

CHAPTER 1

Introduction and Motivation

1.1. Structure of Physical Theories

This section is intended to introduce certain key concepts from physics which we will use later in the book. The treatment will be rather superficial, and is intended for mathematicians without much background in physics. Any physicists reading this book can probably skip ahead to Section 1.2.

1.1.1. Fields and Lagrangians.

1.1.1.1. *Fields.* Since the time of Maxwell and Faraday in the 1860's, most physical theories have been expressed in terms of *fields*¹, e.g., the *gravitational field*, *electric field*, *magnetic field*, etc. Originally, “field” meant (to quote the *Oxford English Dictionary*) “a state or situation in which a force is exerted on any objects of a particular kind (e.g., electric charges) that are present,” but came to be “frequently used as if it denoted an identifiable causal entity.” Fields can be

- scalar-valued functions (physicists often call these *scalars*, which tends to confuse mathematicians because the word “function” is understood but not explicitly stated),
- sections of vector bundles (sometimes simply called *vectors*—again see the comment on the previous item),
- connections on principal bundles² (these are special cases of *gauge fields*), or
- sections of spinor bundles (often called *spinors*).

1.1.1.2. *Lagrangians and Least Action.* In classical physics, the fields satisfy a variational principle — they are extrema, or at least critical points, of the *action* S , which in turn is the integral of a local functional \mathcal{L} called the *Lagrangian*. This is called the *principle of least action*. The Euler-Lagrange equations for critical points of the action are the *equations of motion*.

The idea that physics should be governed by variational principles is very old, and predates the notion of “field” itself. For example, Fermat around 1662 [63, “Synthesis ad Refractiones,” pp. 173–179] proposed a theory of optics based on the

¹In English, the word *field* has two mathematical meanings. One (a commutative ring in which every non-zero element is invertible) is in abstract algebra and is not relevant here. That term is translated in French as *corps* or in German as *Körper*. What we mean here is field as in “vector field,” which is *champs* in French or *Feld* in German.

²A space E is called a principal G -bundle over another space X , for a topological group G , if there is a continuous free G -action on E with quotient space X and a continuous surjective open map $p: E \rightarrow X$ such that X has a covering by open sets U with each $p^{-1}(U)$ G -isomorphic to $U \times G$. A *connection* on a principal bundle, in the case where G is a Lie group and E and X are smooth manifolds, is a consistent way of choosing a “horizontal” subspace H_e of $T_e E$ at each $e \in E$, so that $dp: H_e \rightarrow T_{p(e)} X$ is an isomorphism for all $e \in E$.

principle that light travels along curves which minimize its travel time, which is a functional of the path. This seemingly innocuous assumption turns out to imply Snell’s law of refraction (when light passes from one medium to another, with different speeds in the two media), and also to imply that light does not necessarily travel along straight lines (if it is traveling through a medium with variable index of refraction, such as a fluid with varying density).

The principle of least action in mechanics is due to Lagrange (in his famous *Mécanique Analytique* of 1788) and to Hamilton, ca. 1835 [74]. But what we will need below is the least action principle applied to the theory of fields. The following two examples are the key ones for understanding 20th century physics. Yang-Mills theory (in a slightly more complicated form) is needed for understanding elementary particle physics; the Hilbert-Einstein action is basic to general relativity and thus to the understanding of gravity.

But before we get to the examples, we need to say something about Lorentz *vs.* Euclidean signatures in field theories. Indeed, a point which often confuses mathematicians trying to read the physics literature is a frequent shuttling back and forth between writing things in Lorentz and Euclidean signatures. The basic equations of physics do not treat space and time totally equally, in the sense that the natural metric on spacetime is a Lorentz metric, not a Riemannian one. However, in the Lorentz metric, most of the integrals one needs (such as the one computing the action) do not converge well, since one doesn’t have positivity for the Lagrangian.

Physicists are therefore fond of what’s called *Wick rotation*, replacing t by it and thus “analytically continuing” from Lorentz to Euclidean signature. This results in formulations which are better behaved mathematically but not as realistic physically. Still, one can often use this to some advantage, and we will sometimes do this without further ado.

EXAMPLES 1.1. Let M be a 4-manifold, say compact, representing spacetime.

- (1) Yang-Mills Theory. The field for this theory is a connection A on a principal G -bundle, where G is some compact Lie group. The “field strength” F is the curvature, a \mathfrak{g} -valued 2-form. (Here \mathfrak{g} is the Lie algebra of G .) The action is $S = \int_M \text{Tr } F \wedge *F$. (Here $F \wedge *F$ is a 4-form with values in \mathfrak{g} ; we take its pointwise trace and integrate.) If the bundle is non-trivial, then usually F cannot vanish, since Chern-Weil theory (Theorem 2.3 below) relates F to the characteristic classes of the bundle.
- (2) General Relativity (in Euclidean signature). In this theory, the field is a Riemannian metric g on M . The action is $S = \int_M R \, d\text{vol}$, where R is the scalar curvature of the metric. The associated field equation is Einstein’s equation.

1.1.2. Classical to Quantum Physics.

1.1.2.1. *Quantum Mechanics.* Unlike classical mechanics, quantum mechanics is not deterministic, only probabilistic. The key property of quantum mechanics is the *Heisenberg uncertainty principle*, that observable quantities are represented by *noncommuting operators* A represented on a Hilbert space \mathcal{H} . In the quantum world, every particle has a wave-like aspect to it, and is represented by a wave function ψ , a unit vector in \mathcal{H} . The phase of ψ is not directly observable, only its amplitude, or more precisely, the *state*

$$\varphi_\psi(A) = A(\psi) = \langle A\psi, \psi \rangle$$

which computes the expected (or expectation) value of an observable A for a particle with the given wave function. But the phase is still important, since phenomena such as the Aharonov-Bohm effect (in which the phase of an electron beam is changed without changing the amplitude) and *interference* depend on it. (One of the important consequences of quantum mechanics is that there is no clear distinction between waves and particles. Every particle behaves somewhat like a wave, and two particles can “interfere” with each other, just as waves can. This is responsible for diffraction of electron beams, etc.)

1.1.2.2. *Quantum Fields.* The quantization of classical field theories, called *quantum field theory*, is based on *path integrals*. The idea (not always 100% rigorous in this formulation) is that *all fields contribute*, not just those that are critical points of the action (i.e., solutions of the classical field equations). Instead, one looks at the *partition function*

$$(1.1) \quad Z = \int e^{iS(\varphi)} d\varphi \text{ or } \int e^{-S(\varphi)} d\varphi,$$

depending on whether one is working in Lorentz or Euclidean signature. (Here we’ve taken $\hbar = 1$; if we didn’t do this, since S has units of energy times time, we should divide S by $\hbar \approx 1.054 \times 10^{-34}$ J s so as to get something dimensionless that we can exponentiate.) The problem is that the integration is over all possible fields φ , which live in an infinite dimensional space, so one needs to make sense of the integral through some sort of regularization procedure. By the *principle of stationary phase*, only fields close to the classical solutions should contribute much.

What one wants out of a quantum field theory is more than just the partition function; it is a prediction for the measured values of certain observable quantities. Consider a physical quantity A that takes a scalar value $A(\varphi)$ when the fields of the system are given by φ . Since all quantum theories are probabilistic in nature, the best that we can hope for is to compute the *expectation value* of A , which would be the average measured value of this quantity if we perform an experiment a large number of times. This expectation value $\langle A \rangle$ is also given by a path integral, namely

$$\langle A \rangle = \left(\int A(\varphi) e^{iS(\varphi)} d\varphi \right) / Z.$$

We see from this that the partition function is the universal denominator that comes into the calculation of $\langle A \rangle$ for any observable A .

1.2. Some Basics of String Theory

1.2.1. **Basic Ideas of String Theory.** The basic idea of string theory is to replace point particles (in conventional physics) by one-dimensional “strings.” At ordinary (low) energies these strings are expected to be *extremely* short, on the order of the *Planck length*,

$$l_P = \sqrt{\frac{\hbar G}{c^3}} \approx 1.616 \times 10^{-35} \text{ m}.$$

A string moving in time traces out a two-dimensional surface called a *worldsheet*. The most basic fields in string theory are thus maps $\varphi: \Sigma \rightarrow X$, where Σ is a 2-manifold (the worldsheet) and X is spacetime.

String theory offers [some] hope for combining gravity with the other forces of physics and quantum mechanics.

1.2.1.1. *Strings and Sigma-Models.* Let Σ be a string worldsheet and X the spacetime manifold. String theory is based on the *nonlinear sigma-model*, where the fundamental field is $\varphi: \Sigma \rightarrow X$ and the leading term in the action is

$$(1.2) \quad S(\varphi) = \frac{1}{4\pi\alpha'} \int_{\Sigma} \|\nabla\varphi\|^2 \, d\text{vol},$$

the energy of the map φ (in Euclidean signature). (In the physics literature, (1.2) is called the string sigma model action or Polyakov action.) The constant α' , called the Regge slope parameter, represents (typical string length)², and $1/(2\pi\alpha')$ is the string tension. We have to add to this various gauge fields (giving rise to the fundamental particles) and a “gravity term” involving the scalar curvature of the metric on X . Usually we also require *supersymmetry*; this means the theory involves both bosons and fermions and there are symmetries interchanging the two.³ (But this is a subject for a different course, such as the ones given by Freed [64] or Varadarajan [166].)

1.2.1.2. *The B-Field and H-Flux.* For various reasons, it’s important to add to the action (1.2) another term (sometimes called the *Wess-Zumino term*) of the form

$$(1.3) \quad \frac{1}{4\pi\alpha'} \int_{\Sigma} \varphi^* B,$$

where B is a (locally defined) 2-form on spacetime, X . B is usually called the *B-field*. It need not be closed or even globally defined, just as long as it makes sense locally. (Recall the strings are really “small” in most cases.) But $H = dB$, a 3-form, should always be a well-defined closed 3-form on X , usually called the *H-flux*.

In fact, the 3-form H should correspond to an *integral* cohomology class, which we will also call (by abuse of language) the *H-flux*, even though this overlooks the torsion in $H^3(X)$. The reason for this is the following. Suppose $\varphi(\Sigma)$ is an embedded surface in X bounding two different 3-manifolds (think of solid handlebodies) M and M' . (See Figure 1.)

Since $\partial M = \partial M' = \varphi(\Sigma)$, Stokes’ Theorem gives

$$\int_M H = \int_M dB = \int_{\partial M} B = \int_{\Sigma} \varphi^* B,$$

and similarly with M' in place of M . So

$$\int_M H = \int_{M'} H, \quad \text{and} \quad \int_N H = 0,$$

where N is the closed 3-dimensional submanifold of X obtained by gluing M and M' together along their common boundary. (To make N oriented, reverse the

³For those unfamiliar with the terminology here, a *boson* is a particle like a photon that obeys Bose-Einstein statistics; a *fermion* is a particle like the electron that obeys Fermi-Dirac statistics, or in other words, that satisfies the Pauli exclusion principle. The fact that the universe is as we see it depends very much on the fact that both kinds of particles are present. Fermions are always represented by spinor fields. One way of explaining the difference between bosons and fermions is that bosons transform under true representations of the rotation group $SO(n)$, whereas fermions transform under “genuine” representations of the double covering group $\text{Spin}(n)$ that do not descend to $SO(n)$.

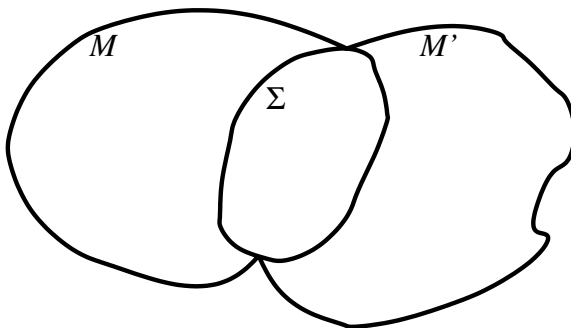


FIGURE 1. A Surface in Spacetime Bounding Two Handlebodies

orientation on M' , so that $N = M \cup_{\varphi(\Sigma)} -M'$.) Thus, for the sake of consistency, we need H to pair to zero against every closed 3-dimensional submanifold N of X .

This seems to contradict the possibility of non-triviality of the de Rham class of H , but we've neglected one important thing. By formula (1.1), it is only e^{iS} , not S itself, that counts. So if we only have

$$\frac{1}{4\pi\alpha'} \int_M H \equiv \frac{1}{4\pi\alpha'} \int_{M'} H \pmod{2\pi\mathbb{Z}},$$

or H integrating over each integral 3-homology class (like $[N]$) to something in $8\pi^2\alpha'\mathbb{Z}$, that's good enough for our purposes, and in physicists' language, the theory is *free of anomalies*.⁴

An interesting model for study, the *Wess-Zumino-Witten model* (WZW model) has $X = G$ a compact semisimple Lie group, say $SU(2)$, and H the canonical 3-form, coming from the trilinear pairing

$$(x, y, z) \mapsto \langle [x, y], z \rangle$$

on the Lie algebra \mathfrak{g} (here $\langle \cdot, \cdot \rangle$ is an invariant inner product on \mathfrak{g} , such as the one coming from the Killing form). This H is closed but not exact, so B cannot be globally defined. (In fact, $\pi_3(X) \cong \mathbb{Z}$ for any compact simple Lie group G , and H , if normalized correctly, has an integral de Rham class dual to the image of the generator under the Hurewicz map.)

1.2.1.3. *D-Branes*. Physicists talk about both *closed* and *open* strings. The terminology doesn't quite match that of mathematicians. Both kinds of strings are given by compact manifolds, but in the "open" case there is a boundary. So to get a reasonable theory one has to impose boundary conditions. Usually, these are Dirichlet or Neumann conditions on some submanifold of X where the boundary of Σ must map. These submanifolds are traditionally called *D-branes*, "D" for *Dirichlet* and *brane* a back-formation⁵ from *membrane*. Sometimes the name D-brane is retained even without Dirichlet boundary conditions. Physicists also talk

⁴In general, an *anomaly* is an inconsistency in a theory, usually due to nontriviality of some topological invariant. Thus physicists often look for topological conditions under which the anomalies will vanish.

⁵The *Oxford English Dictionary* defines this as "The formation of what looks like a root-word from an already existing word which might be (but is not) a derivative of the former." Here of course we are lopping off the *mem* in "membrane."

about Dp -branes or p -branes; that means branes with p space-like dimensions, or dimension $p + 1$ (since usually one also has to allow for one time-like dimension). That leads to the seemingly paradoxical study of (-1) -branes, which just means points in X . Such one-point branes are also called *instantons*, as they are localized at a single “instant” in spacetime. A schematic picture of a D-brane is shown in Figure 2. Note that the open string is free to slide up and down the D-brane but not normal to it.

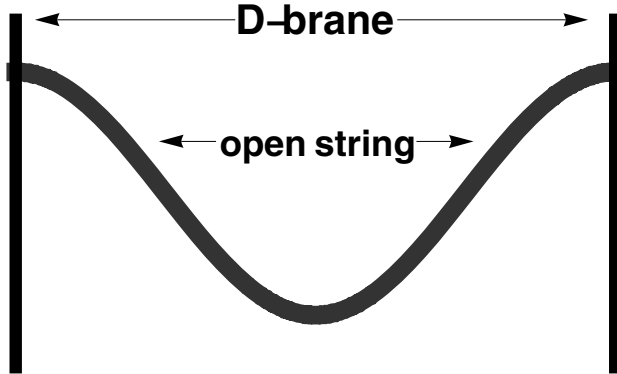


FIGURE 2. Schematic Picture of a D-Brane

In post-1995 string theory, the D-branes play a fundamental role, and are often viewed as fundamental objects in the theory. As we will see, they couple to the (Ramond-Ramond⁶) fields.

1.2.2. The Five Different String Theories. So far we have neglected to mention that there are really *five* different (supersymmetric) string theories, having slightly different fields and Lagrangians. The differences between them have to do with the gauge groups for some of the gauge fields, closed *vs.* open strings, orientation properties of the strings, and *chirality* (left- *vs.* right-handedness). We will not have time to go into the differences between them in detail, but most of the discussion in this book will focus implicitly on types IIA and IIB. In a nutshell, the five theories are:

- Type I. This is the one theory that involves *unoriented* strings, so that the string worldsheet Σ can be a non-orientable surface like a Klein bottle.
- Type IIA. A theory with oriented strings where left-moving and right-moving spinors have opposite handedness.
- Type IIB. A theory with oriented strings where left-moving and right-moving spinors have the same handedness.
- E_8 Heterotic. A theory where left-movers behave as in bosonic theory and right-movers behave as in supersymmetric theory, and the gauge group is the product of two copies of the exceptional Lie group E_8 .

⁶The fields representing particles in string theory lie in different *sectors*, depending on what worldsheet boundary conditions they satisfy (see [14, pp. 122–124], [133, §10.2], or [175, §13.5]). These are called the Ramond (R) and Neveu-Schwarz (NS) sectors. Since one must put the left-moving and right-moving sectors together, that explains terms like “Ramond-Ramond” or “NS-NS.”

- $SO(32)$ Heterotic. A theory where left-movers behave as in bosonic theory and right-movers behave as in supersymmetric theory, and the gauge group is a Lie group locally isomorphic to $SO(32)$.

1.2.3. Textbook References on String Theory. We have deliberately kept our treatment of string theory as short as possible, so the reader is urged to look at some of the standard books on the subject. In particular, this is where to read about the differences between the theories enumerated in Section 1.2.2 above. Here are my own personal favorites.

From the physics point of view:

- J. Polchinski, *String Theory*, 2 vols., Cambridge, 1998. [132, 133]
- B. Zwiebach, *A First Course in String Theory*, Cambridge, 2004. [175]
- K. and M. Becker and J. Schwarz, *String Theory and M-Theory*, Cambridge, 2007. [14]

From a more mathematical point of view:

- *Quantum Fields and Strings: A Course for Mathematicians*, 2 vols., Amer. Math. Soc. and IAS, 1999. [51]
- J. Jost, *Bosonic Strings: a Mathematical Treatment*, Amer. Math. Soc., 2001. [89]
- *Mirror Symmetry*, Amer. Math. Soc. and Clay Math. Inst., 2003. [81]

1.3. Dualities Related to String Theory

1.3.1. The Notion of Duality.

1.3.1.1. *What is a Duality?* A *duality* is a transformation between different-looking physical theories that, rather magically, have the same observable physics. Often, such dualities are part of a discrete group, such as $\mathbb{Z}/2$ or $\mathbb{Z}/4$ or $SL(2, \mathbb{Z})$.

EXAMPLE 1.2 (Electric-magnetic duality). There is a symmetry of Maxwell's equations in free space

$$(1.4) \quad \begin{aligned} \nabla \cdot E &= 0, & \nabla \cdot B &= 0, \\ \frac{\partial E}{\partial t} &= c \nabla \times B, & \frac{\partial B}{\partial t} &= -c \nabla \times E, \end{aligned}$$

given by $E \mapsto -B$, $B \mapsto E$. This is a duality of order 4.

A deep quantum extension of this duality was proposed by Dirac [53], and later generalized by Goddard, Nuyts, and Olive [68] and by Montonen and Olive [115]. Mathematicians who do not follow the physics literature might be interested to learn that when this duality is applied to gauge fields with nonabelian gauge group, such as those appearing in elementary particle theory, then the duality also switches a Lie group G with its Langlands dual G^\vee [93], just as in the Langlands program in representation theory and automorphic forms [32]. But again, this is a subject for another book.

1.3.1.2. Fourier Duality.

EXAMPLE 1.3 (Configuration space-momentum space duality). Another example from standard quantum mechanics concerns the quantum harmonic oscillator (say in one dimension). For an object with mass m and a restoring force with "spring constant" k , the Hamiltonian is

$$(1.5) \quad H = \frac{k}{2}x^2 + \frac{1}{2m}p^2,$$

where p is the momentum. In classical mechanics, $p = m\dot{x}$. But in quantum mechanics (with \hbar set to 1),

$$(1.6) \quad [x, p] = i.$$

We obtain a duality of (1.5) and (1.6) via $m \mapsto \frac{1}{k}$, $k \mapsto \frac{1}{m}$, $x \mapsto p$, $p \mapsto -x$. This is again a duality of order 4, and is closely related to the *Fourier transform*, since in the *Schrödinger representation* on $L^2(\mathbb{R})$, with x corresponding to the usual coordinate on \mathbb{R} , p is represented by the operator $-i\frac{d}{dx}$, whose Fourier transform is x . Recall, incidentally, that if Lebesgue measure is properly normalized, the Fourier transform is indeed a unitary operator on $L^2(\mathbb{R})$ of period 4, with the Hermite functions as eigenvectors.

1.3.2. T-Duality. One of the important dualities in string theory, called *T-duality* (“T” for “target space” or “torus”), will be the main subject of this book. This duality sets up an equivalence of string theories on two very different spacetime manifolds X and X^\sharp . The basic idea is that tori in X are replaced by their *dual tori* in X^\sharp . In the simplest case, that means that X has a circle factor of radius R and X^\sharp has a circle factor of radius $\tilde{R} = \frac{\alpha'}{R}$.⁷ The duality also involves changes in the metric and the B -field, known as the *Buscher rules*, after Buscher, who derived them in 1987–88 [37, 38]. (A similar calculation was also done by Molera and Ovrut [114].)

1.3.2.1. *Derivation of T-Duality, Following Buscher.* Consider the simplest case. Take Σ a closed Riemannian 2-manifold and consider the action (1.2) for a map to a circle with radius R , gotten by integrating a 1-form ω on Σ with integral periods:

$$S(\omega) = \frac{1}{4\pi\alpha'} \int_{\Sigma} \frac{R^2}{\alpha'} \omega \wedge *\omega.$$

Add a new parameter μ , a kind of Lagrange multiplier, and consider instead

$$S(\omega, \mu) = \frac{1}{4\pi\alpha'} \int_{\Sigma} \left(\frac{R^2}{\alpha'} \omega \wedge *\omega + 2\mu d\omega \right).$$

For an extremum of S with respect to variations in μ , we need $d\omega = 0$, so we get back the original theory. But instead we can take the variation in ω .

$$\begin{aligned} \delta S &= \frac{R^2}{4\pi\alpha'^2} \int_{\Sigma} \left(\delta\omega \wedge *\omega + \omega \wedge *\delta\omega + \frac{2\alpha'}{R^2} \mu d\delta\omega \right) \\ &= \frac{R^2}{4\pi\alpha'^2} \int_{\Sigma} \delta\omega \wedge \left(2*\omega + \frac{2\alpha'}{R^2} d\mu \right), \end{aligned}$$

so if $\delta S = 0$, $*\omega = \frac{-\alpha'}{R^2} d\mu$ and $\omega = \frac{\alpha'}{R^2} *d\mu$. If $\eta = *d\mu$, substituting back into $S(\omega, \mu)$ gives

$$\begin{aligned} S'(\eta) &= \frac{1}{4\pi\alpha'} \int_{\Sigma} \left(\frac{R^2}{\alpha'} \left(\frac{\alpha'}{R^2} \right)^2 \eta \wedge *\eta + 2\frac{\alpha'}{R^2} \mu d*d\mu \right) \\ &= -\frac{1}{4\pi\alpha'} \int_{\Sigma} \frac{\alpha'}{R^2} \eta \wedge *\eta \end{aligned}$$

⁷More generally, if V is a finite-dimensional real vector space and Λ is a lattice in V , then the dual torus to $T = V/(2\pi\Lambda)$ is $T^\sharp = V^*/(2\pi\alpha'\Lambda^*)$, where V^* is the dual space and Λ^* consists of elements $\gamma \in V^*$ which take integral values on $\Lambda \subset V$.

which is just like the original action (with η replacing ω , $\tilde{R} = \frac{\alpha'}{R}$ replacing R).

1.3.2.2. *Connection with Theta Functions.* T-duality is also related to the classical theory of θ -functions. Consider a simple theory where $\Sigma = S^1$ and $X = \mathbb{R}/(2\pi R\mathbb{Z})$. (If you like, these are the space-like directions and there is another [inert] time direction, a factor of \mathbb{R} .) A string winding around X is like a wound-up rubber band; the higher the winding number, the greater the energy. For simplicity, let's just sum over the semi-classical states, the harmonic maps $x \mapsto 2\pi nRx: \mathbb{R}/\mathbb{Z} \rightarrow X$, instead of taking the path integral, which involves infinite-dimensional integration over *all* paths. The action (1.2) for this map is

$$\frac{1}{4\pi\alpha'} \int_0^1 \left| \frac{d}{dx} (2\pi nRx) \right|^2 dx = \frac{4\pi^2 n^2 R^2}{4\pi\alpha'} = \frac{\pi n^2 R^2}{\alpha'}.$$

The partition function is therefore:

$$(1.7) \quad Z_R = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 R^2 / \alpha'},$$

a classical θ -function.

Now the Poisson summation formula says that if f is a function in $\mathcal{S}(\mathbb{R})$, the Schwartz space of rapidly decreasing functions, with Fourier transform \hat{f} , then

$$(1.8) \quad \sum_{n=-\infty}^{\infty} f(n) = \sum_{n=-\infty}^{\infty} \hat{f}(n).$$

Take $f(x) = e^{-\pi x^2 R^2 / \alpha'}$. Since $x \mapsto e^{-\pi x^2}$ is its own Fourier transform, rescaling shows that $\hat{f}(s) = \frac{\sqrt{\alpha'}}{R} e^{-\pi s^2 \alpha' / R^2}$. Thus applying the Poisson summation formula (1.8) to f and keeping (1.7) in mind, we get the famous identity (used in the proof of the functional equation of the Riemann ζ -function):

$$(1.9) \quad Z_R = \frac{\sqrt{\alpha'}}{R} Z_{\tilde{R}} \quad \text{or} \quad \sqrt{R} Z_R = \sqrt{\tilde{R}} Z_{\tilde{R}},$$

where $\tilde{R} = \alpha'/R$, which is basically a precise form of T-duality.

1.3.3. Other Dualities in String Theory. An excellent general survey of dualities in string theory may be found in [150]. Here we will be very brief.

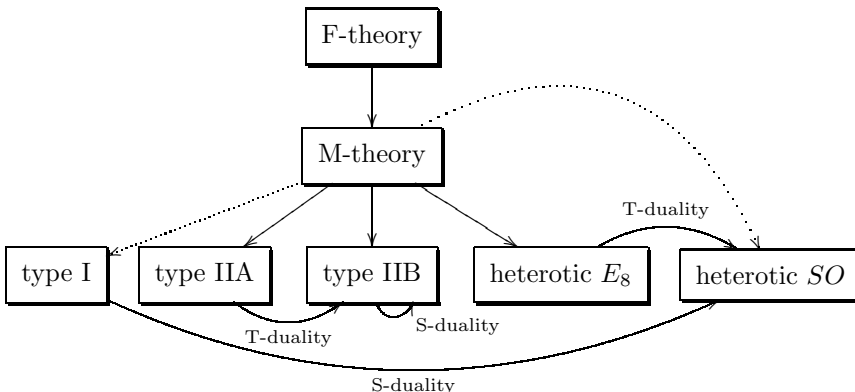
1.3.3.1. *S-Duality.* Another important duality in string theory is *S-duality* (“S” for “strong/weak”). This duality is actually an outgrowth of the classical electric-magnetic duality (see Example 1.2), via a suggestion of Dirac [53] that the charge of a magnetic monopole (which has never been observed) should be $\hbar c/2$ times the reciprocal of the charge of an electron. Since the charge of an electron is small and easily observable, the charge of a magnetic monopole should be large, which perhaps explains why one has never been observed. S-duality interchanges the strong coupling limit of one string theory with the weak coupling limit of another one.

It has been pointed out [75] that S-duality and T-duality are closely linked. S-duality involves the duality between a compact Lie group G and its Langlands dual G^\vee . If we choose a Cartan subalgebra \mathfrak{h} in \mathfrak{g} , then the dual space \mathfrak{h}^* can be identified with a Cartan subalgebra in \mathfrak{g}^\vee , and we get a pair of dual tori, \mathfrak{h}/Λ and \mathfrak{h}^*/Λ^* , where Λ is the coweight lattice for \mathfrak{g} and the weight lattice for \mathfrak{g}^\vee , while Λ^* is the coweight lattice for \mathfrak{g}^\vee and the weight lattice for \mathfrak{g} . The T-duality between

these tori turns out to reproduce S-duality. Mixing S-duality and T-duality gives a broader family of dualities, sometimes called *U-duality* (“U” for “unified”) [83].

1.3.3.2. *AdS/CFT Duality.* Another duality which has attracted a lot of attention recently is often called *AdS/CFT duality*. (Here AdS stands for “anti-de Sitter space,” a spacetime manifold of constant curvature, and CFT stands for “conformal field theory.”) This duality was discovered by Juan Maldacena [105], and in general posits an equivalence between gauge theories in dimension d (usually 4) and string theories in a spacetime of dimension $d + 1$. There is by now a huge literature on this. For more details, those interested can look at Section 1.4 below.

1.3.3.3. *M-Theory and F-Theory.* There are many other dualities connected with string theory, which fit into various patterns which have been schematized by drawings like:



Superstring theories are (to eliminate certain anomalies) required to be 10-dimensional. The dualities between them seem to involve an 11-dimensional theory, called M-theory, which reduces to 11-dimensional *supergravity* in the low energy limit, and a 12-dimensional theory, called F-theory [164].

1.4. ☆ More on S-Duality and AdS/CFT Duality

1.4.1. **S-Duality.** Since quantum field theories in general, and string theories in particular, are quite complicated, very few things can be computed exactly. For that reason, one of the main calculational tools is *perturbation theory*: expanding in a power series in some parameter, and computing the coefficients. Of course, for this to give useful results, one has to be in a realm where this parameter is small, so that there is hope that the series will converge reasonably well.

The most important dimensionless parameter in string theory is the *string coupling constant* g_S , which measures the intensity of interactions between strings, so it is reasonable to consider perturbation expansions in this parameter, but only if $g_S \ll 1$. However, the interesting feature of string theory is that g_S is not a fixed number; rather, it can be expressed as $e^{(\Phi)}$, the exponential of the expectation value of a scalar-valued field Φ called the *dilaton*.

The main feature of S-duality, as opposed to T-duality for instance, is that it exchanges one string theory with a small value of g_S with another with a large value of g_S , by reversing the sign of the dilaton field Φ . T-duality, on the other hand, results in a shift in the dilaton field, but in the form of a translation, so the effect on the coupling constant is less dramatic.

So the main consequence of S-duality is that it makes it possible to probe aspects of string theory not amenable to perturbation theory (i.e., the realm where $g_S \gg 1$) by linking them to the perturbative realm of *another* string theory with $g_S \ll 1$.

1.4.2. AdS/CFT Duality. The discovery of the AdS/CFT duality originally grew out of studies of Yang-Mills theory for $U(N)$ -bundles in the limit as $N \rightarrow \infty$, and the observation that this theory behaves in ways similar to string theory. A more precise conjecture is that, in the large N limit, supersymmetric Yang-Mills theory on 4-dimensional Minkowski space is dual to type IIB string theory on $\text{AdS}^5 \times S^5$, where AdS^5 is anti-de Sitter space, a 5-dimensional Lorentz manifold of constant curvature. (This is (at least up to coverings) the homogeneous space $SO(4, 2)/SO(4, 1)$.)

The fact that string theory is connected to Yang-Mills theory is related to something we will discuss in Section 2.2.2 in the next chapter: that D-branes naturally carry bundles (in type IIB these are $U(N)$ -bundles) and gauge fields. In fact, open string massless states in type IIB string theory (recall that for physicists, “open” strings are not what mathematicians call “open manifolds”—they have worldsheets with *boundary*, the boundary contained in the D-branes) have an effective Lagrangian that looks a lot like that of Yang-Mills theory. This point of view makes it possible to construct the duality correspondence in the reverse direction, from string theory to gauge theory.

In the AdS/CFT correspondence, the string coupling g_S is supposed to correspond (up to a constant factor of 4π) to a coupling constant g_{YM}^2 in Yang-Mills theory. (A factor of $1/g_{\text{YM}}^2$ should have been inserted in front of the Yang-Mills action in Example 1.1 (1); we omitted it there so as not to overburden the reader.) However, the duality is something like S-duality in that it is only supposed to give a reasonable match between the two theories when one is weakly coupled and the other is strongly coupled.

For readers who want to learn more, a good reasonably short survey of the AdS/CFT correspondence, written by its discoverer, may be found in [104]. A more comprehensive survey may be found in [1].