2.4.68}$$

ϕ can be considered to be an ordinary map into N, whereas the odd part ψ represents an (odd) section of the pull-back tangent bundle $\phi^\star TN$.[17] $\langle .,. \rangle$ now denotes a Riemannian metric on the target space; we shall also write $\|v\|^2 := \langle v, v \rangle$ below.

Finally, F is an auxiliary field as before. This time, the algebraic equation for F among the Euler–Lagrange equations is

$$-4 g_{ij}(\phi) F^i + 2 g_{ij,k} \overline{\psi}^k \psi^i - g_{ki,j} \overline{\psi}^k \psi^i = 0,$$

i.e.,

$$F^i = \Gamma^i_{kj} \overline{\psi}^k \psi^j, \tag{2.4.69}$$

so that F can again be eliminated. In particular, when we use Riemann normal coordinates at the point under consideration, F vanishes.

[17] Thus, w.r.t. coordinate changes on the target N, ψ transforms as a vector, whereas on the domain M, it transforms as a spinor. In particular, the setting here is different from the one above in Sect. 1.5.3 for maps between super Riemann surfaces, where the odd field has to transform as a spinor on both domain and target.

In local coordinates, after carrying out the θ-integral, the Lagrangian density becomes

$$\frac{1}{2}\|d\phi\|^2 + \frac{1}{2}\langle\psi, \not{D}\psi\rangle - \frac{1}{12}\epsilon^{\alpha\beta}\epsilon^{\gamma\delta}\langle\psi_\alpha, R(\psi_\beta, \psi_\gamma)\psi_\delta\rangle. \tag{2.4.70}$$

(We get the R-term (the curvature of the target N) after elimination of $\frac{1}{2}\|F\|^2$.) This comes about as follows: According to the rule for the Berezin integral, we need to identify the $\theta^1\theta^2$-term in (2.4.44). For that purpose, we recall that a function of a superfield Y has to be expanded by Taylor's formula as explained in Sect. 1.5.2, see (1.5.20), (1.5.21). In particular, (2.4.44) contains the metric tensor $\langle ., . \rangle$ of the target N. In local coordinates, we have a tensor $g_{ij}(Y)$ whose expansion contains second derivatives $g_{ij,kl}(\phi)$ multiplied with $\theta^1\psi_1\theta^2\psi_2$, which gives the curvature term in (2.4.70). Terms with first derivatives of g_{ij} do not carry an invariant meaning and become 0 in suitable coordinates (Riemann normal coordinates) at the point in N under consideration. In particular, the curvature tensor R has to be evaluated at the point $\phi(x) \in N$. Similarly, the Dirac operator $\not{D}$ contains a covariant derivative at the tangent space $T_{\phi(x)}N$.

We now list the important results for the nonlinear supersymmetric sigma model (for detailed computations see [17]): The Euler–Lagrange equations for the nonlinear supersymmetric sigma model are

$$\tau^m(\phi) - \frac{1}{2}R^m{}_{lij}\overline{\psi}^i(\nabla\phi^l \cdot \psi^j) + \frac{1}{12}g^{mp}R_{ikjl;p}(\overline{\psi}^i\psi^j)(\overline{\psi}^k\psi^l) = 0, \tag{2.4.71}$$

$$\not{D}\psi^m - \frac{1}{3}R^m_{jkl}\psi^k\overline{\psi}^j\psi^l = 0. \tag{2.4.72}$$

The first equation generalizes (2.4.30).

The functional S is invariant under the supersymmetry transformation

$$\begin{cases} \delta\phi^i = \varepsilon\psi^i, \\ \delta\psi^i = \gamma^\alpha\partial_\alpha\phi^i\varepsilon - \Gamma^i_{jk}(\varepsilon\psi^j)\psi^k. \end{cases} \tag{2.4.73}$$

As in (2.4.55), we recognize the F-term, see (2.4.69).

The supercurrent

$$J^\alpha := 2g_{ij}\partial_\beta\phi^i\gamma^\beta\gamma^\alpha\psi^j, \quad \alpha = 1, 2 \tag{2.4.74}$$

is conserved (on-shell), i.e.,

$$D_\alpha J^\alpha \equiv 0. \tag{2.4.75}$$

Again, we wish to consider an interaction Lagrangian with a superpotential W

$$S_{int} = \int W(Y(x, \vartheta))d^2x d^2\theta. \tag{2.4.76}$$

A simple and standard choice is

$$W(Y) = -\frac{1}{3}kY^3 - \lambda Y, \tag{2.4.77}$$

with parameters k, λ. The coefficient of $\theta^1\theta^2$ in the expansion of W is

$$-\lambda F - kF\phi^2 - k\overline{\psi}\psi\phi.$$

We first consider the linear sigma model, that is, we set the curvature tensor $R = 0$. In the Lagrangian $S_4 + S_{int}$, we then have the F terms

$$\frac{1}{2}F^2 + \lambda F + kF\phi^2,$$

leading to the algebraic Euler–Lagrange equation

$$F = -\lambda - k\phi^2.$$

Utilizing this equation, the Lagrangian becomes

$$S_4 + S_{int} = \int d^2x \left(\frac{1}{2}\partial_\mu\phi\partial^\mu\phi + \frac{1}{2}\overline{\psi}\gamma^\mu\partial_\mu\psi - \frac{1}{2}(\lambda + k\phi^2)^2 - k\overline{\psi}\psi\phi \right).$$

A more general interaction term is of the form

$$S_{int} = \frac{1}{2}\int h(Y)d^2xd^2\theta = \int \left(-\frac{1}{2}g^{ij}(\phi)\frac{\partial h}{\partial\phi^i}\frac{\partial h}{\partial\phi^j} - \frac{1}{2}\frac{\partial^2 h}{\partial\phi^i\partial\phi^j}\overline{\psi}^i\psi^j \right)d^2x \tag{2.4.78}$$

where the first term in the integrand arises from eliminating the auxiliary field F in the combined Lagrangian, in the same manner as before.

2.4.4 Boundary Conditions

We start again with the bosonic field ϕ that takes its values in some d-dimensional Riemannian manifold N. We now assume that ϕ is defined on some Riemann surface with boundary. The surface will again be denoted by Σ, and we assume for the moment that its boundary is a smooth curve γ, or a collection of such curves. Boundary conditions for ϕ on $\gamma = \partial\Sigma$ are given by specifying a smooth submanifold B of dimension p of N. In the physics literature [86, 87], this would be called a **brane** or a **D-brane**, with D standing for Dirichlet boundary conditions.[18] Locally, we can choose coordinates on N so that B is given by $x^{p+1} = \cdots = x^d = 0$.

[18] In fact, the physics convention is to consider $d-1$ spatial and one temporal dimensions. A p-brane would then result from fixing p of the $d-1$ spatial dimensions.

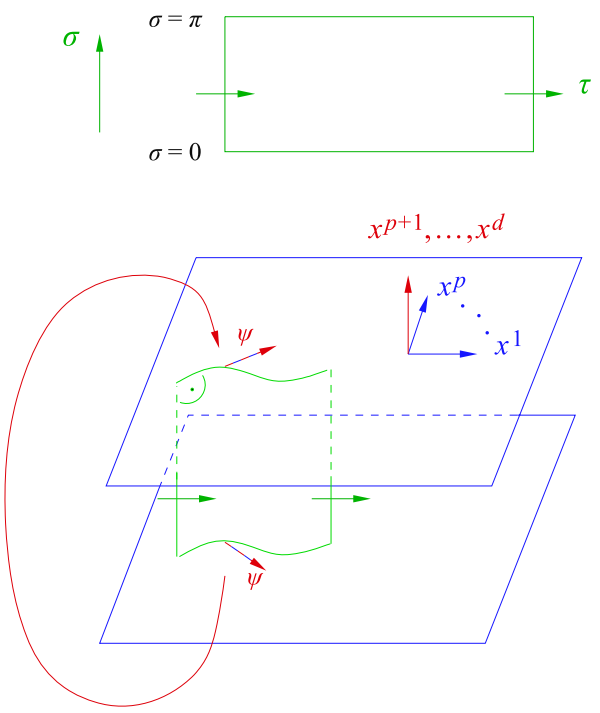


Fig. 2.1 The boundary conditions for bosonic and fermionic fields defined by a D-brane

The boundary conditions are illustrated in Fig. 2.1 and will now be described in formulae.

The first part of the boundary conditions requires that the boundary curve γ be mapped to B. Locally, this therefore means

$$\phi^{p+1} = \cdots = \phi^d = 0 \quad \text{on } \gamma. \tag{2.4.79}$$

This is, of course, a Dirichlet boundary condition for the components $\phi^{p+1}, \ldots, \phi^d$. More generally, the tangential derivative $\frac{\partial}{\partial \tau}$ of these components has to vanish on γ,

$$\frac{\partial \phi^{p+1}}{\partial \tau} = \cdots = \frac{\partial \phi^d}{\partial \tau} = 0 \quad \text{on } \gamma. \tag{2.4.80}$$

The remaining components then should satisfy a Neumann boundary condition, that is, with $\frac{\partial}{\partial \nu}$ denoting the normal derivative on γ,

$$\frac{\partial \phi^0}{\partial \nu} = \cdots = \frac{\partial \phi^p}{\partial \nu} = 0 \quad \text{on } \gamma. \tag{2.4.81}$$

In order to state the boundary condition for the fermionic field ψ, we use the notation of (2.4.48). We first assume that at the point under consideration, the metric

of N is given in normal coordinates, that is, $g_{ij} = \delta_{ij}$. Then the general boundary condition for ψ is in these coordinates:

$$\psi_-^j = \pm\psi_+^j. \tag{2.4.82}$$

A choice of sign in (2.4.82) can be motivated by the following consideration. As our domain, we consider a strip

$$\{\sigma \in [0, 2\pi]\} \times \{\tau \in [0, T]\}, \tag{2.4.83}$$

and we wish to fix boundary conditions for $\sigma = 0, 2\pi$ (since the τ-direction is interpreted as a temporal direction, there might be initial conditions prescribed at $\tau = 0$ and final ones at $\tau = T$, but this is not our concern here). We assume that we have a mapping ϕ defined on this strip, and a vector ψ along ϕ. We look at the simplest situation, where N is Euclidean space $\mathbb{R}^d$, and the brane receiving $\sigma = 0$ is the hyperplane $x^{d+1} = \ldots x^d = 0$. For $\sigma = 2\pi$, we prescribe another brane that is parallel to the first one, say $x^{d+1} = \ldots x^{d-1} = 0, x^d = R$. When we then periodically identify these two branes, that is, dividing $\mathbb{R}^d$ by the translations by R in the x^d-direction, in order to get the fields ψ^j to match on the boundary, we need to require

$$\psi^k(2\pi, \tau) = \psi^k(0, \tau) \quad \text{for } k = 1, \ldots, p, \tag{2.4.84}$$

and

$$\psi^\ell(2\pi, \tau) = -\psi^\ell(0, \tau) \quad \text{for } \ell = p+1, \ldots, d. \tag{2.4.85}$$

When we then put $\psi_+ = \psi$ and define $\psi_-(\sigma, \tau) = \psi_+(2\pi - \sigma, \tau)$ (in which case the field ψ on the range $\sigma \in [0, 2\pi]$ is obtained from the two fields $\psi_\pm$ on half the range, $\sigma \in [0, \pi]$), we then obtain from (2.4.84)

$$\psi_-^j = \psi_+^j \quad \text{for } j = 1, \ldots, p, \quad \text{and} \quad \psi_-^k = -\psi_+^k \quad \text{for } k = p+1, \ldots, d. \tag{2.4.86}$$

Thus, the plus sign in (2.4.82) corresponds to Neumann boundary conditions for the corresponding components of ϕ, and the minus sign to Dirichlet conditions. See also the discussion in Sect. 2.6.3, around (2.6.47).

In general coordinates, the boundary conditions for the ψ-field can be written as

$$\psi_-^j = D_i^j \psi_+^i, \tag{2.4.87}$$

with the tensor D_i^j satisfying

$$D_i^j D_k^i = \delta_k^j \tag{2.4.88}$$

and

$$g_{ij} = D_i^k D_j^\ell g_{k\ell}. \tag{2.4.89}$$

Thus

$$D_{ij} = g_{ik} D_j^k \tag{2.4.90}$$

is symmetric. Again, there is a sign to be determined, according to Neumann or Dirichlet type boundary conditions for ϕ.

The preceding boundary conditions arise as follows. When we consider variations $\delta\phi, \delta\psi$ of the fields in S_4 (neglecting the F-field as this vanishes on-shell anyway), we get a corresponding boundary term in the induced variation δS_4. Employing the version (2.4.52), this boundary term is given by

$$\frac{1}{2}\int g_{jk}\left(\delta\phi^j\frac{\partial\phi^k}{\partial\nu}+i(\delta\psi_+^j\psi_+^k-\delta\psi_-^j\psi_-^k)+i\delta\phi^j\Gamma^k_{\ell m}(\psi_-^\ell\psi_-^m-\psi_+^\ell\psi_+^m)\right)d\xi \tag{2.4.91}$$

where ν as before is the outer normal direction at γ and ξ is a coordinate on γ. These boundary terms then vanish if

1. either

$$\delta\phi^j=0 \quad \text{(Dirichlet)} \tag{2.4.92}$$

or

$$\frac{\partial\phi^k}{\partial\nu}=0 \quad \text{(Neumann)} \tag{2.4.93}$$

2. and

$$\psi_-^j=D_i^j\psi_+^i, \tag{2.4.94}$$

that is, (2.4.87) holds, with the conditions (2.4.88) and (2.4.89).

In order to keep the theory supersymmetric, the brane B should be totally geodesic, that is, every shortest geodesic in N connecting two points in B should already be contained in B, see [1].

2.4.5 Supersymmetry Breaking

The Hilbert space of a quantum field theory can be decomposed as

$$\mathcal{H}=\mathcal{H}^+\oplus\mathcal{H}^-$$

with $\mathcal{H}^+(\mathcal{H}^-)$ being the space of "bosonic" ("fermionic") states. The theory is supersymmetric if there are (Hermitian) supersymmetry operators

$$Q_i:\mathcal{H}\to\mathcal{H},\quad i=1,\dots,N$$

with

$$Q_i(\mathcal{H}^\pm)=\mathcal{H}^\mp. \tag{2.4.95}$$

Witten [104] introduced the operator $(-1)^F$ satisfying

$$(-1)^F\chi=\pm\chi \quad \text{for } \chi\in\mathcal{H}^\pm. \tag{2.4.96}$$

The supersymmetry operators Q_i then anticommute with $(-1)^F$

$$(-1)^F Q_i + Q_i (-1)^F = 0 \quad (i = 1, \ldots, N). \tag{2.4.97}$$

The Q_i must commute with the Hamiltonian H that generates the time translations, i.e.,

$$Q_i H - H Q_i = 0 \quad (i = 1, \ldots, N). \tag{2.4.98}$$

The Q_i are then determined if we require additionally

$$Q_i^2 = H, \qquad Q_i Q_j + Q_j Q_i = 0 \quad \text{for } i \neq j. \tag{2.4.99}$$

As a square of Hermitian operators, H is positive semidefinite.

Let $|b\rangle \in \mathcal{H}^+$ satisfy

$$H|b\rangle = E|b\rangle,$$

i.e., $|b\rangle$ is an eigenvector H with eigenvalue E (≥ 0 as H is positive semidefinite). We consider one of the Q_i, which we simply write as Q for the moment.

We can write

$$Q|b\rangle = \sqrt{E}|f\rangle$$

and get

$$Q|f\rangle = \frac{1}{\sqrt{E}} Q^2 |b\rangle = \frac{1}{\sqrt{E}} H|b\rangle = \sqrt{E}|b\rangle.$$

Thus, if $E \neq 0$, the bosonic and fermionic eigenstates with eigenvalue E are paired in an irreducible multiplet of the supersymmetry algebra. This need not be so any longer if $E = 0$. Since $H = Q^2$, if

$$H|b_0\rangle = 0 \quad \text{for } |b_0\rangle \in \mathcal{H}^+$$

we have

$$0 = \langle b_0|H|b_0\rangle = \|Q|b_0\rangle\|^2 \quad (\text{since } Q \text{ is hermitian}),$$

hence

$$Q|b_0\rangle = 0,$$

and similarly $H|f_0\rangle = 0$ for $|f_0\rangle \in \mathcal{H}^-$ implies

$$Q|f_0\rangle = 0.$$

Thus, the zero eigenvectors of H are supersymmetric, that is, invariant under the supersymmetry operator.

Consequently, for positive energy E, the number of bosonic eigenvectors equals the number of fermionic ones, but this need not be so for zero energy.

Let

$$\nu(0) := \#\text{bosonic} \; - \; \#\text{ fermionic zero eigenvectors.}$$

Regularizing the trace of $(-1)^F$, we then have

$$\mathrm{Tr}(-1)^F = \nu(0). \tag{2.4.100}$$

If $\nu(0) \neq 0$, there must exist at least one—bosonic or fermionic—state with zero energy. Since 0 is the smallest possible value of the energy, such a state furnishes a vacuum that is supersymmetric. If there does not exist a supersymmetric vacuum, i.e., if the smallest eigenvalue of H is positive, one says that supersymmetry is spontaneously broken.

We first consider the supersymmetric point particle with two odd variables with the total Lagrangian (see (2.2.86), (2.2.99), (2.2.101))

$$L_3 + L_{int} = \int dt \left(\frac{1}{2}\dot{\phi}\dot{\phi} + \frac{1}{2}\psi_\alpha \dot{\psi}_\alpha - \frac{1}{2}\left(\frac{dw(\phi)}{d\phi}\right)^2 - \frac{d^2 w}{d\phi^2}\psi_1\psi_2 \right). \tag{2.4.101}$$

The Hamiltonian is

$$H = \frac{1}{2}p^2 + \frac{1}{2}\left(\frac{dw}{d\phi}\right)^2 + \frac{d^2 w}{d\phi^2}\psi_1\psi_2. \tag{2.4.102}$$

As before, see (2.4.77), we choose

$$w(\phi) = -\frac{1}{3}k\phi^3 - \lambda\phi. \tag{2.4.103}$$

We obtain

$$H = \frac{1}{2}p^2 + \frac{1}{2}(k\phi^2 + \lambda)^2 + 2k\phi\psi_1\psi_2. \tag{2.4.104}$$

ψ_1 and ψ_2 are Grassmann valued and odd, and so

$$[\psi_\alpha, \psi_\beta] = \psi_\alpha\psi_\beta + \psi_\beta\psi_\alpha = 0 \quad \text{for } \alpha, \beta = 1, 2. \tag{2.4.105}$$

After quantization, we get, in place of (2.4.105),

$$[\psi_\alpha, \psi_\beta] = \hbar\delta_{\alpha\beta}, \tag{2.4.106}$$

that is, the Grassmann variables become Clifford algebra valued.

We may thus represent the ψ_α by Pauli matrices

$$\psi_\alpha = \sqrt{\frac{1}{2}\hbar}\,\sigma_\alpha \tag{2.4.107}$$

and get

$$H = \frac{1}{2}p^2 + \frac{1}{2}(k\phi^2 + \lambda)^2 + \hbar k\sigma_3\phi. \tag{2.4.108}$$

At the so-called tree level (that is, keeping only the zeroth-order terms (those that are not proportional to $\hbar^n, n > 0$)—the higher order contain corrections of the tree

level), the ground state energy is determined by the potential $(k\phi^2+\lambda)^2$. Hence, supersymmetry is spontaneously broken if $\frac{\lambda}{k}>0$.

The supersymmetry generators here are

$$Q_1=\frac{1}{\sqrt{2}}(\sigma_1 p+\sigma_2(k\phi^2+\lambda)),\qquad Q_2=\frac{1}{\sqrt{2}}(\sigma_2 p-\sigma_1(k\phi^2+\lambda)). \quad (2.4.109)$$

For the sequel, it will be convenient to switch to the operators

$$Q_\pm=\frac{1}{\sqrt{2}}(Q_1\pm i\,Q_2). \quad (2.4.110)$$

Then

$$Q_\pm^2=0,\qquad [Q_+,Q_-]=2H. \quad (2.4.111)$$

In a cohomological interpretation, we call a state $|s\rangle\in\mathcal{H}$ with

$$Q_+|s\rangle=0$$

closed, one that can be written as

$$|s\rangle=Q_+|t\rangle\quad \text{for some } |t\rangle\in\mathcal{H}$$

exact. Since $Q_+^2=0$, exact states are closed. Conversely, if $|s_E\rangle$ is a closed eigenvector of H with eigenvalue $E\neq 0$, i.e.,

$$H|s_E\rangle=E|s_E\rangle,$$

then $|t_E\rangle:=\frac{1}{E}Q_-|s_E\rangle$ satisfies

$$\begin{aligned}Q_+|t_E\rangle&=\frac{1}{E}Q_+Q_-|s_E\rangle=\frac{1}{E}[Q_+,Q_-]|s_E\rangle\quad (Q_-Q_+|s_E\rangle=0 \text{ as } s_E \text{ is closed})\\&=\frac{1}{E}H|s_E\rangle=|s_E\rangle,\end{aligned}$$

and hence $|s_E\rangle$ is exact.

If $E=0$ and if we had again $|s_0\rangle=Q_+|t_0\rangle$, then also

$$H|t_0\rangle=0\quad \text{as } [Q_+,H]=0.$$

However, by (2.4.99), this implies $Q_1|t_0\rangle=Q_2|t_0\rangle=0$ as above, hence also

$$|s_0\rangle=Q_+|t_0\rangle=0.$$

Thus, the nonvanishing eigenstates for $E=0$ are precisely the non-exact closed states.

2.4.6 The Supersymmetric Nonlinear Sigma Model and Morse Theory

We return to the supersymmetric nonlinear sigma model. We let N be a compact Riemannian manifold. We assume that the so-called world sheet, the two-dimensional domain on which the fields are defined, is of the form

$$\{(t,x) : t \in \mathbb{R}, x \in S^1\},$$

i.e., a cylinder, whose circumference we assume to have length L. We also assume that the fields ϕ^i and ψ^i are independent of x. We then get

$$S_5 = \frac{1}{2}L\int dt\left(g_{ij}(\phi)\frac{\partial\phi^i}{\partial t}\frac{\partial\phi^j}{\partial t} + g_{ij}(\phi)\overline{\psi}^i\gamma^0\partial_t\psi^j + \frac{1}{6}R_{ijkl}\overline{\psi}^i\psi^k\overline{\psi}^j\psi^l\right). \tag{2.4.112}$$

After quantization, the spinors ψ^i and their Hermitian conjugates become Clifford algebra valued, i.e.,

$$[\psi^i,\psi^j] = 0 = [\psi^{i^*},\psi^{j^*}], \qquad [\psi^i,\psi^{j^*}] = g_{ij}(\phi).$$

Also, after quantization, supersymmetry is generated by the charges

$$Q_+ = i\psi^{i^*}p_i = \psi^{i^*}D_{\phi^i}, \qquad Q_- = -i\psi^i p_i = -\psi^i D_{\phi^i},$$

with D_{ϕ^i} being a covariant derivative, the momentum conjugate to ϕ^i.

We now recall from Sect. 1.3.2 that we have a representation of the Clifford algebra on the space of spinors given by

$$\psi^{j^*} \sim \varepsilon(dx^j) \quad (\varepsilon(dx^j) \text{ operates as the exterior product with the differential form } dx^j),$$

$$\psi^i \sim i(dx^i) \quad (i(dx^i) \text{ operates as interior contraction with } dx^i).$$

(This representation is obtained from the one in Sect. 1.3.2 by setting the imaginary parts of the differential forms to 0.)

Thus, ψ^{j^*} corresponds to a differential form, ψ^i to a vector field on N (here $x^1, x^2, \ldots$, are local coordinates on N; one should write $\phi^1, \phi^2, \ldots$ in place of $x^1, x^2, \ldots$, but expressions like $d\phi^i$ look a bit awkward).

Moreover, Q_+ then corresponds to the exterior derivative d, Q_- to its adjoint d^*, and the Hamiltonian is

$$H = Q_+Q_- + Q_-Q_+ = dd^* + d^*d, \tag{2.4.113}$$

the Hodge Laplacian (1.1.109). With $Q_1 := d + d^*$, $Q_2 := i(d - d^*)$, we also have

$$H = Q_1^2 = Q_2^2. \tag{2.4.114}$$

On the other hand, we had interpreted ψ^{j^*} as a fermionic creation operator, ψ^i as a fermionic annihilation operator. The states that are annihilated by all ψ^i, i.e., the states with no fermions, are then identified with the functions $f(x)$ on N. Operating on such a state with a ψ^{i^*}, we obtain a state with one fermion, or in the de Rham picture (see the discussion at the end of Sect. 1.1.3), a one-form on N. States with two fermions must be antisymmetric in the fermionic indices, because of the fermion statistics, and can be considered as two-forms.

Thus, we obtain the de Rham complex, with the Hodge Laplacian. The dimension of the space of zero states of this Laplacian, i.e., of harmonic q-forms, is the Betti number b_q.

Equating the two pictures gives Witten's result [104]

$$\mathrm{Tr}(-1)^F = \sum_q (-1)^q b_q(N).$$

We now add our self-interaction term L_{int} with Morse function sh (s here is a parameter) to the Lagrangian S_5. (A smooth (twice continuously differentiable) function h is called a Morse function if at all its critical points the Hessian, that is, the matrix of its second derivatives, is nondegenerate, that is, does not have 0 as an eigenvalue.) This changes d, d^* to

$$d_s = e^{-hs} d e^{hs}, \qquad d_s^* = e^{hs} d^* e^{-hs}. \tag{2.4.115}$$

We have $d_s^2 = 0 = d_s^{*2}$, and we get

$$Q_{1,s} = d_s + d_s^*, \qquad Q_{2,s} = i(d_s - d_s^*). \tag{2.4.116}$$

Moreover,

$$\begin{aligned} H_s &= Q_{1,s}^2 = Q_{2,s}^2 \\ &= d_s d_s^* + d_s^* d_s \\ &= dd^* + d^* d + s^2 g^{ij} \frac{\partial h}{\partial x^i} \frac{\partial h}{\partial x^j} + s \frac{\partial^2 h}{\partial x^i \partial x^j} [\varepsilon(dx^i), i(dx^j)]. \end{aligned} \tag{2.4.117}$$

$s^2 g^{ij} \frac{\partial h}{\partial x^i} \frac{\partial h}{\partial x^j}$ is the potential energy, and it becomes very large for large s, except in the vicinity of the critical points of h. Therefore, the eigenfunctions of H_s concentrate near the critical points of h for large s, and asymptotic expansions in powers of $\frac{1}{s}$ for the eigenvalues depend only on local data near the critical points. This is the starting point of Witten's approach to Morse theory [105], which we shall now discuss.

As mentioned, we assume that h is a Morse function. We let $q_1, q_2, \ldots, q_m$ be the critical points of h. By the Morse lemma (see e.g. [65], p. 311), each critical point q_ν has a neighborhood U_ν with the property that in suitable local coordinates $x = x_\nu = (x_\nu^1, \ldots, x_\nu^n)$ with $x_\nu(q_\nu) = 0$,

$$h(p) - h(q_\nu) = \frac{1}{2} \sum_{k=1}^n \mu_{\nu,k} x_\nu^k(p)^2 \tag{2.4.118}$$

with

$$D^2h(q_\nu) = \mathrm{diag}(\mu_{\nu,1}, \ldots, \mu_{\nu,n}) \tag{2.4.119}$$

(i.e., the Hessian of h at q_ν is diagonalized, and the diagonal elements $\mu_{\nu,1}, \ldots, \mu_{\nu,n}$ are nonzero as h is assumed to be a Morse function).

Also, on U_ν we choose a flat Riemannian metric g_ν for which the $\frac{\partial}{\partial x_\nu^j}$, $j = 1, \ldots, n$, are orthonormal.

Of course, we may assume that the U_ν, $\nu = 1, \ldots, m$, are pairwise disjoint, and moreover that their closures are contained in pairwise disjoint open sets V_ν.

$$V_0 := N \setminus \bigcup_{\nu=1}^{m} \overline{U_\nu}, \quad V_1, \ldots, V_m$$

is then an open covering of N, and we may find a subordinate partition of unity $\{\eta_\nu\}_{\nu=0}^m$, that is, functions $\eta_\nu : N \to \mathbb{R}$ satisfying

$$0 \le \eta_\nu \le 1, \qquad \sum_{\nu=0}^{m} \eta_\nu = 1, \qquad \mathrm{supp}\, \eta_\nu \subset V_\nu,$$

with $\eta_\nu = 1$ on U_ν.

We choose any metric g_0 on V_0 and put

$$g := \sum_{\nu=0}^{m} \eta_\nu g_\nu.$$

g is then a Riemannian metric on N. Since neither the Betti numbers of N nor the critical points of h or their Morse indices depend on the choice of a Riemannian metric on N, we may work with the metric g in the sequel. In this metric, we have on $U_\nu (\nu = 1, \ldots, m)$

$$H_s = \sum_{j=1}^{n} \left(-\left(\frac{\partial}{\partial x^j} \right)^2 + s^2 \mu_{\nu,j}^2 x^{j^2} + \frac{1}{s} \mu_{\nu,j} [\varepsilon(dx^j), i(dx^j)] \right). \tag{2.4.120}$$

In particular, H_s is an operator with separated variables on U_ν. In fact, we have

$$H_s = \sum_{j=1}^{n} \left(\Omega_s^{j,\nu} + \frac{1}{s} \mu_{\nu,j} K^j \right) \tag{2.4.121}$$

with

$$\Omega_s^{j,\nu} := -\left(\frac{\partial}{\partial x^j} \right)^2 + s^2 \mu_{\nu,j}^2 x^{j^2}, \tag{2.4.122}$$

$$K^j := [\varepsilon(dx^j), i(dx^j)]. \tag{2.4.123}$$

The operators $\Omega_s = \Omega_s^{j,\nu}$ are just Hamiltonians for harmonic oscillators, and they have eigenvalues

$$s|\mu_{\nu,j}|(1+2N) \quad (N=0,1,2,\ldots) \tag{2.4.124}$$

with eigenfunctions

$$\phi_N(x) = P_N\big(\sqrt{s|\mu_{\nu,j}|}x\big)e^{-\frac{s}{2}|\mu_{\nu,j}|x^2}, \tag{2.4.125}$$

where the P_N are Hermite polynomials. In particular, for large s, the ϕ_N rapidly decay away from $x=0$, i.e., away from the critical point q_ν of h. Moreover, we have

$$K^j dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r} = \varepsilon_j^{\boldsymbol{\alpha}}(dx^{\alpha_1} \wedge \cdots \wedge dx^{\alpha_r}), \tag{2.4.126}$$

with

$$\varepsilon_j^{\boldsymbol{\alpha}} = \begin{cases} 1, & \text{if } j \in \boldsymbol{\alpha} = (\alpha_1 \ldots \alpha_r), \\ -1, & \text{otherwise} \end{cases} \tag{2.4.127}$$

i.e., K^j has eigenvalues ± 1. H_s thus is a self-adjoint operator with eigenvalues

$$s\sum_{j=1}^{n}((1+2N_{\nu,j})|\mu_{\nu,j}| + \varepsilon_{\nu,j}\mu_{\nu,j}), \tag{2.4.128}$$

$\varepsilon_{\nu,j} = \pm 1$, $N_{\nu,j} = 0,1,2,\ldots$, and orthonormal eigenvectors

$$\begin{aligned}\phi^s_{N_\nu,\alpha_\nu} &= s^{\frac{n}{4}}\prod_{j=1}^{n} P_{N_{\nu,j}}\big(\sqrt{s|\mu_{\nu,j}|}x^j\big) \\ &\quad \times \left[\exp\left(-\frac{s}{2}\sum_{j=1}^{n}|\mu_{\nu,j}|x^{j^2}\right)dx^{\alpha_{\nu,1}} \wedge \cdots \wedge dx^{\alpha_{\nu,r}}\right] \\ &\quad (\text{with } \boldsymbol{\alpha}_\nu = (\alpha_{\nu,1},\ldots,\alpha_{\nu,r})).\end{aligned} \tag{2.4.129}$$

In order for an eigenvalue to vanish, we necessarily have $N_{\nu,j} = 0$ for all j, and moreover $\varepsilon_{\nu,j}$ and $\mu_{\nu,j}$ have opposite signs. Thus, if p_ν has Morse index p, i.e., precisely p of the $\mu_{\nu,j}$ are negative, then p out of the $\varepsilon_{\nu,j}$ must be positive, which means that the corresponding eigenvector is a p-form, as can be seen from (2.4.126) and (2.4.127). Thus, if a_ν is a critical point of Morse index p, it has a one-dimensional contribution to the nullspace of H_s operating on p-forms, while for different Morse index, there is no contribution.

Now this has been a local consideration, and a nulleigenvector on U_ν need not extend to a nulleigenvector on all of N. However, a perturbation argument (see, e.g., [57] or the monograph [15] for details) shows that the other eigenvalues of H_s on $\Lambda^p(N)$ diverge as s tends to ∞, while the global nulleigenvectors concentrate at the critical points and therefore lead to local nulleigenvectors as considered above. We conclude the basic theorem of Morse:

Theorem 2.1

$$m_p \geq b_p, \tag{2.4.130}$$

where m_p is the number of critical points of h of Morse index p and b_p is the pth Betti number of N.

(b_p is the dimension of the kernel of the Hodge Laplacian $dd^* + d^*d$ on $\Omega^p(N)$, and one easily sees that the dimension of the kernel of the perturbed Laplacian $H_s = d_s d_s^* + d_s^* d_s$ is the same for all s.)

Of course, this is an asymptotic argument, for $s \to \infty$, and we only get expansions of the eigenvectors, in contrast to the original case $s = 0$ where we could identify them with harmonic forms. However, here already the classical, i.e., not quantized theory, is not entirely trivial; namely while for $s = 0$, minima of the bosonic part of the action S_5 were simply constants, for $s > 0$ the situation becomes more interesting. In a sense, the Morse function breaks the symmetry that all points of N are equal.

We consider the bosonic part of our action $S_5 + S_{int}$, again on a cylindrical world sheet $\{(t, x) : t \in \mathbb{R}, x \in S^1\}$, and assuming that the fields are independent of x so that we can carry out the x-integration. We then have the total energy or Hamiltonian, see (2.1.7), (2.1.9),

$$H_B(\phi) = \frac{1}{2} L \int dt \left(g_{ij}(\phi) \frac{d\phi^i}{dt} \frac{d\phi^j}{dt} + s^2 g^{ij}(\phi) \frac{\partial h}{\partial \phi^i} \frac{\partial h}{\partial \phi^j} \right). \tag{2.4.131}$$

Obviously, $H_B(\phi) = 0$ if $\phi(t) \equiv q_\nu$, where q_ν is a critical point of h.

These are the classical solutions. We next consider tunneling paths or so-called instanton solutions between such classical solutions.

Given two critical points q_ν, q_μ, we have to find

$$\phi : \mathbb{R} \to N; \qquad \lim_{t \to -\infty} \phi(t) = q_\nu, \qquad \lim_{t \to \infty} \phi(t) = q_\mu \tag{2.4.132}$$

minimizing

$$\begin{aligned} H_B(\phi) &= \frac{1}{2} L \int dt \left(g_{ij}(\phi) \frac{d\phi^i}{dt} \frac{d\phi^j}{dt} + s^2 g^{ij}(\phi) \frac{\partial h}{\partial \phi^i} \frac{\partial h}{\partial \phi^j} \right) \\ &= \frac{1}{2} L \int dt \left\| \frac{d\phi^i}{dt} \pm s g^{ij} \frac{\partial h}{\partial \phi^j} \right\|^2 \mp s L \int dt \frac{d(h \circ \phi)}{dt} \\ &\geq \pm s L \left(\lim_{t \to -\infty} h(\phi(t)) - \lim_{t \to \infty} h(\phi(t)) \right) \\ &= \pm s L (h(q_\nu) - h(q_\mu)), \end{aligned} \tag{2.4.133}$$

(using the simple relation $\| \frac{d\phi^i}{dt} \pm s g^{ij} \frac{\partial h}{\partial \phi^j} \|^2 = g_{ik}(\frac{d\phi^j}{dt} \pm s g^{ij} \frac{\partial h}{\partial \phi^i})(\frac{d\phi^k}{dt} \pm s g^{kl} \frac{\partial h}{\partial \phi^l})$). Equality occurs precisely if

$$\frac{d\phi^i}{dt} = \mp s g^{ij} \frac{\partial h}{\partial \phi^j}, \tag{2.4.134}$$

i.e.,

$$\frac{d\phi}{dt} = \mp s(\nabla h) \circ \phi$$

(2.4.134) means that, up to sign, $\phi(t)$ is a curve of steepest descent for h.

Thus, the minimum action paths between any two critical points are paths of steepest descent, and the action of such a path is

$$sL|h(q_\nu) - h(q_\mu)|.$$

We now let q be a critical point of h of Morse index p, and we let $r_1, \ldots, r_m$ be the critical points of Morse index $p+1$. We put

$$\delta|q\rangle = \sum_{\mu=1}^{m} n(q, r_\mu)|r_\mu\rangle.$$

Here, we associate to each critical point q of index p a basis vector $|q\rangle$ of a vector space V_p. We put

$$n(q, r_\mu) = \sum_{\Gamma(r_\mu, q)} n_\Gamma$$

where $\Gamma(r_\mu, q)$ is the path of steepest descent from r_μ to q, and where n_Γ is ± 1 according to the following rule.

By the above considerations, each critical point q of index p corresponds to a p-form localized near that point, and this p-form yields an orientation of the subspace of $T_q N$ spanned by the p negative eigendirections of the Hessian of h at q. At r_μ, we thus have a $(p+1)$-form, and in fact the direction of steepest descent corresponds to the eigendirection for the smallest eigenvalue of $\nabla^2 h(r_\mu)$. If we thus transport this $(p+1)$-form parallely along Γ and contract it with the tangent direction of Γ, we obtain a p-form at q. Comparing the resulting orientation at q with the one coming from the p-form corresponding to q then determines whether n_Γ is $+1$ or -1, i.e., $n_\Gamma = 1$ if they agree, $n_\Gamma = -1$ else.

The important point is that

$$\delta^2 = 0.$$

This can be verified directly or deduced from representing δ as the limit of d_s for $s \to \infty$. (Note that in any case $d_s : V_p \to V_{p+1}$ also yields a coboundary operator, i.e., $d_s^2 = 0$.) It is a standard result of algebraic topology that once one has such

a coboundary operator, one obtains the strong Morse inequalities encoded in the formula

$$\sum_p m_p t^p - \sum_p b_p t^p = (1+t)Q(t),$$

where $Q(t)$ is a polynomial with nonnegative integer coefficients.

A more general version of the supersymmetry algebra arises if, in addition to the Hamiltonian H, we also have a momentum operator P, and if the supersymmetry operators Q_1, Q_2 satisfy

$$Q_1^2 = H + P, \qquad Q_2^2 = H - P, \qquad Q_1 Q_2 + Q_2 Q_1 = 0.$$

These relations imply

$$[Q_i, H] = 0 = [Q_i, P] \quad \text{for } i = 1, 2.$$

Also,

$$H = \frac{1}{2}(Q_1^2 + Q_2^2)$$

is again positive semidefinite.

A realization of this supersymmetry algebra arises as follows.

Let X be a Killing field on our compact Riemannian manifold N, i.e., an infinitesimal isometry of N. Let L_X be the Lie derivative in the direction of X, and $i(X)$ the interior multiplication with X of a differential form. For $s \in \mathbb{R}$, we consider

$$d_s = d + s i(X).$$

Let d_s^* be the adjoint of d_s. Since X is a Killing field, one computes that

$$d_s^{*2} = -d_s^2.$$

Also

$$d_s^{*2} = -s L_X.$$

The Hamiltonian is

$$H_s = d_s d_s^* + d_s^* d_s.$$

The supersymmetry operators are

$$Q_{1,s} = i^{\frac{1}{2}} d_s + i^{-\frac{1}{2}} d_s^*, \qquad Q_{2,s} = i^{-\frac{1}{2}} d_s + i^{\frac{1}{2}} d_s^*.$$

Defining

$$P = 2is L_X,$$

we then have the above supersymmetry algebra,

$$Q_1^2 = H + P, \qquad Q_2^2 = H - P, \qquad Q_1 Q_2 + Q_2 Q_1 = 0.$$

More generally, one may use a function h invariant under the action of X, i.e.,

$$i(X)dh = 0,$$

and put

$$d_{s_1,s_2} = e^{-hs_2} d_{s_1} e^{hs_2}.$$

Thus, the parameter s_1 corresponds to the Killing field X, whereas s_2 corresponds to the Morse function h. The supersymmetry generators are then

$$Q_{1,s_1,s_2} = i^{\frac{1}{2}} d_{s_1,s_2} + i^{-\frac{1}{2}} d^*_{s_1,s_2},$$
$$Q_{2,s_1,s_2} = i^{-\frac{1}{2}} d_{s_1,s_2} + i^{\frac{1}{2}} d^*_{s_1,s_2}$$

and

$$H_{s_1,s_2} = d_{s_1,s_2} d^*_{s_1,s_2} + d^*_{s_1,s_2} d_{s_1,s_2},$$
$$P = 2is_1 L_X.$$

We return to our supersymmetric nonlinear sigma model with a cylindrical world sheet $\mathbb{R} \times S$, where the space S is a circle of circumference L. Instead of considering maps

$$\phi : \mathbb{R} \times S \to N,$$

we may equivalently consider maps

$$\psi : \mathbb{R} \to \Omega_s(N),$$

where $\Omega_s(N)$ is the loop space of maps from S to N. The loop space $\Omega_s(N)$ will now play the role of our target manifold. Of course, in contrast to what was assumed for our target manifold N, $\Omega_s(N)$ is not compact.

The group $U(1)$ of rotations of S acts on $\Omega_s(N)$ by isometries, simply by mapping a loop $\gamma(t)$ to the loop $\gamma(t + a)$ (the addition in S is the one in $\mathbb{R}$ mod L). As before, we may define the operators

$$d_s = d + si(X), \qquad H_s = d_s d^*_s + d^*_s d_s,$$

where X is the generator of the $U(1)$ action.

Of course, due to the fact that $\Omega_s(N)$ is infinite-dimensional, certain problems of convergence arise when trying to carry over the preceding finite-dimensional analysis.

The approach to Morse theory via the supersymmetric sigma model is due to Witten [104, 105]. This in turn led to Floer's approach to Morse theory that constructs the Morse complex from counting flow lines between critical points, see [37] and the expositions in [65, 96].

The supersymmetric action functional (2.4.112) can also be utilized for a proof of the Atiyah–Singer index theorem [7], as discovered by Alvarez-Gaumé [3], Friedan

and Windey [41, 42] and Getzler [47, 48]. Systematic expositions can be found in [10] and [43].

2.4.7 The Gravitino

The preceding considerations were local insofar as the supersymmetry variation parameter ε was assumed to be constant. It turns out that from a more global perspective, ε has to be considered to be a section of some bundle and cannot in general be taken to be constant. This implies that also derivatives of ε will enter the supersymmetry computations. We address this issue now and see that it will lead us to very interesting geometric structures.

We start with the linear supersymmetric sigma model from Sect. 2.4.3, that is, the extension of (2.4.3) with a supersymmetric partner for the scalar field ϕ, an anticommuting spinor field ψ:

$$S(\phi, \psi, \Sigma) := \frac{1}{2}\int_{\Sigma} (\partial_\alpha \phi^a \partial^\alpha \phi_a + \bar{\psi}^a \gamma^\alpha \partial_\alpha \psi_a) \mathrm{d}^2 z. \tag{2.4.135}$$

Here, the $\gamma^\alpha, \alpha = 1, 2$ are standard Dirac matrices, defined by a representation of $Cl(2, 0)$ as above. (Note: In the physics literature, one usually works with a Minkowski world sheet, that is, one takes an indefinite metric on the underlying surface, and consequently considers $Cl(1, 1)$ instead.)

The equations of motion, that is, the Euler–Lagrange equations for (2.4.135) are simple linear equations ((2.4.58), (2.4.59)):

$$\partial^\alpha \partial_\alpha \phi^a = 0, \tag{2.4.136}$$

$$\gamma^\alpha \partial_\alpha \psi^a = 0 \quad \text{for } a = 1, \ldots, d, \tag{2.4.137}$$

that is, ϕ is harmonic and ψ solves the Dirac equation.

Similarly, one can consider a metric g instead of only a conformal structure and consider the functional

$$S(\phi, \psi, g) := \frac{1}{2}\int_S (g^{\alpha\beta} \partial_\alpha \phi^a \partial_\beta \phi_a + \bar{\psi}^a \gamma^\alpha \partial_\alpha \psi_a) \sqrt{\det g} \mathrm{d}z^1 \mathrm{d}z^2. \tag{2.4.138}$$

One then has the supersymmetry transformations (2.4.65):

$$\delta \phi^a = \bar{\epsilon} \psi^a, \tag{2.4.139}$$

$$\delta \psi^a = \gamma^\alpha \partial_\alpha \phi^a \varepsilon \tag{2.4.140}$$

with an anticommuting ε. (Of course, mathematically, one should consider this as a transformation of the independent variables of an underlying superspace instead of as a transformation of the fields.) The commutator of two such transformations yields a spatial translation:

$$[\delta_1, \delta_2] = \delta_1(\bar{\varepsilon_2} \psi^a) - \delta_2(\bar{\varepsilon_1} \psi^a) = 2\bar{\varepsilon_1} \gamma^\alpha \varepsilon_2 \partial_\alpha \phi^a. \tag{2.4.141}$$

In fact, these are infinitesimal transformations that integrate to local ones, but we also need to consider the global situation. Globally, instead of a translation, we have a diffeomorphism, and so the supersymmetry transformations should generate the superdiffeomorphism group of the underlying supersurface. Also, globally, ε is not a scalar parameter, but transforms as a spin-$1/2$ field, that is, mathematically, a not necessarily holomorphic, anticommuting section of $K^{1/2}$, K being the canonical bundle of Σ (for some choice of a square root of K, that is, of a spin structure). (Even though, w.r.t. its z-dependence, ε transforms as a section of $K^{1/2}$, it also contains an independent odd parameter; therefore, $\varepsilon\psi = -\psi\varepsilon$, but in general, we do not have $\varepsilon\psi = 0$.)

A supersymmetry transformation induces a variation of S; this is computed as (cf. (2.4.56))

$$\delta S = -2 \int \partial_\alpha \bar{\varepsilon} J^\alpha \tag{2.4.142}$$

with the **supercurrent**

$$J_\alpha = \frac{1}{2} \gamma^\beta \gamma_\alpha \psi^a \partial_\beta \phi_a . \tag{2.4.143}$$

Likewise, for a spatial translation, we get the **energy–momentum tensor**:

$$T_{\alpha\beta} = \partial_\alpha \phi^a \partial_\beta \phi_a + \frac{1}{4} \bar{\psi}^a \gamma_\alpha \partial_\beta \psi_a + \frac{1}{4} \bar{\psi}^a \gamma_\beta \partial_\alpha \psi_a - \text{trace}. \tag{2.4.144}$$

Of course, this is the appropriate generalization of (2.4.9). As before, it is traceless, and again, this can be seen as expressing a (super)conformal invariance. Also, as before, both the supercurrent J and the energy–momentum tensor T are divergence-free when the equations of motion hold. With the same implicit identifications as in Sect. 2.4, T is a holomorphic quadratic differential on Σ, that is, a holomorphic section of K^2, while J is a holomorphic section of $K^{3/2}$.

The preceding facts have several important consequences:

- In line with the general concept of supergeometry, the space of independent variables for the ϕ and ψ fields should be a superspace, that is, here it should be a super Riemann surface (SRS). Then, in the same manner that the Dirichlet integral, the action functional $D(\phi, \Sigma)$, yielded a (co)tangent vector to the moduli space M_p when varying Σ, now variations of Σ for $S(\phi, \psi, \Sigma)$ should yield a (co)tangent vector to the moduli space of super Riemann surfaces. From this, we infer that the tangent space to that space should be given by even holomorphic sections of K^2 and odd holomorphic sections of $K^{3/2}$. In particular, the even dimension should be $3p - 3$ as before while the odd one is $2p - 2$, again by Riemann–Roch.
- As before, our action functional is only invariant on-shell, that is, when J is holomorphic. From (2.4.142), we see the obstruction to global invariance, namely the nonvanishing of $\partial_\alpha \bar{\varepsilon}$. As a spin-$1/2$ field, ε is a section of a nontrivial bundle and therefore cannot be taken to be globally constant. Thus, the obstruction to full superdiffeomorphism invariance comes from the global topology of the underlying surface.

In order to understand these issues better, we now make the fundamental observation that the functional S from (2.4.135) or (2.4.138) does not yet constitute a full supersymmetric generalization of the functional $S(\phi, g)$ studied in Sect. 2.4. Namely, we have only given ϕ a supersymmetric partner, but not our other field, namely the metric g. We shall do that now and see that this yields a fully satisfactory theory that gives a profound understanding of the moduli space of super Riemann surfaces.

In place of the metric $(g_{\alpha\beta})$, it is convenient to consider a zweibein e^a_α, from which we can reconstruct the metric as $g_{\alpha\beta} = \delta_{ab} e^a_\alpha \, e^b_\beta$. In other words, we introduce an additional $U(1)$ symmetry which, however, can be easily divided out since that group is compact. The supersymmetric partner of the zweibein is then a gravitino (Rarita–Schwinger field) $\chi_{A\alpha}$ where $A = 1, 2$ is a spinor index that will be suppressed in the sequel, whereas α is a vector index as before. Thus, χ transforms as a spin-$3/2$ field. This might already suggest how to obtain the moduli space of super Riemann surfaces in analogy to 4 of Sect. 1.4.2. Namely, one would take the space of all metrics (equivalently, after dividing out the $U(1)$ symmetry, zweibeins) and gravitinos, and then divide out all the invariances, that is, the superdiffeomorphisms and superconformal scalings. However, although this idea is conceptually insightful, the actual construction of the moduli space of super Riemann surfaces proceeds differently, see [93].[19] In fact, because the spaces involved, like the one of superdiffeomorphisms, are necessarily infinite-dimensional, Sachse had to replace the standard approach of ringed topological spaces by a categorical reformulation of supergeometry, see [94].

The supersymmetry transformations of the fields ϕ, ψ, e, χ are then

$$\delta\chi_\alpha = \partial_\alpha \bar{\varepsilon}, \tag{2.4.145}$$

$$\delta e^a_\alpha = -2\bar{\varepsilon}\gamma^a \chi_\alpha, \tag{2.4.146}$$

$$\delta\phi^a = \bar{\varepsilon}\psi^a, \tag{2.4.147}$$

$$\delta\psi^a = \gamma^\alpha \varepsilon(\partial_\alpha \phi^a - \bar{\psi}^a \chi_\alpha). \tag{2.4.148}$$

The supersymmetric functional is then

$$\begin{aligned} &S(\phi, \psi, g, \chi) \\ &:= \frac{1}{2}\int_S \Big(g^{\alpha\beta}\partial_\alpha\phi^a \partial_\beta \phi_a + \bar{\psi}^a \gamma^\alpha \partial_\alpha \psi_a + 2\bar{\chi}_\alpha \gamma^\beta \gamma^\alpha \psi^a \partial_\beta \phi_a \\ &\quad + \frac{1}{2}\bar{\psi}_a \psi^a \bar{\chi}_\alpha \gamma^\beta \gamma^\alpha \chi_\beta \Big) \sqrt{\det g}\, dz^1 dz^2. \end{aligned} \tag{2.4.149}$$

[19] Using the zweibeins directly would mean taking the phase space of a 2D supergravity theory as the gauge theory for supersymmetry. One would then in addition need a super connection on Σ whose coefficients are the gauge fields. Dividing out the invariances involved becomes very complicated, and therefore, it is better to proceed as in [93].

Here, the first two terms are those from (2.4.138), the third one is introduced to compensate (2.4.142), and the last one is then needed to compensate the terms coming from the variation of $\partial_\beta \phi$ in the third one.

We have thus obtained a functional that is fully supersymmetric even off-shell.

Summary: We see the merging of a profound mathematical concept, namely that of a moduli space of Riemann surfaces and a deep method from theoretical physics, namely the symmetries of action functionals. This suggests a unique concept of a super Riemann surface, for which we have already described the super moduli space. It remains to be seen how the approaches of Sect. 1.4.2 extend to this setting. Ideally, they should as beautifully coincide as in the situation of ordinary Riemann surfaces.

Of course, the preceding formalism can be recast into the mathematical framework of supergeometry.

We have considered only one of the two supersymmetries arising in string theory, namely world-sheet supersymmetry, but not space–time supersymmetry. The latter refers to the target space, which we have taken to be Euclidean space here. For example, while ψ transforms as a spinor on the domain, it transforms as a vector in the target space. For a discussion of space–time supersymmetry, see, e.g., [50]. Here, instead, we replace the Euclidean target space by a Riemannian manifold N. Equation (2.4.138) then becomes the supersymmetric nonlinear sigma model of quantum field theory as treated in the preceding section. The equations for ϕ and ψ then become nonlinear and coupled, and in fact, ψ is a spinor-valued section of $\phi^* TN$, the pull-back of the tangent bundle of N under the map ϕ. Naturally, one can also include the fields g and χ into these considerations, by expanding not only with respect to the map into N, but also with respect to the domain metric.

The supersymmetric action functional with gravitino term is discussed in [26, 50], with more details in [27]. The moduli space of super Riemann surfaces has been constructed from the global analysis perspective advocated here by Sachse [93].

2.5 Functional Integrals

We can now bring the material of the preceding sections together and discuss general (Gaussian) functional integrals. These are formal integrals of the form

$$\int D\varphi \; e^{-S(\varphi)} \tag{2.5.1}$$

where $S(\varphi)$ is some quadratic Lagrangian action as introduced in Sect. 2.2 and we formally integrate w.r.t. to some collection of fields φ. We can, of course, also introduce Planck's constant and replace (2.5.1) by

$$\int D\varphi \; e^{-\frac{1}{\hbar} S(\varphi)}. \tag{2.5.2}$$

When we consider the heuristic limit $\hbar \to 0$, we see that the minimizers of the action S dominate the functional integral more and more, because other fields φ yield exponentially smaller contributions. For physicists, it is then natural to perform an expansion of (2.5.2) in terms of $\frac{1}{\hbar}$, the so-called stationary phase approximation.

Certainly, one can also consider the oscillatory integral

$$\int D\varphi \; e^{\frac{i}{\hbar}S(\varphi)}, \tag{2.5.3}$$

which we may view as a generalization of the Feynman path integral discussed in Sect. 2.1.3.

As before, see Sects. 2.1.2, 2.1.3, we consider Gaussian functional integrals as formal analogs of Gaussian integrals with infinitely many variables. In addition, we shall make use of the invariance considerations in Sect. 2.3.2 to divide out symmetries.

There is one general issue that can be contemplated at this point: It is a general principle of quantum field theory that no arbitrary choices are permitted. Whenever something is selected from some class of possibilities, one should integrate out the possible values of the selection, weighted with some (negative or imaginary) exponential of the underlying action. Thus, we consider (2.5.1) when we have a collection of fields φ. After normalization, we consider $\frac{1}{Z} D\varphi \; e^{-S(\varphi)}$ (where the constant Z has been chosen so that the total integral of the measure becomes 1) as a probability measure on the space of fields (similar to a Gibbs measure in statistical mechanics). For any function $f(\varphi)$ of the field φ, we can then compute its expectation value as

$$\frac{1}{Z} \int D\varphi \; f(\varphi) e^{-\frac{1}{\hbar}S(\varphi)}. \tag{2.5.4}$$

In mathematics, instead of taking a functional integral, in the situation where some underlying structure has to be selected, one attempts to equip the space of all possible choices with some geometric structure. That is then called a moduli space. Above, we have discussed the moduli space of Riemann surfaces.

2.5.1 Normal Ordering and Operator Product Expansions

The following example will bring out the essential aspects. Let (M, g) be a compact Riemannian manifold of dimension d. For a function φ on M, we put

$$\begin{aligned} S(\varphi) &= \frac{1}{4\pi\alpha'} \int_M (\|D\varphi\|^2 + m^2\varphi^2)\, d\mathrm{Vol}_g(M) \\ &= \frac{1}{4\pi\alpha'} (\varphi, (-\Delta_g + m^2)\varphi)_{L^2}, \end{aligned} \tag{2.5.5}$$

where α' is a constant, the so-called Regge slope. Thus, this is essentially the same functional as the one considered in Sect. 2.4, see (2.4.3), with the difference that here

we have an additional mass term and a different normalization factor in front of the integral. $\Delta = \Delta_g$ is the Laplace–Beltrami operator of (M, g), defined in (1.1.103), (2.4.4).

We note some differences here compared to Sect. 2.1.3. There, we had taken functional integrals for paths x (in some Euclidean or Minkowski space), that is, mappings $x : [t', t''] \to \mathbb{R}^d$, say, with fixed boundary conditions $x(t') = x'$, $x(t'') = x''$. Here, we are integrating functions over a more general domain, namely a Riemannian manifold, and we do not impose boundary conditions. In fact, M may be some closed manifold without boundary. If M does have a boundary, we can also impose a boundary condition via an insertion into our functional integral.

According to the general scheme just discussed, the choice of the manifold (M, g) represents an arbitrary choice, and therefore, one should integrate out all such choices, that is, take another functional integral w.r.t. all possible metrics g on M, and perhaps also a sum w.r.t. all diffeomorphism types of M. That is, in fact, done in string theory, where the dimension of M is fixed to be 2 and one then formally integrates w.r.t. all metrics and sums with respect to the different genera of the underlying surface.

We also consider the propagator of the free field of mass m, or, in mathematical terminology, the Green operator

$$G = 2\pi\alpha'(-\Delta + m^2)^{-1}. \tag{2.5.6}$$

Thus,

$$S(\varphi) = \frac{1}{2}(\varphi, G^{-1}\varphi)_{L^2}. \tag{2.5.7}$$

The fundamental object of interest is the partition function (in older texts, this is denoted by the German term *Zustandssumme*)

$$\begin{aligned} Z &:= \int D\varphi \exp(-S(\varphi)) \\ &= \int D\varphi \exp\left(-\frac{1}{2}(\varphi, G^{-1}\varphi)\right) \end{aligned} \tag{2.5.8}$$

with a formal integration over all functions $\varphi \in L^2(M)$.

The analogy with the above discussion of Gaussian integrals (2.1.24), obtained by replacing the coordinate index i in (2.1.24) by the point z in our manifold M, would suggest

$$Z = (\det G)^{\frac{1}{2}}, \tag{2.5.9}$$

when we normalize

$$D\varphi = \prod_i \frac{d\varphi_i}{\sqrt{2\pi}} \tag{2.5.10}$$

to get rid of the factor $(2\pi)^n$ in (2.1.24). Here, the (φ_i) are an orthonormal basis of the Hilbert space $L^2(M)$, for example, the eigenfunctions of Δ.

The idea is then to define $\det G$ as the renormalized product of the eigenvalues λ_n of G. The mathematical construction is based on the Weyl estimates. By these estimates, the λ_n behave as $O(n^{-\frac{2}{d}})$. Motivated by (2.1.26), one then puts

$$\det G := \exp(-\zeta_G'(0)), \tag{2.5.11}$$

where the ζ-function $\zeta_G(s)$ is the meromorphic continuation of $\sum_n \lambda_n^{-s}$, defined for $\mathrm{Re}(s) < -\frac{d}{2}$, to the entire complex plane; it is analytic at 0. This procedure is called zeta function regularization. The determinant defined by (2.5.11) has certain multiplicative properties like the ordinary determinant, see e.g. [111].

Comparing (2.5.8) with (2.1.24), the analogy is then that the coordinate values $x^1, \ldots, x^d$ get replaced by the values of the function φ at the points $y \in M$. That is, we have infinitely many degrees of freedom, corresponding to the points $y \in M$ instead of to the discrete indices $i = 1, \ldots, n$. The values of these degrees of freedom are then assembled into the function φ in place of the vector $x = (x^1, \ldots, x^n)$.

In analogy with (2.1.30), for points $y_1, \ldots, y_m \in M$, we may then define correlation functions

$$\langle \varphi(y_1) \cdots \varphi(y_m) \rangle := \frac{1}{Z} \int D\varphi \varphi(y_1) \cdots \varphi(y_m) \exp(-S(\varphi)). \tag{2.5.12}$$

Note that, in contrast to Sect. 2.1.3, here we are normalizing the integrals by dividing by Z, so that these correlation functions can be interpreted as the expectation values of the product of the evaluations of the fields at the points $y_1, \ldots, y_m$. Again, these vanish for odd m (because a Gaussian integral is quadratic in the fields, hence even), and as in (2.1.31)

$$\langle \varphi(y_1)\varphi(y_2) \rangle = G(y_1, y_2). \tag{2.5.13}$$

Here, the Green function $G(y_1, y_2)$ is the kernel of the operator G, and it has a singularity at $y_1 = y_2$, of order $\log \mathrm{dist}(y_1, y_2)$ for $d = 2$ and $\mathrm{dist}(y_1, y_2)^{2-d}$ for $d > 2$.

Likewise, the analog of Wick's theorem (2.1.32) holds.

We now specialize to the case where the particle is massless, i.e., $m = 0$ in (2.5.5), and M is a Riemann surface Σ. Thus, in complex coordinates, the action is

$$S = \frac{1}{2\pi\alpha'} \int d^2 w \partial\varphi\bar{\partial}\varphi. \tag{2.5.14}$$

We note that the metric g here disappears from the picture. This comes from the fact that S in (2.5.14) is conformally invariant, that is, remains unchanged when the underlying metric is multiplied by some positive function, and therefore depends only on the conformal structure, that is, on the Riemann surface on which it is defined, but not on a particular choice of a conformal metric on that Riemann surface. The issue of conformal invariance plays a fundamental role in conformal field theory and string theory, see [26, 46, 62].

The classical equation of motion is (2.4.21),

$$\partial\bar{\partial}\varphi(z, \bar{z}) = 0. \tag{2.5.15}$$

One writes the argument here as $(z, \bar{z})$ instead of simply z, because the notation $f(z)$ is reserved for a holomorphic function, as explained in Sect. 1.1.2. (2.5.15) implies that $\partial\varphi$ is a holomorphic function $\partial\varphi(z)$, and $\bar{\partial}\varphi$ is an antiholomorphic function $\bar{\partial}\varphi(\bar{z})$.

The complex coordinates

$$z = x^1 + ix^2,$$
$$\bar{z} = x^1 - ix^2$$

admit a Minkowski continuation with $x^0 = -ix^2$. Then, a holomorphic function is a function of $x^0 - x^1$, an antiholomorphic one is a function of $x^0 + x^1$. One calls an (anti)holomorphic function left-(right-)moving.

As before, we wish to compute the expectation values

$$\langle F(\varphi)\rangle = \frac{1}{Z}\int D\varphi \exp(-S)F(\varphi). \tag{2.5.16}$$

We shall now repeat some of the discussion of Sect. 2.1.3 and see how it applies to the present situation. The above analogy between ordinary integrals and path integrals said that the finitely many ordinary degrees of freedom, the coordinate values of the integration variable, are replaced by the infinitely many function values $\varphi(z, \bar{z})$. Therefore, integration by parts should yield that

$$0 = \int D\varphi \frac{\delta}{\delta\varphi(z,\bar{z})}\exp(-S). \tag{2.5.17}$$

This gives

$$0 = -\int D\varphi \exp(-S)\frac{\delta S}{\delta\varphi(z,\bar{z})},$$

and so,

$$0 = -\left\langle \frac{\delta S}{\delta\varphi(z,\bar{z})}\right\rangle = \frac{1}{\pi\alpha'}\langle\partial\bar{\partial}\varphi(z,\bar{z})\rangle. \tag{2.5.18}$$

Thus, the classical equation of motion (2.5.15) becomes an equation for the expectation value of the corresponding operator. Equation (2.5.18) can also be written as

$$\frac{1}{\pi\alpha'}\partial_z\partial_{\bar{z}}\langle\varphi(z,\bar{z})\rangle = 0. \tag{2.5.19}$$

Let us return to (2.5.16). The functional $F(\varphi)$ typically represents certain linear combinations of products of evaluations of φ at points $z_1, \ldots, z_m \in \Sigma$. When none of those points coincides with the point z for which we take the functional derivative $\frac{\delta}{\delta\varphi(z,\bar{z})}$, the preceding computation also goes through for $F(\varphi)$.

Things change when one of those insertion points is allowed to coincide with z. In Sect. 2.1.3, that led us to the temporal ordering scheme for operators. Similarly,

here, from the analog of (2.1.125), we shall be led to the so-called normal ordering scheme.

For example

$$\begin{aligned} 0 &= \int D\varphi \frac{\delta}{\delta\varphi(z,\bar{z})}(\exp(-S)\varphi(\zeta,\bar{\zeta})) \\ &= \int D\varphi \exp(-S)\Big(\delta(z-\zeta,\bar{z}-\bar{\zeta}) + \frac{1}{\pi\alpha'}(\partial_z\partial_{\bar{z}}\varphi(z,\bar{z}))\varphi(\zeta,\bar{\zeta})\Big). \end{aligned} \tag{2.5.20}$$

Thus

$$0 = \Big\langle \delta(z-\zeta,\bar{z}-\bar{\zeta}) + \frac{1}{\pi\alpha'}\partial_z\partial_{\bar{z}}\varphi(z,\bar{z})\varphi(\zeta,\bar{\zeta})\Big\rangle. \tag{2.5.21}$$

Again, this is not affected by other insertions not coincident with z.

We thus interpret (2.5.18) and (2.5.21) as operator equations, that is, as holding for all components of the corresponding quantum mechanical operators, since these are precisely obtained by such insertions.

We thus write the operator equation

$$\frac{1}{\pi\alpha'}\partial_z\partial_{\bar{z}}\varphi(z,\bar{z})\varphi(\zeta,\bar{\zeta}) = -\delta(z-\zeta,\bar{z}-\bar{\zeta}), \tag{2.5.22}$$

as in (2.1.125). When we solve (2.5.22), we therefore obtain a Green function type singularity, $\log|z-\zeta|^2$.

In order to eliminate this contribution, one considers the normal ordered operators

$$\begin{aligned} :\varphi(z,\bar{z}): &= \varphi(z,\bar{z}), \\ :\varphi(z_1,\bar{z}_1)\varphi(z_2,\bar{z}_2): &= \varphi(z_1,\bar{z}_1)\varphi(z_2,\bar{z}_2) + \frac{\alpha'}{2}\log|z_1-z_2|^2. \end{aligned} \tag{2.5.23}$$

This quantum correction will below lead to a central extension of the Lie algebra of the diffeomorphism group of the circle (see (2.5.63), (2.5.66) in Sect. 2.5.3).

We then have

$$\partial_1\bar{\partial}_1 :\varphi(z_1,\bar{z}_1)\varphi(z_2,\bar{z}_2): = 0. \tag{2.5.24}$$

Thus, $:\varphi(z_1,\bar{z}_1)\varphi(z_2,\bar{z}_2):$ is a harmonic function and therefore locally the sum of a holomorphic and an antiholomorphic function. From this, we obtain the Taylor expansion

$$\begin{aligned} &\varphi(z_1,\bar{z}_1)\varphi(z_2,\bar{z}_2) \\ &\quad = -\frac{\alpha'}{2}\log|z_1-z_2|^2 \\ &\qquad + \sum_{\nu=1}^{\infty}\frac{1}{\nu!}\big((z_1-z_2)^\nu :\varphi\partial^\nu\varphi(z_2,\bar{z}_2): + (\bar{z}_1-\bar{z}_2)^\nu :\varphi\bar{\partial}^\nu\varphi(z_2,\bar{z}_2):\big) \end{aligned} \tag{2.5.25}$$

(mixed terms with $\partial\bar{\partial}$ vanish by the equation of motion; note that in general, derivatives need not commute with normal ordering).

Equation (2.5.25) is the prototype of an operator product expansion (OPE). As discussed, the φ's here are considered as quantum mechanical operators.

The transition from functions to operators needs some explanation. In (2.5.16), we can add insertions $I^1, \dots, I^m$, that is, functions of φ evaluated at some points $z_1, \dots, z_m \in M$. Thus, we have expressions of the form

$$\frac{1}{z}\int D\varphi \exp(-S)F(\varphi)I^1(\varphi)(z_1,\bar{z}_1)\cdots I^m(\varphi)(z_m,\bar{z}_m).$$

More generally, we can also have insertions of the form

$$\int I(z,\bar{z};\varphi)\,d\mu(z)$$

for some measure $d\mu(z)$. For example, these insertions can be certain boundary conditions represented by Dirac functionals. When we do not specify these insertions, we simply write

$$\langle F(\varphi)\cdots\rangle.$$

$F(\varphi)$ then determines an operator $\hat{F}(\varphi)$ operating on such insertions. $\langle F(\varphi)\rangle$ is the matrix element $\langle 0|\hat{F}(\varphi)|0\rangle$ of $\hat{F}(\varphi)$ where $|0\rangle$ is the vacuum.

2.5.2 *Noether's Theorem and Ward Identities*

Before proceeding, we need to translate Noether's theorem into the operator setting. The result is a Ward identity.

As in Sect. 2.3.2, we consider a general Lagrangian action

$$S = \int F(\varphi(x), d\varphi(x))\,dx \tag{2.5.26}$$

and transformations

$$\begin{aligned} x &\mapsto x', \\ \varphi(x) &\mapsto \varphi'(x') =: \psi(\varphi(x)). \end{aligned}$$

Infinitesimally,

$$x' = x + s\delta x, \tag{2.5.27}$$

$$\varphi'(x') = \varphi(x) + s\delta\psi(x). \tag{2.5.28}$$

By (2.3.28), the Noether current is

$$j_i^\alpha = \left(-F_{p^\alpha}\frac{\partial\varphi}{\partial x^\beta} + \delta^\alpha_\beta F\right)\frac{\delta x^\beta}{\delta s_i} + F_{p^\alpha}\frac{\delta\psi}{\delta s_i}, \tag{2.5.29}$$

$$\begin{aligned}\delta S &= -\int dx j_i^\alpha \partial_\alpha s_i \\ &= \int dx \partial_\alpha j_i^\alpha s_i.\end{aligned} \tag{2.5.30}$$

According to Noether (2.3.29), invariance implies a conserved current:

$$\partial_\alpha j_i^\alpha = 0. \tag{2.5.31}$$

We now turn to the quantum version, that is, invariance of correlation functions, when action and functional integral measure both are invariant:

$$\langle\varphi(x_1')\cdots\varphi(x_n')\rangle = \langle\psi(\varphi(x_1))\cdots\psi(\varphi(x_n))\rangle \tag{2.5.32}$$

by renaming variables ($\varphi \mapsto \varphi'$) and transforming $D\varphi'$ to $D\varphi$.

Ward identities express symmetries in QFT as identities between correlation functions. According to (2.3.27), the field variations are given by

$$G\varphi := \frac{\delta\psi}{\delta s} - \frac{\delta x}{\delta s}\frac{\partial\varphi}{\partial x}. \tag{2.5.33}$$

For a collection $\Phi = \varphi(x_1)\cdots\varphi(x_n)$ of fields, we have by invariance

$$\begin{aligned}&\frac{1}{Z}\int D\varphi\Phi\exp(-S(\varphi)) \\ &\quad= \langle\Phi\rangle \\ &\quad= \frac{1}{Z}\int D\varphi'(\Phi+\delta\Phi)\exp\left(-\left(S(\varphi) + \int dx\partial_\alpha j_i^\alpha s_i(x)\right)\right).\end{aligned}$$

If the measure is invariant, i.e., $D\varphi' = D\varphi$, then by differentiating w.r.t. s gives

$$\langle\delta\Phi\rangle = \int dx\partial_\alpha\langle j_i^\alpha(x)\Phi\rangle s_i(x) \tag{2.5.34}$$

(note that Φ does not depend explicitly on x, and thus $\partial(j(x))\Phi = \partial(j(x)\Phi)$).

Since

$$\begin{aligned}\delta\Phi &= -\sum_{k=1}^n(\varphi(x_1)\cdots G\varphi(x_k)\cdots\varphi(x_n))s(x_k) \\ &= \int dx s(x)\sum_{k=1}^n(\varphi(x_1)\cdots G\varphi(x_k)\cdots\varphi(x_n))\delta(x-x_k),\end{aligned}$$

we obtain the Ward identity for the current j:

$$\frac{\partial}{\partial x^\alpha}\langle j^\alpha(x)\varphi(x_1)\cdots\varphi(x_n)\rangle = \sum_{k=1}^{n}\langle\varphi(x_1)\cdots G\varphi(x_k)\cdots\varphi(x_n)\rangle\delta(x-x_k). \quad (2.5.35)$$

We now assume that the time $t = x_1^0$ is different from all the times $x_2^0, \ldots, x_n^0$ occurring in Φ. We integrate (2.5.35) between $t-\varepsilon$ and $t+\varepsilon$ for small $\varepsilon > 0$ to obtain

$$\langle Q(t+\varepsilon)\varphi(x_1)\Phi\rangle - \langle Q(t-\varepsilon)\varphi(x_1)\Phi\rangle = \langle G\varphi(x_1)\Phi\rangle \quad (2.5.36)$$

for the charge Q (defined as in (2.3.40)). When we time order the operators, we need to exchange $Q(t-\varepsilon)$ and $\varphi(x_1)$, because $t-\varepsilon < t = x_1^0$. Since (2.5.36) holds for any such Φ, we obtain

$$[Q,\varphi] = G\varphi. \quad (2.5.37)$$

Thus, the conserved change Q is the infinitesimal generator of the symmetry transformations in the operator formalism.

If instead of a Minkowski space–time, we consider Euclidean space, the time ordering is replaced by a radial ordering of the operators as will be discussed in Sect. 2.5.3 below.

2.5.3 Two-dimensional Field Theory

We now compare the preceding with 2-dimensional field theory. We have a spatial coordinate w^1 that may be bounded or periodic,

$$w^1 \sim w^1 + 2\pi, \quad (2.5.38)$$

and a Euclidean time coordinate $\tau = w^2$,

$$-\infty < w^2 < \infty. \quad (2.5.39)$$

We put

$$w = w^1 + iw^2 \quad \text{(equal time coordinates are horizontal lines)} \quad (2.5.40)$$

and

$$z = e^{-iw} \quad \text{(equal time coodinates are concentric circles } C \text{ about origin } z = 0 \text{ which corresponds to the infinite past } w^2 = -\infty). \quad (2.5.41)$$

When going from the Minkowski coordinates w^1, w^2 to the complex coordinates z, temporal invariance $w^2 \to w^2 + t$ then becomes radial invariance $z \to \lambda z$ with

$\lambda \in \mathbb{R}$. This is the starting point of conformal invariance and constitutes one motivation for conformal field theory below.

In the z-coordinates, the charges Q (2.3.40) then become contour integrals of currents j

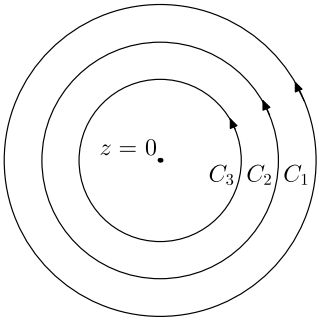
$$Q\{C\} = \oint_C \frac{dz}{2\pi i} j. \tag{2.5.42}$$

Here, we assume that the current j is meromorphic, without poles on the contour C, of course.

We now consider

$$Q_1\{C_1\}Q_2\{C_2\} - Q_2\{C_2\}Q_1\{C_3\}. \tag{2.5.43}$$

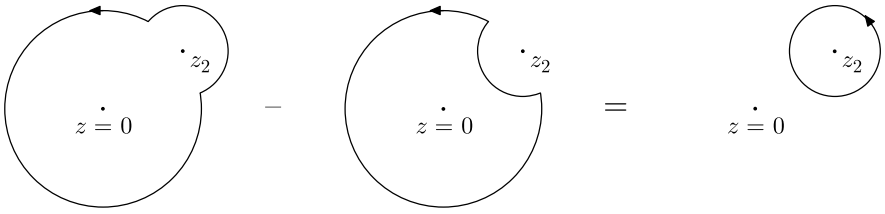
This corresponds to a time ordering $\tau_1 > \tau_2 > \tau_3$.



Therefore, when we time order the operators $\hat{Q}_i$ corresponding to the Q_i, we obtain the expression

$$\hat{Q}_1\hat{Q}_2 - \hat{Q}_2\hat{Q}_1 = [\hat{Q}_1, \hat{Q}_2]. \tag{2.5.44}$$

We now consider a point $z_2 \in C_2$, and we can deform the contours as follows:



When we consider infinitesimal time differences, $\tau_1 = \tau_2 + \varepsilon$, $\tau_3 = \tau_2 - \varepsilon$, $\varepsilon \to 0$, we contract the contour $C_1 - C_3$ to C_2, that is, the small circle about z_2, to the point z_2.

We obtain from (2.5.42)–(2.5.44), leaving out the $\hat{}$ for the operators as usual,

$$[Q_1, Q_2]\{C_2\} = \oint_{C_2} \frac{dz_2}{2\pi i} \mathrm{Res}_{z_1 \to z_2} j_1(z_1)\, j_2(z_2). \tag{2.5.45}$$

This is a fundamental relation. On the l.h.s., we have the commutator algebra of the charges, while on the r.h.s., the singular terms in the operator product expansions (OPEs) of the currents appear.

Instead of the conserved charge $Q_2\{C_2\}$, we can also take an operator $A(z,\bar{z})$ to obtain

$$[Q, A(z_2,\bar{z}_2)] = \mathrm{Res}_{z_1\to z_2} j(z_1)A(z_2,\bar{z}_2). \tag{2.5.46}$$

When Q is the conserved charge for a variation δ, as in Sect. 2.3.2,

$$\varphi(x) \mapsto \varphi(x) + i\varepsilon s(x), \tag{2.5.47}$$

we have, by (2.5.37),

$$[Q, A(z,\bar{z})] = -\frac{1}{i\varepsilon}\delta A(z,\bar{z}). \tag{2.5.48}$$

(2.5.46) and (2.5.48) yield

$$\mathrm{Res}_{z_1\to z_2} j(z_1)A(z_2,\bar{z}_2) = -\frac{1}{i\varepsilon}\delta A(z_2,\bar{z}_2). \tag{2.5.49}$$

We now consider a conserved current j in a two-dimensional field theory. As a conserved current, by (2.3.22) and (2.3.29), it is divergence free, that is

$$\partial_{\bar{z}} j_z + \partial_z j_{\bar{z}} = 0. \tag{2.5.50}$$

Taking as a model the energy–momentum tensor T in Sect. 2.4, we now assume that we have

$$j_z = \delta z j_{zz} + \delta\bar{z} j_{z\bar{z}}, \qquad j_{\bar{z}} = \delta z\ j_{\bar{z}z} + \delta\bar{z}\ j_{\bar{z}\bar{z}} \tag{2.5.51}$$

for some holomorphic variation δz. Equation (2.5.50) then becomes

$$\begin{aligned} \partial_{\bar{z}} j_{zz} + \partial_z j_{\bar{z}z} &= 0, \\ \partial_z j_{\bar{z}\bar{z}} + \partial_{\bar{z}} j_{z\bar{z}} &= 0. \end{aligned} \tag{2.5.52}$$

We also assume that the tensor $(j_{zz}, \ldots)$ is symmetric:

$$j_{z\bar{z}} = j_{\bar{z}z}, \tag{2.5.53}$$

and (noting that $\mathrm{tr}\, j = g^{ab} j_{ba} = g^{z\bar{z}} j_{\bar{z}z} + g^{\bar{z}z} j_{z\bar{z}}$) trace-free:

$$j_{z\bar{z}} = 0, \tag{2.5.54}$$

which it has to be for the theory to be conformally invariant.

These relations imply that it is holomorphic:

$$\begin{aligned} \partial_{\bar{z}} j_{zz} &= 0, \\ \partial_z j_{\bar{z}\bar{z}} &= 0. \end{aligned} \tag{2.5.55}$$

We put

$$j(z) = j_{zz}(z),$$
$$\bar{j}(z) = j_{\bar{z}\bar{z}}(\bar{z}).$$

This implies that $f(z)j(z)$ is conserved as well:

$$\partial_{\bar{z}}(fj) = 0, \tag{2.5.56}$$

for any holomorphic function $f(z)$.

Thus, we obtain infinitely many conserved currents. In two-dimensional field theory, this corresponds to the fact that the local conformal group is infinite-dimensional, as conformal invariance led to the energy–momentum tensor T as our conserved current j in Sect. 2.4.

For each holomorphic f, we therefore obtain a conserved charge

$$Q_f = \oint_C \frac{dz}{2\pi i} f(z)T(z) \tag{2.5.57}$$

which generates the conformal transformation

$$z \mapsto z + \varepsilon f(z). \tag{2.5.58}$$

According to (1.1.89), the induced transformation of a field $\varphi(z, \bar{z})$ is

$$\varphi(z, \bar{z}) \mapsto \varphi(z, \bar{z}) + \delta_{f,\bar{f}}\varphi(z, \bar{z}) \tag{2.5.59}$$

with

$$\delta_{f,\bar{f}}\varphi(z, \bar{z}) = (h\partial_z f + \widetilde{h}\partial_{\bar{z}}\bar{f} + f\partial_z + \bar{f}\partial_{\bar{z}})\varphi(z, \bar{z}) \tag{2.5.60}$$

where h and $\widetilde{h}$ are the conformal weights of φ.

We consider an $(h, 0)$-form $\varphi(z, \bar{z})(dz)^h$. Equations (2.5.48), (2.5.49) and (2.5.57) give

$$\begin{aligned}\delta_f\varphi(z) &= -[Q_f, \varphi(z)] \\ &= -\mathrm{Res}_{z_1 \to z_2} f(z_1)T(z_1)\varphi(z) \\ &= \oint_C \frac{dz_1}{2\pi i} f(z_1)T(z_1)\varphi(z),\end{aligned} \tag{2.5.61}$$

where C is now a small circle about z.

Since $\widetilde{h} = 0$ here, we obtain from (2.5.60) and (2.5.61) that

$$T(z_1)\varphi(z) = \frac{h\varphi(z)}{(z_1 - z)^2} + \frac{\partial_z\varphi(z)}{z_1 - z} + \text{finite terms}. \tag{2.5.62}$$

In particular, for an $(h, 0)$-form, the conformal weight h can be recovered from the operator product with the energy–momentum tensor.

If we take, instead of φ, the energy–momentum tensor T itself in this OPE, we obtain an additional term that essentially comes from the fact that T involves a square of derivatives of fields which induce additional commutator terms:

$$T(z_1)T(z) = \frac{c}{2(z_1-z)^4} + \frac{2T(z)}{(z_1-z)^2} + \frac{\partial_z T(z)}{z_1-z} + \text{finite terms}. \tag{2.5.63}$$

Here, c is some constant, the so-called central charge. Since T is holomorphic, we can Laurent-expand it:

$$T(z) = \sum_{m=-\infty}^{\infty} \frac{L_m}{z^{m+2}}, \tag{2.5.64}$$

that is,

$$L_m = \oint_{C_0} \frac{dz}{2\pi i} z^{m+1} T(z) \tag{2.5.65}$$

for a circle C_0 about the origin $z = 0$.

The L_m are the generators of the Virasoro algebra

$$\begin{aligned}[L_n, L_m] &= \oint_{C_0} \frac{dz}{2\pi i} \oint_{C_z} \frac{dz_1}{2\pi i} z_1^{m+1} z^{m+1} \left[\frac{c}{2(z_1-z)^4} + \frac{2T(z)}{(z_1-z)^2} + \frac{\partial T(z)}{z_1-z} \right] \\ &= \frac{c}{12} n(n-1)(n+1)\delta_{m+n} + (n-m)L_{m+n}. \end{aligned} \tag{2.5.66}$$

To obtain this, one uses

$$\begin{aligned} z_1^{n+1} &= ((z_1-z)+z)^{n+1} \\ &= \frac{n^3-n}{6}(z_1-z)^3 z^{n-2} + \frac{n^2+n}{2}(z_1-z)^2 z^{n-1} \\ &\quad + (n+1)(z_1-z)z^n + z^{n+1} + \cdots . \end{aligned}$$

Summary: The generators of the Virasoro algebra are the Laurent coefficients of the energy–momentum tensor T. The expansion comes from the holomorphicity of T, which in turn follows from the invariance properties of CFT. Since, in contrast to the classical action, the quantum expectation values are not conformally invariant, we obtain a central charge $c \neq 0$ in the commutators of the L_m.

L_0, L_1 and L_{-1} generate an algebra isomorphic to $\mathfrak{sl}(2,\mathbb{R})$, the Lie algebra of $Sl(2,\mathbb{R})$. That Lie algebra is represented here by infinitesimal transformations of the form $\alpha + \beta z + \gamma z^2 = \delta z$, the infinitesimal version at $a = d = 1, b = c = 0$ of $z \mapsto \frac{a\,z+b}{cz+d}$, the operation of $Sl(2,\mathbb{R})$. In fact, for $n, m = -1, 1, 0$, (2.5.66) is the same as (1.3.48), except for the different notation, of course. In general, L_n generates $\delta z = z^{n+1}$. L_n acts on a primary field (*primary* can be defined by this relation) as

$$[L_n, \varphi(z)] = z^n \left(z\partial_z + (n+1)h\right)\varphi(z). \tag{2.5.67}$$

There is one point here that needs clarification, the relationship between the classical energy–momentum tensor as defined in Sect. 2.4, see (2.4.8), and its operator version. According to (2.4.8), the energy–momentum tensor is the Noether current associated with a variation of the inverse metric $\gamma^{\alpha\beta}$:

$$\delta S = \int dx T_{\alpha\beta} \delta\gamma^{\alpha\beta}. \tag{2.5.68}$$

Quantum mechanically, we have

$$\begin{aligned} Z_{\gamma+\delta\gamma} &= \int (D\varphi)_{\gamma+\delta\gamma} \exp(-S(\varphi, \gamma + \delta\gamma)) \\ &= \int (D\varphi)_\gamma \left(1 + \int dx T \delta\gamma^{-1}\right) \exp(-S(\varphi, \gamma)), \end{aligned}$$

assuming that the energy–momentum tensor incorporates both the variation of the action and of the measure,

$$= Z_\gamma + Z_\gamma \int dx \delta\gamma^{-1} \langle T \rangle_\gamma.$$

Thus,

$$\frac{1}{Z_\gamma} \delta Z_\gamma = \int dx \delta\gamma^{-1} \langle T \rangle_\gamma, \tag{2.5.69}$$

or, putting in a factor of 4π for purposes of normalization,

$$\frac{1}{Z_\gamma} \frac{\delta}{\delta\gamma^{-1}(y)} Z_\gamma = \frac{1}{4\pi} \langle T(y) \rangle_\gamma, \tag{2.5.70}$$

and more generally,

$$\begin{aligned} &\frac{1}{Z_\gamma} \frac{(4\pi)^m \delta^m}{\delta\gamma^{-1}(y_1) \cdots \delta\gamma^{-1}(y_m)} (Z_\gamma \langle \varphi(x_1) \cdots \varphi(x_n) \rangle) \\ &\quad = \langle T(y_1) \cdots T(y_m) \varphi(x_1) \cdots \varphi(x_n) \rangle. \end{aligned} \tag{2.5.71}$$

The variation of the measure then induces the central charge c in the expansion (2.5.63) of the operator version of the energy–momentum tensor.

2.6 Conformal Field Theory

2.6.1 Axioms and the Energy–Momentum Tensor

Conformal field theory was introduced by several people. An early paper that was important for the subsequent development of the theory is [9]. A monograph devoted to this topic is [38]. We shall also utilize the treatments in [77] and [46].

In the preceding, we have derived certain formal consequences of the functional integral (2.5.5). In particular, the partition function and the correlation functions satisfy certain relations, and from those, we have obtained the energy–momentum tensor. Its classical version could be identified with a holomorphic quadratic differential in Sect. 2.4. One problem, however, was the definition of the functional integral (2.5.5). There, we briefly discussed the mathematical definition in terms of zeta functions, see (2.5.11), and the spectrum of the Laplace–Beltrami operator. One way to circumvent that problem is to take the indicated algebraic relations and holomorphicity properties as the starting point for an axiomatic theory. This is the idea of conformal field theory.

Thus, abstract conformal field theory specifies for each Riemann surface Σ with a metric g a partition function Z_g and correlation functions $\langle\varphi_1(x_1)\cdots\varphi_n(x_n)\rangle$ for the primary fields with non-coincident $x_1,\ldots,x_n$. These basic data do not need any action or functional integral—although (2.5.5) remains a prime example. The theory is defined in terms of symmetry properties of these correlation functions.

Essentially, these are:

(i) Diffeomorphism covariance: for a diffeomorphism $k:\Sigma\to\Sigma$,

$$Z_g = Z_{k^*g}, \tag{2.6.1}$$

$$\langle\varphi_1(k(x_1))\cdots\varphi_n(k(x_n))\rangle_g = \langle\varphi_1(x_1)\cdots\varphi_n(x_n)\rangle_{k^*g}. \tag{2.6.2}$$

(ii) Local conformal covariance

$$Z_{e^\sigma g} = \exp\left(\frac{c}{96\pi}\left(\|d\sigma\|^2_{L^2_g} + 4\int_\Sigma \sigma(x)R(x)\right)\right)Z_g, \tag{2.6.3}$$

$$\langle\varphi_1(x_1)\cdots\varphi_n(x_n)\rangle_{e^\sigma g} = \prod_{i=1}^n \exp(-h_i\sigma(x_i))\langle\varphi_1(x_1)\cdots\varphi_n(x_n)\rangle_g. \tag{2.6.4}$$

Here, $R(x)$ is the scalar curvature of (Σ,g), and h_i is the conformal weight (see below) of the field φ_i, as introduced in Sect. 1.1.2; c is called the central charge of the theory. (For the conformal field theory defined by (2.5.5), we have $c=1$.)

In particular, and this is the fundamental point, the quantum mechanical partition function is not conformally invariant, but instead transforms with a certain factor that depends on the central charge.

We return to the formula (2.5.9) for the functional (2.5.14) on a Riemann surface for $m=0$. Since $G = 2\pi\alpha'(-\Delta)^{-1}$, we should have, up to a factor,

$$\det G = (\det\Delta)^{-1}.$$

Since, however, Δ has the eigenvalue 0 ($\Delta\varphi_0 = 0$ for a constant function φ_0), we need to restrict it to the orthogonal complement of the kernel of Δ, that is, to the L^2-functions φ with $\int_M \varphi dvol_g(M) = 0$, when defining the determinant by ζ-function regularization. The corresponding determinant is denoted by $\det'$. In fact, one should also normalize it by the volume (area) of M.

Now, however, while the action is conformally invariant, the Laplace operator (1.1.103), (2.4.4)

$$\Delta = \frac{1}{\sqrt{g}} \frac{\partial^2}{\partial z \partial \bar{z}}$$

and therefore also its eigenvalues depend on the metric g and not only on its conformal class. When we consider a variation $g(x) \mapsto e^{\sigma(x)} g(x)$ of the metric, we can compute

$$\frac{\delta}{\delta \sigma(x)} \log\left(\frac{\det'(-\Delta)}{\mathrm{Vol}(M)} \right)\Big|_{\sigma=0} = -\frac{1}{12\pi} R(x), \tag{2.6.5}$$

where $R(x)$ is the scalar curvature of g, see e.g. [38], p. 145ff.

Denoting the partition function for the metric g by Z_g, we then have

$$\frac{\delta}{\delta \sigma(x)} Z_{e^\sigma g}|_{\sigma=0} = \frac{c}{24\pi} R(x) Z_g. \tag{2.6.6}$$

More generally, one defines the energy–momentum tensor as an operator by (2.5.71), that is,

$$\begin{aligned} &\langle T_{\alpha_1\beta_1}(z_1, \bar{z}_1) \cdots T_{\alpha_m\beta_m}(z_m, \bar{z}_m) \varphi(w_1, \bar{w}_1) \cdots \varphi(w_n, \bar{w}_n) \rangle_g \\ &= \frac{1}{Z_g} \frac{(4\pi)^m \delta^m}{\delta g^{\alpha_1\beta_1}(z_1, \bar{z}_1) \cdots \delta g^{\alpha_m\beta_m}(z_m, \bar{z}_m)} \big(Z_g \langle \varphi(w_1, \bar{w}_n) \cdots \varphi(w_n, \bar{w}_n) \rangle_g \big). \end{aligned} \tag{2.6.7}$$

Thus, as an operator, the energy–momentum tensor takes into account the variation of the action S and of the integration measure $D\varphi$, as at the end of Sect. 2.5.3.

In particular

$$\langle T_{\alpha\beta}(z, \bar{z}) \rangle = \frac{4\pi}{Z_g} \frac{\delta}{\delta g^{\alpha\beta}(z, \bar{z})} Z_g. \tag{2.6.8}$$

At a conformal metric $g = \rho^2 |dz|^2$, that is, $g^{z\bar{z}} = 2\rho^{-2}$, $g^{zz} = 0 = g^{\bar{z}\bar{z}}$, we consider the above variation $g \mapsto e^\sigma g$ and obtain

$$\frac{4\pi}{Z_g} \frac{\delta}{\delta \sigma} Z_{e^\sigma g}|_{\sigma=0} = -g^{zz} \langle T_{zz} \rangle_g - 2 g^{z\bar{z}} \langle T_{z\bar{z}} \rangle_g - g^{\bar{z}\bar{z}} \langle T_{\bar{z}\bar{z}} \rangle_g = -4\rho^{-2} \langle T_{z\bar{z}} \rangle_g. \tag{2.6.9}$$

From (2.6.8), (2.6.6), we then conclude

$$4\rho^{-2} \langle T_{z\bar{z}} \rangle_g = -\frac{c}{6} R. \tag{2.6.10}$$

Since the Euclidean metric ($g^{z\bar{z}} = 2$, $g^{zz} = 0 = g^{\bar{z}\bar{z}}$) has vanishing scalar curvature, we have there that

$$\langle T_{z\bar{z}} \rangle = 0, \tag{2.6.11}$$

that is, the energy–momentum tensor is traceless when the metric is Euclidean. For nonvanishing curvature R, however, T is no longer traceless. The trace given by (2.6.10) involves both the curvature and the central charge c.

From Axiom (ii), we obtain

$$\begin{aligned}\langle T_{zz}\rangle_{e^\sigma g} &= \langle T_{zz}\rangle_g + \frac{c}{24}\frac{\delta}{\delta g^{zz}}\left(\|d\sigma\|^2_{L^2_g} + 4\int \sigma R\right)\\ &= \langle T_{zz}\rangle_g - \frac{c}{12}\left(\partial_z^2\sigma - \frac{1}{2}(\partial_z\sigma)^2\right),\end{aligned} \tag{2.6.12}$$

using, for the last step, (2.4.6) and the formula

$$R = -\frac{1}{2}\left(\frac{\partial^2 g^{zz}}{\partial z^2} + \frac{\partial^2 g^{\bar z\bar z}}{\partial \bar z^2}\right) + \text{higher-order terms in } g^{zz}, g^{\bar z\bar z}, \tag{2.6.13}$$

which is valid when we vary the Euclidean metric, that is, when we have $g^{z\bar z} = 2$ (see (1.1.148)). From (2.6.1) and (2.6.12), under a holomorphic transformation $z \mapsto w = f(z)$,

$$\begin{aligned}(f'(z))^2\langle T_{ww}\rangle_{dw\,d\bar w} &= \langle T_{zz}\rangle_{|f'(z)|^2 dz\,d\bar z}\\ &= \langle T_{zz}\rangle - \frac{c}{12}\left(\frac{\partial^2}{\partial z^2}\log f'(z) - \frac{1}{2}\left(\frac{\partial}{\partial z}\log f'(z)\right)^2\right)\\ &= \langle T_{zz}\rangle - \frac{c}{12}\left(\frac{f'''(z)}{f'(z)} - \frac{3}{2}\left(\frac{f''(z)}{f'(z)}\right)^2\right)\\ &= \langle T_{zz}\rangle - \frac{c}{12}\{f; z\},\end{aligned} \tag{2.6.14}$$

where $\{f; z\}$ is the so-called Schwarzian derivative of f. So, we see here an important difference between the classical and the quantum energy–momentum tensor. While the latter is trace-free (2.6.11) for the Euclidean metric (but not in general) and holomorphic (2.6.17) (below) like the former, it no longer transforms as a quadratic differential, but instead picks up an additional term in its transformation rule (2.6.14). That term depends on the central charge c of the theory.

In order to take also variations w.r.t. g^{zz}, we now reconsider (2.6.9), (2.6.10) as

$$-g^{zz}\langle T_{zz}\rangle_g - 2g^{z\bar z}\langle T_{z\bar z}\rangle_g - g^{\bar z\bar z}\langle T_{\bar z\bar z}\rangle_g = \frac{c}{6}R. \tag{2.6.15}$$

Next, applying $\frac{4\pi}{Z_g}\frac{\delta}{\delta g^{z_1 z_1}}Z_g$ to (2.6.15) and recalling that the background metric is flat, that is, $g^{z\bar z} = 2$, $g^{zz} = 0 = g^{\bar z\bar z}$, as well as (2.6.13), and using (2.1.54), we obtain

$$4\pi\delta(z - z_1)\langle T_{zz}\rangle + 4\pi\langle T_{z_1 z_1} T_{z\bar z}\rangle = \frac{\pi c}{3}\partial_z^2\delta(z - z_1). \tag{2.6.16}$$

Next, diffeomorphism invariance implies that $\langle T_{zz}\rangle$ is holomorphic, as in the classical case,

$$\partial_{\bar z}\langle T_{zz}\rangle = 0 = \partial_z\langle T_{\bar z\bar z}\rangle. \tag{2.6.17}$$

Finally, one has the OPE

$$\begin{aligned}\langle T_{zz}T_{z_1z_1}\rangle = {} & \frac{c}{2(z-z_1)^4} + \frac{2}{(z-z_1)^2}\langle T_{z_1z_1}\rangle \\ & + \frac{1}{z-z_1}\partial_{z_1}\langle T_{z_1z_1}\rangle + \text{analytic terms in } z.\end{aligned} \tag{2.6.18}$$

We shall now explain this in more detail.

2.6.2 *Operator Product Expansions and the Virasoro Algebra*

We take up the discussion of Sect. 2.5.3. As before in (2.5.58), we consider $z \mapsto z + \varepsilon f(z)$, f holomorphic.

We apply the general Ward identity (2.5.35) for $j = fT$, T being the energy–momentum tensor in CFT, writing $T(z)$ for T_{zz},

$$\frac{\partial}{\partial\bar z}\langle fT(z)\varphi(z_1)\cdots\varphi(z_n)\rangle = \sum_{k=1}^{n}\langle\varphi(z_1)\cdots\delta\varphi(z_k)\cdots\varphi(z_n)\rangle\delta(z-z_k) \tag{2.6.19}$$

to primary fields with variation (see (2.5.60))

$$\delta\varphi = h\partial_z f\varphi + f\partial_z\varphi. \tag{2.6.20}$$

We also write

$$\delta(z-z_k) = -\frac{1}{\pi}\partial_{\bar z}\left(\frac{1}{z-z_k}\right)$$

and integrate $\partial_z f\delta(z-z_k)$ by parts to obtain, using that (2.6.19) holds for all (holomorphic) f, and neglecting the factor π,

$$\frac{\partial}{\partial\bar z}\langle T(z)\varphi(z_1)\cdots\varphi(z_n)\rangle - \sum_{k=1}^{n}\left(\frac{1}{z-z_k}\partial_{z_k} + \frac{h}{(z-z_k)^2}\right)\langle\varphi(z_1)\cdots\varphi(z_n)\rangle = 0. \tag{2.6.21}$$

Under a holomorphic field f, T has to transform as

$$\delta_f T(z) = f(z)\partial_z T + 2(\partial_z f)T + \frac{c}{12}\partial_z^3 f, \tag{2.6.22}$$

because f transforms like $\frac{\partial}{\partial z}$, and T transforms like $(dz)^2$. As always, c is the central charge.

When we want to use $\varphi(z_1) = T(z_1)$ in the preceding, we therefore have to replace (2.6.20) by (2.6.22) and obtain (2.5.63), that is,

$$\langle T(z)T(z_1)\rangle = \frac{1}{z-z_1}\partial_{z_1}\langle T(z_1)\rangle + \frac{2}{(z-z_1)^2}\langle T(z_1)\rangle + \frac{c}{2}\frac{1}{(z-z_1)^4} + \text{analytic terms.} \tag{2.6.23}$$

This is (2.6.18).

We also recall (2.5.57), saying that the transformation $z \mapsto z + f(z)$ is generated by

$$Q_f = \oint_C \frac{dz}{2\pi i} f(z)T(z).$$

Therefore, in particular,

$$[Q_f, T(w)] = f\partial_w T + 2(\partial_w f)T + \frac{c}{12}\partial_w^3 f.$$

As above, the commutator means that

$$\langle [Q_f, T(w)]\varphi(z_1)\cdots\varphi(z_n)\rangle = \left(\oint_{C_1}\frac{dz}{2\pi i} - \oint_{C_2}\frac{dz}{2\pi i}\right) f(z)\langle T(z)T(w)\varphi(z_1)\cdots\varphi(z_n)\rangle,$$

where z lies inside C_1, but outside of C_2, while the z_k all lie inside C_2.

Integrating this with some function f_2 around the loop C_2 then leads to

$$[Q_{f_1}, Q_{f_2}] = Q_{[f_1,f_2]} + \frac{c}{24}\oint_C \frac{dz}{2\pi i}\big((\partial_z^3 f_1)f_2 - f_1\partial_z^3 f_2\big).$$

This then gives us the Virasoro algebra (2.5.66).

2.6.3 *Superfields*

We recall the basic transformation rules for a family of super Riemann surfaces from Sect. 1.5.3:

$$\begin{aligned} &\tilde{z} = f(z) + \theta k(z), \\ &\tilde{\vartheta} = g(z) + \theta h(z), \\ &f, g, k, h \text{ holomorphic}, \qquad \frac{\partial f}{\partial z} \neq 0. \end{aligned} \tag{2.6.24}$$

We define

$$D_+ := \partial_\theta + \theta\partial_z, \qquad D_+^2 = \partial_z. \tag{2.6.25}$$

(D_+ had been called τ in Sect. 1.5.3, and later on, we shall sometimes write θ_+ in place of θ, and θ_- in place of $\bar{\theta}$.)

The transformation law under holomorphic coordinate changes is

$$D_+ = (D_+\tilde{\theta})\widetilde{D_+} + (D_+\tilde{z} - \tilde{\theta}D_+\tilde{\theta})\widetilde{D_+}^2. \tag{2.6.26}$$

Superconformal means homogeneous transformation law, i.e.,

$$D_+\tilde{z} = \tilde{\theta}D_+\tilde{\theta}. \tag{2.6.27}$$

This is equivalent to

$$\begin{aligned} \tilde{z} &= f(z) + \theta g(z)h(z), \\ \tilde{\theta} &= g(z) + \theta h(z) \end{aligned} \tag{2.6.28}$$

with

$$h^2(z) = \frac{\partial f}{\partial z} + g(z)\frac{\partial g}{\partial z} \quad (g \text{ anticommuting}). \tag{2.6.29}$$

Since $D_+^2 = \partial_z$, (2.6.27) yields

$$\partial_z\tilde{z} + \tilde{\theta}\partial_z\tilde{\theta} = (D_+\tilde{\theta})^2 \tag{2.6.30}$$

as a compact version of the superconformal coordinate transformation rule.

In global terms, θ is a section of $K^{\frac{1}{2}}$, a square root of the canonical bundle of the underlying Riemann surface Σ. Such a square root of K corresponds to the choice of a spin structure on Σ. (To see this transformation behavior, put for example $g = 0$ in (2.6.28). Then from (2.6.29), $\tilde{\theta} = \sqrt{\frac{\partial f}{\partial z}}\theta$.)

We now look at the transformation behavior of conformal (primary) superfields

$$X(z,\theta) = \varphi(z) + \theta\psi(z)$$

of conformal weight h. Since θ as a section of $K^{\frac{1}{2}}$ has conformal weight $\frac{1}{2}$, this means that ψ has weight $h - \frac{1}{2}$, while φ has weight h. According to (2.6.30), we can also express the transformation law as

$$X(z,\theta) = X(\tilde{z},\tilde{\theta})(D_+\tilde{\theta})^{2h} \tag{2.6.31}$$

(since $\tilde{\theta}$ has weight $\frac{1}{2}$ and D_+ has weight $-\frac{1}{2}$, $D_+\tilde{\theta}$ has weight 0, which it should, to make the transformation law consistent). Similarly, (2.5.60) becomes

$$\begin{aligned} \delta_f\varphi(z) &= (h\partial_z f + f\partial_z)\varphi(z), \\ \delta_f\psi(z) &= \left(\left(h - \frac{1}{2}\right)\partial_z f + f\partial_z\right)\psi(z). \end{aligned} \tag{2.6.32}$$

We also have the supersymmetry transformations for an anticommuting holomorphic g,

$$\delta_g \varphi(z) = \frac{1}{2} g\psi, \tag{2.6.33}$$

$$\delta_g \psi(z) = \frac{1}{2} g\partial_z \varphi + h\partial_z g\varphi, \tag{2.6.34}$$

that is,

$$\delta_g X(z,\theta) = \left(\frac{1}{2} g D_+ + h\partial_z g\right) X. \tag{2.6.35}$$

As before, we write this as a commutator with a charge

$$\delta_g X = -[Q_g, X] = \oint_c \frac{dz_1}{2\pi i} g(z_1) T(z_1) X(z) \tag{2.6.36}$$

for a small circle c about z.

Here, T is the (anticommuting) generator of the superconformal algebra. From this, we can draw the same consequences as above. We observe that for two supersymmetry transformations generated by g_1, g_2, if we put

$$f := \frac{1}{2} g_1 g_2, \tag{2.6.37}$$

we have

$$[\delta_{g_1}, \delta_{g_2}]_+ X(z,\theta) = \delta_f X(z,\theta), \tag{2.6.38}$$

where $+$ denotes the anticommutator. Thus, **a supersymmetry transformation is a square root of a conformal transformation**, as it should be according to (2.6.30).

As in Sect. 2.4.3, with $\psi_+ = \psi_1 - i\psi_2$, $\psi_- = \psi_1 + i\psi_2$ and $\theta_+ = \theta_1 + i\theta_2$, $\theta_- = \theta_1 - i\theta_2$ (alternatively, if we wished to conform to the notation in (2.6.24), we could write $\theta, \bar{\theta}$ in place of θ_+, θ_-), we use the operators $D_+ = \partial_{\theta_+} + \theta_+ \partial_z$, $D_- = \partial_{\theta_-} + \theta_- \partial_{\bar{z}}$ and consider a superfield

$$X = \phi + \frac{1}{2}(\psi_+\theta_+ + \psi_-\theta_-) + \frac{i}{2} F\theta_+\theta_- \tag{2.6.39}$$

and obtain the action

$$\begin{aligned} S &= \int \frac{1}{2} D_- X D_+ X d^2x d\theta_- d\theta_+ \\ &= \int \frac{1}{2}\left(4\partial_z\phi\partial_{\bar{z}}\phi - \psi_+ \frac{\partial}{\partial\bar{z}}\psi_+ - \psi_- \frac{\partial}{\partial z}\psi_- + F^2\right) d^2z. \end{aligned} \tag{2.6.40}$$

In Sect. 2.4.3, we derived the equations of motion (2.4.62),

$$D_- D_+ X = 0. \tag{2.6.41}$$

A solution can be decomposed as

$$X(z,\theta_+,\bar{z},\theta_-) = X(z,\theta_+) + X(\bar{z},\theta_-), \tag{2.6.42}$$

and we may write

$$X(z,\theta_+) = \varphi(z) + \theta_+\psi_+(z). \tag{2.6.43}$$

The action is invariant under superconformal transformations and the corresponding energy–momentum tensor is

$$T = -\frac{1}{2}D_+X\partial_z X = T_F + \theta_+ T_B, \tag{2.6.44}$$

with

$$T_F = -\frac{1}{2}\psi\partial_z\varphi, \tag{2.6.45}$$

$$T_B = -\frac{1}{2}(\partial_z\varphi)^2 - \frac{1}{2}\partial_z\psi\cdot\psi. \tag{2.6.46}$$

T_B is a section of K^2, T_F one of $K^{\frac{3}{2}}$.

We consider a complex Weyl spinor ψ_+ on a Riemann surface Σ, that is, a section of a spin bundle $K^{\frac{1}{2}}$, a square root of the canonical bundle K, given by a spin structure on Σ. We let ψ_- be the complex conjugate of ψ_+. Thus, ψ_- is a section of $\bar{K}^{\frac{1}{2}}$ (for the same spin structure).

We now consider the case where Σ is a cylinder, with coordinates $w = \tau + i\sigma$, identifying $\sigma + 2\pi$ with σ and with τ in some interval which is not further specified here. As there are two different spin structures on a cylinder, we have two choices for identifying ψ at $\sigma + 2\pi$ with ψ at σ:

$$\begin{aligned}&\psi_\pm(\tau,\sigma+2\pi) = \psi_\pm(\tau,\sigma), \quad \text{periodic (Ramond)}, \quad \text{or}\\ &\psi_\pm(\tau,\sigma+2\pi) = -\psi_\pm(\tau,\sigma), \quad \text{antiperiodic (Neveu–Schwarz)}.\end{aligned} \tag{2.6.47}$$

These boundary conditions also arise from the following consideration. We consider the half cylinder where σ runs from 0 to π, and we assume boundary relations between the holomorphic field ψ_+ and the antiholomorphic field ψ_-,

$$\begin{aligned}&\psi_+(0,\tau) = \nu\psi_-(0,\tau) \quad \text{with } \nu = \pm 1,\\ &\psi_+(\pi,\tau) = \psi_-(\pi,\tau)\end{aligned} \tag{2.6.48}$$

where the factor $+1$ has been chosen w.l.o.g. in the second equation. We can then combine ψ_+ and psi_- into a single field, defined for $\sigma \in [0, 2\pi]$, by putting

$$\psi_+(\sigma,\tau) = \psi_-(2\pi - \sigma,\tau) \quad \text{for } \pi \le \sigma \le 2\pi. \tag{2.6.49}$$

ψ_+ then is holomorphic, because ψ_- was antiholomorphic. Also,

$$\psi_+(2\pi,\tau)=\psi_-(0,\tau)=\begin{cases}\psi_+(0,\tau) & \text{for } \nu=1,\\ -\psi_+(0,\tau) & \text{for } \nu=-1.\end{cases} \tag{2.6.50}$$

Thus, ψ_+ is periodic (Ramond) in the first and antiperiodic (Neveu–Schwarz) in the second case.

We now map the cylinder to an annulus via

$$z=e^w.$$

Since ψ_+ transforms like $(dw)^{\frac{1}{2}}$, we have

$$\psi_+^{\text{annulus}}(z)(dz)^{\frac{1}{2}}=\psi_+^{\text{cylinder}}(w)(dw)^{\frac{1}{2}},$$

with

$$\left(\frac{dz}{dw}\right)^{\frac{1}{2}}=e^{\frac{w}{2}}.$$

When we now rotate the cylinder by 2π, the factor $e^{\frac{w}{2}}$ changes by a factor -1. Therefore, periodic and antiperiodic identifications are exchanged, and on the annulus, we have

$$\begin{aligned}\text{Ramond:}\quad & \psi_\pm(e^{2\pi i}z)=-\psi_\pm(z) \quad \text{(antiperiodic)},\\ \text{Neveu–Schwarz:}\quad & \psi_\pm(e^{2\pi i}z)=\psi_\pm(z) \quad \text{(periodic)}.\end{aligned}$$

We shall now expand these expressions in terms of

$$\begin{aligned}z_{12}&=z_1-z_2-\theta_1\theta_2,\\ \theta_{12}&=\theta_1-\theta_2.\end{aligned}$$

We obtain

$$\begin{aligned}T(z_1,\theta_1)X(z_2,\theta_2)&=h\frac{\theta_{12}}{z_{12}^2}X(z_2,\theta_2)+\frac{1}{2z_{12}}D_{+,2}X(z_2,\theta_2)\\ &\quad+\frac{\theta_{12}}{z_{12}}\partial_{z_2}X(z_2,\theta_2)+\text{regular terms},\\ T(z_1,\theta_1)T(z_2,\theta_2)&=\frac{c}{6}\frac{1}{z_{12}^3}+\frac{3}{2}\frac{\theta_{12}}{z_{12}^2}T(z_2,\theta_2)+\frac{1}{2z_{12}}D_{+,2}T(z_2,\theta_2)\\ &\quad+\frac{\theta_{12}}{z_{12}}\partial_{z_2}T(z_2,\theta_2)+\text{regular terms}.\end{aligned}$$

In components:

$$T_B(z_1)T_B(z_2)=\frac{c}{6}\frac{1}{(z_1-z_2)^4}+\frac{2}{(z_1-z_2)^2}T_B(z_2)+\frac{1}{z_1-z_2}\partial_{z_2}T_B(z_2)+\cdots,$$

$$T_B(z_1)T_F(z_2) = \frac{3}{2}\frac{1}{(z_1-z_2)^2}T_F(z_2) + \frac{1}{z_1-z_2}\partial_{z_2}T_F(z_2) + \cdots,$$
$$T_F(z_1)T_F(z_2) = \frac{c}{6}\frac{1}{(z_1-z_2)^3} + \frac{1}{2}\frac{1}{z_1-z_2}T_B(z_2) + \cdots.$$

We expand T_B as before and T_F as

$$T_F(z) = \frac{1}{2}\sum_{k\in\mathbb{Z}+a} z^{-k-1-a}G_k \quad \left(G_k = 2\oint_c \frac{dz_i}{2\pi i}T_F(z)z^{k+a}\right),$$

with $a = 0$ corresponding to the Ramond sector and $a = \frac{1}{2}$ corresponding to the Neveu–Schwarz sector.

With $\hat{c} = \frac{2}{3}c$, we obtain the super Virasoro algebra

$$[L_m, L_n]_- = (m-n)L_{m+n} + \frac{\hat{c}}{8}(m^3-m)\delta_{m+n},$$
$$[L_m, G_k]_- = \left(\frac{1}{2}m - k\right)G_{m+k},$$
$$[G_k, G_l]_+ = 2L_{k+l} + \frac{\hat{c}}{2}\left(k^2 - \frac{1}{4}\right)\delta_{k+l}.$$

2.7 String Theory

In conformal field theory, Sect. 2.6, we have kept the Riemann surface Σ fixed and varied the metric on Σ only via diffeomorphisms—which left the partition and correlation functions invariant—and by conformal changes—which, in contrast to the classical case, had a nontrivial effect, the so-called conformal anomaly. In string theory, one also varies the Riemann surface Σ itself. Equivalently, as explained in 7 in Sect. 1.4.2, we permit any variation of the metric γ, including those that change the underlying conformal structure. Here, we can only give some glimpses of the theory. Fuller treatments are given in [50, 77, 87, 88] and, closest to the presentation here, in [62].

In bosonic string theory, one starts with the linear sigma model (Polyakov action) (2.4.7)

$$S(\varphi, \gamma) \tag{2.7.1}$$

and considers the functional integral

$$Z = \sum_{\text{topological types}} \int e^{-S(\varphi,\gamma)} d\varphi d\gamma. \tag{2.7.2}$$

This means that one wishes to average over all fields ϕ and all compact[20] surfaces, described by their topological type (their genus) and their metric, with exponential weight coming from the Polyakov action. Since, as discussed, that action $S(\varphi, \gamma)$ is invariant under diffeomorphisms and conformal changes, that is, possesses an infinite-dimensional invariance group, this functional integral, as it stands, can only be infinite itself. Therefore, one divides out these invariances before performing the functional integral. As described in Sect. 1.4.2, the remaining degrees of freedom are the ones coming from the moduli of the underlying surface, and we are left with an integral over the Riemann moduli space for surfaces of given genus and a sum over all genera. The essential mathematical content of string theory is then to define that integral in precise mathematical terms and try to evaluate it. The sum needs some regularization, that is, one should put in some factor κ_p depending on the genus p that goes to 0 in some appropriate manner as the genus increases. Alternatively, one should construct a common moduli space that simultaneously includes surfaces of all genera. Since lower-genus surfaces occur in the compactification of the moduli spaces of higher-genus ones, this seems reasonable. As discussed above in Sect. 1.4.2, however, the Mumford–Deligne compactification is not directly appropriate for this, as there the lower-genus surfaces that occur in the boundary of the moduli space carry marked points in addition. With each reduction of the genus, the number of those marked points increases by two. When we then consider surfaces of some fixed genus p_0 in a boundary stratum of the moduli space of surfaces of genus p, we have $2(p - p_0)$ marked points, and this number then tends to ∞ for $p \to \infty$. Therefore, we need to resort to the Satake–Baily compactification described in Sect. 1.4.2 which does not need marked points, but is highly singular. We also recall from there that this compactification can be mapped into the Satake compactification of the moduli space of principally polarized Abelian varieties. Again, the compactification of that moduli space for principally polarized Abelian varieties of dimension p contains in its boundary the moduli spaces for the Abelian varieties of smaller dimension. Letting $p \to \infty$ then gives some kind of universal moduli space for principally polarized Abelian varieties of finite dimension, and this space is then stratified according to dimension. Similarly, the analogous universal moduli space for compact Riemann surfaces would then be stratified according to genus. (To the author's knowledge, however, this construction has never been carried through in detail.)

In any case, even the integral over the moduli space for a fixed genus leads to some subtleties. The reason is that while the Polyakov action $S(\varphi, \gamma)$ itself is conformally invariant, the measure $e^{-S(\varphi,\gamma)} d\phi d\gamma$ in (2.7.2) is not. We have seen the reason above from a somewhat different perspective in our discussion of quantization of the sigma model, where we encountered additional terms in the operator expansions. These then led to the nontrivial central charge c of the Virasoro algebra. It then turns out that there are two different sources of this conformal anomaly, one coming from the fields ϕ and the other from the metric γ. The fields are mappings

[20] Since the partition function represents the amplitude of vacuum $\to$ vacuum transitions, only closed surfaces are taken into account.

into some euclidean space $\mathbb{R}^d$, and we get a contribution to the conformal anomaly for each dimension, that is, an overall contribution proportional to d. The conformal anomaly coming from γ is independent of the target dimension d. It then turns out that these two conformal anomalies cancel precisely in dimension $d = 26$. Mathematically, this can be explained in terms of the geometry of the Riemann moduli space, utilizing earlier work of Mumford [84], or with the help of the semi-infinite cohomology of the Virasoro algebra. In conclusion, bosonic string theory lives in a 26-dimensional space.

The same scheme applies in superstring theory. Here, the action is given by (2.4.149),

$$S(\phi, \psi, \gamma, \chi) = \frac{1}{2}\int_\Sigma (\gamma^{\alpha\beta}\partial_\alpha\phi^a\partial_\beta\phi_a + \bar\psi^a\gamma^\alpha\partial_\alpha\psi_a + 2\bar\chi_\alpha\gamma^\beta\gamma^\alpha\psi^a\partial_\beta\phi_a \\ + \frac{1}{2}\bar\psi_a\psi^a\bar\chi_\alpha\gamma^\beta\gamma^\alpha\chi_\beta)\sqrt{\det\gamma}dz^1dz^2, \tag{2.7.3}$$

including also the fermionic field ψ and the gravitino χ. The same quantization principle is applied, and the resulting dimension needed to cancel the conformal anomalies turns out to be $d = 10$.

In order to include gravitational fields, one has to consider more general targets than euclidean space. The appropriate target spaces are Kähler manifolds with vanishing Ricci curvature. The real dimension still has to be 10. In order to make contact with dimension 4 of ordinary space–time, one writes such a target as a product

$$\mathbb{R}^4 \times M \tag{2.7.4}$$

where M now is assumed to be compact (and of such a small scale that it is not directly observable at the macroscopic level). (This vindicates the old idea of Kaluza described in Sect. 1.2.4 above.) The process of making some of the dimensions compact is called compactification in the physics literature. M then has to be a compact Kähler manifold with vanishing Ricci curvature, in order to obtain supersymmetry, of complex dimension 3, a Calabi–Yau space. In fact, by Yau's theorem [109], every compact Kähler manifold with vanishing first Chern class $c_1(M)$ carries such a Ricci flat metric, and this makes the methods of algebraic geometry available for the investigation and classification of such spaces.

In order to describe the physical content of string theory, the basic object is the string, an open or closed curve. As it moves in space–time, it sweeps out a Riemann surface. In contrast to the mathematical framework just described, this Riemann surface will have boundaries, even in the case of a closed string when we follow it between two different times t_1 and t_2. The boundaries will then correspond to the initial position at time t_1 and the final position at time t_2, except when the string only comes into existence after time t_1 and ceases to exist at time t_2. See [62] for the systematic treatment of such boundaries in string theory. For an open string, that is, for a curve with two endpoints moving in space–time, we obtain further boundaries corresponding to the trajectories of these endpoints. More generally, the movement of these endpoints may be confined to lower-dimensional objects in space–time that

carry charges and that can then become objects in their own right, the D-branes[21] first introduced by Polchinski [86]. Symmetries between branes then led to a new relation between string theory and gauge theory, culminating in a conjecture of Maldacena [78].

In any case, when a string moves in space–time, it sweeps out a surface, and the basic Nambu–Goto action of string theory was the area of that surface. Since the area functional is invariant under any reparametrization, it cannot be readily quantized, and therefore, the symmetry was reduced by considering the map that embeds the surface representing the moving string into space–time and the underlying metric of that surface as independent variables of the theory. That led to the Polyakov action (2.7.1), that is, the Dirichlet integral or sigma model action (2.4.7).

According to string theory, all elementary particles are given by vibrations of strings. Gauge fields arise from vibrations of open strings. Their endpoints represent charged particles. For instance, when one is an electron and the other an oppositely charged particle, a positron, the massless vibration of the string connecting them represents a photon that carries the electrical force between them. Collisions between such particles then naturally lead to closed strings. Gravitons, that is, particles responsible for the effects of gravity, arise from vibrations of closed strings. In superstring theory, both bosons and fermions are oscillations of strings. There are only two fundamental constants in string theory, in contrast to the proliferation of such constants in the standard model. These are the string tension, that is, the energy per unit-length of a string, the latter given in terms of the Planck length, and the string coupling constant, the probability for a string to break up into two pieces.

However, superstring theory is far from being unique, and it cannot determine the geometry of the background space–time purely on the basis of physical principles. Thus, there is room for further work in superstring theory, as well as for research on competing theories like loop quantum gravity (that started with Ashtekar's reformulation of Einstein's theory of general relativity [5]) and the development of new ones.

[21] The "D" here stands for Dirichlet, because such types of boundary conditions are called Dirichlet boundary conditions in the mathematical literature. We also recall that the basic action functional (2.4.7), (2.7.1) is called the Dirichlet integral in the mathematical literature. This terminology was in fact introduced by Riemann when he systematically used variational principles in his theory of Riemann surfaces, see [91]. Harmonic functions are minimizers of the Dirichlet integral, and in this sense, string theory is a quantization of the profound ideas of Riemann.