TOPOLOGICAL FIELD THEORY

Danny BIRMINGHAM

CERN, Theory Division, CH-1211, Geneva 23, Switzerland

Matthias BLAU

Centre de Physique Théorique, C.N.R.S., Luminy, Case 907, F-13288 Marseille, Cedex 9, France and NIKHEF-H, P.O. Box 41882, 1009 DB Amsterdam, The Netherlands

Mark RAKOWSKI and George THOMPSON

Institut für Physik, Johannes-Gutenberg-Universität, Staudinger Weg 7, D-6500 Mainz, Germany

TOPOLOGICAL FIELD THEORY

Danny BIRMINGHAM [1]

CERN, Theory Division, CH-1211, Geneva 23, Switzerland

Matthias BLAU

*Centre de Physique Théorique, C.N.R.S., Luminy, Case 907, F-13288 Marseille, Cedex 9, France
and NIKHEF-H, P.O. Box 41882, 1009 DB Amsterdam, The Netherlands* [2]

Mark RAKOWSKI and George THOMPSON

Institut für Physik, Johannes-Gutenberg-Universität, Staudinger Weg 7, D-6500 Mainz, Germany

Contents:

[1] Supported by "Commission des Communautés Européennes" (DG XII-CCR).
[2] Current address.

1. Introduction

The history of the relationship between problems arising in the study of physical systems and the subsequent mathematical developments needed for their analysis is rich and interwoven. Over the past two decades, some of the most fruitful connections have centered around problems arising in gauge theory. Indeed, an understanding of the classical Yang-Mills and instanton equations was called for on physical grounds, and the study of these equations has led to dramatic mathematical advances in the topology and geometry of low dimensional manifolds. Whereas these studies can be

considered to lie within the realm of classical physics, a number of deep and exciting connections have recently emerged which link these developments intimately to quantum theory. The study of these relations has become known as topological quantum field theory. On the one hand, this subject has provided a unifying perspective for many of these mathematical results, while at the same time it has significantly enhanced our understanding of two-dimensional conformal field theory and certain models in statistical mechanics, and promises new insight into string theory. All of these developments underscore the true richness of quantum field theory.

Topological quantum field theories are characterized by observables (correlation functions) which depend only on the global features of the space on which these theories are defined. In particular, this means that the observables are independent of any metric which may be used to define the classical theory. It is an amazing result that one can achieve general covariance in the quantum theory without necessarily integrating over the metric, as one does in quantum gravity. These geometrical and topological invariants, which are computable by standard techniques in quantum field theory, are of prime interest in mathematics. It is natural to hope that a deeper understanding of this special class of field theories, all of which bear a formal resemblance to many systems of longstanding physical interest, will provide new insight into the structure of these more complicated physical systems. Topological quantum field theories are quite generally soluble, and could provide a testing ground for new approaches to quantum field theory. Perhaps the most tantalizing physical conjecture is that topological quantum field theories represent different phases of their more conventional counterparts; in these topological phases general covariance is unbroken. From a purely mathematical point of view, topological quantum field theories provide novel representations of certain global invariants whose properties are frequently transparent in the path integral approach. Although such derivations cannot be considered rigorous, they can be checked by other physical (Hamiltonian) and mathematical methods.

The origin of topological field theories can be traced to the work of Schwarz and Witten. It was Schwarz who showed in 1978 [1.1] that Ray–Singer torsion—a particular topological invariant—could be represented as the partition function of a certain quantum field theory. Quite distinct from this observation was the work of Witten in 1982 [1.2], where a framework was given for understanding Morse theory in terms of supersymmetric quantum mechanics. These two constructions represent the prototypes of all known topological field theories. The model used by Witten also found applications in classical index theorems [1.3], and moreover, suggested generalizations leading to new mathematical results in the form of the holomorphic Morse inequalities [1.4].

The significance of Witten's approach was realized by Floer [1.5], who applied similar techniques in an infinite dimensional setting to obtain new results concerning the topology of three-manifolds. This work was clearly related in some way to the findings of Donaldson [1.6] on the geometry of four-manifolds. In an influential paper [1.7], Atiyah then conjectured that a quantum field theory might provide an understanding of these results, and he produced a nonrelativistic Hamiltonian in three dimensions whose ground states are the Floer groups. A four-dimensional relativistic Lagrangian description of Donaldson's work was supplied by Witten in 1988 [1.8], which established the link between these three- and four-dimensional results.

Quite apart from these developments, a new polynomial invariant of knots was constructed by Jones in 1985 [1.9]. It is noteworthy that this work was strongly influenced by problems in two-dimensional statistical mechanics. As in all previous work on knot theory, the evaluation of these invariants was based on two-dimensional projections. As knots are objects living intrinsically in three dimensions, a longstanding puzzle for knot theorists has been to understand these invariants from a three-dimensional point of view. In a classic paper [1.10], Witten again provided the answer by constructing knot polynomials as correlation functions of Wilson line operators in a

three-dimensional quantum field theory defined by the Chern–Simons action. Moreover, this theory incorporates significant generalizations of the previously known invariants. While these mathematical advances are self-evident, Chern–Simons theory also provides a unifying three-dimensional viewpoint for two-dimensional conformal field theory, as well as new results on quantum gravity in three dimensions [1.11].

Other examples of topological field theories were also given by Witten [1.12, 1.13]. The topological sigma models were used to construct invariants of complex manifolds and are related to other work of Floer [1.14]. Also of importance are the two-dimensional topological gravity models [1.15, 1.16]. There are many tempting conjectures relating two-dimensional topological gravity with string theory. Indeed, it is believed that noncritical string theory with certain matter content is equivalent to topological gravity coupled to topological matter [1.17–1.19].

Given these developments, it was natural to try to understand whether these models were isolated examples, or whether they belong to a larger class of theories which enjoy similar topological properties. Before an answer to this question was provided, it was first necessary to understand the formal field theoretic structure of Witten's actions. An explanation of the origin of these actions was given by several groups [1.20–1.23], and a general prescription for the construction of these and other models was developed [1.24–1.26].

The purpose of this report is to bring together many of these developments. While the subject of topological field theories is still under active research in many directions, a certain body of material is now well understood and can be considered standard technique. It is our aim to explain these field theoretic methods together with the rich mathematical structures which they describe. We assume the reader is versed in standard BRST techniques in quantum field theory, and has as well a basic knowledge of differential geometry. At each stage, we have endeavored to review as much background material as space permits, in an attempt to make this presentation self-contained. Where this has not really been possible, ample references have been given which the reader may consult.

Our plan is as follows. We begin in the next section with a general discussion of topological field theories, their defining properties, and classification. A knowledge of this material will allow the reader to move freely among the other sections. The first model we consider in detail (section 3) is supersymmetric quantum mechanics. This will serve to illustrate many of the generic features of topological field theories in as simple a setting as possible. Topological sigma models, their observables, and the associated mathematics of complex geometry and intersection theory are presented in section 4. Following this, topological gauge theories are discussed in section 5, with particular emphasis on Donaldson theory. The mathematics here is necessarily much more sophisticated than at any other point in this report, and to bridge this gap, a mathematical review of gauge theory and moduli spaces has been included. An analysis of the geometry underlying Donaldson theory gives a general recipe for constructing field theories associated to moduli spaces in arbitrary dimensions, and as an example, we analyze in detail the super BF theories associated with flat connections. Chern–Simons theory and related BF models are the subject of section 6. The connections with knot theory are briefly reviewed and the link with 2D conformal field theory is sketched. We also consider 3D gravity from the Chern–Simons point of view. A thorough discussion of Chern–Simons theory would, however, involve considerable use of conformal field theory, and this lies beyond the scope of this report. This is also the case with two-dimensional topological gravity, but we have nevertheless presented some of its more elementary features in section 7. A presentation of the metric and gauge theory approaches to topological gravity in two dimensions is given, though we stop short of detailed computations involving conformal field theory. As in all quantum field theories, the issue of renormalization needs to be addressed, and one is obliged to show that the

formal topological properties of these theories survive quantization. This point is considered in section 8. We present a detailed analysis of the beta function in certain Witten type theories, and compute one-loop effects in Chern–Simons theory. We will have recourse at many points in this report to apply the Batalin–Vilkovisky quantization procedure, and for convenience, the reader can find the essentials reviewed in appendix A.

One aspect of topological field theory that we do not touch upon in this report is the axiomatic approach initiated by Atiyah [1.27]. Many of the more recent mathematical applications of the ideas of topological field theory rely, to a certain extent, on this formulation, which is similar to that proposed by Segal for conformal field theory [1.28]. Unfortunately, lack of space has prevented us from including this part of the story. For a detailed account, and some interesting applications, we refer to the work of Axelrod [1.29].

It had been our intention in this report to include a complete list of references to topological field theory; however, the subject is still under rapid development and any list at the time of writing is necessarily incomplete. We apologize to those authors whose papers have escaped our attention, and to those whose work we have inadvertently omitted.

2. General aspects of topological field theory

Before embarking on a survey of the various models on the market, it is useful to first present some general definitions and properties shared by all topological theories. Among these are the simple formal arguments which establish, with some exceptions, the topological nature of a given model. In addition, we present a useful classification scheme of the known theories: we characterize models as being either of Witten or Schwarz type; the prototype of the former being Donaldson theory, while Chern–Simons theory is the best known example in the Schwarz class. Finally, we also introduce the important notion of a moduli space, which contains the classical data upon which every topological field theory is built.

2.1. Definitions

Let us begin by recalling the essential ingredients which are present in a conventional gauge field theory, for example Yang–Mills theory. We shall assume that the reader is familiar with BRST quantization of gauge theories [2.1, 2.2]; useful references are [2.3–2.9]. In such a formulation, we denote the collective field content by Φ, which includes the gauge field, ghosts, and multipliers. Corresponding to the local gauge symmetry one constructs a BRST operator Q which is nilpotent, i.e., $Q^2 = 0$. The variation of any functional $\mathcal{O}$ of the fields Φ is denoted by $\delta\mathcal{O} = \{Q, \mathcal{O}\}$, where the bracket notation is used to represent the graded commutator with the fermionic charge Q (see appendix B for this and related conventions). The complete quantum action, denoted by S_q, which comprises the classical action S_c together with the necessary gauge fixing and ghost terms, is by construction Q-invariant.

The physical Hilbert space is defined by the condition $Q|\text{phys}\rangle = 0$; furthermore, a physical state of the form $|\text{phys}\rangle' = |\text{phys}\rangle + Q|\chi\rangle$ is regarded as equivalent to $|\text{phys}\rangle$, for any state $|\chi\rangle$. A state which is annihilated by Q is said to be Q-closed, while a state of the form $Q|\chi\rangle$ is called Q-exact. This equivalence relation thus partitions the physical Hilbert space into what are called Q-cohomology classes, that is, states which are Q-closed modulo Q-exact states [2.3].

Now, from the BRST invariance of the vacuum, it follows immediately that the vacuum expec-

tation value of $\{Q, \mathcal{O}\}$, for any functional $\mathcal{O}$, is zero, i.e.,

$$\langle 0 | \{Q, \mathcal{O}\} | 0 \rangle \equiv \langle \{Q, \mathcal{O}\} \rangle = 0 . \tag{2.1}$$

An operator of the form $\{Q, \mathcal{O}\}$ is called a BRST commutator. Let us now assume that we are defining our theory on some manifold M, with a metric $g_{\alpha\beta}$. In this case the energy–momentum tensor $T_{\alpha\beta}$ is defined by the change in the action under an infinitesimal deformation of the metric

$$\delta_g S_q = \frac{1}{2} \int_M d^n x \sqrt{g} \, \delta g^{\alpha\beta} \, T_{\alpha\beta} . \tag{2.2}$$

Finally, we assume that the functional measure in the path integral is both Q-invariant and metric independent.

We are now in a position to define what we mean by a topological field theory [2.10, 2.11]. Our working definition will be the following: A topological field theory consists of

(a) a collection of fields Φ (which are Grassmann graded) defined on a Riemannian manifold (M, g),

(b) a nilpotent operator Q, which is odd with respect to the Grassmann grading,

(c) physical states defined to be Q-cohomology classes,

(d) an energy–momentum tensor which is Q-exact, i.e.,

$$T_{\alpha\beta} = \{Q, V_{\alpha\beta}(\Phi, g)\} , \tag{2.3}$$

for some functional $V_{\alpha\beta}$ of the fields and the metric.

It turns out that in all examples to date, Q has an identification as a BRST charge, and the Grassmann grading corresponds to ghost number. However, one should remark that such an identification is by no means mandatory. Nevertheless, we shall henceforth refer to Q as the BRST operator. Furthermore, Q is in general metric independent, and this is certainly the simplest situation to deal with, and the only one we shall consider for the moment. However, there are interesting cases where $T_{\alpha\beta}$ is a BRST commutator, although Q fails to be metric independent (supersymmetric quantum mechanics and topological sigma models, for example; see sections 3 and 4). In addition, there are cases where $T_{\alpha\beta}$ fails to be a BRST commutator, while, nevertheless, it is still possible to establish the topological nature of the models (examples are provided by the higher dimensional non-Abelian BF theories of section 6). It is thus clear that the above definition may not be completely adequate in all cases; however, it does, as stated above, provide us with a good working definition, and is general enough to cover most cases.

We now consider the change in the partition function

$$Z = \int [d\Phi] \, e^{-S_q} , \tag{2.4}$$

under an infinitesimal change in the metric. We have

$$\delta_g Z = \int [d\Phi] \, e^{-S_q} \left(-\frac{1}{2} \int_M d^n x \sqrt{g} \, \delta g^{\alpha\beta} T_{\alpha\beta} \right)$$

$$= \int [d\Phi] \, e^{-S_q} \left(-\frac{1}{2} \int_M d^n x \sqrt{g} \, \delta g^{\alpha\beta} \{Q, V_{\alpha\beta}\} \right)$$

$$= \int [d\Phi] \, e^{-S_q} \{Q, \chi\} = \langle \{Q, \chi\} \rangle = 0 , \tag{2.5}$$

where $\chi = -\frac{1}{2} \int_M d^n x \sqrt{g} \, \delta g^{\alpha\beta} \, V_{\alpha\beta}$. We thus see that given the BRST invariance of the vacuum, we have a partition function which is metric independent. That is, the partition function depends not on the local structure of the manifold, but only on global properties: Z is a topological invariant. At this point, however, we should perhaps clarify the use of the terminology "topological". In all cases, our theory is defined with respect to a "base" manifold M. This could be, for example, a Riemannian manifold with metric g, or a more general situation. What we have shown above is that if the conditions (a)–(d) are satisfied, then the partition function takes a constant value on the space of all metrics on M. We shall henceforth use the term "topological" to specify this metric independence.

However, in the mathematics literature the term "topological" is defined in a weaker sense. Two manifolds M and M' are said to be homeomorphic if there exists a homeomorphism f : $M \rightarrow M'$ (i.e., f and f^{-1} are continuous mappings). One can thus partition manifolds into homeomorphism equivalence classes. An object which takes a constant value on each class is called a "topological invariant". However, one can further subdivide each homeomorphism class by specifying diffeomorphisms (i.e. C^∞ mappings) between its members. Each homeomorphism class then comprises a collection of diffeomorphism classes, and an object which takes a constant value on each of the latter is called a differential invariant for the manifold M. An object which is invariant under metric deformations (i.e. topological) is certainly also diffeomorphism invariant and hence corresponds to a differential invariant.

We can now ask the question as to whether there exist other metric independent correlation functions in the theory; does a given theory have a richer set of topological invariants ?

Consider the vacuum expectation value of an observable

$$\langle \mathcal{O} \rangle = \int [d\Phi] \, e^{-S_q} \, \mathcal{O}(\Phi) . \tag{2.6}$$

We wish to determine sufficient conditions for this expectation value to be a topological invariant, i.e., for $\delta_g \langle \mathcal{O} \rangle$ to be zero. Proceeding as before, we find [2.10]

$$\delta_g \langle \mathcal{O} \rangle = \int [d\Phi] \, e^{-S_q} \, (\delta_g \mathcal{O} - \delta_g S_q \cdot \mathcal{O}) . \tag{2.7}$$

Assuming that $\mathcal{O}$ enjoys the properties

$$\delta_g \mathcal{O} = \{Q, R\} , \qquad \{Q, \mathcal{O}\} = 0 , \tag{2.8}$$

for some R, we have that

$$\delta_g \langle \mathcal{O} \rangle = \langle \{Q, R + \chi \mathcal{O}\} \rangle = 0 . \tag{2.9}$$

One should note that the function $V_{\alpha\beta}$ defined earlier contains explicit reference to the metric; nevertheless, it enters the analysis in the form of a BRST commutator and one still has metric independence of Z.

Now, clearly if $\mathcal{O} = \{Q, \mathcal{O}'\}$, for some $\mathcal{O}'$, we automatically have $\langle \mathcal{O} \rangle = 0$. Hence, our real interest is in Q-cohomology classes of operators (i.e., BRST invariant operators which are not Q-exact) which satisfy $\delta_g \mathcal{O} = \{Q, R\}$. In deriving the above relations, we should note that we made essential use of the (assumed) metric independence of the functional measure. To show that this assumption is in fact realized, one needs to check for metric anomalies.

Our aim now is to present a convenient classification scheme for the known topological field theories. The theories that we shall describe in this report can be classified as being of either of

two types: Witten type or Schwarz type [2.12]. Let us first define the Witten type theories [2.10, 2.11]. In this case the complete quantum action S_q, which comprises the classical action plus all the necessary gauge fixing and ghost terms, can be written as a BRST commutator, i.e.,

$$S_q = \{Q, V\},\tag{2.10}$$

for some functional $V(\Phi, g)$ of the fields, and Q is the nilpotent (and in general, metric independent) BRST charge. There is also the freedom to add topological terms to the action (2.10); i.e., terms for which the Lagrangian is locally a total derivative; such terms change neither the equations of motion nor the energy–momentum tensor. Clearly, as a consequence of (2.10), we have

$$T_{\alpha\beta} = \{Q, (2/\sqrt{g})\, \delta V/\delta g^{\alpha\beta}\},\tag{2.11}$$

which ensures us of the topological nature of the model. However, the stronger condition (2.10) allows us to prove that the partition function Z, and the above correlators, are also exact at the semiclassical level. By introducing a dimensionless parameter t (equivalently, $1/\hbar$) and rescaling the action $S_q \to tS_q$, we can consider the variation of Z under a change in t:

$$\begin{aligned}
\delta_t Z &= -\int [d\Phi]\, e^{-tS_q} S_q\, \delta t\\
&= -\int [d\Phi]\, e^{-tS_q} \{Q, V\}\, \delta t = 0.
\end{aligned}\tag{2.12}$$

This shows that Z is independent of t, as long as t is non-zero (one cannot set t to zero, since one needs a damping factor in the path integral) and thus one can evaluate Z in the large-t limit. Such a limit corresponds to the semiclassical approximation, in which the path integral is dominated by fluctuations around the classical minima; such an approximation is exact for Witten type theories. A similar argument applied to (2.6) establishes the semiclassical exactness of the correlation functions.

For the case of Schwarz type theories [2.13, 2.14], one begins with a metric independent classical action $S_c(\Phi)$ which is not a total derivative. Upon gauge fixing, the total quantum action (in certain cases) takes the form

$$S_q(\Phi, g) = S_c(\Phi) + \{Q, V(\Phi, g)\}.\tag{2.13}$$

We should stress that by Schwarz type, we mean a classical action which is nontrivial. Since the classical action is metric independent, the classical energy–momentum tensor vanishes. If (2.13) holds, the complete energy–momentum tensor is given by

$$T_{\alpha\beta} = \{Q, (2/\sqrt{g})\, \delta V/\delta g^{\alpha\beta}\},\tag{2.14}$$

with the entire contribution coming from the gauge fixing and ghost terms. It follows that Z is metric independent. At this point, however, we need to be more precise in our definition of Schwarz type theories. Theories which satisfy the above properties include Chern–Simons theory, the Abelian BF models, and also the two- and three-dimensional non-Abelian BF models of section 6.2. However, the higher dimensional $n > 3$ class of non-Abelian BF theories, discussed in detail there, possess some "non-standard" properties. In particular, while one begins with a metric independent classical action, (2.13) and (2.14) do not hold. The source of the problem lies in the on-shell reducibility of the gauge symmetries involved. More work is then required in order to establish the topological nature of the quantum theory, and we refer to section 6.3.4 for details of how this can be achieved.

As regards the importance of loop corrections in such a theory, a few remarks are in order. Given the fact that Schwarz type theories do not enjoy the property that the quantum action is Q-exact,

we cannot appeal to the general arguments following (2.10) to establish the semiclassical exactness. However, among the known Schwarz type models, it appears that Chern–Simons theory is the only one in which loop corrections to the partition function, and the observables, are non-zero; all other BF models have a partition function (and observables) in which the semiclassical approximation is exact.

We proceed with a few words about the type of gauge symmetries that arise in the two cases. For the Witten type theories, Q is obtained by combining a certain topological shift symmetry ($\delta \Phi = \epsilon$ for certain fields) with any other local symmetry (e.g. conventional Yang–Mills type gauge symmetry). However, for the Schwarz type models, Q corresponds to the usual gauge symmetry, although perhaps of a reducible type.

In addition to our working definition stated above, another essential property of topological field theories is the absence of dynamical excitations. In other words, there are no propagating degrees of freedom. To see this more explicitly in the different classes requires a little more discussion. In the Witten type theories Q is both a supersymmetry and BRST operator. In other words, from the structure of the (topological shift) symmetry, one can see that each bosonic field has a Q-superpartner. In addition, however, we define our theory by the requirement that physical states are annihilated by Q. Hence, the superpartners are interpreted as ghosts, leading to the zero degrees of freedom count.

In Schwarz type theories, Q corresponds to a BRST operator of a gauge symmetry. To establish the absence of degrees of freedom here one can, for example, resort to a standard Dirac analysis of the constraints. This leads to a straightforward determination of the dimension of the reduced physical phase space. The fact that it turns out to be zero is a result of the special structure of the classical action, whereby the number of first class constraints is sufficient to gauge away all degrees of freedom. A more complete discussion of degrees of freedom can be found in section 3.6.

In general, when the conditions (a)–(d) are met, we have

$$
\begin{aligned}
\langle \text{phys}' \mid H \mid \text{phys} \rangle &= \langle \text{phys}' \mid \int T_{00} \mid \text{phys} \rangle \\
&= \langle \text{phys}' \mid \int \{Q, V_{00}\} \mid \text{phys} \rangle = 0 \ ,
\end{aligned}
\tag{2.15}
$$

where H is the Hamiltonian. We thus see that the energy of any physical state is zero, and hence there are no physical excitations.

It is worth pausing for a moment to consider the situation in string theory. Here, the world sheet energy–momentum tensor is also a BRST commutator. This is also true in any theory (for example gravity) where one is integrating over metrics in the path integral with a diffeomorphism invariant action. Indeed, string theory is a topological field theory with respect to the world sheet manifold in the sense of the first criterion. However, it is not topological with respect to the target space–time manifold, and a simple degrees of freedom count shows there are 24 propagating modes for the bosonic string. This can also be seen from the fact that the variation of the action with respect to the target metric is not a BRST commutator.

We should advise the reader that the division of topological field theories into the Schwarz and Witten types, although standard, sometimes goes under different labels [2.15, 2.16]. The Witten type theories are also called "cohomological", while the Schwarz type models are termed "quantum". The term "cohomological" derives from the structure of the observables one encounters in those theories, while "quantum" underscores the non-trivial nature of the Schwarz type quantum theories (semiclassical approximation is not necessarily exact). Essentially, if a topological quantum field theory is not of Witten type ("cohomological"), then it is of Schwarz type ("quantum").

2.2. Moduli space as fields, equations, and symmetries

In the previous section, we outlined the general features which are common to all of the topological field theories which have been studied. One concept that lies at the heart of all these theories is the notion of moduli space. For any given moduli space, there are many different topological field theories (i.e. different fields, equations, and/or symmetries) which describe it. In most cases, those differences may simply be related to the freedom inherent in the quantization program, as we will see once we begin to look at specific models. But there are moduli spaces which have several different classical descriptions, and the associated topological quantum field theories *appear* quite unrelated. Nevertheless, it is a single, unique moduli space that will relate all of those descriptions.

Roughly speaking, a moduli space is the set of equivalence classes of some geometrical object under an equivalence relation. In string theory, for example, the moduli space of Riemann surfaces plays a central role. Two Riemann surfaces M and M' (genus g) are considered equivalent if there exists a diffeomorphism $f : M \to M'$ which is holomorphic in both directions. The moduli space of Riemann surfaces of fixed genus is then the set of equivalence classes in which any two distinct points represent inequivalent Riemann surfaces. In practice, moduli spaces may carry some additional geometrical structure, the moduli space of Riemann surfaces of genus g can be considered as a finite dimensional manifold (modulo singular points) in a natural way.

The moduli space of Riemann surfaces can, like any other moduli space, be described in terms of fields, equations and symmetries [2.16]. One such description is familiar to string theorists; we consider the space of all metrics (fields) and mod-out under the action of diffeomorphisms and Weyl transformations (symmetries). In this case, we do not require any "equations" to further restrict ourselves to the space of interest. Alternatively, we can trade one symmetry for a field equation, by demanding that the metrics have fixed constant scalar curvature. This is possible, since every conformal class of metrics has a unique such representative. The remaining symmetry is then that of diffeomorphisms. We can also change the field content altogether; we take $SL(2, \mathbb{R})$ connections as the fields, require that the connection be flat, and declare the symmetries to be gauge and modular invariance (for a discussion, see sections 5.4.3 and 6.2.7).

This description of a moduli space—in terms of fields, equations, and symmetries—is essentially classical. A topological quantum field theory emerges when one "quantizes" one of those pictures. The interest will then be in studying certain correlation functions of that quantum theory.

Conversely, it is possible to define a Witten type topological field theory by specifying the properties of the physical correlation functions. For instance, one can define a theory by postulating the existence of operators $\mathcal{O}_i$ corresponding to cohomology classes η_i of the moduli space $\mathcal{M}$. One then requires that

$$\langle \mathcal{O}_1 \cdots \mathcal{O}_n \rangle = \int_{\mathcal{M}} \eta_1 \cdots \eta_n \ . \tag{2.16}$$

This leads to the interpretation of the correlation functions as intersection numbers on moduli space; the reader is referred to section 4.5.2 for details. Given the definition (2.16), the task is then to establish that an action with the desired properties can be found, and indeed this has been shown to be possible in some generality [2.17, 2.18]. This description shows, however, that a particular action is not needed to do computations in a Witten type topological field theory.

3. Supersymmetric quantum mechanics

3.1. Introduction

As our first example of a topological field theory we shall consider a relatively simple and tractable model, namely supersymmetric quantum mechanics. Although this model is interesting in its own right [3.1–3.4], the rationale for studying it in detail, in the present context, is to illustrate the fundamental features of Witten type topological field theories. Supersymmetric quantum mechanics was identified as a topological field theory in refs. [3.5, 3.6], and as such, the techniques used in general topological field theories have counterparts in this model. This example allows us to introduce these techniques in a relatively simple setting.

We will use the action of supersymmetric quantum mechanics in the form

$$S = \int d\tau \left[i \left(\frac{d\phi^i}{d\tau} + s g^{ij}(\phi) \frac{\partial V}{\partial \phi^j} \right) B_i + \frac{1}{2} g^{ij}(\phi) B_i B_j + \frac{1}{4} R_{ijkl} \bar{\psi}^i \psi^k \bar{\psi}^j \psi^l \right.$$

$$\left. - i\bar{\psi}_i \left(\delta^i_j \frac{D}{D\tau} + s g^{ik}(\phi) \frac{D^2 V}{D\phi^k D\phi^j} \right) \psi^j \right] . \tag{3.1}$$

Here, ϕ^i are the coordinates of the Riemannian manifold M with metric and curvature denoted, respectively, by g_{ij} and R_{ijkl}; ψ^i and $\bar{\psi}_i$ are the Grassmann odd coordinates of the particle; V is a function on M and s is a parameter. The covariant derivative in eq. (3.1),

$$\frac{D}{D\tau} \psi^i = \frac{d}{d\tau} \psi^i + \Gamma^i_{jk} \dot{\phi}^j \psi^k , \tag{3.2}$$

is the usual pull-back of the covariant derivative on M to the one-dimensional space with Euclidean time coordinate τ. Our conventions for the Riemannian connection and curvature are given by

$$\Gamma^i_{jk} = \frac{1}{2} g^{il} (\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}) ,$$

$$R^i{}_{ljk} = \partial_j \Gamma^i_{lk} - \partial_k \Gamma^i_{lj} + \Gamma^i_{jm} \Gamma^m_{lk} - \Gamma^i_{km} \Gamma^m_{lj} . \tag{3.3}$$

Upon integrating out the auxiliary field B, one recovers the action of refs. [3.1–3.4] with the spinors appearing in the latter decomposed into their components. We have the freedom to choose $\bar{\psi}$ and ψ as either complex conjugates or independent real fields. This is analogous to the situation in gauge theories [3.7]. We choose them to be real.

The supersymmetry of the action is

$$\{Q, \phi^i\} = \psi^i ,$$
$$\{Q, \psi^i\} = 0 ,$$
$$\{Q, \bar{\psi}_i\} = B_i - \bar{\psi}_j \Gamma^j_{ik} \psi^k ,$$
$$\{Q, B_i\} = B_j \Gamma^j_{ik} \psi^k - \frac{1}{2} \bar{\psi}_j R^j{}_{ilk} \psi^l \psi^k , \tag{3.4}$$

and it is straightforward to check that the supersymmetry generator Q is nilpotent: $Q^2 = 0$.

An important ingredient in understanding the nature of this model (3.1), is the existence of a Nicolai map [3.8–3.10]. Not only is this a powerful calculational tool, but more fundamentally it provides us with a variety of ways of reconstructing the action from first principles. We will focus on two of these in sections 3.3 and 3.4, which we shall refer to as the Langevin and Baulieu–Singer

approaches. Our construction of topological field theories in the following sections will be based on these ideas. Since the action (3.1) is already known in the present case, this may seem no more than an unnecessary exercise; however, our motivation for studying these approaches is that they provide a unified way of constructing field theories encoding topological information when the action is not yet known.

From these considerations it will emerge that the operator Q, which was previously called a supersymmetry, is in fact a BRST operator, and that the complete action (3.1) is a BRST commutator. At that point we can appeal to the results of section 2 to establish the topological nature of the model.

Since this interpretation of Q may cause some confusion, in particular, in this low dimensional example, where there is no clearcut distinction between spinors and ghosts, we analyse in some detail, in section 3.6, the definition of physical states in supersymmetric, topological, and BRST quantized gauge theories in general.

In recognizing Q as a BRST operator, one gains the flexibility of choosing different gauge fixing conditions, leading to actions which are quantum mechanically equivalent to, but different from (3.1). This freedom in the choice of gauge can be used to greatly simplify the calculation of the partition function of (3.1), and we illustrate this by explicitly computing the Witten index [3.1] in various supersymmetric theories. In addition, we use this gauge freedom to derive the Gauss–Bonnet and Poincaré–Hopf theorems (relating the Euler number of M to its Riemann curvature, and the number of zeros of a vector field on M, respectively) in this setting. It is, perhaps, worth stressing that we shall use the nomenclature "gauge independence" to refer to independence of the gauge fixing condition.

A proper understanding of topological field theories boils down to an understanding of their zero mode structure, and in section 3.9 we examine the relation between symmetry breaking and the presence of zero modes, as well as the issue of metric and gauge dependence in this situation.

Since the mathematical developments motivating the construction of topological sigma models and Donaldson theory are based on Floer's generalization [3.11] of Witten's supersymmetric quantum mechanics approach to Morse theory [3.2], we review the latter in section 3.10.2.

3.1.1. Toy model

In order that the reader does not get bogged down in all the technicalities of this particular theory, and lose touch with the general ideas, we will deal with a simplified version of this model at the start of each of the sections. This will then be followed by an analysis of the complete theory. It is this toy model, where the target space is one-dimensional, to whose description we now turn. The action (3.1) becomes

$$S = \int d\tau \left[i \left(\frac{d\phi}{d\tau} + s \frac{\partial V}{\partial \phi} \right) B + \tfrac{1}{2} B^2 - i\bar{\psi} \left(\frac{d}{d\tau} + s \frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right] , \tag{3.5}$$

where $\tau \in S^1$ or $\mathbb{R}$; both cases are illustrative. In this section we take $\tau \in S^1$.

We could eliminate B from the action; however, one advantage of retaining it is that the supersymmetry is nilpotent and so reminiscent of a BRST symmetry. For this model, the transformation rules (3.4) take the rather simple form

$$\{Q, \phi\} = \psi , \qquad \{Q, \psi\} = 0 , \qquad \{Q, \bar{\psi}\} = B , \qquad \{Q, B\} = 0 , \qquad \{Q, Q\} = 0 . \tag{3.6}$$

We turn now to an analysis of this theory. The bosonic part of the action is clearly minimized by the first order equation

$$d\phi/d\tau + s\, \partial V/\partial \phi = 0 , \tag{3.7}$$

which means that in the "steepest descent" (one-loop) approximation only such paths will contribute. These classical paths are called "instantons", and the action clearly vanishes for these configurations. In fact these paths are simply points, for taking the square and integrating gives

$$\oint d\tau \, (d\phi/d\tau + s\, \partial V/\partial \phi)^2 = 0 \ . \tag{3.8}$$

Upon integrating by parts and dropping total derivatives (as we are allowed to do having chosen τ to lie on a circle) this becomes

$$\oint d\tau \, [(d\phi/d\tau)^2 + s^2(\partial V/\partial \phi)^2] = 0 \ . \tag{3.9}$$

As this is an integral of a sum of squares it implies

$$d\phi/d\tau = 0 \ , \qquad s\, \partial V/\partial \phi = 0 \ . \tag{3.10}$$

From the first equation we learn that only constant paths are important in this approximation, that is, only an integration over the target manifold $\mathbb{R}$ needs to be performed; while the second equation, if $s \neq 0$, tells us that only those points which correspond to the extrema of the potential contribute. In fact it is possible to show that this semi-classical approximation is exact. The path integral becomes either a sum over the critical points of the potential of signed 1's ($s \neq 0$), or an integral over the whole target manifold. As we have already stated, this will be shown in a variety of ways, thereby elucidating the different approaches to topological field theories.

In the following sections, we will repeatedly encounter the sequence of arguments leading from (3.7) to (3.10); henceforth, we will refer to this as the squaring argument, without explicitly indicating the steps involved.

3.2. Nicolai map

Nicolai has proven [3.8, 3.9] that for theories with a global supersymmetry there exists a non-linear and, in general, non-local mapping of the bosonic fields which trivializes the bosonic part of the action, and whose determinant cancels the Pfaffian (or Salam–Mathews determinant) of the fermionic fields present. We recall that the Pfaffian of an even $2n$-dimensional antisymmetric matrix M_{ij} is defined as $\mathrm{Pf}(M) = \epsilon_{i_1 \cdots i_{2n}} M^{i_1}{}_{i_2} \cdots M^{i_{2n-1}}{}_{i_{2n}}$; with the property that the determinant of the matrix is the square of the Pfaffian. The bosonic part of the action in terms of the new fields is Gaussian and has covariance one[*]. This means that for a globally supersymmetric theory whose partition function, after integrating out the fermion fields, takes the form

$$Z = \int_\phi e^{-S(\phi)} \, \mathrm{Pf}(\mathrm{D}[\phi]) \ , \tag{3.11}$$

where ϕ are the bosonic fields[**], and $\int_\phi$ indicates the path integral over these fields, there exists a map $\phi \to \xi(\phi)$ such that the Jacobian of the transformation compensates the Pfaffian (up to signs). The partition function is then

$$Z = \int_\xi e^{-\frac{1}{2} \int \xi^2} \times (\text{winding number of the mapping}) \ , \tag{3.12}$$

---

[*] By this we mean that the propagator in position space is a delta function, or simply 1 in momentum space. This is not quite the way Nicolai defined the map, but it is the natural definition for topological field theories.

[**] This is *not* the generic situation as can be deduced from (3.1); the curvature term means that integration over the fermions is not simply a Pfaffian.

where the winding number is the number of times ξ runs over its range as ϕ is varied.

Due to the highly non-local character of the map, it has been most often determined perturbatively and there have been few cases where such a map has been given explicitly in closed form. Apart from free theories, complete Nicolai maps were only known for some low dimensional models. Indeed, Nicolai was only able to exhibit a map that has these properties to third order in the coupling for $N = 1$ super Yang–Mills theory in four dimensions [3.10]. In section 5.2.5 we show in which sense that map is complete but for a slightly different theory, in that it trivializes the model introduced by Witten to describe instanton moduli space [3.5, 3.12]! This map has been given a mathematical basis in the recent work of Atiyah and Jeffrey, which we review in section 5.2.6.

From the point of view of topological field theory the existence of Nicolai maps is fundamental [3.5]. This leads to our first categorization of topological field theories:

Witten type topological field theories admit Nicolai maps which trivialize the action and restrict to the moduli space of classical solutions.

3.2.1. Toy model

To show that the "instanton" paths are the only contributions to the path integral we change variables as follows:

$$\phi \to \xi = \mathrm{d}\phi/\mathrm{d}\tau + s\,\partial V/\partial\phi \ . \tag{3.13}$$

The Jacobian that comes from this change of variables is

$$|\det(\delta\phi/\delta\xi)| = |\det(\mathrm{d}/\mathrm{d}\tau + s\,\partial^2 V/\partial\phi\,\partial\phi)|^{-1} \ . \tag{3.14}$$

However, the integration over the fermionic fields in the partition function gives rise to essentially the inverse of this Jacobian. The path integral becomes, after integrating out B,

$$\int\limits_{\xi} \exp\left(-\frac{1}{2}\oint \mathrm{d}\tau\,\xi^2\right)\det(\delta\xi/\delta\phi)\,|\det(\delta\xi/\delta\phi)|^{-1} \ . \tag{3.15}$$

The ratio of determinants is then ± 1, with the result that the path integral itself seems to be (when suitably normalized) ± 1. However, we need to be more precise as we have not yet specified the range of integration of the ξ-field. This involves determining how many times one covers ξ-space as ϕ runs through its range. For this we only need to see how many times ξ goes through zero and in which sense.

To get a feeling for how all of this should work out, consider the much simpler example of a conventional integral (take τ to be a point)

$$\int\limits_{-\infty}^{+\infty} \mathrm{d}\phi\,\mathrm{d}\psi\,\mathrm{d}\bar\psi \exp\left[-\frac{1}{2}\left(\frac{\partial V}{\partial\phi}\right)^2 + \bar\psi\frac{\partial^2 V}{\partial\phi\,\partial\phi}\psi\right] \ . \tag{3.16}$$

Examples

(i) $\partial V(\phi)/\partial\phi = \phi + \phi^2$. The Nicolai map is then $\xi = \phi + \phi^2$ and to determine the range of integration for ξ we must find its turning points with respect to ϕ. There is one at $\phi = -1/2$, so that the ϕ integral must be split into two pieces as

$$\int\limits_{-\infty}^{-1/2} \mathrm{d}\phi + \int\limits_{-1/2}^{+\infty} \mathrm{d}\phi \mapsto \int\limits_{+\infty}^{-1/4} \mathrm{d}\xi + \int\limits_{-1/4}^{+\infty} \mathrm{d}\xi \ . \tag{3.17}$$

The Jacobian and fermionic determinant have cancelled against each other *without* the introduction of a sign. The point to note here is that a relative sign arises between the two contributions in (3.17) only when one declares the limits of integration for ξ to be in a positive sense. On doing this we find that the right hand side becomes

$$\int_{-1/4}^{+\infty} d\xi\,(-1+1) = 0\,, \tag{3.18}$$

where the first sign reflects the fact that the ratio of determinants gives -1 when one runs through the ξ space in a "negative" direction. To ascertain if one is integrating over ξ in a positive or negative sense we only need to find the zeros of ξ; the tangent at those points (with ϕ running from $-\infty$ to $+\infty$) will then give the appropriate sign. Since the tangent is $\partial^2 V/\partial\phi\partial\phi$ we see immediately that the sign is exactly the ratio of determinants previously considered. Evaluating this ratio at the points $\xi = \partial V/\partial\phi = 0$ we recover the result (3.18). The zeros of ξ are at -1 and 0, and the signs at these points are $-$ and $+$, respectively.

(ii) $\partial V(\phi)/\partial\phi = \phi - \frac{1}{3}\phi^3 = \xi$. Following the same analysis as in the previous example, we obtain

$$\int_{-\infty}^{-1} d\phi + \int_{-1}^{1} d\phi + \int_{1}^{+\infty} d\phi \mapsto \int_{+\infty}^{-2/3} d\xi + \int_{-2/3}^{+2/3} d\xi + \int_{+2/3}^{-\infty} d\xi\,. \tag{3.19}$$

We note that in this case, as ϕ runs over its range, ξ begins to run over its domain, backtracks and then proceeds to the end covering its domain once only. In terms of the zeros of ξ, the analysis is the same as before. At the three zeros $(-\sqrt{3}, 0, \sqrt{3})$, the signs of the tangents are $(-, +, -)$, which on addition give the result that we cover ξ-space once, and in a negative sense.

The general result is that for such a model the integration may be expressed, up to a normalization, as

$$\sum_{\{P\}} \text{sign}(V'')\,, \tag{3.20}$$

where P are the turning points of V, so that for $V(\phi) \propto \phi^n$ as $\phi \to \pm\infty$ this sum vanishes for n odd and gives 1 for n even.

Returning to our path integral, we see that had we kept the path integral form of the action and simply dropped the $d\phi/d\tau$ and $d\psi/d\tau$ terms; this would have been the analogue of the finite dimensional integral (3.20); such a limit gives what is known as the ultra local form of the theory. In this limit at least, the partition function is the obvious generalization of (3.20),

$$Z = \sum_{\{P\}} \text{sign det}(\partial^2 V/\partial\phi\partial\phi)\,. \tag{3.21}$$

To determine the range of ξ in general is also straightforward. We can think of the path integral as products of finite dimensional integrals, one for each time instant. Fix a time instant and check the range of ξ. In the derivation of the path integral, at each time instant one has a complete set of position eigenstates $|\phi\rangle$; the question we have posed is: may one insert a complete set of ξ eigenstates at each of these times? For this to be possible, the overlap of the states (the wave functions in the ξ representation) must satisfy from (3.13)

$$\left(-i\frac{\partial}{\partial\phi} \pm is\frac{\partial V(\phi)}{\partial\phi}\right)\langle\xi|\phi\rangle = \xi\langle\xi|\phi\rangle \tag{3.22}$$

(recall that $\dot{\phi}$ type terms appear in the action because of $\partial/\partial\dot{\phi}$ derivatives). The solution is

$$\langle\xi|\phi\rangle = A\,e^{[\pm sV(\phi)+i\xi\phi]}\ ,\tag{3.23}$$

where A is a constant. It remains to be checked whether these states are normalizable (square integrable),

$$\int_{-\infty}^{+\infty}d\phi\,\langle\xi|\phi\rangle^{*}\langle\xi|\phi\rangle = \int_{-\infty}^{+\infty}d\phi\,e^{\pm 2sV(\phi)} < +\infty\ .\tag{3.24}$$

If $V(\phi)\rightarrow\phi^{2n}$ as $\phi\rightarrow\pm\infty$, then there are states which are normalizable[*]; on the other hand, these states are not normalizable if $V(\phi)\rightarrow\phi^{2n+1}$ [3.13]. This agrees with our previous results in examples (i) and (ii) above.

Do we also reproduce these results if we count the number of times that ξ passes through zero? From (3.13) we see that the zeros of ξ are at the classical configurations and (3.10) implies that these correspond to the critical points of the potential $V(\phi)$. We may indeed think of the Nicolai map as leading to a partition function which gives the "degree" of the map, counting the number of distinct configurations of the original fields which are mapped to a given configuration of the Gaussian fields with their algebraic multiplicity. The latter may be calculated by following the zeros of ξ; once more we find that

$$Z = \sum_{\{P\}}\text{sign}(\partial^{2}V/\partial\phi\,\partial\phi)\ .\tag{3.25}$$

The path integral (3.15) calculated about $\xi = 0$ becomes

$$\sum_{\{P\}}\text{sign}\det(d/d\tau + s\,\partial^{2}V/\partial\phi\,\partial\phi)\ .\tag{3.26}$$

This may be given another path integral representation with the action

$$S = \oint d\tau\left[i\left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial\phi}\right)B - i\bar\psi\left(\frac{d}{d\tau} + s\frac{\partial^{2}V}{\partial\phi\,\partial\phi}\right)\psi\right]\ ,\tag{3.27}$$

which is the original action without a B^2 term. In section 3.4 we will explicitly show why these two actions yield equivalent quantum theories. This is one of the important properties of topological field theories; namely, that many of the details of the action are irrelevant. The integral over B gives a delta function restriction to the instanton paths, which, as we have seen, corresponds to the critical points of $V(\phi)$. Expanding about the critical points gives (3.26). To see that (3.26) devolves to (3.25), expand each field in a Fourier series, $\phi = \sum_{n=-\infty}^{\infty}\phi_n e^{in\tau}$, with $\phi_n^* = \phi_{-n}$, etc. For each mode n there is a contribution to the determinant of the form $(in + V_n'')$ from the bosonic variables. From the fermions, on the other hand, we get $\pm(in + V_n'')$. For each $n > 0$ there is an $n < 0$ so that the $\pm$ arising from the fermion integration is always squared. The sign then comes strictly from the $n = 0$ terms, leaving us with (3.25).

3.2.2. General model

For the general theory, the squaring argument shows that the absolute minima of the action are

$$d\phi^{i}/d\tau = 0,\quad s\,\partial V/\partial\phi^{i} = 0\ .\tag{3.28}$$

[*] If s is set to zero there are no normalizable states at all.

Once more, when $s \neq 0$, the relevant points are the critical points of V; on the other hand, when $s = 0$, all the points of the target manifold enter.

We need to distinguish between situations where R_{ijkl} may, or may not, be ignored. When R_{ijkl} is ignorable, as for example, when the target manifold is $\mathbb{R}^n$, the Nicolai map

$$\xi^i = \mathrm{d}\phi^i/\mathrm{d}\tau + sg^{ij}(\phi)\,\partial V/\partial \phi^j \ , \tag{3.29}$$

is, as before, such that the Jacobian of the map cancels the absolute value of the fermionic Pfaffian. The form of the Jacobian may be determined in the following manner. The Jacobian measures the deformation of the (canonical) vector field $\dot{\phi}^i + sg^{ij}(\phi)\,\partial_j V(\phi)$ as the path $\phi(\tau)$ is deformed. To compare the new tangent vector with the old, it must be parallel transported to the original point; in this way one finds that the Jacobian has precisely the same form as the Pfaffian. Alternatively, one may use the method of normal coordinates to prove this result; this is presented in detail for the sigma models of section 4.

The path integral reduces to

$$\sum_{\{P\}} \mathrm{sign}\,\det(H_P V) \ , \tag{3.30}$$

where $H_P V = \partial^2 V/\partial \phi^i\,\partial \phi^j$ is called the Hessian of V at the point P. It is clear that (3.30) is the natural generalization of the toy model result (3.20). This formula is related to the Poincaré–Hopf theorem when the target manifold is taken to be compact and closed, as explained in section 3.8.4.

As we observed in the previous section there are no normalizable modes when the potential is taken to be zero (or $s = 0$). The limit $s \to 0$ needs to be taken with care [3.6]. However, such a choice of potential causes no undue difficulties when the curvature tensor does not vanish. On the other hand, the Nicolai map (3.29) would appear not to trivialize the theory in this case, owing to the very presence of the curvature term in the action (3.1). Related to this is the fact that the Nicolai map in this instance is singular; it is singular as the map vanishes on all the constant paths ϕ_0^i, the space of zeros being the target manifold M itself. In section 3.8.5 it is shown that the curvature term is ignorable for all paths *except* the constant ones. Bearing this in mind, one may perform the Nicolai map (3.29) with the instruction not to include constant paths. The path integral over the non-constant paths simply gives one, leaving only the constant paths to be dealt with. The final result is spelt out in section 3.8.5 and is related to the Gauss–Bonnet theorem.

3.3. Langevin approach

Having shown how to trivialize the theory with the use of the Nicolai map, we would now like to give a method for creating the theory from the same map. This relies on the notion of a Langevin equation, which is connected to much older ideas in field theory. Parisi and Wu [3.14], Parisi and Sourlas [3.15] and also Cecotti and Girardello [3.16, 3.17] introduced some supersymmetric models that are related to classical stochastic equations. Both groups go on to show the connection of these equations to the Nicolai maps which trivialize the respective models. However, the theories that they introduced are non-trivial in *low dimensions* (< 4). They were not able to repeat the construction in high dimensions and this prompted Parisi and Sourlas to remark, "At this stage we feel like wizards who succeed in their first sorcery but are unable to do it again ". Topological field theories may be considered to be the extension to any dimension that these authors were searching for. An equation of the form

$$\xi = \mathrm{d}\phi/\mathrm{d}\tau + s\,\partial V/\partial \phi \tag{3.31}$$

is known as a Langevin equation and the method developed here is called the Langevin approach. The time that enters in (3.31) is a stochastic time variable and would be an "extra" variable in the theory, but in these models it is taken to be real time. In later sections, the Langevin approach will be further elaborated on, but rather than picking out time as a special direction, that is, choosing a manifold of the form $M \times \mathbb{R}$, we will be mainly concerned with a covariant version of the equation.

Our presentation of these ideas is slightly different, but equivalent to that of refs. [3.15, 3.16]. The aim is to run through the trivialization of eqs. (3.11) and (3.12) backwards. We now begin with a trivial Gaussian action

$$S_0 = \frac{1}{2} \oint d\tau \, (G - \zeta(\phi))^2 \, , \qquad (3.32)$$

where $\zeta(\phi)$ is the Nicolai map that was used to trivialize (3.11) and G is an auxiliary field. It is clear that we could easily shift G and eliminate any dependence of the action on ϕ [i.e., define $G' = G - \zeta(\phi)$]. Then we would be left with a Gaussian integration over the G' field, but also an unweighted (infinite) integral over ϕ. This is analogous to the situation which arises in gauge theories; the gauge directions are not weighted and the gauge group volume needs to be factored out to obtain sensible results. In that case one uses the gauge invariance to fix the gauge, thereby factoring out the group volume. In the process the Faddeev–Popov ghosts are engendered; these in turn allow one to reinstate the old gauge invariance in a new guise, namely as a BRST symmetry.

For us then, the problem is to identify the gauge invariance of the action, obtain the corresponding BRST symmetry, and then to choose an appropriate gauge condition. This is surprisingly easy to achieve. First, the gauge invariance is the largest one can possibly have, namely an arbitrary shift symmetry in the ϕ field. This ought not to come as a surprise, as after all we could arrange things in the action so that the ϕ field does not make an appearance. The transformation for G follows on insisting that the action is invariant under this shift. The symmetry reads

$$\delta\phi = \lambda \, , \qquad \delta G = \frac{\partial \zeta}{\partial \phi} \lambda \, . \qquad (3.33)$$

We would like to arrive at the starting action that appears in (3.11); with this in mind we choose the gauge $G = 0$, which can be achieved with the above transformations if $\delta\zeta(\phi)/\delta\phi$ is non-singular (by this we mean that, given an arbitrary G, a gauge transformation can be made which maps it to zero). If we carry out the BRST quantization of (3.33) in this gauge, the resulting partition function can be formally expressed as

$$Z_0 = \int_{\phi} e^{-S(\phi)} \, \Delta_{FP} \, , \qquad (3.34)$$

where Δ_{FP} is the associated Faddeev–Popov ghost determinant or Pfaffian. Indeed Δ_{FP} is precisely the Pfaffian appearing in (3.11), since it represents the inverse to the Jacobian of the map $\phi \to \zeta(\phi)$.

While this technique seems only to reproduce those theories that have fermions entering in the action quadratically, that is, without cubic or quartic interactions among them, this is not the case. A careful analysis of the gauge fixing procedure establishes that this method is *generic*. We will see how this works when we develop the techniques required to deal with the general model. Our next categorization of topological field theories, essentially the inverse of the previous one, is:

Witten type topological field theories are obtained from the quantization of the Langevin equation.

3.3.1. Toy model

Let us now fill in the details of the above analysis. How do we turn the symmetry (3.33) into BRST form? A first attempt is always to eliminate the transformation parameters in favor of the corresponding ghosts, and for this example the substitution suggested works well. So the BRST transformations read

$$\{Q,\phi\} = \psi \ , \qquad \{Q,G\} = (\partial\xi/\partial\phi)\psi \ , \qquad \{Q,\psi\} = 0 \ , \qquad \{Q,Q\} = 0 \ . \tag{3.35}$$

To this one must add an anti-ghost $\bar{\psi}$ and a Lagrange multiplier (or auxiliary) field B, which transform as

$$\{Q,\bar{\psi}\} = B \ , \qquad \{Q,B\} = 0 \ , \tag{3.36}$$

while retaining the nilpotency of the BRST operator Q. The gauge fixed action is then

$$S = \oint d\tau \, [\tfrac{1}{2}(G-\xi)^2 + i\{Q,\bar{\psi}G\}]$$

$$= \oint d\tau \left[\tfrac{1}{2}(G-\xi)^2 + iBG - i\bar{\psi} \left(\frac{d}{d\tau} + s\frac{\partial^2 V}{\partial\phi\partial\phi} \right) \psi \right] \ . \tag{3.37}$$

Integrating over B we get a delta function forcing G to vanish. Alternatively, let us shift G by ξ so that the bosonic terms in the action become

$$\oint d\tau \left[\tfrac{1}{2}G'^2 + iG'B + i \left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial\phi} \right) B \right] \ , \tag{3.38}$$

which on integrating over the G' field yields the action in the familiar form (3.5). Having eliminated the random field G, can we expect the action to remain BRST invariant? We know the answer is yes, but the reason is useful later so we pause to explain it. We eliminated G by performing the Gaussian integration; equivalently we could have eliminated it by its equation of motion,

$$G' = G - d\phi/d\tau - s\,\partial V/\partial\phi = -iB \ . \tag{3.39}$$

Since each term in this equation is invariant under the action of the BRST operator, we can use it without spoiling the symmetry. Thus starting from the Gaussian integration over a random field, we have been able to reproduce the original model, complete with its transformation properties.

3.3.2. General model

The complete action (3.1) may also be derived by gauge fixing a Langevin equation [3.6]. The generalization of (3.32) is the obvious candidate

$$S_0 = \frac{1}{2} \oint g_{ij}(\phi)K^iK^j \ , \tag{3.40}$$

where

$$K^i = G^i - \frac{d\phi^i}{d\tau} - sg^{ij}(\phi)\frac{\partial V(\phi)}{\partial\phi^j} = G^i - \xi^i \ . \tag{3.41}$$

Although the symmetry of this action is more complicated than that of the toy model, due to the necessity of introducing the metric to form the scalar product, it is, nevertheless, invariant under the following transformations:

$$\delta\phi^i = \lambda^i \ , \qquad \delta G^i = \frac{\partial\xi^i}{\partial\phi^j}\lambda^j - \Gamma^i_{jk}K^j\lambda^k \ . \tag{3.42}$$

It is not straightforward to turn this into a nilpotent BRST symmetry as the commutator of two such infinitesimal transformations does not close when acting on the G^i field. Instead one finds

$$[\delta(\lambda_2), \delta(\lambda_1)]G^i = \lambda_1^j \lambda_2^k R^i{}_{ljk} K^l \ . \tag{3.43}$$

As R_{ijkl} was zero in our previous example we did not run into this difficulty. One notices that, if the G^i equation of motion is used, then the right hand side of (3.43) vanishes, that is, the classical gauge symmetry closes to an Abelian algebra on-shell. In such a situation one is able to call upon the Batalin–Vilkovisky algorithm to produce a BRST invariant quantum action, with an on-shell nilpotent BRST operator. This is spelt out in appendix A; here we simply quote the results. The gauge fixed action is

$$S = S_0 + \oint d\tau \{ -i\bar{\psi}_i[(d/d\tau)\delta^i_j + s \partial_j(g^{ik}\partial_k V) - \Gamma^i_{jl}K^l]\psi^j$$

$$+ \tfrac{1}{4} R_{ijkl}\bar{\psi}^i\psi^k\bar{\psi}^j\psi^l + i\bar{B}_i G^i \} \ , \tag{3.44}$$

with the transformation rules

$$\{Q, \phi^i\} = \psi^i \ , \quad \{Q, \psi^i\} = 0 \ , \quad \{Q, \bar{\psi}_i\} = \bar{B}_i \ , \quad \{Q, \bar{B}_i\} = 0 \ ,$$

$$\{Q, G^i\} = d\psi^i/d\tau + s\partial_j(g^{ik}\partial_k V)\psi^j - \Gamma^i_{jk}K^j\psi^k - \tfrac{1}{2}iR^i{}_{jkl}\bar{\psi}^j\psi^k\psi^l \ . \tag{3.45}$$

By integrating over G^i, or equivalently using its equation of motion

$$G^i = -ig^{ij}\bar{B}_j - ig^{ij}\bar{\psi}_k\Gamma^k_{jl}\psi^l + \xi^i \ , \tag{3.46}$$

one eliminates G^i, while at the same time retaining the nilpotency of Q. The form of the action one obtains in this way is not quite that of (3.1); this is mirrored by the fact that the transformation rules (3.45) (the first four, as G^i no longer appears) do not match those of (3.4). But on making the substitution

$$\bar{B}_i = B_i - \bar{\psi}_j\Gamma^j_{ik}\psi^k \ , \tag{3.47}$$

one recovers both (3.1) and (3.4).

The important point is, of course, that we have been able to recover not just the quadratic terms in the fermionic fields but also the quartic couplings, and all of this by simply "quantizing" the Langevin equation.

3.4. Quantizing zero

Soon after Witten introduced his topological field theories [3.18, 3.19], Baulieu and Singer [3.20] and also Brooks, Montano and Sonnenschein [3.21] exhibited that these theories were indeed of BRST type. In conventional gauge theories one adds to the classical gauge invariant action a Q-exact piece, which gives the gauge-fixing and Faddeev–Popov terms. In topological field theories one has the Q-exact pieces but no other terms, leading to the idea that the classical action is zero[*].

Baulieu and Singer [3.22] also applied these ideas to supersymmetric quantum mechanics. We will describe their construction in this section. In particular, the methods give quick derivations of the topological field theory action; however, the idea that one is simply quantizing zero, will be called into question.

---

[*] One may relax this with the addition of a topological invariant to the zero action.

In any event, up to topological terms:
Witten type topological field theories have actions which are Q-exact.

3.4.1. Toy model

The idea here is to obtain (3.5) by taking zero as a starting action and gauge fixing this. Of course, more information is required. Firstly, the field content must be chosen, and this is taken to be, as above, one bosonic field ϕ. The second input is the symmetry, which is taken to be the largest possible, namely a shift in ϕ. This is then transformed into a BRST symmetry so the total field content and transformation rules of (3.6) are adopted.

The next step is to choose a gauge; we pick (none too surprisingly)

$$d\phi/d\tau + s\, \partial V/\partial \phi = 0 \ , \tag{3.48}$$

and this is implemented by adding the gauge fixing and ghost terms

$$\oint d\tau \left\{ Q, \bar{\psi} \left(i\frac{d\phi}{d\tau} + is\frac{\partial V}{\partial \phi} + \tfrac{1}{2}B \right) \right\}$$

$$= \oint d\tau \left[i \left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial \phi} \right) B + \tfrac{1}{2}B^2 - i\bar{\psi} \left(\frac{d}{d\tau} + s\frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right] \ . \tag{3.49}$$

Thus we are once more back to the action we began with. As discussed in section 2, an action that may be expressed in this manner leads to topological information.

This seems to be a satisfactory state of affairs, but let us look a little more closely at the question of gauge fixing. In conventional theories the situation is that any gauge choice is allowed *provided* there is a transformation that takes an arbitrary field to that condition. Here, with the philosophy that one is gauge fixing "zero", the complete action is made up of the gauge fixing and associated ghost terms. Different choices of gauge may well yield quite different theories. As an example, consider the choice of gauge

$$\phi = 0 \ . \tag{3.50}$$

Clearly, there is no way of smoothly deforming this to the classical paths of an arbitrary potential V that was made in choosing eq. (3.48). The philosophy that one is simply quantizing a zero action is then seen to be misleading. More correctly, one is really specifying the same data as we have done in the previous sections.

The advantages of this method should also be apparent. Provided one is careful in specifying the correct information, the action corresponding to the theory is quite quickly derived. Indeed, as long as one wishes to arrive at (3.48), there are other choices of gauge available. We do not wish to change the instanton equation, but that does not preclude us from tampering with the terms involving the B field. Instead of the $\{Q, \tfrac{1}{2}\bar{\psi}B\}$ term in (3.49) one may substitute $\{Q, \tfrac{1}{2}\alpha\bar{\psi}B\}$, where α is some free parameter. The action is

$$\oint d\tau \left[i \left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial \phi} \right) B + \tfrac{1}{2}\alpha B^2 - i\bar{\psi} \left(\frac{d}{d\tau} + s\frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right] \ , \tag{3.51}$$

and only the B^2 term is altered. Irrespective of the value of α, the action is still minimized by the instanton. Better than that, we see that for $\alpha = 0$, we recover the action (3.27), which on integrating over the B field gives a delta function constraint onto the classical trajectory. This parameter α also appears in the Langevin approach if one starts with an action of the form $\tfrac{1}{2}\alpha \oint G'^2$. The reader may now like to check that following the derivation from eq. (3.35) to (3.38), one obtains (3.51).

The particular choice of gauge $\alpha = 0$ will be useful, not only in supersymmetric quantum mechanics, but also in all other Witten type models; we will call it the delta function gauge.

3.4.2. General model

The fact that one needs to specify a whole host of information is brought home when one wishes to reproduce the general action (3.1). Rather than adopting the transformation rules (3.4), Baulieu and Singer [3.22] take the equivalent set [cf. eq. (3.45)]

$$\{Q, \phi^i\} = \psi^i \ , \qquad \{Q, \psi^i\} = 0 \ , \qquad \{Q, \bar{\psi}_i\} = \bar{B}_i \ , \qquad \{Q, \bar{B}_i\} = 0 \ , \tag{3.52}$$

which is obtained directly from (3.4) on the substitution $B_i \to \bar{B}_i + \bar{\psi}_j \Gamma^j_{ik} \psi^k$, cf. eq. (3.47). The action, generalized here by the inclusion of the potential, is taken to be

$$\oint d\tau \left\{ Q, \bar{\psi}_i \left(i\frac{d\phi^i}{d\tau} + i s g^{ij}(\phi) \frac{\partial V(\phi)}{\partial \phi^j} + \tfrac{1}{2} g^{ij}(\phi) \bar{B}_j + \tfrac{1}{2} g^{ij}(\phi) \bar{\psi}_l \Gamma^l_{jk} \psi^k \right) \right\} \ . \tag{3.53}$$

The reader may well see that this is a far from obvious choice. Nevertheless, if one simply shifts back to the original field B_i one sees that (3.53) is the natural *covariant* generalization of (3.49),

$$\oint d\tau \left\{ Q, \bar{\psi}_i \left(i\frac{d\phi^i}{d\tau} + i s g^{ij}(\phi) \frac{\partial V(\phi)}{\partial \phi^j} + \tfrac{1}{2} g^{ij}(\phi) B_j \right) \right\} \ , \tag{3.54}$$

and an application of the transformation rules (3.4) yields our starting action (3.1). Had one not included the affine term in (3.53) then the resulting action would not have been covariant. These non-covariant pieces appear naturally in the Langevin approach; one need only fix on the classical equation of motion that is of interest and the rest follows. However, here one must have some idea of what the outcome should be like; covariance *is* a natural demand, but this means that one must search for a gauge fixing piece which leads to this requirement.

Why do the non-covariant transformation rules (3.4) lead to a covariant action, while the apparently covariant transformations (3.52) do not (directly)? The answer lies in the fact that we may consider the action of Q on ϕ^i as a fermionic diffeomorphism; as such the fields should transform as tensors; they do not in either scheme. However, the affine terms in the transformation rules (3.4) ensure that covariants made from contraction with the metric remain covariant. For example,

$$\{Q, g^{ij}(\phi) B_i \bar{\psi}_j\} = g^{ij}(\phi) B_i B_j + \cdots \ , \tag{3.55}$$

where the affine terms combine with the derivative of the metric to yield the covariant derivative of the metric, which vanishes. We have not included the curvature term in the discussion as it is by itself covariant.

A direct advantage of having the BRST transformation rules as given in (3.52) is that with the new set of fields the BRST operator has no explicit metric dependence. The standard argument that the variation of the partition function with respect to the metric is the expectation value of a Q-exact correlator is now clear. In terms of the original fields this was not so apparent, as Q for this set of fields has an explicit metric dependence.

3.5. Metric independence

So far we have established the coupling constant independence of the theory. That is, we have shown that under small deformations of the potential V and the target metric g_{ij} the partition

function remains fixed. To establish the topological nature of the theory with respect to the base manifold (in this case S^1) we introduce the one-dimensional metric into the action (3.5). Let the einbein be given by

$$e = e(\tau)\,\mathrm{d}\tau \ . \tag{3.56}$$

All the fields that have entered into the theory $\phi, B, \psi, \bar{\psi}$, are taken to be S^1 scalars. $(\mathrm{d}\phi/\mathrm{d}\tau)\mathrm{d}\tau$ is clearly a one-form and so a good volume form; to make other scalars that appear in (3.5) and (3.1) good volume forms we simply multiply by the einbein. In this way we arrive at the covariant action

$$S = \oint \mathrm{d}\tau \left[\mathrm{i}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + se(\tau)\frac{\partial V}{\partial \phi}\right)B + \tfrac{1}{2}e(\tau)B^2 - \mathrm{i}\bar{\psi}\left(\frac{\mathrm{d}}{\mathrm{d}\tau} + se(\tau)\frac{\partial^2 V}{\partial\phi\partial\phi}\right)\psi \right] \ . \tag{3.57}$$

The symmetry transformation rules have not changed, they are still given by (3.6), with the action of Q on the einbein being zero. The BRST operator Q is also a S^1 scalar so that, as before, the action may be expressed as a Q-exact form. The analogue of (3.49) is

$$\oint \mathrm{d}\tau \left\{ Q, \bar{\psi}\left(\mathrm{i}\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + \mathrm{i}se(\tau)\frac{\partial V}{\partial\phi} + \tfrac{1}{2}e(\tau)B\right)\right\} \ , \tag{3.58}$$

with the metric independence of the partition function now being manifest. Moreover, the squaring argument shows that (3.10) is reobtained in the presence of the metric.

All of the manipulations previously performed hold with the one-dimensional metric included in the theory. We have shown that the model does not depend on this metric and hence is topological. For completeness, the reader may like to introduce this metric in the formulae for the general model; needless to say this leads one to the same conclusions as in the previous sections. We will have nothing more to say about the one-dimensional metric and henceforth will ignore it.

3.6. BRST symmetry and physical states

The degrees of freedom in a topological field theory of Witten type and a conventional supersymmetric field theory are quite different. In the former there are no physical degrees of freedom at all. This may seem a little strange since the models that we have been using to describe topological field theories are also supersymmetric theories in their own right. Thinking of them as topological requires that they have no degrees of freedom, while on the other hand thinking of them in conventional terms, one expects to have both bosonic and fermionic states.

There is no contradiction here, and the resolution lies in what Witten calls the "twisting" of the supersymmetry [3.18]. In these low dimensional examples the spin statistics theorem is not as restrictive as in four dimensions, so that the distinction between fermions and ghosts is not clearcut. Nevertheless, in the following we will discuss the properties of topological field theories in general. These general observations are then examined in detail for the toy model. To do this we will need to develop the Hamiltonian framework somewhat. This formalism will also be useful in later sections.

3.6.1. Physical states in supersymmetric theories

The counting of degrees of freedom in conventional supersymmetric theories is straightforward. One simply counts the appropriate number of bosonic states and then doubles so as to include the

fermionic contribution. The supersymmetry charges then give invertible maps between the states of non-zero energy

$$Q|b\rangle = |f\rangle , \qquad Q|f\rangle = |b'\rangle , \tag{3.59}$$

where $|b\rangle$ is any physical bosonic state and $|f\rangle$ a fermionic one. The vacuum, which may not be unique, is defined to be the state annihilated by all the supersymmetry charges.

3.6.2. Physical states in gauge theories

The presence of a BRST invariance in a field theory is the statement that the original model had some gauge symmetry. For Yang–Mills theories this is the conventional invariance under non-Abelian gauge transformations, for gravity it is diffeomorphism invariance, while for strings it corresponds to a combined diffeomorphism and Weyl invariance.

Physical states $|phys\rangle$ in these contexts are required to be gauge invariant. This translates into the statement that these states are annihilated by the (nilpotent) BRST operator [3.23]

$$Q|phys\rangle = 0 . \tag{3.60}$$

Furthermore, states which differ by a Q-exact piece $Q|\chi\rangle$, for any $|\chi\rangle$ are regarded as equivalent.

A further condition which is enforced on the physical states is that they carry zero ghost number; this is the analogue of fermion number in supersymmetric theories. The ghosts are meant to be purely fictitious, as their name suggests. This is in contradistinction to conventional supersymmetric theories where physical states can be bosonic or fermionic. For supersymmetric quantum mechanics, $(-1)^F$ acts as both the fermion and ghost charge. This charge commutes with the Hamiltonian and so states may be chosen to be simultaneous eigenstates of these operators.

For Yang–Mills theory, one counts d degrees of freedom coming from the vector potential (as it is a vector in d dimensions) and -1 for each of the two ghosts (as they are scalars and Grassmann odd); on adding we find $d - 2$ physical degrees of freedom, which is indeed correct for a gauge invariant vector field. For a rank-two antisymmetric tensor field one has the count $d(d - 1)/2 - 2d + 3$, the second term counting the two vector Grassmann ghosts and the third counting the three bosonic ghost-for-ghosts. The total is correctly $(d - 2)(d - 3)/2$.

The previous counting principle holds for second order theories. The counting in first order theories is somewhat different and is explained in the context of BF models in a later chapter. For supersymmetric quantum mechanics the counting is $d/2$ for each of ϕ and B, while the fermions treated as ghosts count $-d/2$ each[*]. The total is zero, implying that only the ground state is a physical state.

3.6.3. Physical states in topological field theories

The counting of degrees of freedom in topological field theories is actually a mixture of the above two. For every bosonic field there is a corresponding fermionic one. But rather than thinking about the fermions as "physical", they should be interpreted as ghosts; their degrees of freedom are then subtracted rather than added to the total. This leaves precisely no degrees of freedom. However, the vacuum—which is all that is left—need not have zero fermion number, i.e., one does not insist that $(-1)^F$ have eigenvalue 1 on vacuum states. This fact, that there may be many vacuum states each with different fermion content, leads to the various topological properties of these theories.

In higher dimensions this situation is actually forced on us. We will see that in all dimensions the objects that naturally appear are differential forms. It is because of this that one is able to

---

[*] This counting is heuristic as this is a zero-dimensional field theory.

formulate these theories on arbitrary base manifolds. Consequently in four dimensions and higher, by the spin statistics theorem, the anti-commuting fields cannot be physical; they are not spinors and so must be interpreted as ghosts.

The models we have been considering then in this chapter are not generic. This situation arises because in one or two dimensions the spin statistics theorem is not so restrictive; and therefore it is up to us to declare whether a fermionic field is deemed physical or ghost. The difference then lies in the conditions one imposes on the states. Using the same Q, one either demands that this maps bosons to fermions to bosons, giving one a conventionally supersymmetric theory, or that it annihilates physical states, as for a gauge theory, leaving only the vacuum, and hence yielding a topological theory.

This facet of the situation, namely that the same theory may well have two distinct interpretations, will be elaborated on in the sequel. It allows for very simple proofs of various otherwise miraculous properties of conventional theories.

3.6.4. The toy model in detail

The details of these distinctions are easy to determine for the toy model. To do this we need to determine the Hamiltonian form of the theory. First notice that when one eliminates B, by integrating it out of the path integral, the bosonic part of the action becomes

$$\frac{1}{2} \oint d\tau \left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial \phi} \right)^2 . \tag{3.61}$$

This is unaltered if one exchanges $-s$ for s, as the difference is $2s \oint d\tau \dot\phi V' = 0$. Furthermore, on making this sign change of s, the fermionic part of the action keeps its form if we simply exchange ψ and $\bar\psi$. We thus see that the action is invariant under the discrete transformations

$$s \to -s , \quad \psi \to \bar\psi , \quad \bar\psi \to \psi , \quad B \to B + 2is\,\partial V/\partial\phi , \tag{3.62}$$

where the B transformation is indicated for the form of the action where the B field has not been eliminated. There is then a secondary BRST symmetry, associated with this model, which is obtained by making the substitutions (3.62) in (3.6). We denote this new BRST operator by Q^*, its action on the fields being

$$\{Q^*,\phi\} = \bar\psi , \quad \{Q^*,\bar\psi\} = 0 , \quad \{Q^*,\psi\} = B + 2is\,\partial V/\partial\phi ,$$
$$\{Q^*,B\} = -2is(\partial^2 V/\partial\phi\partial\phi)\bar\psi , \quad \{Q^*,Q^*\} = 0 . \tag{3.63}$$

The Poisson brackets for the fields are readily read off from this first order action; the non-zero ones are

$$\{B,\phi\} = i , \quad \{\psi,\bar\psi\} = i . \tag{3.64}$$

Our general definition for the momenta, Hamiltonian, and Poisson brackets are

$$\Pi^i = \delta_r L/\delta\Phi_i , \quad -H = \Pi^i\dot\Phi_i - L ,$$
$$\{X,Y\} = \frac{\delta_r X}{\delta\Phi_i}\frac{\delta_\ell Y}{\delta\Pi^i} - (-1)^{XY}\frac{\delta_r Y}{\delta\Phi_i}\frac{\delta_\ell X}{\delta\Pi^i} , \tag{3.65}$$

where the subscripts r and ℓ denote right and left derivatives, respectively, and the peculiar sign for the Hamiltonian is chosen so that the spectrum of H is bounded below. Hence, with $\Phi = (\phi,\psi)$, the conjugate momenta are $\Pi = (iB, -i\bar\psi)$.

The Hamiltonian associated with the above action is

$$
\begin{aligned}
-H &= \mathrm{i}B\frac{\mathrm{d}\phi}{\mathrm{d}\tau} - \mathrm{i}\bar{\psi}\frac{\mathrm{d}\psi}{\mathrm{d}\tau} - L \\
&= -\tfrac{1}{2}B^2 - \mathrm{i}s\frac{\partial V}{\partial \phi}B + \mathrm{i}s\bar{\psi}\frac{\partial^2 V}{\partial \phi \partial \phi}\psi \\
&= \frac{1}{2}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau}\right)^2 - \tfrac{1}{2}s^2\left(\frac{\partial V}{\partial \phi}\right)^2 + \mathrm{i}s\bar{\psi}\frac{\partial^2 V}{\partial \phi \partial \phi}\psi \ ,
\end{aligned}
\tag{3.66}
$$

where the equation of motion for B was used to arrive at the last line. For our purposes, it is more convenient to use the form of the Lagrangian in which B is first eliminated (3.61). On neglecting a total time derivative in the Lagrangian, $\dot{\phi}$ becomes the momentum conjugate to ϕ; the relevant Poisson bracket is $\{\dot{\phi},\phi\} = -1$, and the appropriate form of the Hamiltonian is given in the last line of the above equation.

The BRST charges are obtained on following the Noether prescription, and their action is given on using the Poisson brackets (3.64)

$$
Q = -\mathrm{i}\psi B = -\psi\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + s\frac{\partial V}{\partial \phi}\right) \ ,
\tag{3.67}
$$

$$
Q^* = -\mathrm{i}\bar{\psi}\left(B + 2\mathrm{i}s\frac{\partial V}{\partial \phi}\right) = -\bar{\psi}\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} - s\frac{\partial V}{\partial \phi}\right) \ .
\tag{3.68}
$$

We note that (3.68) also follows directly from (3.67), on making use of the transformations (3.62). It is also worth pointing out the presence of the instanton (anti-instanton) projection operators in the definition of Q and Q^*. The Hamiltonian (3.66) may be rewritten as

$$
H = \tfrac{1}{2}\mathrm{i}\{Q,Q^*\} \ .
\tag{3.69}
$$

Now let us pass to a Schrödinger representation for the quantum mechanical version of the theory, and for definiteness set $V' = \lambda\phi$ and $M = \mathbb{R}$. The Hamiltonian takes the form

$$
2H = -\partial^2/\partial\phi^2 + s^2\lambda^2\phi^2 + s\lambda[\partial/\partial\psi, \psi] \ .
\tag{3.70}
$$

We have used the substitution $\bar{\psi}\psi \to [\bar{\psi},\psi]/2$ and $\dot{\phi} \to \partial/\partial\phi$, which is appropriate for the Euclidean quantum theory. Wave functions will take the form $\Psi(\phi,\psi) = F_0(\phi) + \psi F_1(\phi)$. Simply setting F_1 to be zero we find that F_0 may chosen to be the usual simple harmonic oscillator eigenfunctions, except that with this Hamiltonian their energy is displaced by $s\lambda$. On the other hand, keeping F_1 and setting F_0 to be zero, we see that this time F_1 may be chosen to be the simple harmonic oscillator wave functions with energy eigenvalues displaced by $-s\lambda$.

Let us determine the physical states when treating this theory as a conventional supersymmetric model. Set $F_1 = 0$, and take F_0 to be any one of the simple harmonic oscillator eigenfunctions. This state is clearly bosonic; its supersymmetric partner is

$$
QF_0 = \psi\,\partial F_0/\partial\phi \ .
\tag{3.71}
$$

Providing the eigenvalue is not zero, and as the operator Q commutes with the Hamiltonian, there is always a supersymmetric partner to F_0 which shares the same eigenvalue. By making use of Q^* one may similarly conclude that for all fermionic eigenvectors with non-zero eigenvalue there are

corresponding bosonic eigenstates. There is then a complete tower of bosonic and fermionic energy eigenstates that forms the Hilbert space of this theory.

Taking the attitude that the toy model is really a gauge theory would lead us to define physical states as those that satisfy the conditions

$$Q\Psi(\phi, \psi) = 0 , \qquad [(-1)^F - 1]\Psi(\phi, \psi) = 0 . \tag{3.72}$$

The second of these implies that Ψ does not depend on ψ. The first equation, with the simple harmonic oscillator potential, has a unique solution

$$\Psi(\phi) = e^{-\frac{1}{2}s\lambda\phi^2} , \tag{3.73}$$

which is normalizable if and only if $s\lambda > 0$. If this bound is met then the theory is comprised of just this state, otherwise the theory is empty. The physical Hilbert space has been truncated from the tower of states to only a portion of the zero energy space. We may also see this directly from the form of the Hamiltonian (3.70); the conditions (3.72) imply that the Hamiltonian annihilates the physical states.

Finally, let us analyze the content of this theory when we treat it as a topological field theory. In this case we impose the first of the conditions of (3.72) but not the second. This implies that only F_0 is determined. However, the undetermined part ψF_1 is Q-exact in this example,

$$QF_2 = \psi F_1 , \qquad F_2 = -e^{-sV(\phi)} \int^{\phi} d\sigma \, F_1(\sigma) \, e^{sV(\sigma)} , \tag{3.74}$$

and hence cohomologically trivial. The only allowed state is then essentially the same as we found when treating the theory as a gauge model. On the other hand, when $M = S^1$, it is not possible to show that F_1 is necessarily Q-exact. It is convenient instead to fix on a representative of the equivalence class. To this end, we may also impose the condition that

$$Q^*\Psi(\phi, \psi) = 0 . \tag{3.75}$$

The wavefunction is now completely specified,

$$\Psi(\phi) = A e^{-\frac{1}{2}s\lambda\phi^2} + B\psi \, e^{+\frac{1}{2}s\lambda\phi^2} . \tag{3.76}$$

If $s\lambda > 0$ then the bosonic part of the wave function is normalizable while the fermionic part is not; one must choose $B = 0$. For $s\lambda < 0$, the situation is reversed and one must fix $A = 0$. The "physical" states are now somewhat different. We have seen that in general the path integral of interest boils down to looking at how it behaves about each critical point of the potential V. Around each critical point the potential is essentially a constant plus a simple harmonic oscillator term. The signs of the simple harmonic parts are intimately related to the topology of the target manifold. The construction of the ground states as in (3.76) plays an important role in Witten's generalization of Morse theory, as described in section 3.10. It is important to note that, if we treat these models as gauge theories, then at those points where the simple harmonic potential has a "bad" sign there are *no* ground states.

The fact that we are dealing with only the ground state space for any potential V is also clear. The conditions that Q and Q^* annihilate physical states obviously imply that the Hamiltonian also annihilates these states. The converse will be demonstrated in the next section.

3.7. The Witten index

In this section we will use some of the properties of topological field theories to give easy evaluations of the Witten index $\mathcal{W}$. The possibility of the dynamical breaking of supersymmetry is of considerable importance if supersymmetric theories are to play a role in a description of nature. A reliable method of determining when such a breakdown occurs in a given supersymmetric model, is then required. Witten has given a necessary criterion for this breaking; the index $\mathcal{W}$ must vanish for the supersymmetry to be broken [3.1]. As our aim is to evaluate the index, we will content ourselves here with a rapid description of the ideas leading to it.

The basic idea is to count the difference in the number of bosonic and fermionic ground states; the Witten index is defined to be that difference,

$$\mathcal{W} = (\text{\# of bosonic zero energy states}) - (\text{\# of fermionic zero energy states}) . \tag{3.77}$$

To count the zero energy states it is enough to restrict one's attention to the $P = 0$ states as the energy E is equal to or greater than the magnitude of the momentum $|P|$. The counting of the zero energy states is facilitated by considering the theory in a finite spatial volume with periodic boundary conditions on all the fields. On such a (now compact) manifold the eigenstates of the Hamiltonian will be discrete and hence may be counted in a well defined way. Periodic boundary conditions are imposed on both the fermionic and bosonic fields to ensure that the constant modes of these fields match, preserving the supersymmetry. The Hamiltonian may be expressed as the square of any of the Hermitian supersymmetry charges,

$$H = Q_1^2 = Q_2^2 = \cdots = Q_N^2 , \tag{3.78}$$

where N depends on the dimension of the space-time, and on whether the supersymmetry is an extended supersymmetry or not. Ground states $|0\rangle$ are determined by the condition

$$H|0\rangle = 0 , \tag{3.79}$$

which in turn implies that each of the Q_i annihilate the ground state, since

$$0 = \langle 0|H|0\rangle = \langle 0|Q_i^2|0\rangle = \|Q_i|0\rangle\|^2 . \tag{3.80}$$

As the Q_i are Hermitian operators, the squares only vanish under the condition that

$$Q_i|0\rangle = 0 . \tag{3.81}$$

If the index does not vanish, it implies that there is at least one zero energy state, which is then an appropriate supersymmetric ground state. A necessary criterion for supersymmetry breaking is that $\mathcal{W}$ should vanish; otherwise, there is a supersymmetric invariant ground state and the supersymmetry is preserved.

We now show that all eigenstates of the Hamiltonian of non-zero eigenvalue come in pairs; for each bosonic mode $|b\rangle$ there is a corresponding fermionic mode $|f\rangle$, and vice versa. Consider one of the supersymmetry charges, say Q_1. On a bosonic state with energy eigenvalue E that is not a ground state, we have that

$$Q_1|b\rangle = \sqrt{E}|f\rangle , \qquad Q_1|f\rangle = \sqrt{E}|b\rangle , \tag{3.82}$$

so that $|f\rangle$ is also an eigenmode with eigenvalue E. We will use this fact in the next section.

3.7.1. Path integral representation

Witten [3.1] and Cecotti and Girardello [3.13] have given a path integral representation of W; it is straightforward to do this. First, one rewrites (3.77) as

$$W = \mathrm{Tr}[(-1)^F \, \mathrm{e}^{-\beta H}] \, , \tag{3.83}$$

where the trace is over all the eigenmodes of the Hamiltonian and $(-1)^F$ is plus or minus 1 depending on whether the eigenstate is bosonic or fermionic, respectively. As the positive energy modes come in pairs, but with opposite sign for $(-1)^F$, they cancel in the trace; thus (3.83) and (3.77) agree.

The path integral representation of (3.83) is, using standard techniques, given by

$$W = \int_\Phi \mathrm{e}^{-S(\Phi)} \, , \tag{3.84}$$

where Φ denotes all the fields, $\int_\Phi$ is the integral over the function space of fields and periodic boundary conditions are taken for both the space and time coordinates: $\Phi(t + \beta) = \Phi(t)$ [3.4]. The reason for this condition in the time direction for the fermionic fields is because of the presence of the factor $(-1)^F$ in the trace.

What does this mean for the supersymmetric theory that we have been considering? To calculate the Witten index we consider the fields in the path integral to be periodic in time. To introduce the β parameter, one notes that it represents the upper limit of the time integration range $\int_0^\beta \mathrm{d}t$. If we scale the time by β, we may consider the time to lie on the unit circle. The action (3.1) in this way becomes

$$S = \beta \oint \mathrm{d}\tau \left[\mathrm{i} \left(\frac{1}{\beta} \frac{\mathrm{d}\phi^i}{\mathrm{d}\tau} + s g^{ij}(\phi) \frac{\partial V}{\partial \phi^j} \right) B_i + \tfrac{1}{2} g^{ij}(\phi) B_i B_j + \tfrac{1}{4} R_{ijkl} \bar\psi^i \psi^k \bar\psi^j \psi^l \right.$$
$$\left. - \mathrm{i} \bar\psi_i \left(\frac{1}{\beta} \delta^i_j \frac{\mathrm{D}}{\mathrm{d}\tau} + s g^{ik}(\phi) \frac{\mathrm{D}^2 V}{\mathrm{D}\phi^k \mathrm{D}\phi^j} \right) \psi^j \right] \, . \tag{3.85}$$

As we know that the path integral does not depend on the value of s, following the general arguments of section 2, we scale it also by β to put the action in the form

$$S = \oint \mathrm{d}\tau \left[\mathrm{i} \left(\frac{\mathrm{d}\phi^i}{\mathrm{d}\tau} + s g^{ij}(\phi) \frac{\partial V}{\partial \phi^j} \right) B_i + \frac{\beta}{2} g^{ij}(\phi) B_i B_j + \frac{\beta}{4} R_{ijkl} \bar\psi^i \psi^k \bar\psi^j \psi^l \right.$$
$$\left. - \mathrm{i} \bar\psi_i \left(\delta^i_j \frac{\mathrm{D}}{\mathrm{d}\tau} + s g^{ik}(\phi) \frac{\mathrm{D}^2 V}{\mathrm{D}\phi^k \mathrm{D}\phi^j} \right) \psi^j \right] \, . \tag{3.86}$$

This form of the action will also be useful to us when we consider a field theoretic proof of the Gauss–Bonnet theorem, to be found in section 3.8. For now, we consider the target manifold to be $\mathbb{R}$ so that (3.86) reduces to

$$S = \oint \mathrm{d}\tau \left[\mathrm{i} \left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + s \frac{\partial V}{\partial \phi} \right) B + \frac{\beta}{2} B^2 - \mathrm{i} \bar\psi \left(\frac{\mathrm{d}}{\mathrm{d}\tau} + s \frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right] \, . \tag{3.87}$$

This is identical to (3.51) with $\alpha = \beta$. Since the path integral is insensitive to the value of β, as we have just argued, we see that the theories agree! We have therefore already calculated, in various ways, the Witten index for this theory. It is zero for potentials $V(\phi) \to \phi^{2n+1}$ as $\phi \to \pm\infty$ and non-zero otherwise, for polynomial V. This is in agreement with Witten's calculation [3.24].

3.8. The Euler character

As stated in the introduction, the supersymmetric quantum mechanics action may be used to determine the Euler character of the target manifold (taken to be compact). In this section we would like to exhibit this in some detail. Along the way one is able to prove the Poincaré–Hopf theorem that relates the Euler character to the zeros of vector fields on the manifold, and the Gauss–Bonnet theorem which gives the Euler number as an integral over the Euler class, that is, in terms of the curvature tensor. We are able to do this by making various judicious choices of gauge; since the results are gauge independent the theorems follow.

The definition of the Euler character $\chi(M)$ for a n-dimensional compact manifold M without boundary is

$$\chi(M) = \sum_{i=0}^{n}(-1)^i b_i(M) \;, \tag{3.88}$$

where b_i is the ith Betti number of the manifold. These are defined to be the dimensions of the ith cohomology (homology) groups $H^i(M,\mathbb{R})$ ($H_i(M,\mathbb{R})$). By Poincaré–Hodge duality $b_i = b_{n-i}$, and the Hodge theorem equates the dimension of $H^i(M,\mathbb{R})$ to the number of independent harmonic forms of degree i.

3.8.1. A brief review of de Rham theory and Witten's generalization
To set notation, let

$$d : \Omega^i(M,\mathbb{R}) \to \Omega^{i+1}(M,\mathbb{R}) \tag{3.89}$$

be the exterior derivative which maps smooth differential forms (with values in $\mathbb{R}$) to smooth differential forms of one degree higher and $d^2 = 0$. The inner product of two i-forms is defined with the help of the Hodge star operator $*$,

$$\langle \alpha_i, \beta_i \rangle = \int_M \alpha_i * \beta_i \;, \quad \alpha_i, \beta_i \in \Omega^i(M,\mathbb{R}) \;, \tag{3.90}$$

where the Hodge star operator maps i-forms into $(n-i)$-forms and requires the introduction of a metric on M,

$$(*\alpha_i)_{j_1\ldots j_{n-i}} = \frac{1}{i!}\sqrt{g}\,\epsilon_{j_1\ldots j_{n-i}k_1\ldots k_i}\alpha_i^{k_1\ldots k_i} \;, \qquad (*)^2 = (-1)^{(n-i)i} \;, \tag{3.91}$$

where $g = \det(g_{ij})$ and g_{ij} is the Riemannian metric of the manifold; indices in (3.91) are raised and lowered with this metric.

The adjoint of d, d^* is defined via

$$\langle d\alpha_i, \beta_{i+1} \rangle = \langle \alpha_i, d^*\beta_{i+1} \rangle \;. \tag{3.92}$$

It is thus an operator which maps smooth differential forms to smooth differential forms of one degree less,

$$d^* : \Omega^i(M,\mathbb{R}) \to \Omega^{i-1}(M,\mathbb{R}) \;, \tag{3.93}$$

and satisfies $(d^*)^2 = 0$.

The Laplacian is

$$\Delta = (d + d^*)^2 \ , \tag{3.94}$$

and clearly maps i-forms into i-forms. Harmonic forms—by definition—satisfy the zero eigenvalue Laplace equation

$$\Delta \gamma_i = 0 \ . \tag{3.95}$$

The Hodge theorem asserts that any i-form may be decomposed uniquely as the sum of an exact, a co-exact and a harmonic form,

$$\alpha_i = d\lambda_{i-1} + d^*\beta_{i+1} + \gamma_i \ . \tag{3.96}$$

Note that a harmonic form γ satisfies $d\gamma = d^*\gamma = 0$. This may be proved by noting that, if γ is harmonic, then $\langle \gamma, \Delta\gamma \rangle = 0$, which may be expressed as

$$\langle \gamma, \Delta\gamma \rangle = 0 = \langle d\gamma, d\gamma \rangle + \langle d^*\gamma, d^*\gamma \rangle \ , \tag{3.97}$$

where we have made use of (3.92). Since the right hand side is a sum of squares, the assertion follows. One immediate consequence of this fact is that the decomposition (3.96) is orthogonal.

Let $Z^i = \{\alpha_i : d\alpha_i = 0\}$ be the space of closed i-forms and $B^i = \{\beta_i : \beta_i = d\lambda_{i-1}\}$ denote the space of exact i-forms. Then the de Rham cohomology groups are defined by

$$H^i(M,\mathbb{R}) = Z^i/B^i \ . \tag{3.98}$$

By virtue of the Hodge decomposition, on a compact manifold without boundary, these cohomology groups are isomorphic to the spaces of harmonic forms $\text{Harm}^i(M,\mathbb{R})$

$$H^i(M,\mathbb{R}) \cong \text{Harm}^i(M,\mathbb{R}) \ . \tag{3.99}$$

Witten noticed that one may profitably generalize the above constructions [3.2]. Let d_s be given by

$$d_s = e^{-sV} d\, e^{sV} \ , \tag{3.100}$$

where $V : M \to \mathbb{R}$ is a Morse function on the manifold, that is, V only has isolated critical points and these must be non-degenerate, i.e., $\det(V'') \neq 0$ at these points. The adjoint is defined similarly

$$d_s^* = e^{sV} d^* e^{-sV} \ . \tag{3.101}$$

Both d_s and d_s^* are easily seen to be nilpotent.

The cohomology groups defined by d_s are isomorphic to the de Rham groups and we denote them by $H^i(s)$. The isomorphism follows from the fact that the operators are related by conjugacy. Let $\psi \in H^i(0)$ then $d\psi = 0$ and $\psi \neq d\chi$. Set $\psi_s = e^{-sV}\psi$ so that $d_s\psi_s = e^{-sV}d\psi = 0$. Also $\psi_s \neq d_s\chi_s$ for any χ_s. Namely, suppose the opposite, that $\psi_s = d_s\chi_s$, then $e^{-sV}\psi = e^{-sV}d(e^{sV}\chi_s)$. But then ψ is exact, which is a contradiction. This establishes that $\psi_s \in H^i(s)$. The roles of d and d_s may be interchanged in the above argument and this gives us the isomorphism.

In particular, the dimensions of the cohomology groups match, and once more by the Hodge theorem, the cohomology groups are isomorphic to the space of harmonic forms. The Laplacian in this general case is

$$\Delta_s = (d_s + d_s^*)^2 \ . \tag{3.102}$$

To express this in a compact though explicit form, we introduce creation a^{*i} and annihilation operators a^i at each point p of M. These satisfy the algebra

$$\{a^i, a^{*j}\} = g^{ij} \ , \tag{3.103}$$

and have the following geometric interpretation. The $a^i(p)$ may be thought of as forming an orthonormal basis of tangent vectors at the point p. As operators, they act on the exterior algebra at p by interior multiplication. The $a^{*i}(p)$ are the adjoint operators, acting by exterior multiplication by the one-form dual to $a^i(p)$. In this basis we have

$$d\alpha = a^{*i}\, \partial\alpha/\partial\phi^i \ , \qquad d^*\alpha = -a^i\, \partial\alpha/\partial\phi^i + \Gamma_{ijk} a^i a^{*j} a^k \alpha \ . \tag{3.104}$$

The last of these is easily checked to be of the correct form so that its action on $\alpha_{i_1\ldots i_r} a^{*i_1}\cdots a^{*i_r}|0\rangle$ matches that of d^* on $\alpha_{i_1\ldots i_r}\, d\phi^{i_1}\cdots d\phi^{i_r}$, where the ϕ^i are local charts about p.

With this notation, (3.102) becomes

$$\Delta_s = \Delta + s^2 g^{ij}\frac{\partial V}{\partial\phi^i}\frac{\partial V}{\partial\phi^j} + s\frac{D^2 V}{D\phi^i D\phi^j}[a^{*i}, a^j] \ , \tag{3.105}$$

where D represents the covariant derivative with respect to the metric g_{ij}. The reader will recognize this as the Hamiltonian (3.66), (3.70) when the manifold is taken to be one dimensional.

3.8.2. Path integral representation of the Euler character

Following the analysis of the previous section, the Euler character may now be expressed as

$$\chi(M) = \sum_{i=0}^{n} (-1)^i b_i(s) \ , \tag{3.106}$$

where $b_i(s)$ are the Betti numbers of the $H^i(s)$ cohomology groups, since $b_i(s) = b_i(M)$. It thus makes sense to rewrite this once more as

$$\chi(M) = \text{Tr}_h[(-1)^F] \ , \tag{3.107}$$

where the trace is restricted to be over the harmonic modes of Δ_s; $(-1)^F$ gives $+1$ on even forms and -1 on odd forms, and commutes with the Laplacian $[\Delta_s, (-1)^F] = 0$. It will not have escaped the reader's notice that this construction is almost identical to that for the Witten index of section 3.7.

To put the trace in a more useful form, one would like to relax the restriction that it be only over the harmonic modes; rather, we would like to extend the trace to be over all eigenvalues of the Laplacian. This is possible as the non-zero modes are paired with opposite eigenvalues of $(-1)^F$. To see this, let

$$K_s = d_s + d_s^* \ . \tag{3.108}$$

This operator enjoys the following properties:

$$\Delta_s = K_s^2 \ , \qquad [\Delta_s, K_s] = 0 \ , \qquad \{K_s, (-1)^F\} = 0 \ . \tag{3.109}$$

As $(-1)^F$ commutes with Δ_s we may define simultaneous eigenvectors $|\psi\rangle$,

$$\Delta_s|\psi\rangle = \lambda|\psi\rangle \ , \qquad (-1)^F|\psi\rangle = \pm|\psi\rangle \ . \tag{3.110}$$

Now it is an immediate consequence of (3.109) that $K_s|\psi\rangle$ is also an eigenvector with the same eigenvalue λ, but with opposite eigenvalue for $(-1)^F$; the fact that this eigenvector is not trivial follows from the first equation of (3.109), providing the eigenvalues λ are non-zero.

The trace in (3.107) may now be extended over the complete spectrum of the Laplacian as the eigenvectors of positive and negative eigenvalues of $(-1)^F$ cancel against each other. This, however, requires some regularization as the eigenvalues of Δ_s may become arbitrarily large (though on a compact manifold they are discrete and positive). A term that clearly cuts out the large modes is afforded by the exponential damping factor $e^{-\beta\Delta_s}$, for any $\beta > 0$. The final form for the Euler character as a trace over differential forms is

$$\chi(M) = \mathrm{Tr}[(-1)^F e^{-\beta\Delta_s}] \ . \tag{3.111}$$

Now thinking of Δ_s as a Hamiltonian, call it H_s, we may give a path integral representation for this trace in the standard way. This gives us the desired relationship

$$\chi(M) = \mathrm{Tr}[(-1)^F e^{-\beta\Delta_s}] = \int_\Phi e^{-S} \ , \tag{3.112}$$

where the action is that of the supersymmetric quantum mechanical model (3.1), and because of the $(-1)^F$ in the trace the boundary conditions are periodic for all the fields. β is incorporated on the right hand side as the circumference of the time circle. The trace obviously does not depend on the parameter β, which can be seen from the path integral point of view by the argument that the variation with respect to β gives, on the right hand side of (3.112), the expectation value of a Q-exact term and hence vanishes. We will make use of this property to simplify the calculation of the Euler character as an integral over powers of the curvature tensor.

3.8.3. Supersymmetry and the Laplacian

On calling the Laplacian a Hamiltonian we are anticipating that there is an underlying super-symmetry in the theory. Indeed, this is a straightforward generalization of the one considered in the previous section for the Witten index. We have already denoted Δ_s by H_s; now we relabel d_s by Q and d_s^* becomes iQ^*. It then follows that the properties of the exterior derivative become the standard supersymmetric quantum mechanical relations, namely

$$-2iH_s = \{Q, Q^*\} \ , \qquad Q^2 = 0 \ , \qquad Q^{*2} = 0 \ . \tag{3.113}$$

The supersymmetry in the Lagrangian formulation is of course tied to this generalized de Rham supersymmetry. This subsection is devoted to exhibiting this fact. The following considerations are in fact a generalization of those in section 3.6, and just as in that section, we have here eliminated B^i by its algebraic equation of motion.

Let us denote the charge associated with the supersymmetry transformations (3.4) also by Q. Then following the Noether prescription, just as in section 3.6, we have

$$Q = -\sum_\Phi \frac{\partial_r L}{\partial\dot\Phi}\delta\Phi = -\left(g_{ij}(\phi)\dot\phi^j + s\frac{\partial V}{\partial\phi^i}\right)\psi^i \ . \tag{3.114}$$

We may also read off the following Poisson brackets from the action:

$$\{g_{ik}\dot\phi^k + s\,\partial V/\partial\phi^i, \phi^j\} = -\delta_i^j \ , \qquad \{\psi^i, \bar\psi_j\} = i\delta_j^i \ . \tag{3.115}$$

From (3.114), it appears that Q is metric dependent; however, the Poisson brackets (3.115) show us that this is not the case, since denoting the conjugate momentum to ϕ^i by π_i we may express Q as $\psi^i \pi_i$, with no reference to the metric. If we now make the identifications $\psi^i \rightarrow a^{*i}$ and $\bar{\psi}^k (= g^{kj} \bar{\psi}_j) \rightarrow i a^k$, we see, on making use of eq. (3.104), that Q is identically $d_s = d + sdV$. In particular, Q is nilpotent, $Q^2 = 0$, even though the transformations (3.4), with the B field given by its equation of motion, are nilpotent only on-shell. There is no contradiction here as the Noether theorem makes use of the equations of motion.

To determine Q^* we need to exhibit the dual BRST invariance of the action. Up to a surface term (taken to vanish), the action is invariant under the following combined transformations[*]:

$$\delta \phi^i = \bar{\psi}^i , \qquad \delta \bar{\psi}^i = 0 , \qquad \delta \psi^i = \dot{\phi}^i + sg^{ij}(\phi)\, \partial V / \partial \phi^j - g^{ij}(\phi)\psi_l \Gamma^l_{jk} \bar{\psi}^k . \qquad (3.116)$$

The charge associated with these transformations is Q^*, and from the Noether theorem we have

$$Q^* = -\bar{\psi}^i [g_{ij}(\phi)\dot{\phi}^j - s\, \partial V / \partial \phi^i] + \bar{\psi}^i \bar{\psi}^j \Gamma_{ijk} \psi^k . \qquad (3.117)$$

On comparing this with (3.104), and on substituting the a's for the ψ's, we see that Q^* is precisely d_s^*. A little more work establishes that the Hamiltonian associated to the action of (3.85) is the Laplacian $\Delta_s = -2iH_s = \{Q, Q^*\}$.

3.8.4. The Poincaré–Hopf theorem

Having given a field theoretic form for the Euler character, we will in this and the next section show how different choices of gauge may be used to derive various concrete expressions. The gauge freedom is a consequence of the supersymmetry of the theory, which we have just seen is intimately connected with the de Rham cohomology of the manifold. Here we establish the Poincaré–Hopf theorem, which relates the Euler character to the zeros of a vector field. Our presentation is similar to that of ref. [3.25].

We work in the delta function gauge, so that there is no B^2 term in the action. That is, we take $\beta = 0$ in (3.86), or alternatively we start with the action

$$
\begin{aligned}
S &= i \oint d\tau \left\{ Q, \bar{\psi}_i \left(\frac{d\phi^i}{d\tau} + sg^{ij}(\phi) \frac{\partial V(\phi)}{\partial \phi^j} \right) \right\} \\
&= i \oint d\tau \left[B_i \left(\frac{d\phi^i}{d\tau} + sg^{ij}(\phi) \frac{\partial V(\phi)}{\partial \phi^j} \right) \right. \\
&\quad \left. - \bar{\psi}_i \left(\delta^i_k \frac{D}{D\tau} + sg^{ij}(\phi) \frac{D^2 V(\phi)}{D\phi^j D\phi^k} \right) \psi^k \right] .
\end{aligned}
\qquad (3.118)
$$

An integration over B_i yields a delta function constraint onto

$$d\phi^i / d\tau + sg^{ij}(\phi)\, \partial V(\phi) / \partial \phi^j = 0 , \qquad (3.119)$$

and using the squaring argument leads to

$$d\phi^i / d\tau = 0, \qquad s\, \partial V(\phi) / \partial \phi^i = 0 . \qquad (3.120)$$

There are two distinct possibilities.

---

[*] This symmetry follows, as we saw, from the Q invariance by noting that the action is unaltered by $\psi \rightarrow \bar{\psi}$, $\bar{\psi} \rightarrow \psi$ and $s \rightarrow -s$. This holds up to an integration by parts; if we had an arbitrary vector field V^i, this second symmetry would not be present.

(i) $s = 0$. This implies that $\dot\phi^i = 0$ so that only constant loops are important and one ends up with an integral over the manifold M. But this is not the complete story; we will have more to say on this in the next section.

(ii) $s \neq 0$. This also implies that only constant loops are important; however, the condition that $\partial V(\phi)/\partial \phi^i$ should vanish also implies that only the (by assumption isolated) critical points $\{P\}$ (with local coordinates ϕ_P^i) of V enter. We proceed with this option here.

The path integral does not depend on the metric; therefore we may "deform" the metric so that it is flat in the region of the critical points of the function $V(\phi)$. In this way we see that the curvature and affine terms in the action may be ignored.

Expanding about the classical solutions, the path integral becomes

$$\sum_{\{P\}} \text{sign}\det[\delta_{ij}\, \mathrm{d}/\mathrm{d}\tau + s\, \partial^2 V(\phi)/\partial \phi^i \partial \phi^j|_{\phi = \phi_P}] \ . \tag{3.121}$$

However, the discussion following eq. (3.27) shows that the sign of the determinant comes strictly from the $n = 0$ term. This leaves us with

$$\chi(M) = \sum_{\{P\}} \text{sign}\det[\partial^2 V(\phi)/\partial \phi^i \partial \phi^j] \ . \tag{3.122}$$

This equality is the statement of the Poincaré–Hopf theorem, and the above derivation is a field theoretic proof of this classic result.

To exhibit the power of the functional integral representation we make two observations. Firstly, since the potential is a coupling "constant" we know, by the supersymmetry, that the results will be invariant under generic deformations of V; this shows that there is nothing special about the chosen potential. Secondly, there is a very convenient gauge choice which leads to a quick derivation of (3.122), namely the gauge

$$\partial V/\partial \phi^i = 0 \ . \tag{3.123}$$

We have repeatedly warned that changing the basic field equation may lead to an inequivalent theory. But for $s \neq 0$ (3.123) implies the same conditions as (3.119), that is, only the critical points of $V(\phi)$ are important; as these are isolated only constant paths contribute (namely the critical points themselves). The appropriate action is then

$$\begin{aligned} S &= \mathrm{i} \oint \mathrm{d}\tau \{Q, \bar\psi^i \partial V(\phi)/\partial \phi^i\} \\ &= \mathrm{i} \oint \mathrm{d}\tau \left(B^i \frac{\partial V(\phi)}{\partial \phi^i} - \bar\psi^i \frac{\mathrm{D}^2 V}{\mathrm{D}\phi^i \mathrm{D}\phi^j} \psi^j \right) \ , \end{aligned} \tag{3.124}$$

the partition function of which leads directly to (3.122).

3.8.5. The Gauss–Bonnet theorem

When we take $s = 0$ in the action (3.118) the B field integration directly enforces the constraint $\dot\phi^i = 0$. A glance at the action shows that, on expanding all of the fields in a Fourier series, there are ϕ, B, ψ and $\bar\psi$ constant modes that do not enter. The integration over the constant ϕ yields the volume of M. On the other hand, the integration over the constant B gives an infinity, as B essentially ranges over the tangent space of M. The integrations over the fermion constant modes are zero. Notice that the constant modes do not cause any special problems in the calculation of the previous section as they all enter in the action.

We proceed by making use of the action in the form (3.86) with $s = 0$, even though the B field integration no longer gives a delta function constraint. In this case we use the β independence of the partition function to take the $\beta \to 0$ limit. This calculation has been previously carried out by Alvarez-Gaumé [3.3, 3.4], in a slightly different form by Friedan and Windey [3.26, 3.27] and the details of the mode expansion may be found in ref. [3.28], so we shall be brief here.

First scale all the B and $\bar{\psi}$ modes by $1/\sqrt{\beta}$. Then multiply all the non-constant ϕ and ψ modes by $\sqrt{\beta}$. The path integral measure is invariant under these scalings, the bosonic Jacobian matching precisely the inverse of the fermionic one. We keep only those terms in the action which do not depend on β after these manipulations. The β dependent terms will decay at least as fast as $\sqrt{\beta}$, and treating them as interactions in a perturbation expansion, they will vanish in the limit[*].

In this way we see that the action reduces to

$$\sum_{n=-\infty}^{\infty} [B_{i,n} n \phi^i_{-n} + B_{i,n} g^{ij}(\phi_0) B_{i,-n} - \bar{\psi}_{i,n} n \psi^i_{-n}] - \tfrac{1}{4} R^{ij}{}_{kl}(\phi_0) \bar{\psi}_{i,0} \bar{\psi}_{j,0} \psi^k_0 \psi^l_0 . \qquad (3.125)$$

The integration over the non-constant modes gives 1. To put the constant mode integration in a more conventional form we change variables to $B^i_0 = g^{ij}(\phi_0) B_{i,0}$ and $\bar{\psi}^i_0 = g^{ij}(\phi_0) \bar{\psi}_{i,0}$. Keeping in mind that the path integral measure does not pick up a Jacobian from this change, the partition function becomes

$$\chi(M) = \frac{1}{(2\pi)^{n/2}} \int_M d\phi \int d\bar{\psi}\, d\psi \; e^{\tfrac{1}{4} R_{ijkl} \bar{\psi}^i \bar{\psi}^j \psi^k \psi^l} , \qquad (3.126)$$

where the subscript 0 has been dropped everywhere and the factor of $(2\pi)^{-n/2}$ comes from the B^i_0 integration. This form for the Euler character inspired Mathai and Quillen to give a rather more mathematical basis for this formula [3.29, 3.30]. We briefly discuss that construction in the context of gauge theories in section 5.2.6.

One general property that follows from (3.126) is that for odd dimensional manifolds the Euler character vanishes since the integrals would be over an odd number of ψ, while the exponent has only an even number. This also follows from (3.88) on using Poincaré duality of the Betti numbers.

We would now like to explain from a field theoretic point of view why it is that the curvature tensor survives the limit that we have taken. The crucial point in the above analysis was that we had to be careful with the integral over the zero modes. We could have chosen a delta function gauge for the non-zero modes by not taking a term of the form $\oint \{Q, \bar{\psi}B\}$ in the action, but rather one that only involved the zero mode, $\{Q, \bar{\psi}_0 B_0\}$. In this way we see that the non-zero modes of the fermions and bosons may be ignored; they give rise to determinants whose ratio is 1, leaving only the zero modes. The $\beta \to 0$ limit is singular for the B_0 integration and it is this singularity that ensures a finite contribution from the curvature term.

3.8.6. General properties of the Euler character

There are two further general properties of the Euler character that may be derived directly from the path integral. We sketch the ideas here.

The first property is that

$$\chi(M_1 \# M_2) = \chi(M_1) + \chi(M_2) - \chi(S^n) .$$

[*] The supersymmetry guarantees that the fermionic contributions will cancel the bosonic ones in this expansion in any case.

The symbol # denotes the connected sum for any two manifolds of the same dimension (so that we have in the above formula $\dim M_1 = \dim M_2 = n$). The connected sum is formed by cutting out a disk from each manifold and gluing back on the boundary $(n-1)$-sphere, with an orientation reversing diffeomorphism. This is most readily visualized in the case of Riemann surfaces. The Euler character of the connected sum of a genus g_1 and a genus g_2 manifold is $2[1-(g_1+g_2)]$, the two manifolds are joined along an S^2, so that the formula correctly gives $2(1-g_1)+2(1-g_2)-2$; see ref. [3.31], page 305 for an excellent illustration of this decomposition.

We recover this result in the path integral in the following manner. Split the domain of the path integral measure into two pieces, one covering M_1, the other covering M_2. In this way we have covered $M = M_1 \# M_2$, but have overcounted the intersection of the two, which is S^n, once. Each path integral gives the corresponding Euler character and this is the required result.

The second property that can be "derived" is that for product manifolds

$$\chi(M_1 \times M_2) = \chi(M_1)\chi(M_2).$$

We have shown that the path integral does not depend on the metric, so we choose a product metric on the space. With this metric, the action splits into two pieces, one of which carries the labels of M_1, the other M_2 (and their tangent spaces). The path integral measure may be factorized into two pieces with respect to the labels of the manifolds. Hence, the complete path integral factorizes into two independent path integrals, one over M_1, the other over M_2.

3.9. Symmetry breaking and zero modes

We have found in our evaluation of the path integrals of the previous sections that they reduce to the critical points of the potential V that enters, that is, to spaces of zero dimension. On the other hand, when we determined the Euler character, the Gauss–Bonnet form arose when the integral devolved not to isolated points, but rather to an integral over the original manifold M. The space defined by the instanton equation is called the moduli space, and is denoted by $\mathcal{M}$. When $s \neq 0$ in (3.10) then $\mathcal{M} = \{P \mid dV_P = 0\}$, while for $s = 0$ one finds $\mathcal{M} = M$.

In our discussion so far of topological field theory, we have considered the BRST operator as it acts on all of the fields. The path integral, however, as we have noted, boils down to an integral over the moduli space of instantons (or to a sum over isolated points). One would like then to have explicit formulae on this restricted space as, in the end, it is the only one of interest.

The aim of this section is threefold:

Firstly, we explicitly introduce the moduli space parameters (and their fermionic partners) into the BRST algebra for the toy model, with time taken to lie on the real line $\mathbb{R}$. This has a one-dimensional moduli space parametrized by the "center" of the instanton. By incorporating the moduli space parameters in this way one ends up with rather explicit formulae depending on the moduli space and its (co)tangent bundle. The results so obtained generalize in a straightforward manner to the general model.

Secondly, this instanton calculation is related back to the question of supersymmetry breaking and the Witten index as discussed in section 3.7. One aspect of the supersymmetry breaking mechanism emphasized by Witten is the importance of non-perturbative effects; that is, the symmetry breakdown is "mediated" by instantons. Furthermore, the explicit formulae we obtain provide a clean method for checking when the BRST invariance *itself* is broken[*]. Since one of the outstanding problems

---

[*] It should be made clear that we are looking at the (intrinsic) breakdown of the symmetry in the one dimensional theory and we are not using this as the measure of symmetry breaking of some higher dimensional model.

in this area is how to "liberate" degrees of freedom, we see that symmetry breaking offers us some insight into how dynamics might be reintroduced into topological models.

Our third objective is to establish that in a certain limit ($s \to \infty$) regardless of whether the symmetry is broken or not, the theory remains gauge and coupling constant (target manifold metric) independent. The first two points were addressed in ref. [3.32], and we follow a similar presentation here.

3.9.1. Zero modes of the toy model

The action of interest is

$$S = i \int\limits_{-\infty}^{+\infty} d\tau \left[B \left(\frac{d\phi}{d\tau} + s \frac{\partial V}{\partial \phi} \right) - \bar{\psi} \left(\frac{d}{d\tau} + s \frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right] , \tag{3.127}$$

where $\tau \in \mathbb{R}$ and the target manifold is either the real line or the circle, $\phi : \mathbb{R} \to \mathbb{R}$ or S^1. The path integral in the $\phi(\tau)$ field is taken to be between ϕ_i and ϕ_f, two critical points of the potential $V(\phi)$. We should emphasize here that these boundary conditions are appropriate for "tunneling" configurations and so differ from those considered previously. In particular, when considering vacuum expectation values of operators, it must be borne in mind that the in and out vacua are different. The metric on the target manifold $g(\phi)$ is the standard one on $\mathbb{R}$ and S^1, and is suppressed. It should be apparent from the following presentation that everything could be straightforwardly repeated for arbitrary target manifolds; we will simply quote the general results.

As our aim is to establish that the supersymmetry is broken, it would appear that the addition of a term $\{Q, \bar{\psi}B\} = B^2$ could well lead, if we are successful, to an inequivalent theory. That is, as the symmetry no longer holds, different choices of gauge are *not* gauge equivalent. It turns out that we may add the B^2 term to the action with impunity when calculating the partition function; this is possible since the partition function vanishes due to the presence of a fermionic zero mode. However, only in a certain limit will observables be unaffected by the change of gauge. We will establish this in due course, and proceed with (3.127).

The advantage of this form of the action is that the path integral only takes values on the classical (instanton) paths, since integrating over B leads to the constraint $d\phi/d\tau + s \, \partial V/\partial \phi = 0$. A solution to this equation corresponds to a "tunneling" process. In this case the squaring argument, taking into account the boundary contribution, leads to

$$\int\limits_{-\infty}^{+\infty} d\tau \left[\left(\frac{d\phi}{d\tau} \right)^2 + s^2 \left(\frac{\partial V}{\partial \phi} \right)^2 \right] = 2s[V(\phi_i) - V(\phi_f)] . \tag{3.128}$$

The left hand side is positive semi-definite, so we have a condition for the existence of instantons between ϕ_i and ϕ_f (we take $s > 0$), namely

$$V(\phi_i) - V(\phi_f) \geq 0 . \tag{3.129}$$

Equality holds only when $d\phi/d\tau = 0$ and $\partial V/\partial \phi = 0$ corresponding to the trivial situation of $\phi_i = \phi_f$. To avoid confusion we remind the reader that V is related to the potential energy W of the "particle" by $V(\phi) - V(\phi_i) = \int_{\phi_i}^{\phi} \sqrt{2W}$. Thus we are *not* implying that the potential energy at the two critical points need be different for an instanton to interpolate between them. We take a generic function V which satisfies the strict bound

$$V(\phi_i) - V(\phi_f) > 0 . \tag{3.130}$$

There are zero modes in this theory as the "center" λ of the instanton is a free parameter; the classical solution is a member of the one parameter family

$$\phi_c(\tau - \lambda) \ . \tag{3.131}$$

For infinitesimal λ the zero mode may be expressed as $\phi_0 = \lambda V'(\phi_c)$, and satisfies the linearized equation

$$\mathrm{d}\phi_0/\mathrm{d}\tau + s\, \partial^2 V/\partial\phi\,\partial\phi|_{\phi=\phi_c}\phi_0 = 0 \ . \tag{3.132}$$

Before taking such modes into account we need to know if they are square integrable. No work is required to establish this, for a glance at (3.128) shows that this must be the case, otherwise $V(\phi_i)$ is unbounded. We restrict our attention to those V which are bounded at the critical points. The formally similar linearized equation for a B zero mode B_0

$$\mathrm{d}B_0/\mathrm{d}\tau - s\, \partial^2 V/\partial\phi\,\partial\phi|_{\phi=\phi_c}B_0 = 0 \ , \tag{3.133}$$

has a solution $B_0 = c(V')^{-1}$. This mode, on the other hand, is not square integrable, so we may neglect it. This fact is easily established by making use of the Cauchy-Schwarz inequality, $(\int_{-\infty}^{+\infty} fg)^2 \leq (\int_{-\infty}^{+\infty} f^2)(\int_{-\infty}^{+\infty} g^2)$, which holds for square integrable functions f and g. If we assume that B_0 is indeed normalizable then we must have that $\int_{-\infty}^{+\infty} B_0\phi_0$ is finite. But it clearly diverges, implying that this mode is not normalizable. We can likewise ignore the zero mode of the $\bar{\psi}$ field as it also satisfies (3.133). Alternatively, we see from the supersymmetry transformation $\{Q,\bar{\psi}\} = B$, that B and $\bar{\psi}$ are in one–one correspondence; if there is no zero mode for one of these then there is none for the other[*].

On the other hand, there is a ψ zero mode which may be obtained from (3.131) by a supersymmetry transformation. Setting

$$\{Q,\lambda\} = \sigma \ , \qquad \{Q,\sigma\} = 0 \ , \tag{3.134}$$

the ψ zero mode is given by

$$\psi_c = \sigma\, \mathrm{d}\phi_c(\tau - \lambda)/\mathrm{d}\lambda \ . \tag{3.135}$$

If there was also a $\bar{\psi}$ classical solution that had to be considered, then, just as for the calculation of the Witten Index in section 3.7, the partition function would not necessarily be zero. However, here the partition function will vanish because of the unmatched mode ψ_c, $Z = 0$. Expectation values will also necessarilly vanish unless they are of operators that have ghost degree one, so that the fermion zero mode is saturated.

We evaluate $\langle 0|\mathcal{O}(\phi,\psi)|0\rangle$, where $\mathcal{O}$ has ghost number one, first in a rather formal fashion and then once again taking more care of the zero modes. The result is

$$\langle 0|\mathcal{O}|0\rangle = \int_{\Phi} \mathrm{e}^{-S}\, \mathcal{O}(\phi,\psi)$$

$$= \int \mathrm{D}\phi\, \mathrm{D}\psi\, \delta\left(\frac{\mathrm{d}\phi}{\mathrm{d}\tau} + s\frac{\partial V}{\partial\phi}\right) \delta\left(\frac{\mathrm{d}\psi}{\mathrm{d}\tau} + s\frac{\partial^2 V}{\partial\phi\,\partial\phi}\psi\right) \mathcal{O}(\phi,\psi) \ . \tag{3.136}$$

---

[*] When $\tau \in S^1$ the zero modes, the constant modes of B and $\bar{\psi}$, are normalizable and were incorporated in the calculation following eq. (3.123).

It has been possible to perform the B and $\bar{\psi}$ integrations precisely as these fields do not suffer from zero modes, and we have chosen the "observable" $\mathcal{O}$ not to depend on them. The delta functions ensure that only the classical trajectories will contribute, and so the path integral is restricted to the moduli space and its tangent space (at a point). On expanding about the arbitrary classical solution $\phi_c(\tau - \lambda)$, and the associated fermionic zero mode ψ_c, the path integral becomes

$$\langle 0|\mathcal{O}|0\rangle = \int d\phi_c \, d\psi_c \frac{\det(d/d\tau + sV''(\phi_c))}{|\det(d/d\tau + sV''(\phi_c))|} \mathcal{O}(\phi_c, \psi_c) \ . \tag{3.137}$$

Declaring the sign of the ratio of determinants to be one we find

$$\langle 0|\mathcal{O}|0\rangle = \int d\phi_c \, d\psi_c \, \mathcal{O}(\phi_c, \psi_c) \ . \tag{3.138}$$

As there is only one zero mode, non-zero values for this correlator will be obtained for those observables of the form $\mathcal{O} = \psi f(\phi)$.

This is the final result; however, the above derivation has been somewhat heuristic. We now remedy this by incorporating the zero modes explicitly into the BRST algebra and then into the path integral. Let the fields ϕ and ψ be decomposed as

$$\phi = \phi_q + \phi_c \ , \qquad \psi = \psi_q + \psi_c \ , \tag{3.139}$$

where ϕ_c and ψ_c are the classical configurations given in eqs. (3.131) and (3.135), respectively. As the classical fields interpolate between the critical points of the potential V, the boundary conditions on the quantum fields ϕ_q and ψ_q (as on B and $\bar{\psi}$) are that they vanish at the end points $\tau \pm \infty$. The BRST algebra for the new fields may be determined from the original transformations (3.6) and the transformation rules for the moduli parameter λ and its superpartner σ given in (3.134). The algebra reads

$$\{Q, \phi_q\} = \psi_q \ , \qquad \{Q, \psi_q\} = 0 \ , \qquad \{Q, \bar{\psi}\} = B \ ,$$
$$\{Q, B\} = 0 \ , \qquad \{Q, \lambda\} = \sigma \ , \qquad \{Q, \sigma\} = 0 \ . \tag{3.140}$$

Now that we have explicitly extracted the zero modes, we must ensure that both ϕ_q and ψ_q have no such modes contained in them; otherwise we shall be over counting. This is easily achieved by gauge fixing them to be orthogonal to the zero mode $d\phi_c/d\lambda = -d\phi_c/d\tau$. To implement the gauge fixing we need to introduce two time independent fields $\bar{\sigma}$ of ghost degree -1 and η of ghost degree 0, which transform as

$$\{Q, \bar{\sigma}\} = \eta \ , \qquad \{Q, \eta\} = 0 \ . \tag{3.141}$$

Notice that with all the redefinitions and introduction of new fields the nilpotency property of the BRST operator, $Q^2 = 0$, has been maintained.

The action that we take is (3.127), with additional terms so as to impose the conditions that the fields ϕ_q and ψ_q have no zero mode components. Specifically the action chosen is

$$S = i \int\limits_{-\infty}^{+\infty} d\tau \left\{ Q, \bar{\psi} \left(\frac{d\phi}{d\tau} + s\frac{\partial V(\phi)}{\partial \phi} \right) + \bar{\sigma}\phi_q \frac{d\phi_c}{d\lambda} \right\}$$

$$= i \int\limits_{-\infty}^{+\infty} d\tau \left[B \left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial \phi} \right) - \bar{\psi} \left(\frac{d}{d\tau} + s\frac{\partial^2 V}{\partial \phi \partial \phi} \right) \psi \right.$$

$$\left. + \eta\phi_q \frac{d\phi_c}{d\lambda} - \bar{\sigma}\psi_q \frac{d\phi_c}{d\lambda} - \bar{\sigma}\phi_q \frac{d^2\phi_c}{d\lambda^2}\sigma \right] \ . \tag{3.142}$$

The observable is taken to be a function of the original variables, $\mathcal{O}(\phi, \psi)$. The integration over B implies once more that ϕ satisfies the instanton equation. The boundary conditions $\phi_q(\pm\infty) = 0$ then show that this field is purely a zero mode. However, the integral over η forces ϕ_q to be orthogonal to the zero modes; thus $\phi_q = 0$. The determinant that arises from the delta function constraints is then

$$\delta\left(\frac{\mathrm{d}\phi_q}{\mathrm{d}\tau} + sV''(\phi_c)\phi_q\right)\delta\left(\int \phi_q \frac{\mathrm{d}\phi_c}{\mathrm{d}\lambda}\right) = \frac{\delta(\phi_q)}{|\det'(\mathrm{d}/\mathrm{d}\tau + sV''(\phi_c))|} , \tag{3.143}$$

where the prime on the determinant indicates that the zero eigenvalue is excluded. It follows by supersymmetry that $\psi_q = 0$; explicitly, the $\bar{\psi}$ integral enforces that ψ_q be a pure zero mode, while the integral over $\bar{\sigma}$ (now that the term proportional to $\bar{\sigma}\sigma$ in the action may be neglected) establishes that ψ_q is zero. The determinant that one gets in this way, up to the usual sign ambiguity which is declared to be one, cancels precisely that in (3.143).

The vacuum expectation value of $\mathcal{O}$ becomes

$$\langle 0|\mathcal{O}|0\rangle = \int \mathrm{d}\lambda\,\mathrm{d}\sigma\,\mathcal{O}(\phi_c, \psi_c) , \tag{3.144}$$

which agrees with (3.138).

3.9.2. Symmetry breaking and zero modes

We will now show that there are observables with non-zero expectation value, and that this fact is intimately related to the question of supersymmetry breaking that we addressed in section 3.7.

Let $\mathcal{O} = \psi V'(\phi)$; when this is substituted into (3.144) its expectation value is given by

$$\langle 0|\mathcal{O}|0\rangle = \int \mathrm{d}\lambda\,\mathrm{d}\sigma\,\sigma\frac{\mathrm{d}\phi_c}{\mathrm{d}\lambda}V'(\phi_c) = \int \mathrm{d}\lambda\frac{\mathrm{d}V(\phi_c)}{\mathrm{d}\lambda} = V(\phi_f) - V(\phi_i) \neq 0 . \tag{3.145}$$

This expectation value does not vanish, because of the constraint (3.130); indeed, the right hand side of (3.145) is known as the winding number of the instanton.

Let us pause to ask ourselves what properties an observable should have. Firstly it must have, as we discussed, fermion number 1. The second requirement that we impose is that it be BRST invariant, $\{Q, \mathcal{O}\} = 0$. This is a natural requirement; in a gauge theory we would take expectation values of gauge invariant operators. Clearly we have $\{Q, \psi V'(\phi)\} = 0$, so that it is a good observable. The surprise is that we have calculated a BRST exact correlator

$$\langle 0|\psi\,\partial V(\phi)/\partial\phi|0\rangle = \langle 0|\{Q, V(\phi)\}|0\rangle \neq 0 , \tag{3.146}$$

and obtained a non-zero result; thus *the BRST symmetry is broken*. In a gauge theory, when we evaluate the expectation value of an operator that is pure gauge, we should get zero since the operator is gauge equivalent to the zero operator. This translates into the usual Ward identities in BRST form, namely that the expectation value of the BRST variation of an operator is zero. From (3.146) we see that the Ward identity fails and so we conclude that the symmetry is broken.

In one dimension all the Q-closed observables will in fact be Q-exact. The reason for this is that the moduli space is the real line, $\lambda \in \mathbb{R}$. On the zero mode space Q acts like exterior differentiation. Thus our demand that observables be Q-closed and of fermion degree 1 translates into the requirement that they be closed one-forms on the real line. But $H^1(\mathbb{R}) = 0$, so that all the closed one forms are exact. Letting $\mathcal{O} = \sigma g$ we may then express this as $\mathcal{O} = \{Q, f\}$, where $f = \int^\lambda \mathrm{d}s\, g(s)$.
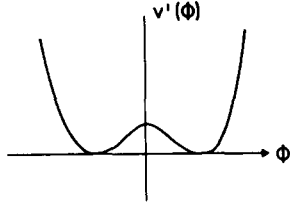
Fig. 3.1. A potential $V'(\phi)$ that tends to plus infinity as $\phi \to \pm\infty$ is displayed. This potential allows for instantons to interpolate between its zeros.
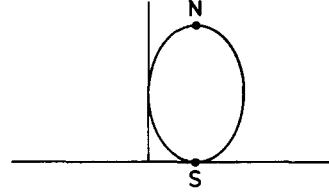
Fig. 3.2. The height function V is displayed for the circle. The north and south poles are indicated; there are two instanton paths between these points.

The reason that Q-exact terms may pick up non-zero expectation values is that, just as in de Rham theory, end point contributions may be important. There is an analogue in gauge theories on non-compact manifolds when there are Higgs fields present (section 5.4.2). From this discussion we see that all observables may be expressed as $\mathcal{O} = \psi f'(\phi)$ and their expectation value follows from (3.145), $\langle 0|\mathcal{O}|0\rangle = f(\phi_f) - f(\phi_i)$, which we may write more suggestively as

$$\int_\Phi Qf = \int_\mathbb{R} df(\phi_c) \ .$$

It is perhaps worth pausing for a second to admire the above equation. In fact, it captures the essence of topological field theories: a path integral over a functional space Φ of a BRST exact observable Qf has been reduced to an integral over a finite-dimensional moduli space $\mathbb{R}$ of a d-exact volume form df.

What do these considerations have to do with the question of supersymmetry breaking as posed in section 3.7? There, the Witten index was introduced to measure the possibility of supersymmetry breaking. However, it should be apparent that, if we find a state with zero energy which is stable under perturbations, then supersymmetry cannot be broken, except by some non-perturbative effect. Consider the form of the potential $V'(\phi)$ in fig. 3.1. There are clearly two critical points; either one would serve as a ground state for the theory. However, the Witten index is zero for this potential allowing for the possibility that the symmetry is broken. The instanton calculation establishes that this is indeed the case, for it shows that the energy degeneracy is lifted and that there is no state with zero energy eigenvalue. When there is only one critical point, there is no instanton calculation to perform and that state remains the ground state. When there are many turning points, we may, by perturbing the potential if need be, deform the problem into either of these two examples; otherwise, proceeding along the lines of this section, one must take into account all the critical points and the instantons that interpolate between them.

3.9.3. Gauge and metric independence

We have exhibited that the BRST symmetry is broken in this theory, although this may not be the case over the complete manifold[*]. In deriving this result we have worked in the delta function gauge. Since the symmetry is broken, different gauge choices are no longer equivalent. However, it is sufficient to establish that the symmetry is broken in any given gauge; the fact that the results may vary from gauge to gauge is itself an indication that the symmetry is broken. We would, nevertheless, like to determine how the results change as we change gauge.

[*] Taking into account all of the critical points and instantons that interpolate between them could restore the symmetry.

We will take the action to be

$$
S = i \int\limits_{-\infty}^{+\infty} d\tau \left\{ Q, \bar{\psi} \left(\frac{d\phi}{d\tau} + s\frac{\partial V(\phi)}{\partial \phi} - \tfrac{1}{2}i\alpha B \right) + \bar{\sigma}\phi_q \frac{d\phi_c}{d\lambda} \right\}
$$

$$
= i \int\limits_{-\infty}^{+\infty} d\tau \left[B\left(\frac{d\phi}{d\tau} + s\frac{\partial V}{\partial \phi} \right) - \tfrac{1}{2}i\alpha B^2 - \bar{\psi}\left(\frac{d}{d\tau} + s\frac{\partial^2 V}{\partial\phi\partial\phi} \right)\psi \right.
$$

$$
\left. + \eta\phi_q \frac{d\phi_c}{d\lambda} - \bar{\sigma}\psi_q \frac{d\phi_c}{d\lambda} - \bar{\sigma}\phi_q \frac{d^2\phi_c}{d\lambda^2}\sigma \right] , \tag{3.147}
$$

and determine the corrections to $\langle 0|\mathcal{O}|0\rangle$ as $s \to \infty$. We begin by scaling ϕ_q and ψ_q by s (the Jacobian is 1). Likewise, multiply $\bar{\sigma}$ and η by s. Let us now keep only those terms in the action that are s independent; all corrections will be of the order $1/s$ or smaller,

$$
S = \int\limits_{-\infty}^{+\infty} d\tau \left(iBV''(\phi_c)\phi_q + \tfrac{1}{2}\alpha B^2 - i\bar{\psi}V''(\phi_c)\psi_q \right.
$$

$$
\left. i\eta\phi_q \frac{d\phi_c}{d\lambda} - i\bar{\sigma}\psi_q \frac{d\phi_c}{d\lambda} - i\bar{\sigma}\phi_q \frac{d^2\phi_c}{d\lambda^2}\sigma + \cdots \right) . \tag{3.148}
$$

In this limit $\mathcal{O}(\phi,\psi) \to \mathcal{O}(\phi_c, \sigma\, d\phi_c/d\lambda)$. A straightforward calculation leads to the result

$$
\langle 0|\mathcal{O}|0\rangle = f(\phi_f) - f(\phi_i) + O(\alpha/s) . \tag{3.149}
$$

The gauge dependence is then suppressed by a factor of $1/s$, with the delta function gauge, $\alpha = 0$, picking out the leading term. All possible metric dependence is therefore also suppressed by this factor. Calculations have been performed in the $\alpha = 1$ gauge in refs. [3.33–3.35], where detailed discussions of supersymmetry breaking are to be found.

3.9.4. The general model

For an n-dimensional target manifold the results are basically unchanged. As the zero mode is associated with the shift invariance of the center of the instanton there remains only one such zero mode. The moduli space is still one-dimensional and diffeomorphic to the real line. An appropriate observable is $\mathcal{O} = \{Q, f(\phi^i)\} = \psi^i \partial f(\phi)/\partial\phi^i$.

A correction to formula (3.144) comes about in general when there is more than just one instanton path. For the circle, when the function V is taken to be the height function, there are two instanton paths from the north pole to the south, fig. 3.2, while for the sphere there are an infinite number of such paths associated to the height function. So generically we have

$$
\langle 0|\mathcal{O}|0\rangle = n(\phi_f, \phi_i)\,[f(\phi_f) - f(\phi_i)] , \tag{3.150}
$$

where $n(\phi_f, \phi_i)$ counts the number of instantons with signs. One method for stipulating these will be presented in the next section; for now we leave it in this symbolic form.

3.10. Morse theory and supersymmetry

We saw in section 3.8 that Witten's generalization of de Rham cohomology provides us with a ready proof of the Gauss–Bonnet theorem. The significance of his work extends beyond this

simple application; rather, he was able to show how Morse theory [3.36] can actually be used to obtain the homology of a manifold [3.2]. Indeed this represents the beginning of topological field theory, for it was those ideas which prompted Floer's work on homology three-spheres, and in turn led Atiyah to establish a Hamiltonian formalism connecting this three-dimensional construct with Donaldson's four-dimensional study. In view of the central role played by these ideas, we include a brief discussion of them here.

Let M be a smooth compact manifold, and consider a smooth function $V : M \to \mathbb{R}$ whose critical points are non-degenerate. By this is meant, that at the necessarily isolated points $\{P\}$ where

$$\mathrm{d}V_P = 0 \;, \tag{3.151}$$

the determinant of the Hessian of V at P,

$$\det H_P V = \|\partial^2 V / \partial \phi^i \, \partial \phi^j\| \;, \tag{3.152}$$

does not vanish. The non-degeneracy of the determinant of the Hessian was required so that the Euler number in section 3.8 was well defined, so this condition arises naturally from the path integral point of view.

The Morse index i_P of the function V at the critical point P is defined to be the number of negative eigenvalues of $H_P V$ at P. Both the non-degeneracy condition and the Morse index are independent of the charts chosen for the local neighborhood of the critical points.

There are two Morse inequalities of interest. The first states that M_i, the number of critical points with the same Morse index i, is greater than or equal to b_i, the Betti number of degree i,

$$M_i \geq b_i \;. \tag{3.153}$$

The second, stronger inequality, states that the polynomial $\mathcal{M}_t = \sum_{i=0}^{n} M_i t^i$ is greater than or equal to the Poincaré polynomial $P_t(M) = \sum_{i=0}^{n} b_i t^i$. Precisely formulated, this condition takes the form

$$\sum_{i=0}^{n} (M_i - b_i) t^i = (1 + t) \sum_{i=0}^{n} Q_i t^i, \qquad Q_i \geq 0 \;. \tag{3.154}$$

The "error" term, $\sum_{i=0}^{n} Q_i t^i$, is not known in general. If we set $t = -1$, however, then the error term drops out of (3.154), and we obtain

$$\chi(M) = \sum_{i=0}^{n} b_i(-1)^i = \sum_{i=0}^{n} M_i(-1)^i \;. \tag{3.155}$$

But we have already established this result by using the Witten complex. In section 3.8 we saw that

$$\chi(M) = \sum_{\{P\}} \text{sign} \det H_P V \;, \tag{3.156}$$

and a moment's thought shows that this sum is identically $\sum_{i=0}^{n} M_i(-1)^i$. For the example of the height function of the circle, fig. 3.3, the numbers displayed at the turning points are the Morse indices. There are two points with $i = 0$ and two with $i = 1$, so that $M_0 = M_1 = 2$, giving $\chi(S^1) = 0$. For the circle of fig. 3.2 there is one point with $i = 0$, the south pole, and one with $i = 1$, the north pole, again giving a vanishing Euler character. Likewise for the torus of fig. 3.4, $M_0 = 1$, $M_1 = 2$ and $M_2 = 1$, so that $\chi(T^2) = 1 - 2 + 1 = 0$.
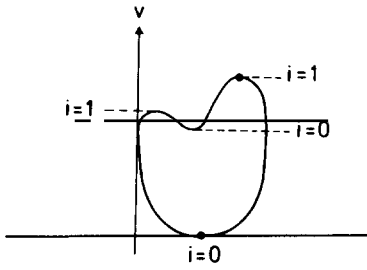
Fig. 3.3. A height function V is displayed for S^1. The Morse index i is indicated at the turning points of V.
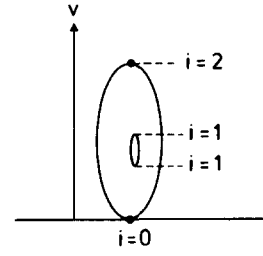
Fig. 3.4. The height function for the torus T^2, as well as the Morse indices of its turning points, are exhibited.

On using the Witten complex we have been able to verify the strong Morse inequalities for a special choice of parameter t. The question that naturally arises is: can one do better? Is it possible to derive both inequalities in generality?

3.10.1. The weak Morse inequalities

The Hamiltonian (or Laplacian) for the Witten complex is

$$2H_s = dd^* + d^*d + s^2(dV)^2 + s\frac{D^2V}{D\phi^i\,D\phi^j}[a^{*i}, a^j] \ . \tag{3.157}$$

Having in mind Morse theory, we see that this is a rather apt operator to consider, for it ties together all the relevant ingredients: the first term is the usual Laplacian $dd^* + d^*d = \Delta$, whose zero eigenfunctions define $b_i(M)$, the second term vanishes precisely at the critical points of V, while the last is proportional to the Hessian of the function.

In the $s \to \infty$ limit the low lying eigenvalues of H_s are centered about the critical points $\{P\}$. Around each such point, the coordinates ϕ^i are chosen so that $\phi^i(P) = 0$, and the metric tensor g_{ij} is taken to be the standard Euclidean metric δ_{ij} up to terms of order ϕ^2. Furthermore, the coordinates may be chosen so that in the vicinity of the critical points, $V(\phi) = V(0) + \frac{1}{2}\sum\lambda_i\phi_i^2 + O(\phi^3)$, for some λ_i. Note that this implies that we have the Hessian in diagonal form as, $V(\phi) = V(0) + \frac{1}{2}H_PV\,\phi_i^2 + \cdots$.

Near the critical point P, H_s may be well approximated by [cf. eq. (3.70)]

$$2H_s = \sum_i(-\partial^2/\partial\phi_i^2 + s^2\lambda_i^2\phi_i^2 + s\lambda_i[a^{i*}, a^i]) \ . \tag{3.158}$$

The term $[a^{i*}, a^i]$ is $2n_i - 1$ where n_i is the ith occupation number for the fermions. As there is either 0 or 1 fermion in each state, this term is given by $n_i^* = 2n_i - 1 = \pm1$. The first two terms in (3.158) correspond to the n-dimensional simple harmonic oscillator. The spectrum is thus

$$E_s = s\sum_i\{|\lambda_i|(1 + 2N_i) + \lambda_i n_i^*\}, \quad N_i \in \mathbb{N} \ . \tag{3.159}$$

The zero energy level is obtained only when the $N_i = 0$ and each n_i^* is chosen to be $-\mathrm{sign}\,\lambda_i$. The fermion number n_i^* in this case measures the presence $(+1)$ or absence (-1) of the ith differential dx^i, so that, if the Hamiltonian is acting on a p-form, then the sum of the positive n_i^* must be p $(+1$ for each $dx^i)$. Since the number of negative λ_i is fixed and is defined to be the Morse index i_P (equal to p, say), then the sum of positive n_i^* must also be this number. Thus there is one ground state at each critical point P, and it must be a p-form.

For the circle [3.37], the height function about a maximum may be expressed as $V(\phi) = c_1 - \frac{1}{2}\phi^2$, and the Hamiltonian about this point takes the form

$$2H_s = -\partial^2/\partial\phi^2 + s^2\phi^2 - s[a^*, a] \ . \tag{3.160}$$

The two possible states are proportional to either $|0\rangle$, which corresponds to functions, or to $|1\rangle = a^*|0\rangle$, which is a basis for one-forms. $n^* = -1$ and $n^* = +1$, respectively, on these states. The spectrum of the Hamiltonian is that of the simple harmonic oscillator (recall that this is $s, 3s, \ldots$) $\pm s$ as it acts on scalars or on one-forms,

$$\mathrm{Spec}(H_s^0) = 2s, 4s, \ldots \ , \qquad \mathrm{Spec}(H_s^1) = 0, 2s, \ldots \ . \tag{3.161}$$

At a minimum, the height function takes the form $V(\phi) = c_2 + \frac{1}{2}\phi^2$, which only changes the above analysis by flipping the signs of n^*. The spectra of the Hamiltonian on the functions and forms are therefore exchanged,

$$\mathrm{Spec}(H_s^0) = 0, 2s, \ldots \ , \qquad \mathrm{Spec}(H_s^1) = 2s, 4s, \ldots \ . \tag{3.162}$$

For this example, we see that the one-form is a ground state at the critical point where the Morse index is 1, while at critical points of Morse index 0, the ground states are functions.

On a general manifold, taking into consideration all of the critical points leads to the weak Morse inequalities. The approximate ground states that we have constructed are not necessarily annihilated by H_s, but it is clear that there are no more ground states available as all the other states have energies that go like s for large s. Hence, the number of approximate harmonic p-forms is at most given by the number of critical points with Morse index p. The number of true harmonic p-forms being b_p, we have established that $M_p \geq b_p$.

3.10.2. The Witten complex and Morse theory

The information that has been derived so far has been quite local, being centered about each critical point. As we saw previously for the calculation of the Euler character, this was quite adequate. In fact, as long as we consider paths in the path integral that are loops, then the series of steps in section 3.8 establishes that *only* information about each individual critical point is relevant. However, in the last section, paths that connected the critical points, that is to say the instantons, were considered and these can be used to give "relative" information. In this way, Witten establishes that the Betti numbers may be determined from Morse theory.

Suppose we have two critical points P_1 and P_2 that are joined by an instanton,

$$d\phi^i/d\tau + sg^{ij}(\phi)\,\partial V/\partial\phi^j = 0 \ . \tag{3.163}$$

Then we know that there is exactly one fermionic zero mode, and this forces the partition function to vanish. How do we interpret such a mode? The ψ are equated with fermionic creation operators and then in turn identified with a basis of differential forms. An unmatched mode (there is no $\bar{\psi}$ zero mode) may be interpreted as saying that the ket $|\ \rangle$ has form degree one less than the bra $\langle\ |$. The inner product of these two then naturally vanishes.

This means that if we wish to calculate the transition between one of the approximate ground states $|P_1\rangle$ at P_1 with Morse index $i_{P_1} = p_1$ (so it must be a p_1-form), and one of the approximate ground states $|P_2\rangle$ at P_2, then the Morse index at P_2 must equal $p_1 + 1$. A potentially non-vanishing expectation value of interest to calculate is $\langle P_2|d_s|P_1\rangle$ (recall that $d_s \leftrightarrow Q$). This we have already done in section 3.9, eq. (3.150),

$$\langle P_2|d_s|P_1\rangle = n(P_1, P_2) \ , \tag{3.164}$$

where $n(P_1, P_2)$ is an integer counting the number of instantons with appropriate signs between the two critical points. Determining the sign is somewhat more difficult, and we will only sketch the method here. Let ϕ_c be an instanton interpolating between P_1 and P_2, and let V_1 and V_2 be the $(p + 1)$- and p-dimensional vector spaces of negative eigenvectors of $H_{P_1} V$ and $H_{P_2} V$, respectively. The orientation of these vector spaces is naturally given by the states $|P_1\rangle$ and $|P_2\rangle$, as they are forms of the appropriate degree. The tangent vector to ϕ_c at P_1 is denoted by v, and $\tilde{V}_1$ is the subspace of V_1 orthogonal to v. By interior multiplication of v with the $(p + 1)$-form state $|P_1\rangle$, the orientation of $\tilde{V}_1$ is fixed.

The vector space $\tilde{V}_1$ is p-dimensional and so its orientation may be compared with that of V_2. The comparison may be made by parallel transporting the vector space $\tilde{V}_1$ along ϕ_c. n_{ϕ_c} for the path is defined to be $+1$ if the orientations agree, and -1 otherwise.

Now as $n(P_1, P_2) = \sum_{\phi_c} n_{\phi_c}$, the transition (3.164) is determined and this allows us to define a new "twisted" cohomology. Let X_p be a vector space of dimension M_p which is generated by the critical points and define the coboundary operator $\delta : X_p \to X_{p+1}$ by

$$\delta|R\rangle = \sum_P n(R, P)|P\rangle \ , \tag{3.165}$$

where the sum is over all the basis elements $|P\rangle$ of X_{p+1}. From eqs. (3.164) and (3.165) we see that the matrix elements of δ are given by the action of d_s or the BRST operator Q. Now as Q is nilpotent, then so is δ on these spaces, so that with $\delta^2 = 0$, it is a coboundary operator and hence defines a cohomology.

Denoting by Y_p the Betti numbers associated with this cohomology, Witten conjectured that $Y_p = b_p$. The instanton calculation establishes that, if $(\delta\delta^* + \delta^*\delta)|\Psi\rangle = \lambda|\Psi\rangle$ with non-zero eigenvalue λ, then $|\Psi\rangle$ has non-zero energy. The problem is to show that when $\lambda = 0$ the approximate eigenstate $|\Psi\rangle$ really has zero energy so that it is a true eigenstate. Intuitively, this is the case as instanton calculations frequently eliminate approximate degeneracies that exist in perturbation theory.

Further reading

The Langevin approach to topological field theories, as considered in the text, was introduced by Labastida and Pernici [3.38] in the context of Donaldson theory, and is reviewed in section 5. The general application of this approach, as well as its connection to Nicolai maps, was established in refs. [3.5, 3.6]. The stochastic approach to quantizing topological theories, with a stochastic time, was developed in ref. [3.39].

A variant of the model considered here was used to determine the index of the Dirac and Dolbeault operators coupled to gauge fields [3.3, 3.4, 3.26, 3.27]. The methods we have used to prove metric and coupling constant independence of the Euler character extend to these cases as well. In ref. [3.40] the Euler character for manifolds with boundary was determined using the supersymmetry model with a special choice of potential.

The question of BRST symmetry breaking was considered by Fujikawa [3.41]. He gave criteria analogous to those of Witten for the breaking of conventional supersymmetry, which coincide in one dimension. The models he considered we would now identify as being topological field theories, though they are topologically trivial. This work also has relevance to the question of the triviality of observables in Donaldson theory, which we discuss in section 5. A discussion pertaining to the present model can be found in ref. [3.42].

A treatment of zero modes in theories with instantons and solitons may be found in refs. [3.43–3.46]. A systematic BRST treatment for topological field theories was given in refs. [3.47, 3.32].

The question of the value of the Witten index for potentials which are not Morse functions has been addressed in refs. [3.48, 3.49].

Other works on supersymmetric quantum mechanics as a topological field theory include refs. [3.50–3.52].

4. Topological sigma models

4.1. Introduction

Topological sigma models can be studied in a manner analogous to the approach we took in discussing supersymmetric quantum mechanics. These models are related to some mathematical results of Floer [4.1] on the number of fixed points of certain symplectic diffeomorphisms, as well as to related work of Gromov [4.2]. In one of Floer's constructions, a chain complex, analogous to the Witten complex, is defined where the boundary operator is associated with a certain type of instanton. The Morse theoretic information in this case deals with fixed points of exact diffeomorphisms of a symplectic manifold. From the field theory point of view, once one has the appropriate instanton equation in hand, it is a straightforward procedure to "quantize" that classical equation, as we did in supersymmetric quantum mechanics, and so construct a topological quantum field theory. In this case, the instantons that enter Floer's work are certain types of maps from a two-dimensional domain into a target space which has a symplectic structure. It is quite natural then to reformulate these data in terms of a sigma model where we consider maps from a Riemann surface into some target space. This approach allows for certain extensions in the original scenario, and one can construct models where the target space has only an almost complex structure. These models will, however, simplify greatly in the Kähler case.

In section 4.2, we undertake a brief review of some mathematical concepts which are unavoidable in any presentation of the topological sigma models. With this machinery in hand, we will then proceed in section 4.3 to review some of the key results of refs. [4.1, 4.2], which served as the motivating factors in Witten's construction [4.3]. Following this, we construct these models from various points of view in section 4.4, and describe the topological data encoded in their observables in section 4.5.

4.2. Review of complex manifolds

It is our goal here to review some of the mathematics associated with complex manifolds. Our presentation will be spartan; we will review only those aspects of the subject which enter in the succeeding discussions. As usual, we assume that the reader is familiar with the basics of real manifold geometry and topology.

Let us begin with the definition of a complex manifold. A *complex manifold* of (complex) dimension m is a paracompact Hausdorff space, together with a covering by open sets each homeomorphic to $\mathbb{C}^m$. In addition, we require that the coordinate transformations (or transition functions) which are defined in the overlaps of two of these open sets (called coordinate patches) are given by holomorphic functions. The collection of local neighborhoods together with the transition functions is called the *atlas*.

The definition of a complex manifold differs from that of a real $2m$-dimensional manifold only in the requirement that the transition functions are holomorphic, and not merely C^∞ smooth. Clearly, every complex manifold can be considered as a real manifold. It is a natural and important problem

to determine which real even dimensional manifolds contain a subatlas consisting of holomorphic transition functions. Two complex manifolds M and N will be considered equivalent if there exists a diffeomorphism $f : M \rightarrow N$ which is also holomorphic in both directions.

We will of course be interested in tensors on complex manifolds, and their definitions are analogous to those encountered on real manifolds. At each point p of the complex manifold M of dimension m, there are the tangent and cotangent spaces, denoted by $T_p(M)$ and $T_p^*(M)$ respectively, which are complex vector spaces of dimension m. If we let $(z^1, \ldots, z^m)$ denote the complex coordinates in some coordinate patch, then the tangent space is spanned by the collection of vectors

$$\frac{\partial}{\partial z^1}, \ldots, \frac{\partial}{\partial z^m} , \tag{4.1}$$

while the cotangent space is spanned by the one-forms

$$dz^1, \ldots, dz^m . \tag{4.2}$$

We said before that any complex manifold can be considered as a smooth real manifold in a natural way; we simply map the complex coordinates $\mathbb{C}^m$ into $\mathbb{R}^{2m}$,

$$(z^1, \ldots, z^m) \rightarrow (x^1, y^1, \ldots, x^m, y^m) , \tag{4.3}$$

where $z^a = x^a + iy^a$. In contrast to the tangent space defined above, the tangent space to the underlying real manifold has $2m$ real dimensions, and is spanned by the collection of all partial derivatives with respect to both the x^a and y^a. If we introduce the complex conjugates of the coordinates, $\bar{z}^a = x^a - iy^a$, then we can exchange the x, y description with the $z, \bar{z}$ notation. It is then natural to define

$$\frac{\partial}{\partial z^a} = \frac{1}{2}\left(\frac{\partial}{\partial x^a} - i\frac{\partial}{\partial y^a}\right) , \qquad \frac{\partial}{\partial \bar{z}^a} = \frac{1}{2}\left(\frac{\partial}{\partial x^a} + i\frac{\partial}{\partial y^a}\right) . \tag{4.4}$$

Equivalently, we can write the derivatives in the x and y directions as linear combinations of the z and $\bar{z}$ derivatives. It is conventional to denote the tangent space $T_p(M)$ to the complex manifold M at point p by $T_p^{(1,0)}(M)$ and the analogous complex vector space spanned by the $\bar{z}$ derivatives by $T_p^{(0,1)}(M)$. We should emphasize that $T_p(M)$, as we have defined it for a complex manifold, is quite different from the tangent space to the underlying real manifold, which is given by the real linear combinations of the vectors $\partial/\partial x^a$ and $\partial/\partial y^a$.

A key observation is that the barred and unbarred vectors do not transform into each other under a holomorphic change of coordinates. The tensor algebra then has a finer decomposition than in the case of a real manifold. It is now meaningful—in the global sense—to discuss tensors with some definite number of holomorphic and anti-holomorphic indices of either covariant or contravariant type. For example, a two-form ω of type $(1,1)$ is defined to be a tensor which in each coordinate patch takes the form

$$\omega_{a\bar{b}} \, dz^a \wedge d\bar{z}^b . \tag{4.5}$$

Notice that we place a bar over an index on a tensor component if it is to be contracted with an anti-holomorphic vector or covector. It is also sometimes customary to place a bar over the index on the conjugate of a holomorphic coordinate vector or covector, e.g. $d\bar{z}^a$, or equivalently $dz^{\bar{a}}$. Since there is really no possibility of confusion, we will omit the extra bars.

If $M_\mathbb{R}$ denotes the underlying real manifold of a complex manifold M, then we know that the exterior operator $d : \Omega^r(M_\mathbb{R}) \to \Omega^{r+1}(M_\mathbb{R})$ is defined and it plays an important role, where $\Omega^r(M_\mathbb{R})$ is the space of r-forms on $M_\mathbb{R}$. In terms of the real x and y coordinates, this operator takes the form

$$d = \mathrm{d}x^a \, \partial/\partial x^a + \mathrm{d}y^a \, \partial/\partial y^a \ . \tag{4.6}$$

Rewriting this in complex coordinates, we have the equivalent expression

$$d = \mathrm{d}z^a \, \partial/\partial z^a + \mathrm{d}\bar{z}^a \, \partial/\partial \bar{z}^a \ . \tag{4.7}$$

Our observation that holomorphic coordinate transformations do not mix the holomorphic vectors $\partial/\partial z^a$ with the anti-holomorphic vectors $\partial/\partial \bar{z}^a$ means that the exterior operator d decomposes into the sum of two globally defined operators,

$$d = \partial + \bar{\partial} \ , \tag{4.8}$$

where $\partial = \mathrm{d}z^a \, \partial/\partial z^a : \Omega^{(p,q)}(M) \to \Omega^{(p+1,q)}(M)$, and similarly for $\bar{\partial}$. The nilpotency of d translates into the set of conditions

$$0 = d^2 = \partial^2 = \bar{\partial}^2 = \partial\bar{\partial} + \bar{\partial}\partial \ . \tag{4.9}$$

The operator $\bar{\partial}$ is called the Dolbeault operator.

An important property of the exterior operator is expressed in the Poincaré lemma, which states that, if a form $\alpha \in \Omega^p(U)$, $p \geq 1$, satisfies $d\alpha = 0$ (α is closed), then there exists a $\beta \in \Omega^{p-1}(V)$ for some open set $V \subset U$ such that $\alpha = d\beta$ (α is locally exact). It is no surprise that this lemma has a refinement on a complex manifold. The Dolbeault–Grothendieck lemma states that, if $\alpha \in \Omega^{(p,q)}$, $q \geq 1$ and $\bar{\partial}\alpha = 0$, then locally there exists $\beta \in \Omega^{(p,q-1)}$ such that $\alpha = \bar{\partial}\beta$.

We have already seen that the structure of a complex manifold allows one to define intrinsic objects such as ∂ and $\bar{\partial}$, which are not common to every real manifold. Another such structure is a globally defined tensor field J with one covariant and one contravariant index (i.e., J is an endomorphism of the space $T^{(1,0)} \oplus T^{(0,1)}$) whose square is -1. Such a tensor can be defined by

$$J^a{}_b = \mathrm{i}\delta^a{}_b \ , \qquad J^{\bar{a}}{}_{\bar{b}} = -\mathrm{i}\delta^{\bar{a}}{}_{\bar{b}} \ , \tag{4.10}$$

and where the other components vanish, in each coordinate neighborhood. These piece together to define a global tensor field since holomorphic changes of coordinates preserve this structure. We will return to a further discussion of J in due course; here we merely cite this construction as an observation.

With these remarks in hand, we will now address the issue of whether a given real manifold can be viewed as a complex manifold by restricting the atlas in such a way as to select a subatlas whose transition functions are holomorphic. It might well be the case that there are many such subatlases which are not holomorphically equivalent. In any case, we now see that the real manifold in question must at least admit an *almost complex structure*, i.e., a globally defined tensor field with one covariant and one contravariant index whose square is -1. It is a trivial exercise in linear algebra to show that the existence of such a matrix requires that the tangent space be of even (real) dimension. Moreover, one can also show that the existence of an almost complex structure implies that the manifold is orientable (a global, nowhere vanishing form of maximal degree is defined), so in particular, all complex manifolds are orientable. It is convenient to call a real manifold *almost complex* if it admits an almost complex structure. Our problem can now be refined: which almost complex manifolds contain a holomorphic subatlas ?

The answer to this question is the content of a theorem of Newlander and Nirenberg, which states that a given almost complex structure arises from a complex manifold if the Nijenhuis tensor, defined by

$$N^k_{ij} = J^l{}_i(\partial_l J^k{}_j - \partial_j J^k{}_l) - J^l{}_j(\partial_l J^k{}_i - \partial_i J^k{}_l) \ , \tag{4.11}$$

is zero. One can show that N is a tensor by first defining an analogous quantity with covariant derivatives (choose some metric and the associated Christoffel connection) and then noting that the terms proportional to the connection drop out; this follows from the fact that the connection coefficients Γ^k_{ij} are symmetric in the two lower indices. In other words, an almost complex manifold has a holomorphic subatlas which gives rise to the given almost complex structure if the Nijenhuis tensor vanishes. That this is a necessary condition is obvious; one need only look to see that our construction of J from a holomorphic atlas was in terms of constant tensor components. The proof that this condition is in fact sufficient is difficult, and we will say no more about that here.

A convenient condition that guarantees the vanishing of the Nijenhuis tensor—as we have already noted—is when the almost complex structure is covariantly constant with respect to some Christoffel connection,

$$D_k J^i{}_j = 0 \ . \tag{4.12}$$

It is important to emphasize that we have a torsion-free connection; we cannot conclude that N^k_{ij} vanishes if we merely have covariant constancy with respect to some connection with non-zero torsion.

A metric h (positive definite as always) on an almost complex manifold is called *Hermitian* if it is compatible with the almost complex structure J in the sense that $h(JX, JY) = h(X, Y)$. In coordinates, this condition takes the form

$$h_{ij} = J^m{}_i J^n{}_j h_{mn} \ . \tag{4.13}$$

When the manifold is actually complex, where it is possible to put the complex structure in the canonical form defined above, then this condition simply means that $h_{ab} = h_{\bar{a}\bar{b}} = 0$. The existence of Hermitian metrics is not really an issue, since we can construct a Hermitian metric from any given Riemannian metric g by defining

$$h(X, Y) = \tfrac{1}{2}(g(X, Y) + g(JX, JY)) \ , \tag{4.14}$$

where positive definiteness is clearly preserved. Now, if we are given a Hermitian metric h, there is a natural way to construct a two-form K; just define

$$K(X, Y) = h(X, JY) \ . \tag{4.15}$$

Since h is symmetric, and $J^2 = -1$, it is trivial to check that $K(X, Y) = -K(Y, X)$. In components, we have that $K_{ij} = h_{ik} J^k{}_j$, so we see that K is also non-degenerate (invertible) since both h and J are non-degenerate.

The constructions we have just considered are set within the framework of an almost complex manifold, which is by definition real. If the Nijenhuis tensor vanishes, then it is possible to find a subatlas in which the transition functions are holomorphic. It this case, we would like to extend the tensors we have defined to act on complex vector spaces. The complexification of a real vector space V (for example, V could be the tangent space to the real manifold) is defined to be $V^c = V \otimes \mathbb{C}$, the tensor product of V with the complex numbers. This vector space decomposes into the $+i$ and

–i eigenvalue subspaces of the operator J, $V^c = V^{(1,0)} \oplus V^{(0,1)}$. It is straightforward to verify that h extends uniquely to a complex symmetric bilinear form which satisfies

(1) $h(\overline{X}, \overline{Y}) = \overline{h(X,Y)}$;

(2) $h(X, \overline{X}) > 0$, for $X \neq 0$;

(3) $h(X, \overline{Y}) = 0$, for $X \in T^{(1,0)}(M)$, $Y \in T^{(0,1)}(M)$. (4.16)

Similarly, the two-form K extends to a two-form of type $(1,1)$, $K \in \Omega^{(1,1)}(M)$.

A complex manifold with a Hermitian metric h is said to be *Kähler*, if the associated two-form K—which we defined above—is closed, $dK = 0$. The metric in this case is called the *Kähler metric*. Since $\partial K = \overline{\partial} K = 0$, the Dolbeault lemma, which we quoted earlier, implies that—at least locally—there exists a function ϕ such that

$$K = i\partial\overline{\partial}\phi \; , \tag{4.17}$$

and this function ϕ is called the *Kähler potential*. This Kähler situation will arise if we have an almost complex structure which is covariantly constant with respect to the Christoffel connection of some Riemannian metric, i.e., $D_k J^i{}_j = 0$. If the Riemannian metric is denoted by g, then we first form the Hermitian metric h and the associated two-form K,

$$K_{ij} = h_{ik} J^k{}_j \; , \tag{4.18}$$

where $h_{ij} = \frac{1}{2}(g_{ij} + J^m{}_i J^n{}_j g_{mn})$ as before. Since both g and J are covariantly constant, so are h and K. Hence, $D_k K_{ij} = 0$ implies $D_{[k} K_{ij]} = \partial_{[k} K_{ij]} = 0$, and we see that K is a closed two-form which is, moreover, non-degenerate. Any real manifold which has a closed, non-degenerate two-form is called a *symplectic manifold*, and this particular two-form is called the *symplectic structure*; all Kähler manifolds are therefore symplectic.

4.3. Mathematical motivation

In this section, we will highlight some of the mathematical results that predated Witten's field theoretic formulation. Some results of Floer [4.1] on the number of fixed points of certain symplectic diffeomorphisms are noteworthy in this regard. In his Morse theoretic approach, Floer constructs a chain complex, analogous to the Witten complex (which we reviewed in section 3.8.1), where the boundary operator is defined in terms of an instanton-like equation. Given such an equation, it is straightforward to construct a suitable quantum field theory based on the space of its solutions. Let us begin by setting the stage for Floer's fixed point theorems.

Let P denote a symplectic manifold with symplectic structure ω (ω is a closed, nondegenerate two-form), and consider the vector field X_t generated by a Hamiltonian H_t, where t labels any explicit time dependence that may be present. That is, to each smooth function $H : P \times \mathbb{R} \rightarrow \mathbb{R}$, $H(x,t) = H_t(x)$, we naturally have a one-parameter family of vector fields X_t defined by the condition

$$\omega(\cdot, X_t) = dH_t \; . \tag{4.19}$$

The uniqueness of this vector field is simply a consequence of the nondegeneracy of ω. The integral curves of this vector field satisfy (by definition)

$$\frac{d}{dt}\phi_{H,t}(x) = X_t(\phi_{H,t}(x)) \; , \qquad \phi_{H,0}(x) = x \; , \tag{4.20}$$

and are a one-parameter family of diffeomorphisms when P is compact, which we assume. Compactness of P guarantees that the integral curves are complete [4.4]. Moreover, this one-parameter family of diffeomorphisms, which is clearly homotopic to the identity, also preserves the symplectic structure in the sense that $\phi_t^* \omega = \omega$, for all $t \in \mathbb{R}$. To see this, it is clearly sufficient to show that $\phi_t^* \omega$ is independent of t,

$$(d/dt)\phi_t^* \omega = -L_{X_t}\omega = 0 \ . \tag{4.21}$$

The last equality is easily established by writing the Lie derivative in coordinates, and using the defining property of the vector field X_t. We are interested in the set $\mathcal{D}$ of diffeomorphisms which arise in this way,

$$\mathcal{D} = \{\phi_{H,t} \mid H \in C^\infty(P \times \mathbb{R}) \text{ and } t \in \mathbb{R}\} \ , \tag{4.22}$$

and these are called the *exact diffeomorphisms*.

The theorem we wish to discuss involves nondegenerate fixed points of exact diffeomorphisms. If x is a fixed point of some diffeomorphism ϕ, i.e., $\phi(x) = x$, then one says that the fixed point is nondegenerate [4.5] if the Jacobian $J_{\phi(x)}$ satisfies

$$\det[J_{\phi(x)} - \mathrm{id}] \neq 0 \ . \tag{4.23}$$

The theorem of Floer [4.1] can now be easily stated:

Let P be a compact, closed symplectic manifold with $\pi_2(P) = 0$. Consider an exact diffeomorphism $\phi : P \to P$ all of whose fixed points are non-degenerate. Then the number of fixed points is greater than or equal to the sum of the Betti numbers of P with respect to Z_2 coefficients.

The key to this theorem is the construction of a chain complex based on the set of all fixed points of ϕ. The boundary operator is defined in terms of an instanton-like equation, analogous to the one we encountered in supersymmetric quantum mechanics. For $\phi \in \mathcal{D}$, let

$$\Omega(\phi) = \{z \in C^\infty([0,1], P) \mid z(1) = \phi(z(0))\} \ . \tag{4.24}$$

If we choose an almost complex structure J on P such that $g = \omega(J\cdot, \cdot)$ is a metric, then an instanton u is a one-parameter family in $\Omega(\phi)$, $u : \mathbb{R} \times [0,1] \to P$ which satisfies

$$\partial u(\tau,t)/\partial \tau + J(u)\,\partial u(\tau,t)/\partial t = 0 \ , \tag{4.25}$$

and converges to fixed points of ϕ as $\tau \to \pm\infty$. Given this equation, and our experience in supersymmetric quantum mechanics, it is most natural to consider a quantum field theory, defined more generally on an arbitrary Riemann surface, whose classical action is given by the square of the Langevin equation. We will carry out this construction in the next section.

4.4. Construction and properties of the model

4.4.1. The Langevin approach

We shall now demonstrate that the topological sigma model action of Witten [4.3] can be obtained upon BRST quantization of the following classical action [4.6, 4.7]:

$$S_c = \int_{\Sigma} d^2\sigma \, h_{\alpha\beta} g_{ij} K^{\alpha i} K^{\beta j} \ , \tag{4.26}$$

where

$$K^{\alpha i} = G^{\alpha i} - \tfrac{1}{2}(\partial^\alpha u^i + \epsilon^\alpha{}_\beta J^i{}_j \partial^\beta u^j) \, , \tag{4.27}$$

and $G^{\alpha i}$ is the random Gaussian field; as usual we shall suppress $\sqrt{h}$ factors.

The above action describes a theory of maps $u^i(\sigma)$ from a Riemann surface Σ to (in our case) an almost complex manifold M. The coordinates on Σ are denoted by σ^α ($\alpha = 1, 2$), while those on the target manifold M are denoted by u^i ($i = 1, ..., \dim M$). $h_{\alpha\beta}$ and $\epsilon^\alpha{}_\beta$ are the metric and complex structure of Σ, respectively. They obey the relations $\epsilon^\alpha{}_\beta \epsilon^\beta{}_\gamma = -\delta^\alpha{}_\gamma$ and $\epsilon_{\alpha\beta} = h_{\alpha\gamma}\epsilon^\gamma{}_\beta$. g_{ij} and J_{ij} are the metric tensor and almost complex structure of M and obey analogous relations to the above. At present we will be completely general, treating the case of almost complex manifolds. Following this we then specialize to various complex and Kähler manifolds.

Our first problem is to establish the local symmetries of the action (4.26). As before, we postulate the shift symmetry

$$\delta u^i = \epsilon^i \, . \tag{4.28}$$

The invariance of the action then determines the transformation for $G^{\alpha i}$. As there are some subtleties involved here, we present a few steps in the derivation. It is first useful to define the following self-dual and anti-self-dual projection operators

$$P_\pm^{\alpha i}{}_{\beta j} = \tfrac{1}{2}(\delta^\alpha{}_\beta \delta^i{}_j \pm \epsilon^\alpha{}_\beta J^i{}_j) \, . \tag{4.29}$$

The fields $G^{\alpha i}$ and $K^{\alpha i}$ both satisfy the self-duality constraint

$$G^{\alpha i} = P_+^{\alpha i}{}_{\beta j} G^{\beta j} \, , \qquad K^{\alpha i} = P_+^{\alpha i}{}_{\beta j} K^{\beta j} \, . \tag{4.30}$$

Now in deriving the transformation for G and K, it is important to ensure that this constraint is maintained; in other words we have

$$K'^{\alpha i} = P_+'^{\alpha i}{}_{\beta j} K'^{\beta j} \, , \tag{4.31}$$

and similarly for G. Rewriting (4.31) as

$$P_-^{\alpha i}{}_{\beta j} \delta K^{\beta j} = \tfrac{1}{2}\epsilon^\alpha{}_\beta \delta J^i{}_j K^{\beta j} \, , \tag{4.32}$$

we see that the variation of K must contain an anti-self-dual part, in order to maintain the self-duality of the original field K. The total variation of K can now be expressed as

$$\delta K^{\alpha i} = P_+^{\alpha i}{}_{\beta j} A^{\beta j} + \tfrac{1}{2}\epsilon^\alpha{}_\beta \delta J^i{}_j K^{\beta j} \, , \tag{4.33}$$

and similarly for G,

$$\delta G^{\alpha i} = P_+^{\alpha i}{}_{\beta j} B^{\beta j} + \tfrac{1}{2}\epsilon^\alpha{}_\beta \delta J^i{}_j G^{\beta j} \, , \tag{4.34}$$

where $A^{\alpha i}$ and $B^{\alpha i}$ are arbitrary tensors. The idea now is to use (4.27) to establish a relation between A and B. Invariance of the action fixes A, and from this we can obtain the transformation of G. From (4.27) we have

$$\delta K^{\alpha i} = \delta G^{\alpha i} - P_+^{\alpha i}{}_{\beta j} \partial^\beta \epsilon^j - \tfrac{1}{2}\epsilon^\alpha{}_\beta \delta J^i{}_j \partial^\beta u^j \, . \tag{4.35}$$

The third term on the right hand side of (4.35) can be decomposed into self-dual and anti-self-dual components as follows:

$$\epsilon^\alpha{}_\beta \delta J^i{}_j \partial^\beta u^j = P_+^{\alpha i}{}_{\beta j} \epsilon^\beta{}_\gamma \delta J^j{}_k \partial^\gamma u^k + P_-^{\alpha i}{}_{\beta j} \epsilon^\beta{}_\gamma \delta J^j{}_k \partial^\gamma u^k \, . \tag{4.36}$$

This leads to the relation

$$A^{\beta j} = B^{\beta j} - \partial^\beta \epsilon^j - \tfrac{1}{2} \epsilon^\beta{}_\gamma \delta J^j{}_k \partial^\gamma u^k \ . \tag{4.37}$$

If we now examine the variation of the action under the transformation (4.28) we find that S_c is invariant if

$$A^{\alpha i} = -\Gamma^i_{jk} K^{\alpha i} \epsilon^k \ , \tag{4.38}$$

leading to

$$\delta G^{\alpha i} = P_+^{\alpha i}{}_{\beta j} (D^\beta \epsilon^j + \tfrac{1}{2} \epsilon^\beta{}_\gamma \epsilon^l (D_l J^j{}_k) \partial^\gamma u^k) + \tfrac{1}{2} \epsilon^\alpha{}_\beta \epsilon^k (D_k J^i{}_j) G^{\beta j} - \Gamma^i_{lk} \epsilon^k G^{\alpha l} \ . \tag{4.39}$$

Having determined the classical symmetries of the model, we can now proceed with the quantization, the details of which can be found in appendix A. The result is that the complete quantum action can be written as a BRST commutator,

$$S_q = - \int \mathrm{d}^2\sigma \, \{Q, \overline{C}_{\alpha i} (\partial^\alpha u^i - \tfrac{1}{4} \alpha B^{\alpha i})\} \ , \tag{4.40}$$

where α is a gauge fixing parameter, and Q is the off-shell nilpotent BRST operator defined by $\delta = -\epsilon \{Q, \ \}$,

$$\begin{aligned}
\delta u^i &= -\epsilon C^i \ , \qquad \delta C^i = 0 \ , \\
\delta \overline{C}_{\alpha i} &= \epsilon (B_{\alpha i} + \tfrac{1}{2} \epsilon_\alpha{}^\beta (D_k J^j_i) \overline{C}_{\beta j} C^k + \Gamma^k_{ij} \overline{C}_{\alpha k} C^j) \ , \\
\delta B^{\alpha i} &= \tfrac{1}{4} \epsilon C^k C^l (R_{kl}{}^i{}_t + R_{klrs} J^{ri} J^s{}_t) \overline{C}^{\alpha t} - \tfrac{1}{2} \epsilon \epsilon^\alpha{}_\beta (D_k J^i{}_j) C^k B^{\beta j} \\
&\quad + \tfrac{1}{4} \epsilon (C^k D_k J^i{}_s)(C^l D_l J_t{}^s) \overline{C}^{\alpha t} + \epsilon \Gamma^i_{jk} C^j B^{\alpha k} \ ,
\end{aligned} \tag{4.41}$$

where

$$B_{\alpha i} = B'_{\alpha i} - P_{+\alpha i}{}^{\beta j} \overline{C}_{\beta l} \Gamma^l_{jk} C^j \ , \tag{4.42}$$

and $B'_{\alpha i}$ is the original multiplier field in the theory. We have, of course, the freedom to choose the value of α, and different values expose different facets of the theory. In the sequel, we shall be mainly concerned with the values $\alpha = 0$ and 1.

For example, when $\alpha = 1$, the action upon integration over B is [4.3, 4.7]

$$\begin{aligned}
S_q = \int \mathrm{d}^2\sigma \Big[&\tfrac{1}{2} h^{\alpha\beta} g_{ij} \partial_\alpha u^i \partial_\beta u^j + \tfrac{1}{2} \epsilon^{\alpha\beta} J_{ij} \partial_\alpha u^i \partial_\beta u^j + \overline{C}_{\alpha i} [D^\alpha C^i + \tfrac{1}{2} \epsilon^\alpha{}_\beta (D_j J^i{}_k) \partial^\beta u^k C^j] \\
&+ \tfrac{1}{8} \overline{C}^m_\alpha \overline{C}^{\alpha k} R_{mkjr} C^j C^r + \tfrac{1}{16} \overline{C}_{\alpha i} \overline{C}^{\alpha k} (D_j J^{li})(D_r J_{lk}) C^j C^r \Big] \ .
\end{aligned} \tag{4.43}$$

As explained for the case of supersymmetric quantum mechanics, the presence of the quartic ghost coupling terms have their origin in the fact that the classical gauge algebra only closes on-shell. This in turn is reflected by the cubic terms in the BRST transformations (4.41). Obviously, here we must also confront the issue of the metric and complex structure independence of the BRST charge defined in (4.41). This leads us naturally to discuss the derivation of Witten's action as presented by Baulieu and Singer [4.8].

4.4.2. The Baulieu–Singer approach

An alternative derivation of Witten's topological sigma model action, for the Kähler case, was presented by the above authors. This involves taking the classical action to be

$$S_c = \frac{1}{2} \int_\Sigma d^2\sigma \, \epsilon^{\alpha\beta} J_{ij} \partial_\alpha u^i \partial_\beta u^j = \frac{1}{2} \int_\Sigma J \; , \tag{4.44}$$

where J is a two-form on the target manifold M. Now for the case of a symplectic target space, this two-form J is closed, $dJ = 0$. It then follows that the action (4.44) is a topological invariant, depending only on the particular homology class of Σ. The classical invariance of this action is an arbitrary shift symmetry $\delta u^i = \epsilon^i$.

As for the case of quantum mechanics, the basic aim of the Baulieu–Singer construction is to write down a simple geometrical set of transformation rules, and then to choose what would conventionally be regarded as an unusual gauge fixing condition. In this way one can recover Witten's action (not only for the Kähler case, but also when the target space is almost complex). In the present case the BRST transformations take the form

$$\delta u^i = -\epsilon C^i \; , \qquad \delta C^i = 0 \; , \qquad \delta \overline{C}^{\alpha i} = \epsilon \overline{B}^{\alpha i} \; , \qquad \delta \overline{B}^{\alpha i} = 0 \; , \tag{4.45}$$

and the quantum action is expressed as

$$S_q = -\int d^2\sigma \, \{Q, \overline{C}^{\alpha i} P_{+\alpha i, \beta j} (\partial^\beta u^j - \tfrac{1}{4}\Gamma^j_{rs} P^{\beta r}_{+\gamma t} \overline{C}^{\gamma t} C^s - \tfrac{1}{4}\overline{B}^{\beta j})\} \; . \tag{4.46}$$

Again, the points which need to be stressed here are the following. The BRST rules (4.45) are conventional in the sense that the anti-ghost transforms into the multiplier, while the multiplier transforms into zero. In order to be able to generate quartic ghost coupling terms, one must then choose a gauge fixing condition which depends quadratically on the ghosts. The exact form of this condition is determined by the requirement that the final action be covariant. As before, (4.45) and (4.46) follow immediately from (4.40) and (4.41) by a simple shift in the fields

$$\overline{B}^{\alpha i} = B^{\alpha i} + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_k J^i{}_j) \overline{C}^{\beta j} C^k - \Gamma^i_{jk} \overline{C}^{\alpha j} C^k.$$

We note that the term proportional to $D_k J^i{}_j$ does not contribute in (4.46) because of the self-duality constraint.

The second point is that the BRST charge Q defined by (4.45) is independent of the metric and complex structure of both the base and target space. It thus follows trivially that the variations of the action with respect to these parameters are also BRST commutators, thereby ensuring similar invariances for the partition function. However, the crucial difference in the present sigma model case is the self-duality constraint on the anti-ghost and multiplier field. It is important to realize that the transformations given in (4.45) are defined for unconstrained fields. For self-dual fields one requires the following modification:

$$\delta \overline{C}^{\alpha i} = \delta (P^{\alpha i}_{+\beta j} \overline{C}^{\beta j}) = (\delta P^{\alpha i}_{+\beta j}) \overline{C}^{\beta j} + P^{\alpha i}_{+\beta j} \delta \overline{C}^{\beta j} = \tfrac{1}{2}\epsilon \epsilon^\alpha{}_\beta (\partial_k J^i{}_j) \overline{C}^{\beta j} C^k + \epsilon \overline{B}^{\alpha i} \; , \tag{4.47}$$

and similarly for $\overline{B}^{\alpha i}$. In other words, although one notices the presence of the complex structure on the right hand side of these transformation rules, this arises solely through the variation of the self-dual projection operator with respect to the original—metric and complex structure independent—Q. It is also worth noting that the offending term on the right hand side of (4.47) vanishes

when the target space is Kähler. In presenting the Baulieu–Singer construction, we have used the transformations for unconstrained fields (4.45), and inserted by hand the self-dual projection operators in the action (4.46).

4.4.3. The Nicolai map

Here, we shall show explicitly that a complete Nicolai map exists for this theory, as expected. The simplest way to see this is to work in the delta function gauge, by choosing $\alpha = 0$. The action in this case is given by

$$
\begin{aligned}
S_q &= -\int d^2\sigma \, \{Q, \overline{C}_{\alpha i} \partial^\alpha u^i\} \\
&= \int d^2\sigma \left(B_{\alpha i} \partial^\alpha u^i + \overline{C}_{\alpha i} [D^\alpha \delta^i{}_j + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k) \partial^\beta u^k] C^j \right) \ .
\end{aligned}
\tag{4.48}
$$

We first note that

$$
B_{\alpha i} \partial^\alpha u^i = B_{\alpha i} P_+^{\alpha i}{}_{\beta j} \partial^\beta u^j \ .
\tag{4.49}
$$

The Nicolai map is then defined by

$$
u^i \rightarrow \xi^{\alpha i} = P_+^{\alpha i}{}_{\beta j} \partial^\beta u^j \ .
\tag{4.50}
$$

Such a transformation obviously trivializes the bosonic part of the above action. Our task now is to determine the Jacobian of such a change of variables.

This is achieved as follows: We first write

$$
u^i \rightarrow u^i + \hat{u}^i \ ,
\tag{4.51}
$$

where $\hat{u}^i$ is a small fluctuation. Expanding $\xi^{\alpha i}$ to first order in the fluctuation then allows us to read off the Jacobian determinant. The problem, however, is that, since $\hat{u}^i$ is the difference between two coordinate values (u^i and $u^i + \hat{u}^i$), it does not transform simply under target space reparametrizations (except when M is flat). The general method for dealing with this situation is well known [4.9], and involves choosing a geodesic $\lambda^i(t)$ with $\lambda^i(0) = u^i$ and $\lambda^i(1) = u^i + \hat{u}^i$, and defining $\xi^i = \dot{\lambda}^i(0)$. ξ^i is then a contravariant vector on M, and all fields can be expanded covariantly in powers of ξ^i.

A general covariant tensor field $T_{k_1 \cdots k_m}(u^i)$ has an expansion of the form

$$
T_{k_1 \cdots k_m}(u^i + \hat{u}^i) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{\partial}{\partial \xi_{i_1}} \cdots \frac{\partial}{\partial \xi_{i_n}} \right) T_{k_1 \cdots k_m}(u^i) \xi^{i_1} \cdots \xi^{i_n} \ ,
\tag{4.52}
$$

where the coefficients are tensors which can be expressed in terms of covariant derivatives of T and the Riemann tensor of M. To obtain the manifestly covariant form of these coefficients, it is simplest to use a normal coordinate system. For our purposes, it suffices to state the following results:

$$
g_{ij}(u + \hat{u}) = g_{ij}(u) + O(\xi^2) \ ,
\tag{4.53}
$$

$$
T_{ij}(u + \hat{u}) = T_{ij}(u) + (D_k T_{ij}(u)) \xi^k + O(\xi^2) \ ,
\tag{4.54}
$$

$$
\partial_\alpha(u^i + \hat{u}^i) = \partial_\alpha u^i + D_\alpha \xi^i + O(\xi^2) \ ,
\tag{4.55}
$$

where the linear term in (4.53) is absent since the metric is covariantly constant. Using (4.53) and (4.54) we find

$$J^i{}_j(u + \hat{u}) = J^i{}_j(u) + (D_k J^i{}_j(u))\xi^k + O(\xi^2) \ . \tag{4.56}$$

The expansion of $\xi^{\alpha i}$ now follows,

$$\xi^{\alpha i}(u + \hat{u}) = \xi^{\alpha i}(u) + P_+^{\alpha i}{}_{\beta j} D^\beta \xi^j + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k(u))\partial^\beta u^k \xi^j + O(\xi^2) \ . \tag{4.57}$$

We can now read off the Jacobian operator, and we notice that it is precisely the ghost operator defined in the action (4.48). Hence, the partition function (for the case of zero-dimensional moduli space) reduces to a signed ratio of determinants,

$$Z = \frac{\det[P_+^{\alpha i}{}_{\beta j}(D^\beta \delta_l^j + \tfrac{1}{2}\epsilon^\beta{}_\gamma (D_l J^j{}_k)\partial^\gamma u^k)]}{|\det[P_+^{\alpha i}{}_{\beta j}(D^\beta \delta_l^j + \tfrac{1}{2}\epsilon^\beta{}_\gamma (D_l J^j{}_k)\partial^\gamma u^k)]|} \ , \tag{4.58}$$

where we have not included the trivialized action $B_{\alpha i}\xi^{\alpha i}$.

4.4.4. A more general model

Having discussed some features of the model due to Witten, it is interesting to ask whether this model can be generalized. In analogy with supersymmetric quantum mechanics, one wonders whether there is the freedom to add a potential term to the action. As we will now see, this is indeed possible, and the implications are in fact quite interesting.

Firstly, it allows us to identify a simple flat space model due to Cecotti and Girardello [4.10] as a topological field theory. Secondly, the addition of a potential term allows for the possibility of studying supersymmetry breaking, along the lines discussed for the case of quantum mechanics; and finally, it gives us another interpretation of the Nicolai map, namely, that it corresponds to a Bäcklund transformation for the system [4.11], as we shall describe in the following section. Without further ado, let us present the classical action,

$$S_c = \int d^2\sigma \sqrt{h}\, h_{\alpha\beta} g_{ij} K^{\alpha i} K^{\beta j} \ , \tag{4.59}$$

where

$$K^{\alpha i} = G^{\alpha i} - \tfrac{1}{2}(\partial^\alpha u^i + \epsilon^\alpha{}_\beta J^i{}_j \partial^\beta u^j) + \tfrac{1}{2}(\partial^i V^\alpha + \epsilon^\alpha{}_\beta J^i{}_j \partial^j V^\beta) \ . \tag{4.60}$$

Here, $V(u)$ is the potential depending on the target space coordinates.

There is no obstruction to quantizing this general model; however, for our purposes we wish only to consider a specific example. Let us take the base manifold to be flat, and the target space to be a two-dimensional flat Kähler manifold, namely the complex line $\mathbb{C}$. In this case the action (4.59) takes the simple form

$$S_c = 2\int d^2\sigma \sqrt{h}\, h^{+-} g_{I\bar{J}} K_+{}^I K_-{}^{\bar{J}} = 2\int d^2\sigma\, K_+{}^u K_-{}^{\bar{u}} \ , \tag{4.61}$$

where the non-zero components of $K_\alpha{}^i$ are

$$K_+{}^u = G_+{}^u - \partial_+ u + \partial_{\bar{u}} V^- \ , \qquad K_-{}^{\bar{u}} = G_-{}^{\bar{u}} - \partial_- \bar{u} + \partial_u V^+ \ . \tag{4.62}$$

Our conventions here are the following: The target space coordinates are

$$u^I = u = u^1 + iu^2 \ , \qquad u^{\bar{I}} = \bar{u} = u^1 - iu^2 \ , \tag{4.63}$$

while the base manifold coordinates are denoted by

$$\sigma^\pm = \tfrac{1}{2}(\sigma^1 \mp i\sigma^2) \ , \tag{4.64}$$

with

$$\partial_\pm = \partial_1 \pm i\partial_2 \ , \tag{4.65}$$

and the potential is an analytic function of one variable with

$$\begin{aligned}
V^+ &= V^1 + iV^2 = V(u) \ , &\quad \partial_{\bar{u}}V^+ &= 0 \ , \\
V^- &= V^1 - iV^2 = V^*(\bar{u}) \ , &\quad \partial_u V^- &= 0 \ .
\end{aligned} \tag{4.66}$$

The classical symmetries of this action are

$$\delta u = \lambda \ , \qquad \delta\bar{u} = \bar{\lambda} \ , \qquad \delta G_+{}^u = \partial_+\lambda - (\partial_{\bar{u}}^2 V^-)\lambda \ , \qquad \delta G_-{}^{\bar{u}} = \partial_-\bar{\lambda} - (\partial_u^2 V^+)\bar{\lambda} \ , \tag{4.67}$$

where $\lambda^I = \lambda$ and $\lambda^{\bar{I}} = \bar{\lambda}$. To quantize this system, it is sufficient to invoke the standard Faddeev-Popov prescription, which yields the following quantum action:

$$\begin{aligned}
S_q &= S_c + \int d^2\sigma \{Q, \rho_-{}^{\bar{u}}G_+{}^u + \rho_+{}^{\bar{u}}G_-^{\bar{u}}\} \\
&= S_c + \int d^2\sigma \, [B_-{}^{\bar{u}}G_+{}^u + B_+{}^u G_-{}_{\bar{u}} \\
&\quad + \rho_-{}^{\bar{u}}(\partial_+ c - (\partial_{\bar{u}}^2 V^-)\bar{c}) + \rho_+{}^u(\partial_- \bar{c} - (\partial_u^2 V^+)c)] \ ,
\end{aligned} \tag{4.68}$$

where B are the multipliers enforcing the gauge constraints $G = 0$, and we are using the notation ρ for the anti-ghosts.

Our aim now is to establish contact with a well known $N = 2$ supersymmetric model in two dimensions. The action, given by Cecotti and Girardello [4.10], is

$$S = \int d^2\sigma \left[\partial_\alpha\phi\partial_\beta\phi^*\delta^{\alpha\beta} + \frac{\partial V}{\partial\phi}\frac{\partial V^*}{\partial\phi^*} \right. \\
\left. + \overline{\psi}\left(\gamma^\alpha\partial_\alpha - \tfrac{1}{2}(1 + \gamma_3)\frac{\partial^2 V}{\partial\phi\,\partial\phi} - \tfrac{1}{2}(1 - \gamma_3)\frac{\partial^2 V^*}{\partial\phi^*\,\partial\phi^*}\right)\psi\right] \ , \tag{4.69}$$

where the gamma matrices are defined by

$$\gamma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \qquad \gamma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \ , \qquad \gamma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \ , \tag{4.70}$$

and

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \ , \tag{4.71}$$

is a Dirac spinor.

Now the fermionic determinant arising from an integration over ψ and $\overline{\psi}$ is

$$\det[\gamma^\alpha\partial_\alpha - \tfrac{1}{2}(1 + \gamma_3)V'' - \tfrac{1}{2}(1 - \gamma_3)V^{*''}] \ . \tag{4.72}$$

Using the definition of the gamma matrices, it is easy to verify that this fermionic determinant is precisely the Faddeev–Popov ghost determinant in (4.68), with the identification

$$\psi = \begin{pmatrix} c \\ \bar{c} \end{pmatrix} \ , \qquad \bar{\psi} = \begin{pmatrix} \rho_+^{\ u} & \rho_-^{\ \bar{u}} \end{pmatrix} \ . \tag{4.73}$$

Together with the identifications $u = \phi$, $\bar{u} = \phi^*$, $V = V^+$ and $V^* = V^-$, we thus see that we can interpret the above supersymmetric model as arising from the BRST quantization of the topological field theory (4.61).

There is one piece of evidence, already provided by Cecotti and Girardello, which suggests that this should be the case, namely, the existence of a Nicolai map for the model. This is defined in ref. [4.10] to be

$$\xi = \partial_+\phi - \partial V^*(\phi^*)/\partial\phi^* \ , \qquad \xi^* = \partial_-\phi^* - \partial V(\phi)/\partial\phi \ . \tag{4.74}$$

Given such a Nicolai map, it is by now clear how to reconstruct the action of Cecotti and Girardello by BRST quantizing the square of the corresponding Langevin equation. While we have treated only the case where both metrics are flat, it is clear from the starting action (4.59) that indeed this model is topological. The ability to "twist" an $N = 2$ supersymmetric model in two dimensions into a topological model has been shown quite generally in ref. [4.12]. The difference between the two theories then lies in the interpretation of the physical states, as discussed in section 3.6. We also remark here on the interpretation of topological field theories (in general) as representing phases of unbroken general covariance. This interpretation arises due to the absence of physically propagating degrees of freedom. The question remains as to how one may effect a satisfactory breaking of this topological symmetry, and so liberate degrees of freedom. One of the main problems in this regard is to decide on a suitable order parameter which could distinguish the different phases of such a model. In the case under study in this section, i.e. the topological sigma model, we have the possibility of describing a phase of string theory in which general covariance is unbroken.

4.4.5. Nicolai maps and Bäcklund transformations

Having studied the simple flat space example above, we can now use our knowledge of this model to establish a connection between Nicolai maps and Bäcklund transformations [4.11]. In fact we shall show that, when the potential is either the Liouville or sine-Gordon potential, then the Nicolai map of the above system is precisely the Bäcklund transformation for the corresponding equation. Let us first define what is meant by a Bäcklund transformation, and illustrate it by way of a few examples.

Suppose we have two uncoupled partial differential equations, in two independent variables x and t, for the two functions f and g. The two equations are expressed as

$$P(f) = 0 \ , \qquad Q(g) = 0 \ , \tag{4.75}$$

where P and Q are, in general, non-linear operators. Let R_i $(i = 1, 2)$ be a pair of first order relations,

$$R_i(f, g; f_x, g_x; f_t, g_t; x, t) = 0 \ . \tag{4.76}$$

Then $R_i = 0$ is called a Bäcklund transformation for the system (4.75) if, given a solution f with $P(f) = 0$, it is integrable for g, and vice versa. If $P = Q$, the transformation is called an auto Bäcklund transformation.

Examples

(i) *Laplace's equation.* An auto Bäcklund transformation for Laplace's equation,

$$f_{xx} + f_{tt} = 0 \ , \qquad g_{xx} + g_{tt} = 0 \ , \tag{4.77}$$

is provided by the Cauchy–Riemann equations

$$f_x - g_t = 0 \ , \qquad f_t + g_x = 0 \ . \tag{4.78}$$

Thus, given the solution $g(x,t) = xt$, we can use (4.78) to generate another solution via $f_x = x$ and $f_t = -t$, namely $f(x,t) = \frac{1}{2}(x^2 - t^2)$.

(ii) *Liouville's equation.* The Liouville equation is

$$f_{xt} = e^f \ . \tag{4.79}$$

To this, we append the equation

$$g_{xt} = 0 \ . \tag{4.80}$$

The Bäcklund transformation for this system is given by

$$(f + g)_x = \sqrt{2} \, e^{(f-g)/2} \ , \qquad (f - g)_t = \sqrt{2} \, e^{(f+g)/2} \ . \tag{4.81}$$

(iii) *Sine-Gordon equation.* The sine-Gordon equation is

$$f_{xt} = \sin f \ , \qquad g_{xt} = \sin g \ , \tag{4.82}$$

with Bäcklund transformation

$$\tfrac{1}{2}(f + g)_x = a \sin((f - g)/2) \ , \qquad \tfrac{1}{2}(f - g)_t = \frac{1}{a} \sin((f + g)/2) \ . \tag{4.83}$$

Recall now the Nicolai map of eq. (4.74). This is simply a rewriting of the above Bäcklund transformations, as can be seen from the following change of variables: If we identify

$$\partial_x = \partial_+ \ , \qquad \partial_t = \partial_- \ , \qquad \phi = f + g \ , \qquad \phi^* = f - g \ , \tag{4.84}$$

then the result follows.

4.4.6. The $O(3)$ supersymmetric sigma model

As our final example, let us consider the well-known $O(3)$ supersymmetric sigma model [4.13–4.16], with action

$$\begin{aligned} S = \frac{2}{g^2} \int \mathrm{d}^2\sigma \frac{1}{\rho^2} &\left(\partial_\alpha \phi \partial^\alpha \phi^\dagger - \tfrac{1}{2}\mathrm{i}[\psi^\dagger \gamma^\alpha \partial_\alpha \psi - (\partial_\alpha \psi^\dagger) \gamma^\alpha \psi] \right. \\ &\left. + \frac{\mathrm{i}}{\rho} \psi^\dagger \gamma^\alpha \psi (\phi^\dagger \partial_\alpha \phi - \phi \partial_\alpha \phi^\dagger) + \frac{1}{2\rho^2}(\psi\psi)^\dagger(\psi\psi) \right) \ . \end{aligned} \tag{4.85}$$

Here $\gamma^\alpha = \sigma^\alpha$ ($\alpha = 1,2$) are the Pauli matrices defined in (4.70); $\rho = (1 + \phi^\dagger \phi)$, $\phi = \phi^1 + \mathrm{i}\phi^2$, and

$$\psi = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}$$

is a Dirac spinor.

We wish to identify this model as a topological model with a two-dimensional Kähler target space, and for particular choices of base and target metric. To do this let us recall from (4.43) the general action in the Feynman $\alpha = 1$ gauge,

$$S_q = \int d^2\sigma \sqrt{h} \left(\tfrac{1}{2} h^{\alpha\beta} g_{ij} \partial_\alpha u^i \partial_\beta u^j + \tfrac{1}{2} \epsilon^{\alpha\beta} J_{ij} \partial_\alpha u^i \partial_\beta u^j \right.$$

$$\left. + \overline{C}_{\alpha i} D^\alpha C^i + \tfrac{1}{8} \overline{C}_\alpha{}^m \overline{C}^{\alpha k} R_{mkjr} C^j C^r \right) , \tag{4.86}$$

where we have integrated over the multiplier field, thereby enforcing the gauge constraint $G = 0$. In Kähler form this becomes

$$S_q = 2 \int d^2\sigma \sqrt{h} \, [h^{+-} g_{I\overline{J}} \partial_+ u^I \partial_- u^{\overline{J}} + \tfrac{1}{2} \overline{C}_+{}^I (D_- C^{\overline{J}}) h^{+-} g_{I\overline{J}}$$

$$+ \tfrac{1}{2} \overline{C}_-{}^{\overline{I}} (D_+ C^J) h^{+-} g_{\overline{I}J} + \tfrac{1}{4} h^{+-} \overline{C}_+{}^I \overline{C}_-{}^{\overline{I}} R_{I\overline{I}J\overline{J}} C^J C^{\overline{J}}] . \tag{4.87}$$

We want to consider this action when the target space is the two-sphere. The base line element is

$$ds^2 = 4 \, d\sigma^+ \, d\sigma^- , \qquad h_{+-} = 2 , \qquad h^{+-} = \tfrac{1}{2} . \tag{4.88}$$

The metric on the target space is taken to be the usual S^2 metric written in complex coordinates,

$$ds^2 = \frac{1}{\rho^2} du \, d\bar{u} , \qquad \rho = (1 + \bar{u}u) , \qquad g_{u\bar{u}} = \frac{1}{2\rho^2} , \qquad g^{u\bar{u}} = 2\rho^2 ,$$

$$\Gamma^u_{uu} = -\frac{2\bar{u}}{\rho} , \qquad \Gamma^{\bar{u}}_{\bar{u}\bar{u}} = -\frac{2u}{\rho} , \qquad R_{u\bar{u}u\bar{u}} = -\frac{1}{\rho^4} . \tag{4.89}$$

The action now simplifies to

$$S_q = \frac{1}{2} \int d^2\sigma \left[\frac{1}{\rho^2} \partial_+ u \partial_- \bar{u} + \frac{1}{2\rho^2} \overline{C}_+{}^u \left(\partial_- C^{\bar{u}} - \frac{2u}{\rho} (\partial_- \bar{u}) C^{\bar{u}} \right) \right.$$

$$\left. + \frac{1}{2\rho^2} \overline{C}_-{}^u \left(\partial_+ C^u - \frac{2\bar{u}}{\rho} (\partial_+ u) C^u \right) - \frac{1}{2\rho^4} \overline{C}_+{}^u \overline{C}_-{}^{\bar{u}} C^u C^{\bar{u}} \right] . \tag{4.90}$$

In order to make the identification manifest, a certain amount of reshuffling needs to be performed on the original action (4.85). However, integration by parts yields the result

$$\int d^2\sigma \rho^{-2} \left(-\tfrac{1}{2} i (\psi^\dagger \gamma^\alpha \partial_\alpha \psi - (\partial_\alpha \psi^\dagger) \gamma^\alpha \psi) + \frac{i}{\rho} \psi^\dagger \gamma^\alpha \psi (\phi^\dagger \partial_\alpha \phi - \phi \partial_\alpha \phi^\dagger) \right)$$

$$= \int d^2\sigma \frac{-i}{\rho^2} \left[\psi_2^* \left(\partial_+ \psi_1 - \frac{2}{\rho} (\phi^+ \partial_+ \phi) \psi_1 \right) + \psi_2 \left(\partial_- \psi_1^* - \frac{2}{\rho} (\phi \partial_- \phi^\dagger) \psi_1^* \right) \right] . \tag{4.91}$$

If we make the identifications

$$\phi = u , \quad \phi^\dagger = \bar{u} , \quad \psi_1 = C^u , \quad \psi_1^* = C^{\bar{u}} , \quad \psi_2 = \tfrac{1}{2} i \overline{C}_+{}^u , \quad \psi_2^* = \tfrac{1}{2} i \overline{C}_-{}^{\bar{u}} , \tag{4.92}$$

we see that the two actions are identical, up to the overall scale $2/g^2$ of the action. We can thus recognize the $O(3)$ supersymmetric model as being topological in nature. The spinors of this model are simply the BRST ghosts, the identification in this case being explicitly given by (4.92). We

should note that the supersymmetric action (4.85) differs from the topological action (4.87) only in the addition of the topological term $\int J$, which does not affect any of the above considerations.

Having identified this O(3) model as being topological, we can now use this freedom to express the action in different forms. In the above we have established this connection explicitly in the Feynman gauge. However, as we know, we are free to write the action in different gauges, and as we shall show in section 8.3 the Landau $\alpha = 0$ gauge is particularly useful for studying the renormalization properties of the theory; the result being that the β-function vanishes in this gauge. Indeed, it had already been established that the beta function for this model, in the form (4.85), was exact at one loop order [4.16]. It can now be seen that such a property arises because of the underlying topological nature of the model. However, it should be stressed that the presence of a non-zero β-function, in the first instance, is a gauge artifact. A discussion of this point can be found in section 8.3.

4.4.7. Generalizations

There are various generalizations of the above models which are interesting to consider [4.3], the principal ones being those which incorporate $N = 1$ or $N = \frac{1}{2}$ world sheet supersymmetry; in addition it is possible to incorporate gauge fields into the theory. We shall now briefly describe two of these models.

$N = 1$ *supersymmetric model.* In addition to the bosonic coordinates of Σ, we introduce two fermionic partners $\theta^A, A = 1, 2$, which are spinors of the Lorentz group. We also have the anti-symmetric invariant ϵ_{AB}, as well as the gamma matrices $\gamma^\alpha{}_{AB}$, with the superspace covariant derivative being defined as

$$D_A \doteq \partial/\partial\theta^A - \mathrm{i}(\gamma^\alpha\theta)_A \, \partial/\partial\sigma^\alpha \ . \tag{4.93}$$

One can then construct, in the usual way, superfields $\Phi(\sigma^\alpha, \theta^A)$ and actions which are manifestly invariant with respect to this supersymmetry. However, the aim here is to construct an action which is also invariant under a topological symmetry. This is quite straightforward to achieve, and involves replacing the formulae (4.40)–(4.42) with the corresponding superfield versions.

Specifically, $\hat{u}^i$ and $\hat{C}^i$ are, respectively, commuting and anti-commuting superfields. One replaces $\overline{C}^{\alpha i}$ and $\pi^{\alpha i}$ with spinor superfields $\widehat{\overline{C}}^{Ai}$ and $\hat{\pi}^{Ai}$, which satisfy the self-duality constraint $\widehat{\overline{C}}^{Ai} = \epsilon^A{}_B J^i{}_j \widehat{\overline{C}}^{Bj}$. The quantum action can then be constructed, together with the BRST transformations, by simply making the above superfield replacements. In the Langevin approach, for example, we begin with the classical action

$$S_\mathrm{c} = \int \mathrm{d}^2\sigma \, \mathrm{d}^2\theta \, \sqrt{h} \, \delta_{AB} g_{ij} \hat{K}^{Ai} \hat{K}^{Bj} \ , \tag{4.94}$$

where $\hat{G}^{Ai}$ is the random Gaussian field which is now a self-dual superfield. The rest of the analysis proceeds in direct analogy with the main model discussed in this section. To achieve the $N = \frac{1}{2}$ generalization, one can impose a further chirality condition on the superfields, namely $\widehat{\overline{C}}^{Ai} = \pm\mathrm{i}\epsilon^A{}_B\widehat{\overline{C}}^{Bi}$.

Incorporation of gauge fields. The possibility of including gauge fields in the theory arises from the fact that the σ^α derivatives do not appear anywhere in the BRST transformations (4.41). As a result, we can allow the target space metric and complex structure to depend explicitly on the world sheet coordinates. The generalization which ensues is the following.

Consider a family of almost complex manifolds fibered over the Riemann surface Σ. By this is meant a manifold X fibered over Σ, with an almost complex structure on X that reduces for each fiber to an almost complex structure on the fiber. However, it is now necessary to regard the coordinates $u^i(\sigma^\alpha)$ as a section of the bundle X, and replace the derivatives $\partial_\alpha u^i$ by suitable covariant derivatives; this involves the introduction of a gauge field. Let Diff M be the diffeomorphism group of M, and diff M its associated Lie algebra with generators V^i_a, where the index a runs over the infinite dimensional basis of the algebra, and the index i indicates that each generator corresponds to a vector field on M. The appropriate covariant derivatives are

$$
\begin{aligned}
D_\alpha u^i &= \partial_\alpha u^i + A^a_\alpha V^i_a \ , \\
D_\alpha C^i &= \partial_\alpha C^i + \Gamma^i_{jk} \partial_\alpha u^j C^k + A^a_\alpha \partial_j V^i_a C^j \ ,
\end{aligned}
\tag{4.95}
$$

where A^a_α is the gauge field. Again, the analysis from this point is straightforward.

4.5. Construction of observables

In section 2, we reviewed the construction of a generic topological field theory and found that the interesting observables—at least from the topological point of view—represent BRST cohomology classes. An observable $\mathcal{O}$, in the BRST sense, is one that is invariant under the BRST symmetry,

$$
\{Q, \mathcal{O}\} = 0 \ .
\tag{4.96}
$$

If the infinitesimal change in the operator $\mathcal{O}$ under a perturbation in the base metric h is also BRST exact, so that

$$
\delta_h \mathcal{O} = \{Q, R\}
\tag{4.97}
$$

for some R, then the vacuum expectation value $\langle \mathcal{O} \rangle$ is independent of the base metric. Furthermore, in the case at hand, one can prove that the partition function and its observables are independent of the metric and complex structure associated with both the base and target manifolds; one can establish this by following the same line of argument as presented in section 2. The reader may wish to review that discussion at this time. Since we are assuming that the vacuum is Q-invariant, we can add anything of the form $\{Q, \mathcal{O}'\}$ to $\mathcal{O}$ without affecting this matrix element, and we are thus led to the BRST cohomology classes of operators.

In the case of topological sigma models, an interesting class of observables has been described by Witten [4.3]. In this construction, we first associate an operator $\mathcal{O}^{(0)}_A$ to each p-form $A = A_{i_1 \cdots i_p} \, \mathrm{d} u^{i_1} \wedge \cdots \wedge \mathrm{d} u^{i_p}$ on the target space M, given by

$$
\mathcal{O}^{(0)}_A = A_{i_1 \cdots i_p} C^{i_1} \cdots C^{i_p} \ ,
\tag{4.98}
$$

where C^i is the ghost field we encountered in section 4.4. As a field on Σ, $C^i(u(\sigma))$ is a section of the pull-back $u^*(\mathrm{T}^*(M))$ of the cotangent bundle of M, where $u : \Sigma \to M$ is any smooth map (we also use u^i to denote the coordinates on M). If we define the evaluation map $f_\sigma : \mathrm{Map}(\Sigma, M) \to M$ by $f_\sigma(u) = u(\sigma)$, then $C^i = f_\sigma^*(\mathrm{d} u^i)$. Under a Q transformation, we see that

$$
\{Q, \mathcal{O}^{(0)}_A\} = -\partial_{i_0} A_{i_1 \cdots i_p} C^{i_0} \cdots C^{i_p} = -\mathcal{O}^{(0)}_{dA} \ ,
\tag{4.99}
$$

since these ghosts are BRST invariant. Hence, $\mathcal{O}^{(0)}_A$ is BRST invariant if and only if A is a closed p-form. Similarly, if A is an exact p-form, then the corresponding operator is Q-exact. Hence, the

BRST cohomology classes of these operators are in one-one correspondence with the de Rham cohomology classes on M. The reason for assigning the peculiar superscript to the operator $\mathcal{O}^{(0)}$ will become clear at the end of this construction.

Notice that operators of the form $\mathcal{O}_A^{(0)}$ can be used as building blocks for constructing new observables. If we consider a set of closed forms $A_1, \ldots, A_k$, then the product of the associated operators $\mathcal{O}_{A_1}^{(0)} \cdots \mathcal{O}_{A_k}^{(0)}$ is clearly Q-invariant as well.

Now, when we consider the vacuum expectation values of operators which are polynomials in the fields, there is some implicit dependence on the points where the operators are located. In the case at hand, however, the operator $\mathcal{O}_A^{(0)}(\sigma)$ at the point σ has a vacuum expectation value which is a topological invariant, and so the VEV cannot depend on the chosen point. To see this explicitly, we consider all fields defined over Σ, and differentiate the operator with respect to some local coordinates σ^α,

$$
\frac{\partial}{\partial \sigma^\alpha} A_{i_1 \cdots i_p} C^{i_1} \cdots C^{i_p} = (\partial_{i_0} A_{i_1 \cdots i_p}) \frac{\partial u^{i_0}}{\partial \sigma^\alpha} C^{i_1} \cdots C^{i_p}
$$

$$
+ p \, A_{i_1 \cdots i_p} (\partial_{i_0} C^{i_1}) \frac{\partial u^{i_0}}{\partial \sigma^\alpha} C^{i_2} \cdots C^{i_p} \ . \tag{4.100}
$$

In terms of exterior derivatives, this takes the form,

$$
d\mathcal{O}_A^{(0)} = \partial_{i_0} A_{i_1 \cdots i_p} \, du^{i_0} \, C^{i_1} \cdots C^{i_p} + p A_{i_1 \cdots i_p} \, dC^{i_1} \, C^{i_2} \cdots C^{i_p}
$$

$$
= \{Q, \mathcal{O}_A^{(1)}\} \ , \tag{4.101}
$$

where $\mathcal{O}_A^{(1)} = -p A_{i_1 \cdots i_p} \, du^{i_1} \, C^{i_2} \cdots C^{i_p}$, and we have used the fact that A is a closed p-form. If we let γ represent any path between two arbitrary points P and P', then this expression has the integral form,

$$
\mathcal{O}_A^{(0)}(P) - \mathcal{O}_A^{(0)}(P') = \{Q, \int_\gamma \mathcal{O}_A^{(1)}\} \ , \tag{4.102}
$$

and we see that the VEV of $\mathcal{O}_A^{(0)}$ is point independent by the BRST invariance of the vacuum. The same remark applies to any product of operators of the form we are considering.

To continue our construction, consider a one-dimensional homology cycle γ $(\partial \gamma = 0)$, and define

$$
W_A^{(1)}(\gamma) = \int_\gamma \mathcal{O}_A^{(1)} \ . \tag{4.103}
$$

This new operator $W_A^{(1)}(\gamma)$ is then BRST invariant by inspection,

$$
\{Q, W_A^{(1)}(\gamma)\} = \int_\gamma \{Q, \mathcal{O}_A^{(1)}\} = \int_\gamma d\mathcal{O}_A^{(0)} = 0 \ . \tag{4.104}
$$

Moreover, if γ happens to be the boundary of a two-dimensional surface $(\gamma = \partial \beta)$, so that γ is trivial in homology, then this new operator is likewise trivial in Q cohomology,

$$
W_A^{(1)}(\gamma) = \int_\gamma \mathcal{O}_A^{(1)} = \int_\beta d\mathcal{O}_A^{(1)} = \{Q, \int_\beta \mathcal{O}_A^{(2)}\} \ , \tag{4.105}
$$

where $\mathcal{O}_A^{(2)} = -\frac{1}{2}p(p-1)A_{i_1\cdots i_p}\,du^{i_1} \wedge du^{i_2}\,C^{i_3}\cdots C^{i_p}$.

As before, let us now associate to each homology two-cycle β $(\partial\beta = 0)$, another BRST invariant operator $W_A^{(2)}$ defined by

$$W_A^{(2)}(\beta) = \int_\beta \mathcal{O}_A^{(2)} \, . \tag{4.106}$$

The BRST invariance follows trivially as in (4.105).

In summary, we have produced three operators $\mathcal{O}_A^{(0)}$, $\mathcal{O}_A^{(1)}$, and $\mathcal{O}_A^{(2)}$ from any given closed form A, which satisfy the relations

$$0 = \{Q, \mathcal{O}_A^{(0)}\} \, , \qquad d\mathcal{O}_A^{(0)} = \{Q, \mathcal{O}_A^{(1)}\} \, , \qquad d\mathcal{O}_A^{(1)} = \{Q, \mathcal{O}_A^{(2)}\} \, , \qquad d\mathcal{O}_A^{(2)} = 0 \, . \tag{4.107}$$

The BRST observables are then given by arbitrary products of the integrated operators $W_A^{(i)}(\gamma) = \int_\gamma \mathcal{O}_A^{(i)}$, where γ is any i-cycle in homology.

4.5.1. Moduli space and the ghost number anomaly

The quantum field theory of topological sigma models that we have been discussing deals fundamentally with maps from a Riemann surface Σ into an almost complex target space M. In the functional integral, we integrate over all maps $\Sigma \to M$ in a fixed homotopy class. The crucial feature of these models, as we have repeatedly emphasized, lies in the metric independence of certain correlation functions, including the partition function. If we replace any chosen metric g_{ij} on the target space with tg_{ij}, then a quick check of the quantum action shows that $S_q[tg_{ij}] = tS_q[g_{ij}]$; hence any of these correlation functions

$$\langle \mathcal{O} \rangle = \int_\Phi e^{-tS_q} \mathcal{O} \tag{4.108}$$

are independent of t, and we can evaluate those functional integrals in the large-t limit where the contributions are dominated by those configurations in which the action S_q vanishes (see section 2.1). The classical action is minimized by the instanton configurations which satisfy

$$0 = \partial_\alpha u^i + \epsilon_{\alpha\beta}J^i{}_j\partial^\beta u^j \, , \tag{4.109}$$

and it is these field configurations that we expand about in a semi-classical approximation. We will have more to say about the space of these solutions shortly.

There are, in addition, a variety of ghost fields in the quantum action, and whether or not they possess zero modes is an important issue related to a ghost number anomaly. All of the points we wish to make here are most transparent in a one-loop background field analysis (which is exact anyway). To compute, say, the partition function at this order, we expand about a background instanton field and consider the part of the quantum action which is quadratic in the quantum fields. We expand the quantum fields into eigenfunctions of the operators that appear there, and do a functional integral over the modes. If there are fermion zero modes, then those modes might not enter in the action, and the fermionic integrals ($\int d\chi = 0$) over those modes will cause a correlation function to vanish unless that function has the right fermion content; the zero modes must be absorbed. In our case, a look at the quantum action indicates that we should concern

ourselves with C^i and $\overline{C}_{\alpha i}$ zero modes. A C^i zero mode is clearly in the kernel of the operator $\overline{D} : \Gamma(u^*T^*(M)) \to P_+ [\Gamma(u^*T^*(M)) \otimes \Omega^1(\Sigma)]$, where

$$\overline{D}^i_{\alpha j} = D_\alpha \delta^i{}_j + \epsilon_{\alpha\beta} J^i_j D^\beta + \epsilon_{\alpha\beta} (D_j J^i{}_k) \partial^\beta u^k \ , \tag{4.110}$$

and a $\overline{C}_{\alpha i}$ zero mode is a zero eigenfunction of its adjoint

$$\overline{D}^* : P_+ [\Gamma(u^*T(M)) \otimes \Omega^1(\Sigma)] \to \Gamma(u^*T^*(M)) .$$

Recalling that C^i and $\overline{C}_{\alpha i}$ have ghost number $+1$ and -1, respectively, it is therefore apparent that the VEV of any observable will vanish unless that observable has a ghost number equal to the number of $\overline{D}$ zero modes minus the number of $\overline{D}^*$ zero modes. This difference is called the index of the operator,

$$\text{Index}[\overline{D}] = \dim \text{Ker}(\overline{D}) - \dim \text{Ker}(\overline{D}^*) \ , \tag{4.111}$$

and appears in many applications.

Let us now return to our discussion of the moduli space of instantons. That is, we are considering the space of maps $\Sigma \to M$ in a specified homotopy class, which satisfy eq. (4.109). A natural question that arises here is whether these instantons are isolated or form a continuous family. We can examine the constraint that arises in this latter case, by considering an instanton u, and another neighboring solution $u + \hat{u}$, where $\hat{u}$ is some infinitesimal deformation. Looking to first order in $\hat{u}$, or equivalently the tangent vector ξ^i, as in section 4.4.3, we see that ξ^i must be a zero mode of the operator $\overline{D}$. This is no coincidence, and we can interpret the ghost fields C^i as cotangent vectors to instanton moduli space. Clearly then, the dimension of moduli space is at most given by the dimension of $\text{Ker}(\overline{D})$. Although we might naively expect that the number of these zero modes will give the actual dimension of the instanton space, there is an obstruction of a global nature, and this is related to the zero modes of $\overline{D}^*$. The problem is that not all of these infinitesimal solutions can be integrated. One can prove that the index of $\overline{D}$ actually gives the virtual dimension of the moduli space (see section 4.1.4) since the dimension may not be well defined at all points, though we will not be able to show that here.

4.5.2. Observables and intersection theory

It is possible to interpret some of the observables that we have described in terms of intersection theory applied to the moduli space of instantons. In particular, one can show that all correlation functions of the form

$$\langle \mathcal{O}^{(0)}_{A_1} \cdots \mathcal{O}^{(0)}_{A_s} \rangle \tag{4.112}$$

are intersection numbers of certain submanifolds of moduli space. We do not assume the reader is fluent in intersection theory so we will first review the key ideas and theorems that are relevant to this application. We begin by discussing Poincaré duality and the relationship between cohomology and homology. Concepts associated to transversal intersection are then reviewed, and finally related to the observables. In this section, all manifolds (and submanifolds) will be taken to be compact and oriented without special mention.

We have already seen that de Rham cohomology and Hodge theory are important ideas which underlie even the simplest topological field theories. The de Rham cohomology groups $H^i(M)$ of a manifold M (see section 3.8.1) was defined as the quotient

$$H^i(M) = Z^i/B^i \ , \tag{4.113}$$

where Z^i is the space of d-closed i-forms on M, and B^i denotes the space of i-forms which are d-exact, $B^i = d\Omega^{i-1}(M)$. One version of Poincaré duality, which can be stated entirely in terms of cohomology, is simply that the pairing

$$H^i(M) \otimes H^{n-i}(M) \to \mathbb{R} \ , \quad ([\phi], [\psi]) \to \int_M \phi \wedge \psi \ , \tag{4.114}$$

is nondegenerate. It is trivial to check that this map is actually well defined on cohomology. The non-trivial content is in the assertion that the inner product is non-degenerate; a good source for a proof is ref. [4.17].

There is another formulation of Poincaré duality which is expressed as a relationship between de Rham cohomology (as we have defined in terms of closed differential forms), and homology (which can be defined in terms of subspaces of M [4.18]). Our aims here are modest, and it will not be necessary to launch into a complete discussion of this subject. The restatement of this duality principle can, however, be appreciated with the machinery at hand. In one direction, the assertion of the theorem is that we can associate to each boundaryless submanifold N of codimension k, a cohomology class $[\phi] \in H^k(M)$, such that

$$\int_M \phi \wedge \psi = \int_N \psi \ , \tag{4.115}$$

for all $[\psi] \in H^{n-k}(M)$. By ψ on the right hand side of this equation, we mean the pull-back $i^*\psi$ under the inclusion $i : N \to M$. Conversely, to each closed k-form ϕ on M, we can associate an $(n-k)$-cycle N (it is in general a chain of subspaces), unique up to homology, such that the previous relation is satisfied. One can also show [4.17] that the Poincaré dual to N can be chosen in such a way that its support is localized within any given open neighborhood of N in M (essentially delta function support on N).

Let us leave this discussion of duality for the moment, and move on to intersections of submanifolds. For simplicity, we will first consider the intersection of two submanifolds M_1 and M_2 contained in M. We will say that these two submanifolds have *transversal intersection* if the tangent spaces satisfy

$$T_x(M_1) + T_x(M_2) = T_x(M) \tag{4.116}$$

for all $x \in M_1 \cap M_2$. It is a theorem that a submanifold of codimension k can be locally "cut-out" by k smooth functions, i.e., the submanifold is locally specified by the zeros of this set of functions. It is a worthwhile exercise to convince oneself that the definition of transversal intersection is equivalent to the statement that the functions which cut out M_1 are independent of those which cut out M_2 [4.19]; in symbols we have

$$\mathrm{codim}(M_1 \cap M_2) = \mathrm{codim}(M_1) + \mathrm{codim}(M_2) \ . \tag{4.117}$$

More generally, we can consider the transversal intersection of any collection of submanifolds, and we will say that the intersection $M_1 \cap \cdots \cap M_s$ of s submanifolds is transversal if the intersection of every pair of them is transversal. It then follows trivially by the previous argument, that the codimensions must satisfy

$$\mathrm{codim}(M_1 \cap \cdots \cap M_s) = \sum_{i=1}^{s} \mathrm{codim}(M_i) \ . \tag{4.118}$$

A special case which will be important for us occurs when the intersection of submanifolds is a collection of points, i.e., when the codimension of the intersection is equal to the dimension of M. Since these points are isolated, the compactness of M guarantees that they are finite in number.

Now, we would like to assign "intersection numbers" to those points in $M_1 \cap M_2$ when the dimension of this intersection is zero. Let x be one of those points, and consider the ordered basis for $T_x(M_1)$ and $T_x(M_2)$ given, respectively, by $(v_1, \ldots, v_k)$ and $(w_1, \ldots, w_{n-k})$, which define the orientations of those submanifolds. Now, $(v_1, \ldots, v_k, w_1, \ldots, w_{n-k})$ is a basis for the tangent space to M at x. We define the *intersection number* of M_1 and M_2 at this point to be $+1$ if this ordered basis gives the orientation of $T_x(M)$ and -1 otherwise; we write $\#_x(M_1 \cap M_2) = \pm 1$. Notice that the order of M_1 and M_2 can be important. The intersection number of M_1 and M_2 is now simply defined by summing these numbers over all points in the intersection,

$$\#(M_1 \cap M_2) = \sum_{x \in M_1 \cap M_2} \#_x(M_1 \cap M_2) \ . \tag{4.119}$$

The idea of assigning intersection numbers to two submanifolds extends naturally to the general case where $\sum_{i=1}^{s} \mathrm{codim}(M_i) = \dim(M)$, and we write the sum over all intersection numbers as $\#(M_1 \cap \cdots \cap M_s)$.

We would now like to describe in what sense correlation functions of the form $\langle \mathcal{O}_{A_1}^{(0)} \cdots \mathcal{O}_{A_s}^{(0)} \rangle$ determine intersection numbers in the moduli space $\mathcal{M}$ of instantons [4.12]. By definition, this moduli space is the set of maps from Σ to M which satisfy (4.109). For convenience, let us begin by choosing the forms A_i which represent de Rham cohomology classes on M, together with their Poincaré duals M_i, such that the forms have essentially delta function support on their respective submanifolds. Since each of the operators in the correlation function depends on some fixed point σ_i, it is meaningful to define the submanifolds $L_i \equiv \{u \in \mathcal{M} \mid u(\sigma_i) \in M_i\} \subset \mathcal{M}$. Now, the correlation function represents a functional integral over the space of maps $\mathrm{Map}(\Sigma, M)$, and we have argued that this integral only receives contributions from the instanton configurations. Since the operators $A_i(u(\sigma_i))$ vanish unless $u \in L_i$ by our choice of the Poincaré duals, we see that the only contribution to the functional integral can be from those maps which lie in the intersection $L_1 \cap \cdots \cap L_s$. By ghost number considerations, this correlation function must vanish unless $\dim(\mathcal{M}) = \sum_{i=1}^{s} \mathrm{codim}(L_i)$, meaning that this intersection is simply a finite number of points. In the sigma model case that we are considering here, the intersection number assigned to each point in the intersection is always $+1$, since the ratio of determinants that arises is always $+1$ due to the nature of the complex geometry. This simplification will not hold when we interpret correlation functions in Donaldson theory, in the next section.

Further reading

Our review of topological sigma models has certainly not been exhaustive. We have not dealt with the equivariant/superspace approaches to this theory, and we refer the reader to the original papers [4.20, 4.21]. Other work that we have not discussed may be found in refs. [4.22–4.28].

5. Topological gauge theories of Witten type

In this section we are going to give a detailed account of perhaps the richest branch of topological field theories, namely topological gauge theories. In section 5.1 we present the necessary mathematical background, including a description of a number of more advanced results. This

includes the instanton deformation complex, the relation between gauge theory and the topology of four-manifolds, and the construction of the Donaldson polynomials.

This is indispensable for an appreciation of the subsequent sections, and in particular section 5.2. There we describe at length Witten's original topological field theory [5.1] (which we shall refer to as Donaldson theory) as the prototype of a topological gauge theory of Witten type, and its most important representative. In that section we follow closely Witten's paper, and we have chosen this historical route because it makes it clear that one can *derive* the most important properties of Donaldson theory by very elementary physical manipulations. However, in order to *understand* why the action has these remarkable properties, a deeper understanding of the geometry of topological gauge theories, and the principles behind the construction of this action, is required. Moreover, to *appreciate* the significance of these properties and their consequences, it is necessary to know something about the mathematics underlying this theory. Our treatment in this section will be guided by the attempt to illuminate these different facets and levels of Donaldson theory.

In order to gain a better understanding of this theory we then explore the geometry underlying topological gauge theories in general (section 5.3), and—based on that—clarify a number of issues which had arisen in section 5.2. The main results of that section will be the completion of the proof that the observables of Donaldson theory constructed in section 5.2 are the Donaldson polynomials, as well as the emergence of a geometrical framework for constructing topological gauge theories associated with arbitrary moduli spaces of connections. To show how that works in practice we explicitly construct the quantum actions for topological gauge theories based on the moduli spaces of flat and Yang–Mills connections in any dimension in section 5.4. There we also discuss moduli spaces of flat connections and their deformation complex in general, as well as the Casson invariant and its relation to the partition function of a three-dimensional gauge theory. A more detailed summary of the contents can be found at the beginning of each section.

5.1. Mathematical background

5.1.1. Introduction

While the mathematics underlying the theories we have discussed so far (quantum mechanics, sigma models) is, roughly speaking, that of spaces of maps (thus falling into the realms of differential topology and algebraic geometry), gauge theories are deeply rooted in the differential geometry of fiber bundles and spaces of connections. We feel that we can safely assume a basic understanding of the dictionary (P is a principal G-bundle)

gauge potential A_μ^a	$\leftrightarrow$	connection A on P
field strength $F_{\mu\nu}^a$	$\leftrightarrow$	curvature F_A of A
gauge group G	$\leftrightarrow$	structure group of P
gauge transformations	$\leftrightarrow$	vertical automorphisms of P

between physical and mathematical terminology. We will nevertheless give a short exposition of the geometry of principal bundles in section 5.1.2, mainly to establish our notation and terminology. The reader desiring a more detailed treatment of these matters and their relation to gauge theories may wish to consult refs. [5.2, 5.3]. In order to give a flavor of the more advanced mathematical developments which were the original motivation behind the construction of topological field theories (we are thinking here in particular of Donaldson's [5.4, 5.5] and Floer's [5.6, 5.7] work) we will then have to briefly recall the most important features of (moduli) spaces of connections (section 5.1.3), in particular those related to the existence of reducible connections. Turning to instantons we will need a rough understanding of the so-called deformation complex of instanton

moduli space (section 5.1.4), since that is directly related to the appearance of fermionic zero modes and the construction of observables in Donaldson theory (section 5.2.7). A good deal is known (in the case of SU(2) at least) about the structure of the singularities of instanton moduli space, but since the field theoretic point of view has so far not advanced our understanding of these singularities (and since the validity of formal field theoretic manipulations becomes doubtful in the presence of these singularities) we will explain that part of the theory only to the extent that we know which conditions are sufficient to ensure smoothness of the moduli space. The standard references for these results are ref. [5.8] and the monograph [5.9]. Our presentation has been influenced by the lectures of Freed [5.10].

After having discussed spaces of connections in general and instanton moduli space in particular, we shall then attempt to explain what the latter has to do with the topology of four-manifolds (section 5.1.5). We begin with a brief overview of the subject, recalling why—from the point of view of smoothing theory—four dimensions are special for topology, and summarizing some of the most important "classical" and new results. Following refs. [5.4, 5.9] we then outline the proof of Donaldson's theorem establishing the existence of a large number of non-smoothable topological four-manifolds. The remainder of the section is devoted to Donaldson's recent work on polynomial invariants [5.5]. We explain Donaldson's μ-map, which expresses the cohomology of the moduli space of connections in terms of the homology of the underlying four-manifold and show how this map —which has a perfect counterpart in Witten's construction of observables in Donaldson theory—can be used to define polynomial rational cohomology classes which—when evaluated on the moduli space—lead to the Donaldson invariants.

In the following section we explain Floer's idea of applying an infinite dimensional version of Morse theory to the Chern–Simons functional on the space of connections. We will be guided by the beautiful and influential paper of Atiyah [5.11], whose Hamiltonian version of Donaldson theory we describe at the end of section 5.1.6.

Our presentation in sections 5.1.5 and 5.1.6 is necessarily incomplete and cannot possibly do justice to the importance and depth of the mathematical results. We have nevertheless attempted to sketch at least the main ideas, in the hope that this may make the original literature somewhat more accessible.

In order to make mathematically precise statements we will have to be fairly specific in section 5.1 about the topological conditions under which the quoted results hold. As a rule, however, these conditions (like the simple connectivity of four-manifolds in section 5.1.4 or the restriction to homology three-spheres in section 5.1.6) will not enter directly into our subsequent discussion of topological gauge theories. The reader who feels uneasy about this is invited to add these conditions explicitly in the appropriate sections. Throughout we have also avoided to work with the Sobolev completions of the infinite dimensional spaces and groups appearing, and refer to the literature [5.9] for the confirmation of the fact that this can always be done in a satisfactory and essentially routine manner.

5.1.2. Geometry of gauge theories

The arena for gauge theories in general and topological field theories in particular is the space $\mathcal{A} = \mathcal{A}_P$ of connections on a principal G-bundle $P \xrightarrow{\pi} M$, and the associated quotient space $\mathcal{C} = \mathcal{A}/\mathcal{G}$ of gauge equivalence classes of connections, as well as various subspaces thereof. Let us start by making precise what we mean by $\mathcal{A}$ and $\mathcal{G}$.

Given a principal bundle there is a natural notion of verticality for tangent vectors to P: a vector $X_p \in T_pP$ is vertical if it is in the kernel of the projection $\pi_* : T_pP \to T_{\pi(p)}M$. However, in

order to connect neighboring fibers (i.e., to have a notion of parallel transport) one also needs to know what "horizontal" means. This is a concept not canonically associated with P and requires additional structure for its definition—a connection. From this point of view a connection is then a decomposition of the tangent space at every point $p \in P$ into a vertical part $V_p = \text{Ker}(\pi_*)_p$ and a horizontal part H_p,

$$T_p P = V_p \oplus H_p \ ,$$

which should moreover be compatible with the right action R_g of G on P in the sense that the family of subspaces $\{H_{pg}, g \in G\}$ is G-invariant, i.e.,

$$R_{g*} H_p = H_{pg} \ .$$

Such a G-invariant decomposition can, for instance, be performed with the help of a G-invariant metric on P by declaring H_p to be the orthogonal complement to V_p with respect to that metric. For an example of this cf. section 5.3.1.

While this point of view on connections is extremely useful for certain purposes, there is a dual description in terms of differential forms on P which is more commonly used and which makes obvious the relation to the formalism of gauge theories. One equivalently defines a connection to be a one-form A on P with values in the Lie algebra $\mathbf{g}$ of G with the properties that

$$A(\xi_P) = \xi \ , \tag{5.1}$$

$$A(R_{g*} X) = \text{ad}(g^{-1}) A(X) \ , \tag{5.2}$$

where ξ_P is the (vertical) fundamental vector field on P generating the right action of $\exp \xi \in G$, and X is an arbitrary vector field on P. Horizontal vectors are now defined to be those annihilated by A. By the second condition (5.2) above, this definition is indeed G-invariant, as required.

Note that the difference between any two connections A and A' can be identified with a Lie algebra valued form on the base manifold M, since $A' - A$ is horizontal and G-equivariant. The space $\mathcal{A}_P$ of all connections is thus an affine space modeled on $\Omega^1(M, \mathbf{g})$ [more precisely, any two connections differ by a one-form taking values in the bundle $\text{ad} P$ of Lie algebras, which we will define below; let us agree to denote the space of such forms by $\Omega^1(M, \mathbf{g})$]. Two connections (or families of horizontal subspaces) should, however, be regarded as equivalent if they are related by a diffeomorphism $\varphi : P \to P$ (via pull-back), which is compatible with the structure of P in the sense that it preserves the base points of the fibers and commutes with the right action of G, i.e.,

$$\pi(\varphi(p)) = \pi(p) \quad \text{and} \quad \varphi(pg) = \varphi(p)g \ . \tag{5.3}$$

The set of all such φ's forms a group called the (vertical) automorphism group, more commonly known as the group $\mathcal{G}$ of gauge transformations.

Since this is not the way physicists tend to think about gauge transformations, let us pause to explain the relation to the more common point of view in which gauge transformations are—at least locally—regarded as maps from the base manifold M to the structure group G. Since φ preserves base points of fibers we can write it as $\varphi(p) = p\tilde{\varphi}(p)$, where $\tilde{\varphi}$ is a map from P to G. The compatibility condition with the right action of G then requires $\tilde{\varphi}(pg) = g^{-1}\tilde{\varphi}(p)g$. Thus we can alternatively think of gauge transformations as Ad-equivariant functions on P. In turn, every such function defines a section $\hat{\varphi}$ of the group bundle $\text{Ad} P = P \times_{\text{Ad}} G$ associated to P via the adjoint action of G on itself, given by

$$\hat{\varphi}(m) = [(p, \tilde{\varphi}(p))] \ , \tag{5.4}$$

where $m = \pi(p)$. $[\ ,\ ]$ denotes the equivalence class in $P \times G$ under the projection $P \times G \to P \times_{\mathrm{Ad}} G$, and equivariance of $\tilde{\varphi}$ ensures that the right hand side of (5.4) does not depend on the choice of $p \in \pi^{-1}(m)$. Thus locally, a gauge transformation can now indeed be regarded as a map from M to G. From this point of view it is also almost evident that the Lie algebra of $\mathcal{G}$ (locally given by maps from M to $\mathbf{g}$) is the space of sections of the bundle of Lie algebras $\mathrm{ad}\, P = P \times_{\mathrm{ad}} \mathbf{g}$. When talking of Lie algebra valued functions or forms on M, one usually means (recall our discussion above of the affine structure of $\mathcal{A}$) ad-equivariant horizontal forms on P, or, equivalently, forms taking values in $\mathrm{ad}\, P$, and in keeping with that terminology we will in the following refer to these as elements of $\Omega^*(M, \mathbf{g})$.

In whatever way we choose to look at gauge transformations, they act on connections via pull-back,

$$\varphi : A \to \varphi^* A = \varphi^{-1} A \varphi + \varphi^{-1} \mathrm{d}\varphi \ . \tag{5.5}$$

Writing $\varphi = \exp t\xi$ with $\xi \in \Omega^0(M, \mathbf{g})$ one derives the infinitesimal version of (5.5) to be the familiar

$$A \to A + d_A \xi \ .$$

Here $d_A \xi : \Omega^0(M, \mathbf{g}) \to \Omega^1(M, \mathbf{g})$ is the *covariant exterior derivative* defined by

$$d_A \xi = \mathrm{d}\xi + [A, \xi] \ . \tag{5.6}$$

One easily checks that with this definition $d_A \xi$ is indeed horizontal and ad-equivariant if ξ is, and that d_A extends to an operator $\Omega^k(M, \mathbf{g}) \to \Omega^{k+1}(M, \mathbf{g})$ on all of $\Omega^*(M, \mathbf{g})$.

In contrast to the ordinary exterior derivative on M or P, d_A no longer squares to zero, and the failure to do so is measured by multiplication by an element F_A of $\Omega^2(M, \mathbf{g})$. Indeed for any $\xi \in \Omega^k(M, \mathbf{g})$ one finds

$$(d_A)^2 \xi = [F_A, \xi] \ ,$$

where

$$F_A := \mathrm{d}A + \tfrac{1}{2}[A, A] \tag{5.7}$$

is the *curvature* of the connection A. It transforms homogeneously under gauge transformations,

$$F_{\varphi^* A} = \varphi^{-1} F_A \varphi \ ,$$

and can locally be regarded as a $\mathbf{g}$-valued two-form on M. Note that F_A satisfies the Bianchi identity

$$d_A F_A = 0 \ . \tag{5.8}$$

The covariant derivative and the curvature allow us to write down gauge invariant equations for $A \in \mathcal{A}$ like the Yang–Mills equation

$$d_A * F_A = 0 \ . \tag{5.9}$$

Here $*$ is the Hodge duality operator with respect to some metric on M, extended to $\mathbf{g}$-valued forms. Of interest to us in the following sections (in particular sections 5.3, 5.4 and 6) will be the condition $F_A = 0$ defining the moduli space of flat connections. In section 5.4.3 we have collected some of the mathematical results we need concerning (moduli) spaces of flat connections.

Special to (Euclidean) four dimensions is the *instanton* equation

$$*F_A = \pm F_A$$

(this makes sense in four dimensions, as $*$ is then a map from the space of two-forms to itself and satisfies $*^2 = 1$ there if a Euclidean metric is used). Among other things the interest in these equations lies in the fact that by virtue of the Bianchi identity (5.8) solutions to the (first order) instanton equations are automatically solutions of the (second order) Yang–Mills equations (5.9), and that, moreover, these solutions are the absolute minima of the Yang–Mills action functional. We will discuss these equations and the rich topological structure associated with the instanton moduli space in sections 5.1.4 and 5.1.5.

What makes a connection an interesting additional structure on P is the fact that not all connections are gauge equivalent. The moduli space $\mathcal{C} := \mathcal{A}/\mathcal{G}$ of gauge equivalence classes of connections is on the contrary infinite dimensional and (as opposed to the contractible space $\mathcal{A}$) topologically quite complicated. By gauge invariance the equations mentioned above determine (moduli) subspaces of $\mathcal{C}$, which are, however, (if M is compact) finite dimensional, due to ellipticity of the corresponding operators (this will become clear from the deformation complex approach we will discuss in sections 5.1.4 and 5.4.3 for the case of instantons and flat connections, respectively). We will now first take a closer look at $\mathcal{C}$ itself.

5.1.3. Spaces of connections

In this section we will study the action

$$A \to \varphi^* A = \varphi^{-1} A \varphi + \varphi^{-1} \, \mathrm{d}\varphi \tag{5.10}$$

of $\mathcal{G}$ on $\mathcal{A}$, and in particular the solutions to the equation

$$\varphi^{-1} A \varphi + \varphi^{-1} \, \mathrm{d}\varphi = A \tag{5.11}$$

defining the isotropy subgroup I_A of $\mathcal{G}$. Infinitesimally (5.11) reads $d_A \xi = 0$. The center $Z(G)$ of G is contained in I_A for every A, since the (global) gauge transformation

$$\varphi_z : p \mapsto \varphi_z(p) = pz , \quad z \in Z(G) ,$$

evidently satisfies (5.11) (note that, in general, the right action of G on P is *not* a gauge transformation). Connections A with $I_A = Z(G)$ are called irreducible, and this is the generic case. The quotient of the space $\mathcal{A}^*$ of irreducible connections by the group $\mathcal{G}/Z(G)$ is then a smooth manifold $\mathcal{C}^*$, the moduli space of gauge equivalence classes of irreducible connections. Of interest to us later will be the fact that at an irreducible connection the Green's function $G_A = (d_A * d_A)^{-1}$ of the scalar Laplacian $\Delta_A = d_A * d_A$ exists, since there are no non-trivial solutions to the equation $d_A \xi = 0$.

To obtain information about the non-generic points of $\mathcal{A}$ we proceed as follows (cf. ref. [5.9]). In the subgroup $\mathcal{G}'$ of $\mathcal{G}$ consisting of gauge transformations which are the identity $\varphi(p) = p$ for some (and thus all) points on the fiber $\pi^{-1}(m)$ above an arbitrary but fixed base point $m \in M$, the only solution to (5.11) is the identity $\varphi = \mathrm{id}$, since (5.11) is a first order differential equation for φ (here we have tacitly assumed that M is connected and we shall continue to do so in the following). Thus $\mathcal{G}'$ acts freely on $\mathcal{A}$ and $\mathcal{C}' = \mathcal{A}/\mathcal{G}'$ turns out to be a smooth manifold. Since $\mathcal{G}$ is an extension of $\mathcal{G}'$ by G the above remark allows us to conclude that I_A is isomorphic to a subgroup of G. Equation (5.11) shows that elements of I_A are precisely those commuting with parallel transport by A, whence we can alternatively view I_A as the centralizer of the holonomy group in G. The last piece of information we can obtain in this generality is the fact that the isotropy groups of gauge equivalent connections are conjugate to each other,

$$I_{\psi^* A} = \psi^{-1} I_A \psi .$$

This follows from the fact that if $\varphi^* A = A$, then $(\psi^{-1}\varphi\psi)^*\psi^* A = \psi^* A$.

Let us now specialize to $G = \mathrm{SU}(2)$. This is a tractable example since the only possibilities for I_A are now [apart from $Z(\mathrm{SU}(2)) = \mathbb{Z}_2$] $\mathrm{U}(1)$ and $\mathrm{SU}(2)$. The latter occurs for flat connections and since we are —in the following sections on instantons—mainly interested in non-trivial bundles, connections with $I_A = \mathrm{SU}(2)$ will not appear. In the sections 5.3, 5.4, and 6 dealing with topological field theories of flat connections, however, the problems with reducibility will haunt us in various guises. It fortunately turns out that, although from the point of view of C, flat connections can be quite singular objects, the moduli spaces of flat connections themselves are nevertheless reasonably nice spaces. This leaves us with the case $I_A = \mathrm{U}(1)$. In that case the connection A can be thought of as coming from a $\mathrm{U}(1)$ bundle on M, and it is fairly easy to see (using the fact that there is a covariantly constant section of $\mathrm{ad}\,P$) that this defines a *splitting* of the covariant derivative d_A and the complex two-plane bundle

$$E = P \times_{\mathrm{SU}(2)} \mathbb{C}^2 \ ,$$

into a sum of line bundles with connections, i.e., $E = E_1 \oplus E_2$, $d_A = d_1 + d_2$. We will be more precise about this splitting and how to count the number of possible splittings in the next section. The inclusion of the space of gauge equivalence classes of these $\mathrm{U}(1)$ connections into C leads to the singular nature of C at these points. The tangent space to C at an irreducible connection A (near which C is smooth) is simply the infinite dimensional Hilbert space

$$T_A C = \Omega^1(M,\mathbf{g})/d_A\Omega^0(M,\mathbf{g}) \ .$$

Splitting the Lie algebra $\mathbf{g}$ of $\mathrm{SU}(2)$ into the $\mathrm{U}(1)$ part $\mathbf{t}$ and the rest $\mathbf{k}$, $\mathbf{g} = \mathbf{t} \oplus \mathbf{k}$, the tangent space at a reducible connection on the other hand has the form

$$T_A C = \frac{\Omega^1(M,\mathbf{t})}{d\Omega^0(M,\mathbf{t})} \oplus \left(\frac{\Omega^1(M,\mathbf{k})}{d_A\Omega^0(M,\mathbf{k})} \right) \Big/ \mathrm{U}(1) \ , \tag{5.12}$$

where the first summand in the above corresponds to directions in the space of reducible [i.e. $\mathrm{U}(1)$] connections, and the second is a cone on $\mathbb{C}P^\infty$.

In summary, we have seen that near an irreducible connection C is smooth, whereas reducible connections lead to cone-like singularities in C. Similar results are known to hold for $G = \mathrm{SO}(3)$. For higher dimensional gauge groups, however, the singularity structure of C and its associated moduli spaces will be much more intricate (due to the larger number of possibilities for I_A), and so far not much is known about this case.

5.1.4. Instanton moduli space

$\mathrm{SU}(2)$ bundles P over a closed four-manifold M are classified by the second Chern class $c_2(E) \in H^4(M,\mathbb{Z})$ of the associated complex two-plane bundle $E = P \times_{\mathrm{SU}(2)} \mathbb{C}^2$, or alternatively by the first Pontrjagin class $p_1(E) = (c_1^2 - 2c_2)(E) = -2c_2(E)$, which can be represented by the four-form $-(2\pi i)^{-2}\,\mathrm{tr}\,F_A^2$. The topological charge (or quantum number) associated with P is the integer

$$k = c_2(E)[M] = \frac{1}{8\pi^2} \int_M \mathrm{tr}\,F_A^2 \ , \tag{5.13}$$

which can take positive or negative values. Note that k is independent of the connection A on P. Note also that these sign conventions may look unfamiliar: we have chosen tr to be positive definite

and thus *minus* the Killing–Cartan form on $\mathbf{g} = \mathrm{su}(2)$, and we have arranged for k to be *positive* for bundles supporting connections with anti-self-dual curvature.

The Yang–Mills functional (action) is

$$S(A) = \int_M \mathrm{tr}(F_A * F_A) \equiv \|F_A\|^2 \ , \tag{5.14}$$

and the variational equations following from it are the Yang–Mills equations (5.9), $d_A * F_A = 0$. We now want to show that solutions to the (anti-)self-duality equation

$$*F_A = -F_A \tag{5.15}$$

[which are solutions to the Yang–Mills equations by the Bianchi identity (5.8)] are absolute minima of (5.14). To show this, we prove that the absolute value $|k|$ of k gives a (topological) lower bound on $S(A)$,

$$S(A) \geq 8\pi^2 |k| \ . \tag{5.16}$$

It is then clear from (5.13) and (5.14) that this bound is saturated by connections A satisfying $*F_A = F_A$ ($*F_A = -F_A$) for $k < 0$ ($k > 0$).

Introducing the projection operators $P_\pm$ on $\Omega^2(M, \mathbf{g})$,

$$P_\pm = \tfrac{1}{2}(1 \pm *) \ , \qquad P_\pm^2 = P_\pm \ , \qquad P_+ P_- = 0 \ ,$$

any two-form α can be decomposed into the sum of a self-dual and an anti-self-dual part,

$$\alpha = \alpha^+ + \alpha^- \ , \qquad \alpha^\pm = P_\pm \alpha \in \Omega^2_\pm(M, \mathbf{g}) \ , \qquad \alpha^\pm = \pm * \alpha^\pm \ .$$

Applying this decomposition to the curvature two-form F_A we find

$$\mathrm{tr}\, F_A F_A = \mathrm{tr}\, F_A^+ F_A^+ + \mathrm{tr}\, F_A^- F_A^- = \mathrm{tr}\, F_A^+ * F_A^+ - \mathrm{tr}\, F_A^- * F_A^- \ , \tag{5.17}$$

$$\mathrm{tr}\, F_A * F_A = \mathrm{tr}\, F_A^+ * F_A^+ + \mathrm{tr}\, F_A^- * F_A^- \ , \tag{5.18}$$

and therefore, as claimed,

$$\|F_A\|^2 = \|F_A^+\|^2 + \|F_A^-\|^2 \geq |\, \|F_A^+\|^2 - \|F_A^-\|^2\, | = 8\pi^2 |k| \ , \tag{5.19}$$

with equality iff $k > 0$ and $F_A^+ = 0$ or $k < 0$ and $F_A^- = 0$ (excluding the case of flat connections $k = 0$, $F_A = 0$).

We will now take a look at the space $\mathcal{A}^+ := \{A \in \mathcal{A} \mid F_A^+ = 0\} \subset \mathcal{A}$ of anti-self-dual connections, and the instanton moduli space $\mathcal{M} = \mathcal{A}^+ / \mathcal{G} \subset \mathcal{C}$. $\mathcal{M}$ of course depends on the base manifold M, the isomorphism class k of the bundle P, and the (conformal class of the) metric used in the definition of the Hodge duality operator, but we will indicate this dependence explicitly only where needed.

Assuming that $\mathcal{A}^+$ is non-empty, a tangent vector $\tau \in \Omega^1(M, \mathbf{g})$ to $\mathcal{A}^+$ at a connection A has to satisfy the linearized instanton equation

$$\frac{\delta F_A^+}{\delta A}[\tau] = (d_A \tau)^+ = 0 \ . \tag{5.20}$$

If τ is of the form $\tau = d_A \Lambda, \Lambda \in \Omega^0(M, \mathbf{g})$, and therefore tangent to the gauge orbit of $\mathcal{G}$ through A, eq. (5.20) is satisfied identically, since then $(d_A \tau)^+ = [F_A^+, \Lambda] = 0$. This is a general fact about

gauge invariant equations of motion which we will encounter again in section 5.3.4: if $\mathcal{F}(A)$ is some gauge invariant functional of A, then $(\delta\mathcal{F}/\delta A)[d_A\Lambda]$ is zero identically if $\mathcal{F}(A) = 0$.

Turning now to $\mathcal{M}$ we have thus seen that a one-parameter family (curve) of instantons in $\mathcal{M}$ defines an element of $\operatorname{Ker} P_+ d_A / \operatorname{Im} d_A$, and therefore an element of the first cohomology group $\mathbf{H}_A^1$ of the *instanton deformation complex*

$$0 \to \Omega^0(M,\mathbf{g}) \xrightarrow{d_A} \Omega^1(M,\mathbf{g}) \xrightarrow{P_+d_A} \Omega_+^2(M,\mathbf{g}) \to 0 \qquad (5.21)$$

of Atiyah, Hitchin and Singer [5.8]. This complex is elliptic (since we have quotiented away the action of $\mathcal{G}$), whence its cohomology groups are finite dimensional, and $h_A^1 = \dim \mathbf{H}_A^1$ should give the dimension of (the tangent space at A of) $\mathcal{M}$.

Now h_A^1 can alternatively be written as $h_A^1 = \dim(\operatorname{Ker} P_+ d_A \cap \operatorname{Ker} d_A^*)$, the number of linearly independent solutions to

$$P_+ d_A \tau = 0 \ , \qquad d_A^* \tau = 0 \qquad (5.22)$$

(i.e., instead of modding out by the action of $\mathcal{G}$ we "fix the gauge" $d_A^* \tau = 0$). This way of looking at h_1^A amounts to replacing the deformation complex (5.21) by the single elliptic operator

$$D_A = P_+ d_A + d_A^* : \Omega^1(M,\mathbf{g}) \to \Omega^0(M,\mathbf{g}) \oplus \Omega_+^2(M,\mathbf{g}) \ . \qquad (5.23)$$

This is a standard trick in index theory. In this way the Euler character of the de Rham complex

$$0 \to \Omega^0(M) \xrightarrow{d} \Omega^1(M) \xrightarrow{d} \cdots$$

can for instance be calculated as the index of the operator $d + d^* : \Omega^{\text{even}}(M) \to \Omega^{\text{odd}}(M)$. The latter identifies this as the Euler character of M.

The Atiyah–Singer index theorem can now be used to compute the index

$$\operatorname{Ind} D_A = \dim \operatorname{Ker} D_A - \dim \operatorname{Ker} D_A^* = h_A^1 - h_A^0 - h_A^2 \ , \qquad (5.24)$$

where $h_A^k = \dim \mathbf{H}_A^k$,

$$\mathbf{H}_A^0 = \{\xi \in \Omega^0(M,\mathbf{g}) \mid d_A\xi = 0\} \qquad (5.25)$$

is the space we have already encountered in our discussion of reducible connections, and

$$\mathbf{H}_A^2 = \{\chi \in \Omega_+^2(M,\mathbf{g}) \mid d_A^*\chi = 0\} \ . \qquad (5.26)$$

Atiyah, Hitchin and Singer used (5.23) to compute the dimension of the moduli space of irreducible ($\mathbf{H}_A^0 = 0$) instantons on a self-dual Riemannian manifold M with positive scalar curvature. The latter two conditions allowed them to prove a vanishing theorem for $\mathbf{H}_A^2$ and therefore they could use the index formula for D_A directly to compute the dimension of $\mathcal{M}$ as $d(\mathcal{M}) = h_A^1 = \operatorname{Ind} D_A$. Vanishing of $\mathbf{H}_A^2$ also enabled them to prove the smoothness of the moduli space of irreducible instantons (more on that below).

In general $\operatorname{Ind} D_A$ computes what is known as the formal (or virtual) dimension of $\mathcal{M}$,

$$d(\mathcal{M}) = h^1 - h^0 - h^2 \ , \qquad (5.27)$$

and the Atiyah–Singer index theorem determines this to be

$$d(\mathcal{M}) = p_1(\operatorname{ad} P) - \tfrac{1}{2}\dim G \left[\chi(M) + \sigma(M)\right] \ , \qquad (5.28)$$

where $\chi(M) = \sum_{i=0}^{n}(-)^{i} b_i(M)$ and $\sigma(M) = b_2^+(M) - b_2^-(M)$ are the Euler characteristic and the signature of M. For $G = \mathrm{SU}(2)$ and M one-connected (we shall assume this for the remainder of this and the following section) this is

$$d(\mathcal{M}) = 8k - 3[1 + b_2^+(M)] \; . \tag{5.29}$$

In particular for $M = S^4$, one recovers the familiar $(8k - 3)$-parameter family of instantons $[b_2(S^4) = 0]$. Heuristically, these correspond to k $(k = 1)$-instantons with $5k$ parameters indicating their position and size, and $\dim G(k - 1) = 3k - 3$ parameters specifying the relative orientation of the instantons in $\mathbf{g}$. The explicitly known five-parameter family of solutions shows that this heuristic picture is correct in the case $k = 1$. It also allows one to understand the non-compactness of $\mathcal{M}$ in general, where—to be precise—by "in general" we mean the five-dimensional moduli spaces of simply connected manifolds with $b_2^+ = 0$. It is these moduli spaces (of simply connected manifolds with negative definite intersection form—cf. the next section) which feature prominently in Donaldson's first applications of instantons to the topology of four-manifolds [5.4]. For these manifolds Taubes [5.12] has shown by an ingenious "grafting" procedure that the S^4 instantons can be transplanted to M, thus establishing that the moduli space is not empty in that case. The above description then shows (or rather, suggests) that there are instantons which are highly concentrated around points of M. As the scale of the instantons approaches zero they approximate instantons with a delta function support which (because of their singular nature) are not included in the original moduli space $\mathcal{M}$. It is then possible to compactify $\mathcal{M}$ by adding a "collar" consisting of highly concentrated instantons, the limiting boundary configurations being in one-to-one correspondence with points of M. Therefore the compactified moduli space $\overline{\mathcal{M}}$ will have one boundary component equal to M. We will take a look at "the other end" of $\mathcal{M}$ below.

It should perhaps be emphasized at this point that a formula like (5.28) or (5.29) by no means proves that instantons exist if $d(\mathcal{M}) > 0$. The correct statement is, that if instantons exist, then the formal dimension of the moduli space is given by (5.28).

If either of h^0 or h^2 is non-zero one meets obstructions [5.8, 5.9] when trying to extend the above infinitesimal (tangent space) analysis to the local (every element of $\mathbf{H}_A^1$ is defined by a one-parameter family—the converse to the above) and global (the local moduli spaces of h^1-dimensional families give local coordinates on $\mathcal{M}$, $\mathcal{M}$ is Hausdorff) level. Put crudely, the moduli space $\mathcal{M}$ will then have singularities. Conversely, one of the main results of this analysis is that $d(\mathcal{M})$ gives the actual dimension of $\mathcal{M}$, if $\mathcal{M}$ is smooth.

Let us note the following facts about the cohomology groups $\mathbf{H}_A^0$ and $\mathbf{H}_A^2$ [5.9, 5.10]:

(i) As we have seen above, $\mathbf{H}_A^0$ is non-zero iff A is reducible. As such $\mathbf{H}_A^0$ is of course metric independent. However, the answer to the question of whether or not there are reducible instantons, depends on the metric. It is known that, regardless of the metric, reducible instantons—if they exist—are isolated in $\mathcal{M}$, and that no reducible instantons exist for an open dense set (or: generic choice) of metrics if $b_2^+(M) > 0$.

(ii) Near an irreducible instanton A, $\mathcal{M}$ is the kernel of the operator

$$\mathcal{D}_A : \tau \mapsto d_A^* \tau \oplus P_+ F_{A+\tau} \; , \tag{5.30}$$

whose linearization D_A, eq. (5.23), is surjective if the cokernel $\mathbf{H}_A^0 \oplus \mathbf{H}_A^2 = \mathbf{H}_A^2$ vanishes. In that case the implicit function theorem can be used to deduce that the kernel of $\mathcal{D}$ is smooth near A. Again it can be shown that $\mathbf{H}_A^2$—which is clearly metric dependent, eq. (5.26)—is zero for a generic choice of metric.

We will make use of these results in section 5.2.6., where we will re-encounter the cohomology groups H^0_A, H^1_A and H^2_A in the guise of fermionic η, ψ and χ zero modes in the path integral of Donaldson theory.

(iii) If $b^+_2 (M) = h^2 = 0$, then reducible instantons are generically unavoidable and a neighborhood in $\mathcal{M}$ of a reducible instanton is modeled on $H^1_A/U(1)$, which is now [cf. (5.12)] a cone on a complex projective space of dimension $h^1_A/2$. Alternatively one can remove small neighborhoods of these singular points. At that end the moduli space (for $k = 1$, say) will then be a smooth manifold with boundary a disjoint union $\amalg CP^2$ of complex projective spaces. Combined with our previous observation on the structure of compactified moduli space $\overline{\mathcal{M}}$ this suggests that a simply connected compact oriented smooth four-manifold with $b^+_2 = 0$ is cobordant to a disjoint union of $\mathbb{C}P^2$'s. Donaldson shows that this is correct, and this is the basic observation allowing an application of Yang–Mills theory to the topology of four-manifolds [5.4], a subject to which we turn now.

5.1.5. Topology of four-manifolds and Donaldson invariants

The purpose of this section is twofold: We first wish to explain the significance and the proof of the following theorem of Donaldson (actually a corollary of his main theorem):

Theorem. *Let M be a simply connected closed oriented topological four-manifold with non-trivial negative-definite even intersection form. Then M admits no smooth structure.*

Then we will indicate the construction of the Donaldson, invariants which are able to distinguish inequivalent smooth structures on a topological manifold.

To set the stage for this we begin with a lightning review of four-dimensional topology, trying to describe briefly why four dimensions is special not only for physics but also for topology. Regrettably but unavoidably, such a review cannot be but incomplete. Readable introductions to four-dimensional topology can be found in refs. [5.13–5.15] and, in particular, in the monographs [5.16] and [5.17], to which we refer for details.

Recall that an n-dimensional topological manifold is a topological (Hausdorff) space locally homeomorphic to $\mathbb{R}^n$, and that a smooth manifold is locally diffeomorphic to $\mathbb{R}^n$. Evidently every smooth manifold is a topological manifold, but the converse need not be the case. Moreover, a smooth structure on a manifold (provided by a smooth atlas) is not necessarily unique, in the sense that two smooth manifolds with the same underlying topological manifold need not be diffeomorphic.

Now the situation concerning smooth structures in dimensions other than four can be roughly summarized as follows:

– in less than four dimensions every topological manifold has a unique smooth structure;

– in more than four dimensions the homotopy type and the Pontrjagin classes of a manifold determine the smooth structure (if it exists) up to a finite ambiguity; in fact, in these dimensions smoothing theory (the standard reference is [5.18]) reduces to obstruction theory, whence to problems involving characteristic classes; as examples of manifolds with non-standard smooth structures we mention the exotic spheres of Kervaire and Milnor [5.19] (27 in 7 and 991 in 11 dimensions), familiar to physicists from Witten's discussion of global gravitational anomalies [5.20];

– finally let us mention that the contractible flat spaces $\mathbb{R}^n$ do not share this bizarre property with the spheres S^n; it is a famous result that, for $n \neq 4$, $\mathbb{R}^n$ has a unique smooth structure.

Naively one would expect the situation in four dimensions to be somewhere "in between" that in less and that in more than four dimensions, and as far as topological four-manifolds are concerned

this is indeed true to a certain extent. Smoothing theory in four dimensions, however, turns out to be vastly different, a far cry from the discrete and finite situation encountered in higher dimensions.

The most important invariant of a manifold M is its fundamental group $\pi_1(M)$, and in two (three) dimensions it was (is) one of the main objects of interest. In four dimensions, however, $\pi_1(M)$ is not a good starting point for a classification of manifolds, since virtually "anything" (more precisely: any finitely presentable group) can appear as the fundamental group of a smooth compact four-manifold. Parenthetically it may be worth remarking that as a consequence of this the classification problem of smooth structures is non-algorithmic in $n \geq 4$ (cf. ref. [5.21] and references therein)! Interest has therefore until recently mainly centered around simply connected four-manifolds. Let us then assume for the time being that $\pi_1(M) = 0$. The fundamental invariant of a simply connected four-manifold M is its intersection form ω_M, a symmetric bilinear form on $H^2(M,\mathbb{Z})$ [note that $H^2(M,\mathbb{Z})$ is torsion free], defined by

$$\omega_M : H^2(M,\mathbb{Z}) \times H^2(M,\mathbb{Z}) \to \mathbb{Z}, \quad (a,b) \mapsto (a \cup b)[M] \ . \tag{5.31}$$

Here $a \cup b$ is the cup-product of the two cohomology classes, and $(a \cup b)[M]$ denotes evaluation on the fundamental class of M (this requires a choice of orientation). As a consequence of Poincaré duality ω_M is non-degenerate and unimodular. If M is smooth there is a less fancy way of defining ω_M via de Rham cohomology: if a and b now denote closed forms representing de Rham cohomology classes $[a], [b] \in H^2(M,\mathbb{R})$, then

$$\omega_M([a],[b]) = \int_M a\, b \tag{5.32}$$

(as always, the wedge product is understood). The basic invariants of ω are its rank $\rho(\omega) = \dim H^2(M,\mathbb{Z}) = b_2(M)$, and its signature $\sigma(\omega)$ [5.22]. This is the number of positive minus the number of negative eigenvalues of ω. Equation (5.32) shows that in the case of a smooth manifold this is the same as the number of self-dual minus the number of anti-self-dual harmonic two-forms and as such coincides with the signature of M introduced in (5.28), i.e.,

$$\sigma(\omega_M) = \sigma(M) = b_2^+(M) - b_2^-(M) \ .$$

ω is called even if all its diagonal entries $\omega(a,a)$ are even. The most prominent example of an even positive definite form is the Cartan matrix E_8 of rank 8. Finally, to prove Donaldson's theorem we need to know the intersection form of $\mathbb{CP}^2$, which is—since $H^2(\mathbb{CP}^2,\mathbb{Z})$ has a single generator—just the (1×1) matrix (1) [or (-1) if the orientation of $\mathbb{CP}^2$ is reversed].

Now it has been known for a long time [5.23] that the intersection form ω_M determines the homotopy type of M, and that every (non-degenerate symmetric, bilinear and unimodular) form is realized as ω_M for some simply connected homotopy four-manifold. Note that, unlike the situation in higher dimensions, the Pontrjagin numbers of a manifold provide no independent information in four dimensions, since Hirzebruch's famous signature theorem expresses $p_1(M)$ in terms of ω as $3\sigma(\omega_M) = p_1(M)$.

However, classification by homotopy type is rather coarse, and so it was a significant step forward when Freedman [5.24] showed in 1982 that also the homeomorphism type of a manifold is uniquely determined by ω if ω is even, and that there are precisely two non-homeomorphic topological manifolds for a given odd ω. In this way the classification of topological four-manifolds is essentially reduced to the algebraic classification of quadratic forms.

Turning now to smooth manifolds, two results which have emerged as a consequence of the investigations initiated by Donaldson may suffice to illustrate the radically different character the classification of smooth manifolds has in four dimensions:

(i) There are closed topological four manifolds with a countably infinite number of distinct smooth structures [5.25, 5.26].

(ii) There is an uncountable family of distinct smooth structures on $\mathbb{R}^4$ [5.27] (cf. also ref. [5.28] for a countably infinite two-parameter family).

We will not be able to indicate the proof of either of these results, since even the simplest argument [5.9] establishing the existence of at least one exotic $\mathbb{R}^4$ requires some knowledge of surgery. We thus turn our attention to two other facts (mentioned at the beginning of this section) which have emerged from these investigations, namely

(iii) many topological manifolds admit no smooth structure at all,

(iv) there are rational cohomology invariants which are able to distinguish inequivalent smooth structures (in marked contrast with, say, the rational Pontrjagin classes, which were shown by Novikov to be topological invariants).

Later on in this section we shall sketch how these invariants can be constructed. As regards (iii), we will see now that we have acquired almost enough information already to prove the theorem quoted above (and more).

Recall from the previous section that the (compactified) moduli space $\overline{\mathcal{M}}$ of instantons provides a cobordism between M and $\amalg\mathbb{CP}^2$ if $b_2^+(M) = 0$ (i.e., if ω_M is negative definite). As a simple consequence of Poincaré duality for manifolds with boundary, the signature of an oriented boundary is zero [5.29, theorem 8.2.1]. Applied to $\partial\overline{\mathcal{M}}$ (the orientability of $\overline{\mathcal{M}}$ has been established in refs. [5.4, 5.9] and more generally in ref. [5.30]) this means that the signature of M equals that of $\amalg\mathbb{CP}^2$. Now we have not said anything about the relative orientation of the $\mathbb{CP}^2$'s, but regardless of that we can certainly deduce from $\sigma(\omega_M) = \sigma(\amalg\mathbb{CP}^2)$ that

$$-n(\omega) \leq \sigma(\omega) \leq n(\omega) \;, \tag{5.33}$$

where $n(\omega)$ is the number of $\mathbb{CP}^2$'s which we need to determine. Recall from section 5.1.3 that U(1)-reducible connections (responsible for the $\mathbb{CP}^2$'s) arise from a splitting $E = E_1 \oplus E_2$ of E into line bundles. Since the structure group is SU(2) we obtain

$$c_1(E) = c_1(E_1) + c_1(E_2) = 0 \;,$$

which implies that $E_2 = E_1^{-1}$ since line bundles are uniquely determined by their first Chern class. Then we find for $c_2(E)$,

$$c_2(E) = -c_1(E_1) \cup c_1(E_1) \;.$$

Thus the number $m(\omega)$ of splittings of the $k = c_2(E)[M] = 1$ bundle E is equal to half the number of solutions to

$$\omega_M(a,a) = -1 \;, \quad a \in H^2(M,\mathbb{Z})$$

(half because a and $-a$ determine the same splitting). By associating to each such a its unique anti-self-dual harmonic representative in $H^2(M,\mathbb{R})$, one sees that a gives rise to an instanton. Therefore $m(\omega) = n(\omega)$, and some elementary linear algebra [5.4, 5.9] shows that for a negative definite intersection form $n(\omega) \leq \rho(\omega)$ with equality iff ω is diagonalizable over $\mathbb{Z}$, i.e., iff it is equivalent to the standard form $(-1) \oplus \cdots \oplus (-1)$. Combining $n(\omega) \leq \rho(\omega)$ with the above

inequality (5.33) and recalling that $\sigma(\omega) = -p(\omega)$, we see that necessarily $p(\omega) = n(\omega)$. We have thus proved Donaldson's theorem:

Theorem. *Let M be a smooth simply connected closed four-manifold with negative definite intersection form ω_M. Then*

$$\omega_M \simeq (-1) \oplus \cdots \oplus (-1) \ .$$

Donaldson's theorem thus determines many symmetric bilinear forms ω (realized as the intersection form of some topological four-manifold by Freedman's classification) which cannot arise as the intersection form of any smooth four-manifold. From the above, the result quoted at the beginning of this section follows immediately.

Subsequent work concentrated on relaxing one or the other of the conditions in Donaldson's theorem. In refs. [5.31, 5.32] it was shown that analogous conclusions can still be drawn under less stringent conditions on the fundamental group of M, and Donaldson was then able to show that the above theorem is true for arbitrary $\pi_1(M)$ [5.30]. Similar results are also available [5.33] for small non-zero values of b_2^+, and it was then another major breakthrough when Donaldson was able to prove a number of powerful theorems for manifolds with arbitrary odd $b_2^+ \geq 3$ [5.5].

Responsible for these developments was a shift in emphasis from cobordism to homology, the basic idea being to try to use $\mathcal{M}$ (or rather $\mathcal{M}^*$, the moduli space of irreducible instantons) to define a cycle $[\mathcal{M}]$ in the homology of C^*. Of course this is anything but straightforward: $\mathcal{M}$ depends on the metric, and one has to make sure that it only varies within its homology class as the metric is changed. Moreover $\mathcal{M}$ is usually non-compact, with quite a complicated structure at the "ends". It turns out, however, that this difficulty can be overcome for suitably generic metrics, provided that one is in a "stable range" of k, $|k| > k_0$ for some k_0 depending on b_2^+. The latter condition ensures that the lower dimensional strata of the compactified moduli space $\overline{\mathcal{M}}_k$, $k > 1$, are of high enough codimension so as not to contribute to the evaluation of (compactly supported) cohomology classes of C^* on $[\mathcal{M}]$. It is this cohomological point of view to which we turn now.

As a prerequisite for this approach to make sense, one has of course to make sure that (the compactification of) $\mathcal{M} = \mathcal{M}_k$ is orientable. To see what this requirement amounts to, recall from the previous section that generically the tangent space to $\mathcal{M}$ at a point $[A]$ is $H_A^1 = \mathrm{Ker}\,D_A$. An orientation of the index bundle $\mathrm{ind}\,D = \mathrm{Ker}\,D - \mathrm{Coker}\,D$ over C will then define an orientation of the tangent bundle of $\mathcal{M}$. It thus needs to be shown that the *determinant line bundle* [5.34] of D is orientable, and the latter has been established by Donaldson [5.30].

Next one needs to know something about the cohomology of C^*. The rational cohomology ring of C^* is generated by cohomology classes in two and four dimensions (in particular, all rational cohomology classes lie in even dimensions) and the key ingredient in the construction of cohomology classes which can be evaluated on $[\mathcal{M}]$ is a map

$$\mu : H_i(M) \to H^{4-i}(C^*) \ , \tag{5.34}$$

which (for $i = 0, 2$) expresses these generators in terms of the homology of M. Since this map will play an important role in the following we will now give two descriptions of it [5.33].

The first one is in terms of determinant line bundles: If Σ is an embedded surface in M, there is a Dirac operator ∂_Σ which can be coupled to the restriction $r_\Sigma : C^* \to C_\Sigma^*$ of the gauge fields to Σ. This family of Dirac operators defines a determinant line bundle [5.34] L_Σ on C_Σ^* which can be pulled back to a line bundle $\mathcal{L}_\Sigma = r_\Sigma^* L_\Sigma$ on C^*. We now define $\mu([\Sigma]) = c_1(\mathcal{L}_\Sigma^{-1})$. By Poincaré

duality this characteristic class can also be represented by a codimension two-subspace of C^*, and this is essentially the definition Donaldson adopts in ref. [5.5].

Alternatively μ can be defined in terms of the *universal bundle* $\mathcal{Q}$ [5.35] over $M \times C^*$, which we discuss in detail in sections 5.3.1 and 5.3.2: Let $E_\mathcal{Q}$ be the associated two-plane bundle, $c_2(E_\mathcal{Q})$ (a representative of) its second Chern class, and set $\mu([\Sigma]) = \int_\Sigma c_2(E_\mathcal{Q})$. This equation is to be understood as follows: As a four-form on $M \times C^*$, c_2 decomposes into a sum of $(i, 4 - i)$-forms, where an $(i, 4 - i)$-form is an i-form on M and a $(4 - i)$-form on C^*. In the above equation only the $(2,2)$-part contributes and, integrated over Σ, this leaves us, as desired, with a two-form on C^*. The precise way of saying this is that $\mu([\Sigma])$ is the *slant product* $c_2(E_\mathcal{Q})/[\Sigma]$.

The relation between these two definitions is provided by the family index theorem of Atiyah and Singer [5.35, 5.36], which expresses the characteristic classes of the index bundle in terms of those of the universal bundle. Explicitly one has (denoting by ch_i the term of degree $2i$ in the expansion of the Chern character ch)

$$c_1(\mathcal{L}_\Sigma^{-1}) = -ch_1(\text{ind}\,\partial_\Sigma) = -ch_2(E_\mathcal{Q})/[\Sigma] = c_2(E_\mathcal{Q})/[\Sigma] = \mu([\Sigma]) \ , \tag{5.35}$$

confirming the equivalence of the two definitions given above.

Let us now see formally (i.e., ignoring questions of transversality, genericity, and reducibility) how this map can be used to define the desired cohomology classes (Donaldson polynomials). μ can of course be extended to a map

$$\mu : \underbrace{H_2(M) \times \cdots \times H_2(M)}_{d \text{ times}} \to H^{2d}(C^*)$$

via the cup product in $H^*(C^*)$. This map gives an injection from the polynomial algebra on $H_2(M)$ into $H^{\text{even}}(C^*)$, and it is this polynomial which—when evaluated on the homology cycle $[\mathcal{M}] \in H_*(C^*)$—defines the Donaldson invariants. For this to lead to non-trivial results, $\mathcal{M}$ should of course be even dimensional, and a look at (5.29) reveals that for simply connected M this is the case precisely when b_2^+ is odd. Writing $b_2^+ = 2p + 1$, we then have $d(\mathcal{M}) = 2d$ where $d = 4k - 3(1 + p)$, and we define

$$q_k([\gamma_1], \ldots, [\gamma_d]) = (\mu([\gamma_1]) \cup \cdots \cup \mu([\gamma_d]))[\mathcal{M}_k] \ , \tag{5.36}$$

where $[\gamma_i] \in H_2(M)$. Using the equations of section 5.3.1 or 5.3.2 for the curvature of the universal connection on $\mathcal{Q}$, (5.36) can now be written explicitly as an integral of a product of closed differential forms over $\mathcal{M}$, and this is the form in which we are going to obtain (5.36) in the next section.

Alternatively, in the Poincaré dual picture, (5.36) is (as the observables of the topological sigma model, section 4.5.2) an intersection number, determined by the intersection of the codimension-two subspaces of C^* and $\mathcal{M}$ and the orientation of $\mathcal{M}$. Generically these intersections will be transverse and consist of isolated points.

If the formal dimension of $\mathcal{M}$ is zero and $\mathcal{M}$ itself consists of isolated points, the Donaldson invariant will just be the number of these points counted with signs, the signs being determined by the relative orientations (i.e., the relative orientations of the determinant lines of D) at these points. This situation does not occur for $SU(2)$ bundles in the stable range of k mentioned above, but is possible for $SO(3)$ bundles. We mention it here since we will see in section 5.2.5 that the partition function is non-zero and equal to this "first" Donaldson invariant precisely in that case. An interpretation of this invariant as the Euler number of an infinite dimensional vector bundle has been provided by Atiyah and Jeffrey [5.37], and we will explain this in section 5.2.6.

It can now be verified that the numbers associated to M in this way are independent of the metric on M. The proof of this fact can be found in ref. [5.5], and a formal argument based on the standard equations of topological field theory (section 2) is sketched at the end of section 5.2.7. This metric independence means that the Donaldson invariants are differential invariants of M, and while this can be proved directly, the real mystery concerning these invariants is why they are *not* topological invariants, i.e., why they are able to distinguish inequivalent smooth structures. The latter fact was discovered in ref. [5.25] and expanded in several directions in refs. [5.26, 5.5]. For a discussion of the application of these invariants (which are very hard to compute explicitly in general) cf. the references mentioned above and refs. [5.11, 5.17].

5.1.6. Floer homology and Morse theory

Another important contribution to low dimensional topology—which then led directly to the development of topological field theories—is due to Floer [5.6, 5.7, 5.38]. He developed a new infinite dimensional version of Morse theory (*relative* Morse theory), which permitted him to successfully tackle a number of difficult problems in symplectic geometry and (relevant for us here) the study of three-manifolds.

This relative Morse theory is an infinite dimensional generalization of Witten's tunneling approach to classical Morse theory [5.39]. In its homological version (the cohomological version has been explained in detail in section 3.10.2), the Witten complex for a Morse function f consists of chain groups W_q having one generator for each critical point P with Morse index $\mu(P) = q$, and a boundary operator $\partial_W : W_q \to W_{q-1}$,

$$\partial_W |P\rangle = \sum_{Q \in W_{q-1}} n(P, Q)|Q\rangle , \qquad (\partial_W)^2 = 0 , \qquad (5.37)$$

where the $n(P, Q)$ are integers counting the gradient lines between P and Q with appropriate signs. The homology groups of this complex coincide with the homology groups of the manifold M, which of course (section 3) also arise as the ground states of the Hamiltonian obtained from the Laplace operator by convoluting it with the Morse function f.

A trivial but perhaps helpful remark may be that this definition of homology groups immediately implies (almost tautologically) the weak Morse inequalities, since

$$b_q = \dim H_q = \dim \frac{\text{Ker} \, \partial_{W_q}}{\text{Im} \, \partial_{W_{q+1}}} \leq \dim \text{Ker} \, \partial_{W_q} \leq \dim W_q = N_q .$$

As it turns out this approach to defining (co)homology groups has a generalization to certain infinite dimensional situations, where it defines what Atiyah [5.11] calls a middle dimensional cohomology which is expected to reveal information inaccessible by more classical methods. This cohomology is not unlike the semi-infinite cohomology familiar from string theory [5.40].

In the context of three-manifolds, the desire for such a generalization arises as follows. One of the main objects of interest is the fundamental group $\pi_1(Y)$, which is conveniently studied by means of its representations in some Lie group, say $\text{SU}(2)$. One is thus (identifying conjugate representations) interested in the space

$$\mathcal{M}(Y, \text{SU}(2)) = \text{Hom}(\pi_1(Y), \text{SU}(2))/\text{SU}(2) ,$$

or rather—due to its somewhat singular nature in general—in the component $\mathcal{M}^*(Y, \text{SU}(2))$ consisting of irreducible representations. If Y is a homology three-sphere [i.e. a closed three-manifold with $H_1(Y, \mathbb{Z}) = 0$], this is automatically taken care of, i.e.,

$$\mathcal{M}^*(Y, \text{SU}(2)) = \mathcal{M}(Y, \text{SU}(2)) \backslash \{1\} ,$$

since any reducible representation would factor through to a representation in $U(1)$ and thence to a representation of the abelianization of $\pi_1(Y)$ which is trivial by assumption.

Recognizing $\mathcal{M}(Y, SU(2))$ as the moduli space of flat $SU(2)$ connections on Y (we will explain this identification in section 5.4.3), one is therefore motivated to study the Chern–Simons functional

$$CS(A) = \frac{1}{4\pi} \int tr(Ad\,A + \tfrac{2}{3}A^3) \tag{5.38}$$

of a, necessarily trivial, $SU(2)$ bundle on Y as a Morse function on the space $C^* = \mathcal{A}^*/\mathcal{G}$ of gauge equivalence classes of (irreducible) connections on Y, since its critical points are precisely the flat connections.

A minor problem arising at this point is that $CS(A)$ is not invariant under large gauge transformations (this fact and its implications will be reviewed in section 6.2). As a consequence, $CS(A)$ is only well defined modulo $2\pi\mathbb{Z}$ as a function on C^*.

Much more serious, however, is the fact that the Hessian of CS at the gauge equivalence class of a flat connection (critical point) A is the operator $H_A = *d_A$ acting on the space $\Omega^1(Y, su(2))/d_A\Omega^0(Y, su(2))$ of Lie algebra valued one-forms modulo gauge transformations. This operator (being of Dirac type) has a spectrum which is unbounded from above and below, so formally its Morse index is $\mu = \infty$ at every critical point. But although this seems to be disastrous for a potential Morse theoretic treatment, it is nevertheless possible to make sense of a *relative* Morse index $\mu(a,b) = \mu(a) - \mu(b)$ for two critical points a and b, which is—in the light of the previous discussion of the Witten complex—really all that we need to know to define a homology theory. Note that in this case we are forced to consider the information contained in the gradient flow between two critical points: ordinary Morse theory does not exist for CS, but nevertheless we can use this functional to define a homology. This illustrates clearly the power of Witten's and Floer's approach.

In finite dimensions, the relative Morse index can be determined by extending the Hessian at the critical points a and b to a one-parameter family $H(t)$ of matrices with $H(0) = H_a$ and $H(1) = H_b$. One then counts the number of positive eigenvalues of $H(t)$ changing to negative values as t increases from zero to one, minus those crossing the eigenvalue zero in the opposite direction. The result—the relative Morse index—is clearly independent of the one-parameter family chosen to interpolate between H_a and H_b.

This method of determining the relative Morse index has an infinite dimensional generalization in the *spectral flow* of a family of operators (this concept has been introduced by Atiyah, Patodi and Singer in ref. [5.41]). The relative Morse index $\mu(a,b)$ can then be defined as the spectral flow (modulo 8) of a family of operators $H(t)$ along a path in C from the flat connection a to the flat connection b with $H(0) = H_a, H(1) = H_b$.

The necessity for the "modulo 8" can be understood as follows: As a discrete (integer) invariant, the spectral flow is certainly invariant under smooth deformations (homotopies) of the chosen path. Thus in order to check that the above definition of the relative Morse index is well defined (independent of the chosen path) we only have to check this on homotopy classes of paths. Now the difference between the spectral flows along two non-homotopic paths γ and γ' is the same as the spectral flow along the closed non-contractible loop $\gamma \cup \gamma'^{-1}$, which can be computed [5.41] as the index of an operator on $M = Y \times S^1$. The presence of these non-contractible loops is due to the existence of "large" gauge transformations on Y, since

$$\pi_1(C^*) = \pi_0(\mathcal{G}) = \mathbb{Z} \; . \tag{5.39}$$

Such large gauge transformations can be used as patching data (or clutching functions) to construct

non-trivial SU(2) bundles on M. In technical terms these are the SU(2) bundles over the mapping torus of the gauge transformation. Explicitly [5.41, p. 95], the operator on M in question is [replacing $*d_A$ acting on $\Omega^1(Y,\mathbf{g})/d_A\Omega^0(Y,\mathbf{g})$ by the operator $B = *d_A - d_A*$ acting on all of $\Omega^1(M,\mathbf{g})$] $\mathrm{d}t\,\partial/\partial t + B$, which is precisely the (dual of the) deformation operator D_A, eq. (5.23), of the instanton deformation complex (5.21) (in the $A_0 = 0$ gauge). Therefore we can apply eq. (5.28) for the index, and using the fact that $b_1(M) = 1$ and $b_2(M) = 0$ (since Y is a homology three-sphere), we see that this index is a multiple of 8, the integer k corresponding to $k \in \mathbb{Z}$ labeling the winding number (or homotopy sector) in (5.39). Thus the spectral flow is indeed well defined modulo 8, and consequently the chain groups W_q as well as the (Floer) homology groups HF_q will also be indexed modulo 8. The definition of the boundary operator ∂_W involves, as before, information contained in the gradient flow between critical points (for details cf. refs. [5.7] and [5.42–5.44]).

One of the reasons why we have gone through all this and why Floer homology is relevant for Donaldson's work, is that this gradient flow is determined by the equation

$$(\mathrm{d}/\mathrm{d}t)A = -*F_A ,\qquad\qquad(5.40)$$

which, when interpreted as an equation on $Y \times \mathbb{R}$, is precisely the (anti-)self-duality equation in the $A_0 = 0$ gauge. Thus we really have an instanton tunneling from the flat connection a at $t = -\infty$ to the flat connection b at $t = +\infty$. This also explains the discovery of Donaldson (which originally led to the interest in Floer's work) that the definition of his instanton invariants on a four-manifold M with boundary Y involves the Floer homology groups, since near the boundary M looks like $Y \times \mathbb{R}_+$. The reason why this observation is useful is that, as mentioned in the previous section, the Donaldson invariants are very difficult to compute in general. Given the above result, however, one could imagine computing the Donaldson invariants for a four-manifold M by writing M as the sum of two manifolds joined along a homology three-sphere (the analog of the Heegard splitting of three-manifolds along Riemann surfaces), $M = M_1 \#_Y M_2$. The computation is then reduced to one in Floer homology [5.11], which may be more tractable. For some progress along these lines see ref. [5.45]. We will recover this result from a path integral point of view in section 5.2.9.

In analogy with the case of supersymmetric quantum mechanics (section 3), the Floer (co)homology groups we have defined in this way are (formally) the ground states of the Hamiltonian [5.11, 5.39, 5.1]

$$H = \tfrac{1}{2}(\delta_t\delta_t^* + \delta_t^*\delta_t) ,\qquad\qquad(5.41)$$

where δ is the exterior derivative on $\mathcal{C}$, δ^* its adjoint, and

$$\delta_t = \mathrm{e}^{-2\pi t\mathrm{CS}(A)}\,\delta\,\mathrm{e}^{2\pi t\mathrm{CS}(A)} ,\qquad \delta_t^* = \mathrm{e}^{2\pi t\mathrm{CS}(A)}\,\delta^*\,\mathrm{e}^{-2\pi t\mathrm{CS}(A)} .$$

Introducing one-forms $\delta A_i^a(x) \equiv \psi_i^a(x)$ and vector fields $\chi_i^a(x)$ on $\mathcal{A}^3$ satisfying the anti-commutation relations

$$\{\psi_i^a(x),\psi_j^b(y)\} = 0 , \quad \{\chi_i^a(x),\chi_j^b(y)\} = 0 , \quad \{\psi_i^a(x),\chi_j^b(y)\} = g_{ij}\delta^{ab}\delta^{(3)}(x-y) ,\quad(5.42)$$

such that

$$\delta = \int \mathrm{d}^3x\,\psi_i^a(x)\frac{\delta}{\delta A_i^a(x)} , \qquad \delta^* = -\int \mathrm{d}^3x\,\chi_i^a(x)\frac{\delta}{\delta A_i^a(x)} ,$$

this Hamiltonian is more explicitly

$$H = \int \mathrm{d}^3x \left[\frac{1}{2} \sum_{i,a} \left(\frac{1}{\mathrm{i}} \frac{\delta}{\delta A_i^a(x)} \right)^2 + \frac{1}{2} t^2 \operatorname{tr} B_i B^i + t \epsilon^{ijk} \operatorname{tr} \psi_i D_j \chi_k \right] , \tag{5.43}$$

with $B_i = \frac{1}{2} \epsilon_{ijk} F^{jk}$. Note that the first two terms are the usual Yang–Mills Hamiltonian, while the third term is a Lorentz non-invariant coupling to anti-commuting spin one fields. Atiyah therefore conjectured [5.11] that there should exist a relativistic four-dimensional field theory with the following features:
– it should be related to Donaldson's instanton invariants,
– in a Hamiltonian treatment on a four-manifold of the form $M = Y \times \mathbb{R}$ it should reproduce the picture sketched above; in particular the Floer homology groups of Y should emerge as the ground states of the theory.
A field theory meeting these requirements was soon thereafter constructed by Witten [5.1] and we shall now turn to a detailed discussion of its properties.

5.2. Donaldson theory

5.2.1. Fundamental properties

For reasons explained at the beginning of the previous section, we start by considering the action and its properties as presented by Witten [5.1]. As a first step towards a better understanding of this theory, we then show in sections 5.2.2 and 5.2.3 how Witten's action can be derived from the Baulieu–Singer (Brooks–Montano–Sonnenschein) and the Labstida–Pernici points of view. These derivations—although perhaps not truly fundamental—make obvious certain of the properties of Donaldson theory, like the absence of degrees of freedom and the role played by instanton configurations. After a group theoretic interlude in section 5.2.4 which shows that Donaldson theory is a "twisted" $N = 2$ super Yang–Mills theory, we then gradually make contact with the mathematical results of the previous section. In section 5.2.5 we show that the partition function equals the first Donaldson invariant, described towards the end of section 5.1.5, and in section 5.2.6 we explain the interpretation of the partition function as the Euler number of an infinite dimensional vector bundle over $\mathcal{A}/\mathcal{G}$, due to Atiyah and Jeffrey. In section 5.2.7 we relate the counting of fermionic zero modes and the ghost number violation to the index of the instanton deformation complex. In the quest for suitable (BRST invariant, metric independent) observables, a field theoretic analogue of Donaldson's μ-map appears naturally, and in section 5.2.8 the resulting observables are identified with closed differential forms on $\mathcal{M}$. In section 5.3 we will show in detail that the geometry described by the zero mode sector of Donaldson theory is that of the universal bundle $\mathcal{Q}$. Anticipating this result allows us to complete the identification of these observables with the Donaldson polynomials. In section 5.2.9 we will take a look at the theory from the Hamiltonian point of view, making contact with the features of Donaldson theory described in section 5.1.6.

In large parts of this section we follow closely Witten's original paper, referring to it wherever necessary for the details we have omitted. Our notation (we start off in components and gradually converge to an index free differential form notation) reflects the growing influx of mathematical ideas from the previous section. All in all, however, the mathematical level here is considerably lower—and our treatment more formal—than that in section 5.1.

The action

$$S = \int_M \sqrt{g}\,\mathrm{d}^4 x\ \mathrm{tr}(\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta} - 2\chi_{\alpha\beta}D^\alpha\psi^\beta + \eta D_\alpha\psi^\alpha$$

$$+ \bar\phi[\psi_\alpha, \psi^\alpha] - \bar\phi D_\alpha D^\alpha\phi - \tfrac{1}{2}\phi[\chi_{\alpha\beta}, \chi^{\alpha\beta}]) \tag{5.44}$$

is of the expected form Yang–Mills + $\cdots$, and is constructed from fields having an additional (internal) Grassmann grading which we refer to as ghost number, anticipating the BRST interpretation of these fields below. Up to some numerical factors, which we have absorbed by simple field redefinitions, this action is the one used by Witten [5.1]. A_α is an SU(2) gauge potential, $F_{\alpha\beta}$ its field strength, $(\phi, \bar\phi)$ are even scalar fields with ghost numbers $(2, -2)$, $\chi_{\alpha\beta} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}\chi^{\gamma\delta}$, ψ_α and η are odd self-dual, vector and scalar fields with ghost numbers $(-1, 1, -1)$, respectively. All fields take values in the Lie algebra su(2) of the structure group SU(2). And from now on, the trace tr in integrals of Lie algebra valued forms will always be understood. We will also abbreviate the volume element $\sqrt{g}\,\mathrm{d}^4 x$ to $\mathrm{d}x$.

This action is invariant under the usual Yang–Mills symmetry as well as the following BRST-like transformations:

$$\delta A_\alpha = \psi_\alpha\ , \qquad \delta\psi_\alpha = -D_\alpha\phi\ , \qquad \delta\phi = 0\ ,$$

$$\delta\chi_{\alpha\beta} = F_{\alpha\beta}^+\ , \qquad \delta\bar\phi = \eta\ , \qquad \delta\eta = [\phi, \bar\phi]\ , \tag{5.45}$$

where $F_{\alpha\beta}^\pm = \tfrac{1}{2}(F_{\alpha\beta} \pm \tilde F_{\alpha\beta})$ and $\tilde F_{\alpha\beta} = \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}F^{\gamma\delta}$. We also note that in writing $\delta\Phi$, we shall now mean $\{Q, \Phi\}$, where Q is the (BRST) charge operator.

The energy–momentum tensor of (5.44), defined by $\delta S = \tfrac{1}{2}\int \mathrm{d}x\, T_{\alpha\beta}\delta g^{\alpha\beta}$,

$$T_{\alpha\beta} = \mathrm{tr}\{(F_{\alpha\gamma}F_\beta{}^\gamma - \tfrac{1}{4}g_{\alpha\beta}F_{\gamma\delta}F^{\gamma\delta})$$

$$+ [(D_\alpha\psi_\gamma - D_\gamma\psi_\alpha)\chi_\beta{}^\gamma + (D_\beta\psi_\gamma - D_\gamma\psi_\beta)\chi_\alpha{}^\gamma - g_{\alpha\beta}(D_\gamma\psi_\delta)\chi^{\gamma\delta}]$$

$$+ [D_\alpha\phi D_\beta\bar\phi + D_\beta\phi D_\alpha\bar\phi - g_{\alpha\beta}D_\gamma\phi D^\gamma\bar\phi] - [(D_\alpha\eta)\psi_\beta + (D_\beta\eta)\psi_\alpha - g_{\alpha\beta}(D_\gamma\eta)\psi^\gamma]$$

$$+ 4[\bar\phi\psi_\alpha\psi_\beta - \tfrac{1}{2}g_{\alpha\beta}\bar\phi\psi_\gamma\psi^\gamma]\}\ , \tag{5.46}$$

can be written in the form

$$T_{\alpha\beta} = \{Q, V_{\alpha\beta}\}\ , \tag{5.47}$$

with

$$V_{\alpha\beta} = \mathrm{tr}[(F_{\alpha\gamma}\chi_\beta{}^\gamma + F_{\beta\gamma}\chi_\alpha{}^\gamma - \tfrac{1}{2}g_{\alpha\beta}F_{\gamma\delta}\chi^{\gamma\delta}) + (\bar\phi D_\alpha\psi_\beta + \bar\phi D_\beta\psi_\alpha - g_{\alpha\beta}\bar\phi D_\gamma\psi^\gamma)]\ .$$

Equation (5.47) is an immediate consequence of the following considerations: adding the metric independent term $\int F_{\alpha\beta}\tilde F^{\alpha\beta}$ to the action (5.44) (this changes neither the energy–momentum tensor nor the equations of motion) we can write it (upon using the χ equation of motion) as

$$S' = S + \frac{1}{4}\int_M \mathrm{d}x\, F_{\alpha\beta}\tilde F^{\alpha\beta} = \{Q, V\}\ , \tag{5.48}$$

where

$$V = \int \mathrm{d}x\, (\tfrac{1}{2}F_{\alpha\beta}^+\chi^{\alpha\beta} + \bar\phi D_\alpha\psi^\alpha)\ . \tag{5.49}$$

For certain purposes (e.g. path integral considerations) this form of the action is more convenient. Varying (5.48) with respect to the metric one establishes (5.47), with

$$V_{\alpha\beta} = \frac{2}{\sqrt{g}} \frac{\delta V}{\delta g^{\alpha\beta}} \; .$$

By our general discussion of section 2 we have thus established the topological nature of the model with all its consequences (metric independence of the partition function, ...).

Moreover, the preceding equations suggest the following interpretation: $\int F_{\alpha\beta} \tilde{F}^{\alpha\beta}$ is the classical action supplemented with gauge fixing terms which have the usual structure anti-ghost $(\chi_{\alpha\beta}, \bar{\phi})$ times "gauge condition" $F_{\alpha\beta}^+ = D_\alpha \psi^\alpha = 0$. This observation underlies the approach of Baulieu and Singer [5.46] and Brooks, Montano, and Sonnenschein [5.47], which we will discuss shortly.

The fact that the gauge constraint $F_{\alpha\beta}^+ = 0$ arises, suggests that (anti-)instantons indeed play, as required, an important role in the theory. This is brought out more clearly by the following observations:

If we examine the χ and η equations of motion (in the small coupling limit)

$$D_\alpha \psi_\beta - D_\beta \psi_\alpha + \epsilon_{\alpha\beta\gamma\delta} D^\gamma \psi^\delta = 0 \; , \qquad D_\alpha \psi^\alpha = 0 \; , \tag{5.50}$$

we find that these are precisely the equations (5.22) we have encountered above in our discussion of deformations of instanton moduli. Thus the zero modes of ψ are (co)tangent vectors to the instanton moduli space. Furthermore, one sees that the absolute minima of the action S' are the (anti-)instanton configurations, which are thus also the vacua of this theory. These are BRST invariant because of $\delta\chi_{\alpha\beta} = F_{\alpha\beta}^+$. Thus we expect that in a weak coupling expansion (which is legitimate in view of the arguments of section 2) the ψ integration reduces to an integral over the moduli space. We will return to these matters in our discussion of observables below.

In light of the preceding discussion it is now conceivable that there are (at least) two approaches to constructing the action (5.48). One [5.46, 5.47] is to regard the instanton equation as a gauge fixing condition associated with the BRST-like shift symmetry $\delta A_\alpha = \psi_\alpha$, eq. (5.45), while in the other (pioneered by Labastida and Pernici [5.48]) it arises as a *classical* equation of motion of a bosonic action. We will now discuss in turn these constructions, which are the analogues of those given for supersymmetric quantum mechanics and the topological sigma model in sections 3 and 4. Historically, however, the case of Yang–Mills was treated first and suggested the application to other models. In section 5.3 we will pursue yet another approach—the most transparent from the geometric point of view: there we start off with a non-trivial *classical* action with a BRST-like supersymmetry [5.49].

5.2.2. The approach of Baulieu–Singer and Brooks–Montano–Sonnenschein

Motivated by the desire to interpret the symmetry $\delta A_\alpha = \psi_\alpha$ as the BRST version of the topological shift symmetry $A_\alpha \to A_\alpha + \epsilon_\alpha$, one is led to look for a classical action which is invariant under such a large local symmetry. Two obvious candidates are zero and the Pontrjagin number $\int F\tilde{F}$. Taking the latter as a starting point we now wish to quantize the theory. Since the action is a constant number which does not provide a good measure for the path integral, it requires some rethinking as to what one means by quantizing this theory. In this section we will, however, proceed naively, while we will address some of the technical and conceptual problems inherent in this approach below.

In order to quantize the theory we have to expose the full symmetry of the action, which is

$$A_\alpha \to A_\alpha + \epsilon_\alpha + D_\alpha \epsilon \; . \tag{5.51}$$

This description of the symmetry [5.46] is redundant, since the ordinary Yang–Mills gauge transformation part $D_\alpha \epsilon$ can be absorbed into the shift ϵ_α by a field redefinition. This reducibility of the symmetry will—according to the general prescription of [5.50]—lead to cubic ghost terms like those appearing in Witten's action (5.44). Indeed, the above transformation law—which keeps the ordinary gauge symmetry separate from the shift symmetry—will lead directly to Witten's action in a form where the remaining Yang–Mills symmetry has also been gauge fixed.

At this point we can anticipate the field content of the gauge fixed theory. Firstly we have the usual Yang–Mills triplet $(c, \bar{c}, b)$ consisting of the ghost, anti-ghost and multiplier fields needed to enforce the gauge constraint $\partial.A = 0$ [or its background version $d_{A_0} * (A - A_0) = 0$]. Analogously we introduce the set $(\psi_\alpha, \chi_{\alpha\beta}, B_{\alpha\beta})$ (χ and B are self-dual), which allows us to (partly!) gauge fix the shift symmetry by imposing the instanton equation $F^+_{\alpha\beta} = 0$ as the gauge constraint, as suggested in the discussion following (5.49). However, the reducibility mentioned above implies that ψ has its own gauge invariance. We thus have to introduce one further triplet of scalars $(\phi, \bar{\phi}, \eta)$ with ghost numbers $(2, -2, -1)$, respectively. The appearance of the Grassmann even ghost-for-ghost ϕ is characteristic of a first order reducible gauge symmetry [5.50], and η is the multiplier for the gauge fixing condition $D_\alpha \psi^\alpha = 0$ on ψ.

The complete set of off-shell nilpotent BRST transformations is then

$$\delta A_\alpha = D_\alpha c + \psi_\alpha \ , \qquad \delta\psi_\alpha = -[c, \psi_\alpha] - D_\alpha \phi \ , \qquad \delta\phi = -[c, \phi] \ ,$$
$$\delta\chi_{\alpha\beta} = B_{\alpha\beta} \ , \qquad \delta B_{\alpha\beta} = 0 \ , \qquad \delta\bar{\phi} = \eta \ , \qquad \delta\eta = 0 \ ,$$
$$\delta c = -\tfrac{1}{2}[c, c] + \phi \ , \qquad \delta\bar{c} = b \ , \qquad \delta b = 0 \ . \tag{5.52}$$

The structure of the transformations of the geometrical (A, ψ, ϕ, c) sector—for instance the appearance of ϕ in the c transformation law—finds a natural explanation within the framework of the universal bundle with connection of Atiyah and Singer [5.35]. This will be discussed in section 5.3.

We are now in a position to write down the complete quantum action,

$$S = -\frac{1}{4} \int dx \, F_{\alpha\beta} \tilde{F}^{\alpha\beta} + \{Q, V\} \ , \tag{5.53}$$

where

$$V = \int dx \, \chi^{\alpha\beta} (F^+_{\alpha\beta} - \tfrac{1}{2}\alpha B^{\alpha\beta}) + \bar{\phi}(D_\alpha \psi^\alpha - \tfrac{1}{2}\beta\eta) + \bar{c}(\partial.A - \tfrac{1}{2}\gamma b) \ . \tag{5.54}$$

Upon choosing the gauge parameters α, β, γ to have the values 1, 0, 0 and integrating out the field $B_{\alpha\beta}$ we arrive at Witten's action (5.44) supplemented by the Yang–Mills gauge fixing terms (up to simple field redefinitions, e.g., to generate the cubic term $\phi[\chi, \chi]$, shift B to $B - [c, \chi]$). Note that we are also free to choose the gauge $\alpha = 0$, in which the B integration enforces the delta function constraint $F^+_{\alpha\beta} = 0$, reducing the A integral to one over anti-self-dual configurations. This delta function gauge, which we are already familiar with from section 3, will play an important role in our discussion of renormalization of topological field theories in section 8. We will also make use of it in sections 5.3 and 5.4 to construct other topological gauge theories.

A variant of the above derivation was discovered independently by Brooks, Montano, and Sonnenschein [5.47]. They also began with $\int F\tilde{F}$ as the classical action, the gauge symmetry being (5.51) without the Yang–Mills part $D_\alpha \epsilon$. This necessitated a second stage of gauge fixing (due to the fact that the initial gauge fixed action had a residual "ghostly" local symmetry) leading to the action (5.48) and (5.49).

5.2.3. The Labastida–Pernici approach

The basic idea here is to regard the instanton equation $F^+_{\alpha\beta} = 0$ as arising from a suitable classical action, which in the case at hand is

$$S = \frac{1}{2} \int_M dx \, (G_{\alpha\beta} - F^+_{\alpha\beta})^2 \,, \tag{5.55}$$

where $G_{\alpha\beta}$ is an auxiliary self-dual field. The G equation of motion is $G_{\alpha\beta} - F^+_{\alpha\beta} = 0$, which is simply the Langevin equation for the system. As we will now show there is enough local symmetry to set $G_{\alpha\beta} = 0$, thereby recovering the instanton equation.

We see that (5.55) is invariant under the transformations

$$\delta A_\alpha = D_\alpha \epsilon + \epsilon_\alpha \,, \qquad \delta G_{\alpha\beta} = D_{[\alpha}\epsilon_{\beta]} + \tfrac{1}{2}\epsilon_{\alpha\beta\gamma\delta}D^\gamma\epsilon^\delta - [\epsilon, G_{\alpha\beta}] \,. \tag{5.56}$$

The important point to notice here is that, when $\epsilon_\alpha = -D_\alpha\epsilon$, these transformations possess the on-shell redundancy

$$\delta A_\alpha = 0 \,, \qquad \delta G_{\alpha\beta} = [G_{\alpha\beta} - F^+_{\alpha\beta}, \epsilon]|_{\text{on-shell}} = 0 \,. \tag{5.57}$$

Quantizing the theory with this on-shell reducibility requires us to make use of the Batalin–Vilkovisky procedure; this is straightforward and the result is that the quantum action is that of Witten (5.48), with the Yang–Mills gauge symmetry also gauge fixed, or—equivalently—that of Baulieu and Singer described in the previous section. Relevant steps leading to this result are explained in appendix A.

5.2.4. Other approaches

There is yet another way to understand the origin of the action (5.44). The motivation here is to obtain the (scalar) BRST supercharge by "twisting" a set of conventional (spinorial) supercharges.

We are of course free to add any BRST exact term to the action provided that it respects gauge invariance and power counting renormalizability. Adding $\{Q, \eta[\phi, \bar\phi]\}$ we obtain an action which bears a formal similarity to that of usual $N = 2$ supersymmetric Yang–Mills theory (on $\mathbb{R}^4$) [5.51].

This resemblance can be made more precise as follows: the rotation group of $\mathbb{R}^4$ is locally $SU(2)_L \times SU(2)_R$, while the global internal symmetry group of $N = 2$ super-Yang–Mills is $SU(2)_I \times U(1)$. Replacing $SU(2)_R$ by $SU(2)_{R'} = \mathrm{diag}(SU(2)_R \times SU(2)_I)$ the supercharges—which originally transform as $(\tfrac{1}{2}, 0, \tfrac{1}{2}, -1) \oplus (0, \tfrac{1}{2}, \tfrac{1}{2}, 1)$ under $SU(2)_L \times SU(2)_R \times SU(2)_I \times U(1)$—now transform as $(\tfrac{1}{2}, \tfrac{1}{2}, -1) \oplus (0, 0, 1) \oplus (0, 1, 1)$ under $SU(2)_L \times SU(2)_{R'} \times U(1)$. In this way we have obtained a scalar (singlet) supercharge which we identify with the BRST charge Q, with $U(1)$ labeling the ghost number. It is crucial to realize that—as a consequence of the scalar nature of the BRST charge—the resulting theory is supersymmetric not just on $\mathbb{R}^4$ but on an arbitrary four-manifold.

The above procedure can obviously be applied to other $N \geq 2, d = 4$ supersymmetric theories, and some examples have been worked out by Yamron [5.52] and Karlhede and Roček [5.53], while the general procedure in two dimensions—based on a clever modification of the original energy–momentum tensor—has more recently been explained by Witten [5.54] in his discussion of topological 2D gravity (cf. also refs. [5.55, 5.56]).

Finally we mention the beautiful interpretation of Atiyah and Jeffrey [5.37]. The latter we shall describe briefly below (section 5.2.6), although lack of space will not permit us to develop the required mathematics (of equivariant cohomology, based on ref. [5.57]) to the extent required for a full appreciation of the elegance of their approach.

5.2.5. Evaluation of the partition function

We will now—after having investigated the origin of the action of Donaldson theory from several different points of view—turn to some applications and explicit computations. Let us recall from the general discussion in section 2, that in theories of Witten type the partition function and certain correlation functions are topological invariants.

If there are fermionic zero modes the partition function will—as in the toy-model of supersymmetric quantum mechanics in section 3—be zero, and this will lead us back to the issue of observables below. Here we shall assume that the moduli space consists of isolated instantons. In that case the partition function Z will in general be non-zero and coincides—as we will now show—with the first Donaldson invariant described at the end of section 5.1.5. We will now present two ways of evaluating the partition function.

As a consequence of the coupling constant independence of Z, we can compute it in the weak coupling (semi-classical) limit. We can therefore write it as a sum of contributions from the neighborhoods of the isolated instantons. The calculation is performed in section 8 in the context of renormalization of Donaldson theory, and the result is that the contribution from one isolated instanton is $Pf(D_A)/\det^{1/2} D_A^* D_A$ where $Pf(D_A)$ is the Pfaffian of the real skew-symmetric deformation operator D_A, eq. (5.23), of the instanton deformation complex (5.21) acting on the fields (χ, η, ψ). Up to a sign, this ratio is of course 1 (as it should be by supersymmetry), and it thus only remains to determine the relative signs of the contributions from different instantons. Choosing one and declaring its contribution to be $+1$, we can determine the sign of any other contribution by studying the spectral flow along a curve in $\mathcal{A}$ connecting the two instantons. In view of our considerations in section 5.1.5, this is precisely the prescription for comparing the relative orientations of these isolated instantons. The final result,

$$Z(M) = \sum_{\text{instantons}} \pm 1 \ , \tag{5.58}$$

therefore coincides precisely with what we called the first Donaldson invariant, provided that we declare the choice $+1$ for our reference instanton to fix the overall orientation of $\mathcal{M}$. The above procedure is consistent (i.e. independent of the path chosen) since $\mathcal{M}$ is orientable. In field theoretic terms this is equivalent to the absence of a global anomaly in Donaldson theory.

An alternative demonstration of (5.58) is based on the observation that the Langevin equation $G_{\alpha\beta} - F_{\alpha\beta}^+ = 0$ defines a *complete* Nicolai map for the theory [5.58, 5.59].

As we are considering isolated instantons the terms $\bar{\phi}[\psi_\alpha, \psi^\alpha]$ and $\phi[\chi_{\alpha\beta}, \chi^{\alpha\beta}]$ in the action may be ignored. One way of seeing this is to assign the charges $(a, -a, b, -b, -b)$ to the fields $(\phi, \bar{\phi}, \psi, \chi, \eta)$. In this way the action is chargeless except for the cubic terms; expanding the path integral in these and noting that due to the absence of fermionic zero modes the measure is chargeless we see that only the zero order term contributes.

Now let us define a map

$$\xi(A) = F_A^+ \ , \qquad \lambda(A) = d_{A_0} * (A - A_0) \ , \tag{5.59}$$

around each isolated instanton A_0. The Jacobian of this map matches the inverse of the Pfaffian of the (ψ, χ, η) system up to sign if we impose the same background gauge fixing on ψ. Then the $\bar{c}-c$ ghost kinetic term also cancels against that of the $\bar{\phi}-\phi$ system. The sign obviously has the same source as that in (5.58), and since the remaining integral over ξ is just a Gaussian we recover the previous result.

On a four-manifold of the form $M = Y \times \mathbb{R}$ another natural choice is the temporal gauge $A_0 = 0$. The change of variables

$$\xi_{0i}^+ = \mathrm{d}A_i/\mathrm{d}t + (*F_A)_i \ , \qquad \lambda(A) = A_0 \ , \tag{5.60}$$

trivializes the partition function in that case (i.e., reduces it to a sum of contributions from the instantons). The zeros of this map are precisely the solutions to the gradient equation (5.40) for the Chern–Simons functional we discussed in connection with the relation between Floer's and Donaldson's work in section 5.1.6.

5.2.6. The Atiyah–Jeffrey interpretation

The above eq. (5.58), expressing the topological invariant $Z(M)$ as a sum of ± 1's, is reminiscent of similar formulae in differential geometry, expressing, e.g., the Euler character of a manifold in terms of the signed sum of zeros of a vector field. That there is more, indeed much more, to this analogy, has been shown by Atiyah and Jeffrey [5.37]. Using a formalism developed by Mathai and Quillen [5.57] they have not only identified (5.58) as the Euler number (character) of a vector bundle over the space $\mathcal{A}/\mathcal{G}$ of gauge equivalence classes of connections (cf. section 5.1.3); they have moreover been able to reproduce Witten's action (5.48), (5.49) term by term from purely differential geometric considerations.

Explaining the latter would unfortunately lead us too far astray, and in the following we will therefore explain only the first assertion. It will nevertheless be necessary to digress briefly on an observation made in ref. [5.57] concerning integral expressions for the Euler number. This digression will reveal a close resemblance between the Mathai–Quillen formalism and that of supersymmetric quantum mechanics. And although we will not go into this in any detail, it may be helpful in the following to keep in mind section 3.8, where we discussed the Euler number from the quantum mechanics point of view.

We start with some classical material (for details cf. ref. [5.60]). Recall that an oriented $2m$-dimensional real vector bundle E over a manifold X has an Euler class $e(E) \in H^{2m}(X, \mathbb{Z})$. If $\dim X = 2m$, this class can be evaluated on (the fundamental class $[X]$ of) X to give the Euler number $\chi(E) = e(E)[X]$. In particular, if $E = TX$, the tangent bundle of X, $\chi(TX) \equiv \chi(X)$ is the Euler number of X. There are two concrete ways of thinking about $\chi(E)$. On the one hand, the Gauss–Bonnet–Chern theorem provides one with an explicit differential form representative $e_\nabla(E)$ of $e(E)$ constructed from the curvature Ω of a connection ∇ on E, such that

$$\chi(E) = \int_X e_\nabla(E) \ . \tag{5.61}$$

On the other hand, $\chi(E)$ can be computed as the number of zeros of a generic section s of E (counted with signs),

$$\chi(E) = \sum_{x:s(x)=0} \pm 1 \ . \tag{5.62}$$

A more general formula,

$$\chi(E) = \int_X e_{s,\nabla}(E) \ , \tag{5.63}$$

obtained by Mathai and Quillen, interpolates between the two quite different descriptions (5.61) and (5.62). Here $e_{s,\nabla}$ is a closed $2m$-form on X, depending on both a section s and a connection

∇, with the following properties: if s is the zero section of E, then $e_{s,\nabla} = e_\nabla$ and (5.63) reduces to (5.61); if one replaces s by ts, with $t \in \mathbb{R}$, and evaluates (5.63) in the limit $t \to \infty$ using the stationary phase approximation, (5.62) is reproduced. Moreover $e_{s,\nabla} \equiv e_s$ (we will suppress the dependence on the connection ∇ in the following) is the pull-back to X via s of a closed form U on the total space E of the vector bundle, $e_s = s^*U$. U is a representative of the Thom class [5.60] of E but, unlike the classical Thom class, which has compact support in the fiber directions, U is Gaussian shaped along the fibers [cf. eq. (5.64) below].

At this point it will be necessary to introduce some more notation: we let ξ^a denote fiber coordinates of E, χ_a corresponding Grassmann odd variables, and Ω^{ab} the curvature two-form of E. Then the Mathai–Quillen form U can be written as a fermionic integral over the χ's,

$$U = N\,\mathrm{e}^{-\xi^2} \int \mathrm{d}\chi\ \mathrm{e}^{\chi_a \Omega^{ab}\chi_b/4 + \mathrm{id}\xi^a\chi_a} \tag{5.64}$$

(N is a normalization factor). s^*U is obtained from U simply by replacing ξ by $s(x)$. We see that, if we take s to be the zero section, (5.64) coincides with the Gauss–Bonnet integral expression (3.126) for $X = M$,

$$N' \int \mathrm{d}\bar{\psi}\,\mathrm{d}\psi\ \mathrm{e}^{-R_{ijkl}\bar{\psi}^i\bar{\psi}^j\psi^k\psi^l/4}$$

($\Omega^{ab} = \Omega^{ab}{}_{ij}\psi^i\psi^j$), derived in section 3.8, and therefore with (5.61), provided that we convert the space–time indices on the $\bar{\psi}$'s to internal indices ($\bar{\psi}^i = e^i_a\chi^a$), and remember that the ψ integral serves to pick out the top form part of that expression, something that is implicit in (5.64).

In finite dimensions (5.64) may perhaps be regarded as an unnecessary complication, since one has the simple classical formula (5.61) at one's disposal. But, as Atiyah and Jeffrey have pointed out, (5.64) is of a definite advantage when dealing with infinite dimensional bundles, where (5.61) is not terribly well defined, but where it may be possible to give a meaning to (5.63) for a suitable choice of section s. (5.63) can then be regarded as defining a regularized Euler number $\chi_s(E)$, which is, however, no longer necessarily independent of s. If s is a section canonically associated with E, $\chi_s(E)$ may nevertheless carry interesting topological information.

As an example consider, instead of $X = M$, its loop space $X = LM = \{x(t) : S^1 \to M\}$. A natural section of the tangent bundle $T(LM)$ is $s(x)(t) = \dot{x}(t)$. With this choice of section the exponent in (5.64),

$$\xi^2 - \chi_a \Omega^{ab}\chi_b/4 - \mathrm{id}\xi^a\chi_a \tag{5.65}$$

(summation over the fiber indices now includes an integration over t) is precisely the action (3.1) of supersymmetric quantum mechanics without the potential V [the complete action can be obtained by choosing, instead of the above section, $s(x)(t) = \dot{x}(t) + V'(x(t))$]. We conclude that the regularized Euler number $\chi_s(LM)$ of the loop space is equal to the partition function of supersymmetric quantum mechanics, i.e.,

$$\chi_s(LM) = \int_{LM} s^*U = \chi(M)\ .$$

This has admittedly been somewhat sketchy, and we will come back to these matters (and related issues, cf. the remarks at the end of section 5.4.4) in a future publication [5.61].

We now return to Donaldson theory, and our aim is to show that its action can also be brought into the form (5.65) (up to relative numerical factors; these can be reconciled by simple rescalings

of the fields). This is the (less elegant but simpler) converse of what has been done by Atiyah and Jeffrey, who derived the action from (5.64) and (5.65). The action we will use is (5.44) with the topological term $\frac{1}{4} \int_M dx \, F_{\alpha\beta} \tilde{F}^{\alpha\beta}$ added, i.e.,

$$S = \int\limits_M dx \{ \tfrac{1}{2}(F_{\alpha\beta}^+)^2 - 2\chi_{\alpha\beta} D^\alpha \psi^\beta + \eta D_\alpha \psi^\alpha + \bar{\phi}[\psi_\alpha, \psi^\alpha] - \bar{\phi} D_\alpha D^\alpha \phi - \tfrac{1}{2}\phi[\chi_{\alpha\beta}, \chi^{\alpha\beta}] \} \ . \quad (5.66)$$

Since we are working equivariantly, the A integration is understood to be over the space $X = \mathcal{A}/\mathcal{G}$, while the η integral forces ψ to be (co)tangent to X. The $\bar{\phi}$ integral leads to the delta function constraint

$$\phi = (D_\alpha D^\alpha)^{-1}[\psi_\beta, \psi^\beta] \ , \quad (5.67)$$

which incidentally shows that the vacuum expectation value $\langle \phi \rangle$ of ϕ is also given by the above expression. We will need this result for our discussion of observables in Donaldson theory in section 5.2.8. Plugging (5.67) back into (5.66) one arrives at

$$\tilde{S} = \int\limits_M dx \, \tfrac{1}{2}(F_{\alpha\beta}^+)^2 - 2\chi_{\alpha\beta} D^\alpha \psi^\beta + \tfrac{1}{2}[\chi_{\alpha\beta}, (D_\gamma D^\gamma)^{-1}[\psi_\delta, \psi^\delta]]\chi^{\alpha\beta} \ . \quad (5.68)$$

Modulo rescalings of the fields this is already of the desired form (5.65), with $X = \mathcal{A}/\mathcal{G}$. Comparison of χ_a with $\chi_{\alpha\beta}(x)$ shows that the standard fiber of the vector bundle $\mathcal{E}$ in question is the infinite dimensional vector space $\Omega_+^2(M, \mathbf{g})$ of self-dual two-forms. $\mathcal{E}$ has the canonical section $s(A_\alpha) = F_{\alpha\beta}^+$ giving rise to the first two terms of $\tilde{S}$, since $ds(A_\alpha) = (D_\alpha \psi_\beta)^+$. It remains to identify the curvature term in (5.68) in order to determine $\mathcal{E}$. In section 5.3.1 we will show that $(D_\alpha D^\alpha)^{-1}[\psi_\beta, \psi^\beta]$ (with $D_\alpha \psi^\alpha = 0$) is a curvature form on the principal $\mathcal{G}$-bundle $\mathcal{A} \to \mathcal{A}/\mathcal{G}$. Thus the third term in (5.68) is a curvature form of the vector bundle $\mathcal{E} = (\mathcal{A} \times \Omega_+^2(M, \mathbf{g}))/\mathcal{G}$ associated to $\mathcal{A}$ through the adjoint action of $\mathcal{G}$ on $\Omega_+^2(M, \mathbf{g})$. Thus putting all this together we have derived the result (obtained in ref. [5.37]) that the partition function of Donaldson theory (the first Donaldson invariant) can be interpreted as the Euler number of the vector bundle $\mathcal{E}$,

$$Z(M) = \chi_s(\mathcal{E}) \ . \quad (5.69)$$

5.2.7. Construction of observables

If the dimension of the moduli space $\mathcal{M}$ is non-zero, there will be non-trivial solutions to the deformation eqs. (5.22). Since—as pointed out above—these are precisely the equations (5.50) that ψ has to satisfy, we will have to soak up these fermionic zero modes in such a way that the topological properties of the theory are preserved. We have already seen the general strategy (discussed in section 2) at work in the topological sigma model (section 4.5).

Zero modes of the fields $\chi, \eta, \phi, \bar{\phi}$ will be present, in addition to the above ψ (and corresponding number of A) zero modes, if there are non-trivial solutions to the equations

$$D_\alpha \chi^{\alpha\beta} = 0 \ , \qquad D_\alpha \phi = D_\alpha \bar{\phi} = D_\alpha \eta = 0 \ .$$

Written in differential form notation,

$$d_A^* \chi = 0 \ , \qquad d_A \phi = d_A \bar{\phi} = d_A \eta = 0 \ ,$$

we recognize these as the equations defining $\mathbf{H}_A^2$ and $\mathbf{H}_A^0$, respectively, eqs. (5.25) and (5.26), while the equations obeyed by ψ, $(d_A \psi)^+ = d_A^* \psi = 0$, are of course precisely those characterizing $\mathbf{H}_A^1$.

Noticing that χ and η have opposite (-1) ghost number to ψ, we see that the formal dimension of $\mathcal{M}$ [the index (5.24) of the deformation complex (5.21)], $d(\mathcal{M}) = h^1 - h^0 - h^2$, is equal to the *net* ghost number violation of Donaldson theory. In analogy with 't Hooft's treatment of instantons [5.62], we thus expect to have to insert $d(\mathcal{M})$ fermionic ghost zero modes into the path integral (compensating for the non-invariance of the naive measure under the ghost number symmetry) in order to get a non-zero result.

If $\mathcal{M}$ is smooth [and whence $d(\mathcal{M})$ equals the actual dimension of $\mathcal{M}$] this procedure has the obvious and attractive interpretation of turning the scalar Z into a $d(\mathcal{M})$ form which—as a volume form—can be integrated over $\mathcal{M}$ (this will be explained in more detail below). It is this case (where $\mathbf{H}^0 = \mathbf{H}^2 = 0$ and where there will be neither χ zero modes nor reducible connections to worry about) which we shall consider in the following. This simplifying assumption can be justified to a certain extent by recalling from section 5.1 that [for $SU(2)$] $\mathbf{H}^2$ is zero for a "generic" choice of metric, and that reducible instantons [which are in any case isolated for $SU(2)$] do not exist for an open dense set of metrics if $b_2^+(M) > 0$ (this also being the relevant regime for the Donaldson polynomials). We are then only left with the A and ψ zero modes.

If the dimension of G is greater than three, reducible connections cannot so easily be avoided and lead to serious mathematical problems since the singularity structure of $\mathcal{M}$ becomes much more intricate. At present it is therefore not known to what extent Donaldson's work can be generalized in that direction. And although the field theory point of view may offer some insights into this question, further input from mathematics seems to be required, to learn which way of handling the zero modes associated with reducible connections corresponds to the way the singularities of $\mathcal{M}$ are dealt with on the mathematical side.

Let us now return to our discussion of observables. Recall that these are BRST equivalence classes of gauge invariant and metric independent functionals of the fields. In order to construct such functionals and to discuss their properties, it will be convenient to resort to a differential form notation (see appendix B for details), in which the BRST transformations in the geometrical sector read

$$\delta A = \psi - d_A c \ , \qquad \delta \psi = -[c, \psi] - d_A \phi \ ,$$
$$\delta c = -\tfrac{1}{2}[c, c] + \phi \ , \qquad \delta \phi = -[c, \phi] \ , \qquad \delta^2 = 0 \ . \tag{5.70}$$

These equations imply

$$(d + \delta)\,\mathrm{tr}(F_A + \psi + \phi)^n = 0 \tag{5.71}$$

as a consequence of the "Bianchi identity"

$$(d + \delta)(F_A + \psi + \phi) + [A + c, F_A + \psi + \phi] = 0 \ , \tag{5.72}$$

whose geometrical origin will become clear within the framework of section 5.3, and on which we will (for $n = 2$) base our construction of observables leading to the Donaldson polynomials (other observables, related to 2D Yang–Mills theory, have been proposed in ref. [5.63]). In principle it is of course possible to consider the equations with $n > 2$ as well, which may lead to new results in the case of higher rank gauge groups. Their mathematical relevance has, however, not yet been established, due to the problems with higher rank gauge groups alluded to above.

Writing

$$\tfrac{1}{2}\,\mathrm{tr}(F_A + \psi + \phi)^2 = \sum_{i=0}^{4} W_i \ , \tag{5.73}$$

where the W_i are are i-forms on M with ghost number $4 - i$ given by

$$W_0 = \tfrac{1}{2}\operatorname{tr}(\phi^2) \;, \qquad W_1 = \operatorname{tr}(\psi\phi) \;, \qquad W_2 = \operatorname{tr}(F_A\phi + \tfrac{1}{2}\psi\psi) \;,$$
$$W_3 = \operatorname{tr}(F_A\psi) \;, \qquad W_4 = \tfrac{1}{2}\operatorname{tr}(F_A^2) \;, \tag{5.74}$$

we can expand (5.71) in terms of ghost number and form degree as

$$\delta W_0 = 0 \;, \qquad dW_0 + \delta W_1 = 0 \;, \qquad dW_1 + \delta W_2 = 0 \;,$$
$$dW_2 + \delta W_3 = 0 \;, \qquad dW_3 + \delta W_4 = 0 \;, \qquad dW_4 = 0 \;. \tag{5.75}$$

Thus picking a k-homology cycle γ on M we can construct a functional

$$W_k(\gamma) = \int_\gamma W_k \;, \tag{5.76}$$

which is clearly metric independent and gauge invariant. As a consequence of $\delta W_k = -dW_{k-1}$ it is BRST-closed,

$$\delta W_k(\gamma) = -\int_\gamma dW_{k-1} = -\int_{\partial\gamma} W_{k-1} = 0 \;,$$

and therefore an observable. Moreover, it is topological in the sense that its BRST cohomology class only depends on the homology class of γ, since

$$W_k(\partial\beta) = \int_\beta dW_k = -\delta \int_\beta W_{k+1} \;.$$

This implies, for instance, that the ghost number 4 observable $W_0(P)$ is independent of the chosen point $P \in M$ in a connected component of M, as it should be, since in a topological theory individual points have no intrinsic meaning.

Actually, the transformations (5.70) appear to imply that all the observables constructed above are not only BRST-closed but actually BRST-exact. For instance, $\operatorname{tr}\phi^2$ can be written as $\{Q, \operatorname{tr}(\phi c - 1/3c^3)\}$. This and related observations had given rise to some controversy in the literature [5.64–5.66] (a good discussion of this can be found in ref. [5.67]), but the situation is now well understood, and there are several ways of establishing the non-triviality of these observables and their counterparts in other topological field theories. We will explain these in section 5.3.3. Suffice it to say here that the above relation says as much (or little) about the triviality of W_0 as the (formally identical) relation $\operatorname{tr}F_A^2 = d\operatorname{tr}(AF_A - 1/3A^3)$ says about the triviality of the Pontrjagin number $\int \operatorname{tr}F_A^2$.

Accepting the non-triviality of the above observables, we have thus found an assignment of k-homology classes of M to BRST equivalence classes of observables with ghost number $4 - k$. Thus, for any given (formal) dimension $n = d(\mathcal{M})$ of $\mathcal{M}$, we can now define new topological invariants $Z(\gamma_1, \gamma_2, \ldots, \gamma_r)$ by choosing k_i-homology cycles γ_i such that the superselection rule

$$\sum_{i=1}^r (4 - k_i) = n \tag{5.77}$$

is satisfied , and setting

$$Z(\gamma_1, \gamma_2, \ldots, \gamma_r) = \left\langle \prod_{i=1}^{r} W_{k_i}(\gamma_i) \right\rangle \; . \tag{5.78}$$

5.2.8. Observables as differential forms on moduli space

Our aim now is to show that these correlation functions reduce to integrals of products of closed differential forms on $\mathcal{M}$. More precisely, we will be able to assign closed differential forms $\hat{W}_k(\gamma)$ on $\mathcal{M}$ to the $W_k(\gamma)$ in such a way that

$$Z(\gamma_1, \ldots, \gamma_r) = \int_{\mathcal{M}} \hat{W}_{k_1}(\gamma_1) \cdots \hat{W}_{k_r}(\gamma_r) \; . \tag{5.79}$$

This form of the correlation functions is already quite reminiscent of the equation defining the Donaldson invariant $q_k([\gamma_1], \ldots, [\gamma_r])$, and what remains to be shown then is that the cohomology class of $\hat{W}_k(\gamma)$ can be identified with $\mu([\gamma])$.

The key property responsible for allowing us to go from (5.78) to (5.79) is again the coupling constant independence of (5.78). As in the case of the partition function Z we can therefore compute these correlation functions exactly in a weak coupling (or semi-classical) limit. Alternatively, we could of course use the delta function gauge directly to express (5.78) as an integral over $\mathcal{M}$.

Assuming for simplicity, as in our discussion above, that the only zero modes are those of A and its superpartner ψ, the steps leading to (5.79) are the following:

– one integrates out all the non-zero modes to obtain as the remaining measure $\int da^1 \cdots da^n \times d\psi^1 \cdots d\psi^n$, where the a^i are coordinates on $\mathcal{M}$; this measure is canonical since the da^i transform inversely to the $d\psi^i$;

– in order to get a non-zero result one needs to insert an observable $\mathcal{O}$ of ghost number n into the path integral, which—upon "integrating out" the non-zero modes—reduces to

$$\mathcal{O}' = \frac{1}{n!} \mathcal{O}'_{i_1 \cdots i_n}(a^k) \psi^{i_1} \cdots \psi^{i_n} \; ;$$

we will be more precise about what this amounts to in practice below;

– one then arrives at

$$\langle \mathcal{O} \rangle = \int da^1 \cdots da^n \, d\psi^1 \cdots d\psi^n \mathcal{O}' \int \mathcal{O}'_{12 \cdots n}(a^k) \, da^1 \cdots da^n \equiv \int_{\mathcal{M}} \hat{\mathcal{O}} \; ;$$

– the final step involved in proving a formula like (5.79) is to show that if an observable is of the form $\mathcal{O} = \mathcal{O}_1 \mathcal{O}_2 \cdots \mathcal{O}_k$ its associated differential form is

$$\hat{\mathcal{O}} = \hat{\mathcal{O}}_1 \hat{\mathcal{O}}_2 \cdots \hat{\mathcal{O}}_k \; , \tag{5.80}$$

at least to lowest order in e^2 and modulo Q-exact terms.

Since on $\mathcal{M}$, δ reduces to the exterior derivative (for a detailed explanation of this point cf. section 5.3.2), the above procedure assigns closed differential forms on $\mathcal{M}$ to observables. In the cases of interest to us $[\mathcal{O} = W_k(\gamma)]$ this assignment moreover descends to a map $\nu : H_k(M) \to H^{4-k}(\mathcal{M})$ since the arguments of the previous section show that the cohomology class of $\hat{W}_k(\gamma)$ only depends on the homology class of γ.

Explicitly, the $\hat{W}_k(\gamma)$ are obtained from the $W_k(\gamma)$, which are functionals of F_A, ψ and ϕ, by replacing

- F_A by its instanton value,
- ψ by its zero mode part,
- ϕ by the zero mode part of its VEV (5.67),

$$\langle \phi \rangle = -G_A[\psi, *\psi] \ , \tag{5.81}$$

where G_A is the Green's function for the operator $(d_A * d_A)$. Equation (5.81) can also be derived directly from the superpartner of the classical equation of motion $d_A * \psi = 0$ (section 5.3.2). Note also that (5.81) expresses the ghost number 2 field ϕ explicitly as a two-form on $\mathcal{M}$—in accordance with our general identification of ghost number and form degree. Furthermore the product relation (5.80) holds by virtue of the fact that all the Wick contractions involved in the difference between the left and right hand side lead (on purely dimensional grounds) to higher powers of e^2. This completes the derivation of (5.79).

In section 5.3.2 we will show in detail that the zero mode part of $F_A + \psi + \phi$ is the curvature of the universal connection on $\mathcal{Q}$. Whence the map ν is equal to Donaldson's map μ, eq. (5.34), by the definitions (5.73) and (5.76) and the slant product description of μ. Thus the field theoretic approach of this section has precisely reproduced the construction of the Donaldson polynomials explained in section 5.1.5. For the sake of comparison with section 5.1.5 let us note that Donaldson only considers homology two-cycles in M [for these the superselection rule (5.77) reduces to $2r = d(\mathcal{M})$]. If, as in section 5.1.5, M is simply connected, there are no non-trivial one- and three-cycles, and four-cycles just give rise to constant functions on $\mathcal{M}$. However, Donaldson's techniques can certainly be extended to include the zero-cycles (points) of M leading to additional invariants involving the generator of dimension 4 in the cohomology of C^*.

Although we have derived these equations for the Donaldson polynomials in a non-rigorous field theoretic way, their requisite properties can now be checked by completely classical arguments. The computations establishing that

- the $\hat{W}(\gamma)$ are closed,
- their cohomology class does not depend on the representative of the homology class $[\gamma]$ of γ,

can, for instance, be done in such a way as to mimic the quantum field theory argument at every step, the crucial difference, however, being that these are now finite-dimensional and whence—at least as long as one ignores questions related to the singularities of $\mathcal{M}$—perfectly rigorous manipulations. Of course these properties of the Donaldson polynomials also follow straightforwardly from their slant product definition and the descent equations (5.75) associated with the second Chern class of the universal bundle $\mathcal{Q}$. However, in the proof of the fact that

- they only vary within their cohomology class under variations of the metric,

the field theoretic approach does give some new insight and provides a formula that is not completely obvious from the point of view of section 5.1.5. In order to display this formula, let us consider a one-parameter family of metrics and the corresponding one-parameter family of moduli spaces, which we call the parametrized moduli space Z. Assume for simplicity that we can split the exterior derivative d_Z into the exterior derivative $d_{\mathcal{M}}$ on $\mathcal{M}$ and the exterior derivative d_g along the curve of metrics. We are interested in computing $d_g \hat{\mathcal{O}}$, where $\mathcal{O}$ is a p-form on $\mathcal{M}$. By our standard arguments we obtain, using the fact that $\mathcal{O}$ is metric independent,

$$d_g \hat{\mathcal{O}} = \widehat{\mathcal{O} d_g S} = \mathcal{O}\{\widehat{Q, d_g V}\} = d_{\mathcal{M}} \widehat{\mathcal{O} d_g V} \equiv d_{\mathcal{M}} T_{\mathcal{O}} \ , \tag{5.82}$$

which shows roughly that $\hat{\mathcal{O}}$ only varies within its cohomology class as the metric is varied. A better way of phrasing the above result is the following: decomposing forms on Z according to their

form degree on $\mathcal{M}$ and their form degree on the space of metrics, $\mathcal{O}$ is a $(p,0)$-form and $T_\mathcal{O}$ is a $(p-1,1)$-form. Then (5.82) shows that $\mathcal{O}$ extends to a d_Z-closed p-form $\mathcal{O}_Z$ on Z, in our case $\mathcal{O}_Z := \mathcal{O} - T_\mathcal{O}$, which restricts to $\mathcal{O} = \mathcal{O}(g)$ on each fiber $\mathcal{M} = \mathcal{M}(g)$; thus the field theoretic approach gives an explicit expression for the required extension term $T_\mathcal{O}$.

In the Poincaré dual situation considered by Donaldson [5.5], the analogous statement is that the intersections are one-dimensional manifolds with boundary, which give the desired homology between the (boundary) intersection points corresponding to two different metrics.

5.2.9. The Hamiltonian point of view

The purpose of this section is to recover the results on the relation between Floer homology and Donaldson theory anticipated in the non-relativistic treatment at the end of section 5.1.6. Since Witten's original work no significant (published) progress seems to have been made in the Hamiltonian approach (some aspects have been investigated in refs. [5.68–5.70]), and indeed the most fundamental problems (related to questions of unitarity and the presence of zero modes) in the rigorous construction of the Hilbert space of the theory still remain to be overcome. We shall therefore be content with indicating briefly how the Floer cohomology groups arise perturbatively as the theory's ground states in the simplest of all situations (when the flat connections are irreducible and isolated—this is also the situation considered by Floer and the only one in which Floer homology has so far been constructed rigorously). We will then show how the fact that Donaldson invariants on a manifold with boundary should be defined as taking values in the dual of the Floer homology groups of the boundary, arises quite naturally from the path integral point of view.

Before turning to these issues, let us comment on the existence of an anti-BRST operator $\overline{Q}$ in the Hamiltonian version of Donaldson theory. For the time being we assume that the four-manifold M is of the form $M = Y \times \mathbb{R}$. In this case the fundamental equation $T_{\alpha\beta} = \{Q, V_{\alpha\beta}\}$ implies the relation

$$H = \tfrac{1}{2}\{Q, \overline{Q}\} \,, \tag{5.83}$$

familiar from supersymmetric quantum mechanics (section 3.1), where $\overline{Q} = 2\int_Y V_{00}$. Nilpotency of $\overline{Q}$ [which corresponds to the second supercharge δ_t^* of Atiyah's Hamiltonian (5.41), (5.43)] can be established either by direct computation or—more elegantly but indirectly—by appealing to a time reversal symmetry possessed by the action (5.44) augmented by certain Q-exact higher order terms.

The usefulness of $\overline{Q}$ in the present context is unfortunately severely limited by the fact that the natural Hilbert space scalar product is not positive definite, which prevents us from making direct use of the powerful machinery of Hodge theory to investigate the physical sector of Donaldson theory. And while ref. [5.1] contains some suggestions on how to overcome this problem, we believe that to date no truly satisfactory answer has been found. We should, however, point out that this is an important point and that it is to be expected that a resolution of this problem will require, or lead to, major progress in our understanding of Floer homology. We will therefore make no further reference to $\overline{Q}$-cohomology in the following.

Let us now look at the ground states of the theory. For small coupling these are obtained by expanding around the classical minima of the potential. The Yang–Mills part of the action contributes a term proportional to $F_{ij}F^{ij}$, which tells us that classical minima are flat connections. The scalar contribution to the energy is proportional to $D_i\bar\phi D^i\phi$ and we thus require $D_i\phi = D_i\bar\phi = 0$, which is compatible with $F_{ij} = 0$ and—by our assumption that flat connections are irreducible—implies $\phi = \bar\phi = 0$. If the flat connections are isolated, there will be no A and ψ zero modes to

worry about and whence, setting all the other fields to zero, we will get a unique ground state for each flat connection.

In perturbation theory it is easy to see that all these states are BRST invariant, since an isolated flat connection is annihilated by Q according to the transformation laws (5.45), and in expanding around such a flat connection Q is quadratic in the oscillators and will therefore also annihilate the state.

The BRST invariance of these ground states would have followed from (5.83) if the Hilbert space scalar product were positive definite, since then (by standard arguments, cf. section 3.7)

$$H\Psi = 0 \Leftrightarrow Q\Psi = \overline{Q}\Psi = 0 \ . \tag{5.84}$$

In the present case the $\Rightarrow$ implication no longer holds automatically, and we therefore had to check by hand that the ground states are really BRST invariant.

The converse—that Floer homology classes have zero energy—follows from our general arguments of section 2: H is zero on physical states—the latter being defined as the cohomology vector space of the BRST operator acting on the (unconstrained) Hilbert space of the theory. The existence of a non-BRST invariant (i.e., non-physical) zero-energy state would thus have signalled a kind of spontaneous breakdown of BRST symmetry (which we have, however, ruled out perturbatively above).

In the above discussion we have assumed the absence of reducible flat connections, not because the generalization is straightforward, but, on the contrary, because it is simply not known at present how to deal with other situations. Let us comment briefly on the nature of the problems that arise in this context [5.71]. On the one hand, reducible flat connections lead to singularities of the moduli space (cf. sections 5.1.4 and 5.4.3), which in turn casts some doubt on the validity of the semi-classical approximation and the formal arguments establishing its exactness. On the other hand, reducible connections give rise to non-compact directions in the moduli space (space of solutions) of the theory, since $d_A\phi = 0$ is a linear equation for ϕ. Then the convergence of the various integrals over moduli space we have encountered in this section is no longer guaranteed, and further analysis is required. While certain ad hoc resolutions of these problems are conceivable, a satisfactory treatment of these matters is still lacking. Analogous remarks apply to all other topological gauge theories.

Let us now see how Floer cohomology groups make their appearance in Donaldson theory on a four-manifold M with boundary $\partial M = B$ [5.11]. In order to compute correlation functions $\langle \mathcal{O} \rangle$ of local operators [like the $W_k(\gamma)$] on such a manifold one needs to specify boundary conditions on B. Computing the partition function or a correlation function with that boundary condition, and varying the unspecified fields on B, one obtains a functional of the boundary values which is a state in the Hilbert space of the theory defined on $B \times \mathbb{R}$. Conversely any such state may be used to specify a boundary condition.

Upon integrating over the fields at the boundary one gets a number, and the question arises under which circumstances this number will be a topological invariant. If we are computing correlation functions of BRST invariant (and metric independent) operators, the by now standard arguments imply that all we have to require is BRST invariance of the boundary condition. In terms of Hilbert space states Ψ on B this translates into the requirement $Q\Psi = 0$, which implies that Ψ represents a Floer cohomology class. Likewise, had we chosen the boundary conditions in such a way that the corresponding state were Q-exact, the Q Ward identities would have told us that the correlation function vanished. Thus we have shown that they are independent of the representative of a Floer cohomology class.

Moreover, the fact that the observables $\mathcal{O}$ representing the Donaldson polynomials give us a

number $\langle \mathcal{O} \rangle_\Psi$ once we have specified a state $[\Psi] \in \mathrm{HF}^*(B)$ shows that on a manifold with boundary the Donaldson invariants $\langle \mathcal{O} \rangle$, regarded as maps from HF^* to the complex numbers, $\Psi \mapsto \langle \mathcal{O} \rangle_\Psi$, take values in the dual of the Floer cohomology groups. As mentioned in section 5.1.6, this observation of Donaldson—which was one of the main motivations behind Atiyah's suggestions leading to the construction of topological field theories—may lead to a powerful new method for calculating Donaldson invariants [5.11, 5.45].

5.3. Geometry of topological gauge theories

In the previous sections we have analyzed Donaldson theory from several different points of view. We now wish to obtain a clearer picture of the geometrical structures underlying this theory.

In section 5.1 we have already seen that the Donaldson polynomials can be constructed from the characteristic classes of the universal bundle with connection of Atiyah and Singer. We shall therefore start with a description of this bundle. This will (as had been noticed by a number of groups [5.46, 5.72, 5.73]) also allow us to understand the origin and geometrical significance of the BRST transformations of Donaldson theory. It will also suggest a number of ways of resolving the issue of "triviality of observables" encountered above. Moreover, the underlying geometry will turn out to be so general that it immediately provides us with topological gauge theories associated with arbitrary moduli spaces of connections [5.49, 5.74]. As an illustration of how this works in practice, we are going to construct topological gauge theories of flat and Yang–Mills connections in section 5.4. There we have also included a section on moduli spaces of flat connections, which we will make use of in our subsequent discussion of observables in these models.

5.3.1. The universal bundle

In ref. [5.35] Atiyah and Singer introduced a certain universal bundle with connection $(\mathcal{Q}, \hat{A})$ in order to compute characteristic classes of the index bundle of families of Dirac operators coupled to gauge fields. This bundle can be described as follows:

Let $P \xrightarrow{\pi} M$ be a principal G-bundle over M, $\mathcal{A}$ the affine space of connections on P [modeled on $\Omega^1(M, \mathbf{g})$] and $\mathcal{G}$ the group of vertical automorphisms (gauge transformations) of P. Then there is a natural action of $\mathcal{G}$ on $P \times \mathcal{A}$ which has no fixed points, and therefore $P \times \mathcal{A} \to (P \times \mathcal{A})/\mathcal{G} = \mathcal{Q}$ is a principal $\mathcal{G}$-bundle over $\mathcal{Q}$. Since the G action on $P \times \mathcal{A}$ commutes with that of $\mathcal{G}$, G acts on $\mathcal{Q}$. If one chooses either $\mathcal{A}$ to be the space of irreducible connections or $\mathcal{G}$ to be the group of pointed gauge transformations, $\mathcal{Q}$ is the total space of a principal G-bundle,

$$\mathcal{Q} \to \mathcal{Q}/\mathcal{G} = M \times \mathcal{A}/\mathcal{G} , \tag{5.85}$$

over $M \times \mathcal{A}/\mathcal{G}$. A G-invariant metric on $\mathcal{Q}$ defines (cf. section 5.1.2) a connection for (5.85) by declaring horizontal vector fields to be those orthogonal to the fundamental vector fields of G. Given metrics g on M and tr on G, such a metric is obtained naturally from the $G \times \mathcal{G}$-invariant metric $\hat{g}$ on $P \times \mathcal{A}$ defined by $[X_i \in T_p P, \tau_i \in T_A \mathcal{A} = \Omega^1(M, \mathbf{g})]$

$$\hat{g}_{(p,A)}((X_1, \tau_1), (X_2, \tau_2)) = g_{\pi(p)}(\pi_* X_1, \pi_* X_2) + \mathrm{tr}\, A(X_1) A(X_2) + \int_M \tau_1 * \tau_2 . \tag{5.86}$$

If we realize $\mathcal{A}/\mathcal{G}$ locally by a section of $\mathcal{A} \to \mathcal{A}/\mathcal{G}$, then evidently the $(1,0)$ part of this connection $\hat{A}$ with respect to the decomposition of forms on $M \times \mathcal{A}/\mathcal{G}$ is $\hat{A}|_{M \times \{[A]\}} = A$, where A is the representative of the equivalence class $[A]$ in the gauge chosen. The curvature $\hat{F}$ of $\hat{A}$ has

components of form degree (2,0), (1,1), and (0,2), and evaluated on horizontal vectors X_i and τ_i [i.e., $A(X_i) = d_A * \tau_i = 0$]; these are

$$\hat{F}_{(2,0)}(X_1, X_2) = F_A(X_1, X_2) \ , \tag{5.87}$$

$$\hat{F}_{(1,1)}(X_1, \tau_1) = \tau_1(\pi_* X_1) \ , \tag{5.88}$$

$$\hat{F}_{(0,2)}(\tau_1, \tau_2) = -G_A[\tau_1, *\tau_2] \ , \tag{5.89}$$

where $G_A = (d_A * d_A)^{-1}$. The non-locality of (5.89) has its origin in the fact that the horizontal projector H in $\mathcal{A}$ has the form

$$H(\tau) = \tau - d_A G_A d_A * \tau \ . \tag{5.90}$$

Indeed, one verifies easily that

$$d_A * H(\tau) = 0 \ , \qquad H(H(\tau)) = H(\tau) \ .$$

Equation (5.90) shows that $G_A d_A *$ is a connection on the principal $\mathcal{G}$-bundle $\mathcal{A}$, which immediately implies (5.89).

5.3.2. Geometry of Donaldson theory

Let us study the connection and curvature on this bundle in some more detail, but from a different point of view. Pulling the G-bundle $\mathcal{Q}$ back to (a trivial G-bundle on) $P \times \mathcal{A}$ (the reason for doing this will become apparent below) we can write the connection as the sum of a (1,0)- and a (0,1)-form on $P \times \mathcal{A}$, $\hat{A} = A + c$. Likewise we split the exterior derivative $\hat{d}$ on $P \times \mathcal{A}$ as $\hat{d} = d + \delta$. We then find that the curvature

$$\hat{F} = \hat{d}\hat{A} + \tfrac{1}{2}[\hat{A}, \hat{A}] = \hat{F}_{(2,0)} + \hat{F}_{(1,1)} + \hat{F}_{(0,2)} \tag{5.91}$$

is given by

$$\hat{F}_{(2,0)} = dA + \tfrac{1}{2}[A, A] \ , \tag{5.92}$$

$$\hat{F}_{(1,1)} = \delta A + dc + [A, c] \ , \tag{5.93}$$

$$\hat{F}_{(0,2)} = \delta c + \tfrac{1}{2}[c, c] \ . \tag{5.94}$$

As $\delta^2 = 0$, (5.93) and (5.94) also imply

$$\delta \hat{F}_{(1,1)} = -[c, \hat{F}_{(1,1)}] - d_A \hat{F}_{(0,2)} \ , \tag{5.95}$$

$$\delta \hat{F}_{(0,2)} = -[c, \hat{F}_{(0,2)}] \ . \tag{5.96}$$

We therefore see that, if we identify

$$\psi \equiv \hat{F}_{(1,1)} \ , \qquad \phi \equiv \hat{F}_{(2,0)} \ , \tag{5.97}$$

eqs. (5.93)–(5.96) are *formally* identical to the BRST transformations (5.70),

$$\delta A = \psi - d_A c \ , \tag{5.98}$$

$$\delta c = \phi - \tfrac{1}{2}[c, c] \ , \tag{5.99}$$

$$\delta \psi = -[c, \psi] - d_A \phi \ , \tag{5.100}$$

$$\delta \phi = -[c, \phi] \ , \tag{5.101}$$

of the fields in the geometrical sector of Donaldson theory. Furthermore, the descent equations (5.75) are then nothing but the Bianchi identity for $\hat{F}$. These observations provided the basis for the claim that the universal bundle $\mathcal{Q}$ adequately describes the geometry underlying this theory.

However, this is far from being the whole story yet, since, e.g., the identification (5.97) is purely formal so far. To put this on a firmer footing and to establish the relation between the geometry described here and that of the universal bundle discussed above, we have to understand in particular
- how ϕ ends up being given by $\phi(\tau_1,\tau_2) = -G_A[\tau_1,*\tau_2]$, eq. (5.89),
- the apparent discrepancy between (5.93), $\hat{F}_{(1,1)} = \psi = \delta A + d_A c$, and (5.88), which can be read as $\hat{F}_{(1,1)} = \delta A$.

In the course of resolving these issues we will also be led to understand
- why eqs. (5.98)–(5.101) have necessarily to be regarded as (global) equations on $P \times \mathcal{A}$ rather than as (local) equations on $M \times \mathcal{A}/\mathcal{G}$,
- how triviality of δ-cohomology can be reconciled with the non-trivial topology of $\mathcal{A}/\mathcal{G}$ (and $\mathcal{M}$),
- how the covariant gauge fixing condition on ψ, $d_A * \psi = 0$, is compatible with a background gauge fixing $d_{A_0} * (A - A_0) = 0$ of A.

Let us start by analyzing the consequences of the gauge fixing condition $d_A * \psi = 0$. This appears as a δ-function constraint in the path integral treatment of Donaldson theory, and therefore the following equations [5.75] can be read either as (classical) equations of differential geometry or as relations holding at the level of expectation values in Donaldson theory.

The first thing to notice is that, with $d_A * \psi = 0$, (5.98) describes a decomposition of δA into two pieces which are orthogonal with respect to the natural scalar product on $\mathcal{A}$ inherited from a metric on M. Since this is the metric we used to define a connection on $\mathcal{A}$, regarded as the total space of a principal bundle, we see that (5.98) gives a decomposition of δA into its horizontal (ψ) and vertical ($-d_A c$) part; and this is the reason why we have to regard δ as being the exterior derivative on $\mathcal{A}$ and not on $\mathcal{A}/\mathcal{G}$, and why (5.98) only makes sense as an equation on the former.

One may therefore now be tempted to declare that one obtains the exterior derivative on $\mathcal{A}/\mathcal{G}$ (a necessary preliminary step if we want to end up with δ as the exterior derivative on $\mathcal{M}$ and thus with the de Rham cohomology on $\mathcal{M}$) by taking the horizontal ($c = 0$) part of (5.98), (5.100) and (5.101). But this is wrong since, by definition, the horizontal part δ^H of the exterior derivative is the *covariant* exterior derivative, whose square is the curvature of the bundle; in contrast to this the vertical part of δ, δ^V, is nilpotent. It is exactly the BRST operator of ordinary Yang–Mills theory, and as such its interpretation as the exterior derivative along the gauge orbits is of course well known.

The truncated equations

$$\delta^H A = \psi \ , \qquad \delta^H \psi = -d_A \phi \ , \qquad \delta^H \phi = 0 \ , \tag{5.102}$$

are identical to the BRST transformations (5.45) used by Witten, and the statement that his BRST operator is nilpotent only up to a gauge transformation generated by ϕ is another way of saying that ϕ is a curvature form. The equation $\delta^H \phi = 0$ is then the Bianchi identity for ϕ, analogous to $d_A F_A = 0$, while $\delta \phi = -[c, \phi]$ is the analogue of $dF_A = -[A, F_A]$. However, on invariant polynomials of the curvature tensor, for instance, the exterior covariant derivative can be identified with the exterior derivative on the base space, and this is the way the exterior derivative on $\mathcal{M}$ arises in Witten's approach.

The condition $d_A * \psi = 0$ has yet another important consequence (in either the δ picture developed here or the δ^H picture chosen by Witten): the exterior (covariant) derivative of this

equation is

$$[\psi, *\psi] + d_A * d_A \phi = 0 \ , \tag{5.103}$$

which implies that ϕ is indeed the curvature tensor (5.89) determined previously,

$$\phi = -G_A[\psi, *\psi] \ . \tag{5.104}$$

At the level of expectation values we have already derived this equation in section 5.2.6.

Now that we have gauge fixed ψ to be orthogonal to gauge transformations, it remains to choose a gauge (slice in $\mathcal{A}$) for A itself, in order to get back from the trivial bundle on $P \times \mathcal{A}$ to the non-trivial bundle $\mathcal{Q}$ over $M \times \mathcal{A}/\mathcal{G}$. This we do as usual by demanding $d_{A_0} * (A - A_0) = 0$ for some background connection A_0. The exterior derivative of this equation (with respect to δ) gives

$$d_{A_0} * (\psi - d_A c) = 0 \ , \tag{5.105}$$

or

$$c = K_A d_{A_0} * \psi \ , \quad K_A = (d_{A_0} * d_A)^{-1} \ . \tag{5.106}$$

These equations again have a number of interesting consequences:

(i) The first (trivial) remark is that (5.106) is of course again satisfied at the level of expectation values, as follows from the action (5.53). As a consistency check it can also be verified that the equations

$$\phi = \delta c + \tfrac{1}{2}[c,c] \ , \quad \delta \phi = -[c,\phi] \ ,$$

hold with c and ϕ given by (5.104) and (5.106), respectively.

(ii) One illuminating way of looking at (5.105) [5.75] is to regard c as a gauge transformation taking ψ (gauge fixed at A) to a cotangent vector of A (gauge fixed at A_0), namely $\psi - d_A c$. This also explains why the relation $\hat{F}_{(1,1)} = \delta A$ of Atiyah and Singer is replaced by $\hat{F}_{(1,1)} = \delta A + d_A c$ in the present context. The former relation is valid on tangent vectors $\tau \in \Omega^1(M, \mathbf{g})$ satisfying $d_A * \tau = 0$. With the gauge fixing chosen for A here this is only achieved by adding $d_A c$ to δA, where c is given by (5.106).

(iii) Related to this is the role of c as the $(0,1)$ part of the connection on $\mathcal{Q}$ (or its pull-back). From (5.106) we see that it is gauge fixed to vanish on tangents to $\mathcal{A}$ horizontal with respect to A_0. In this respect it differs from the more conventional connection on $\mathcal{Q}$, which is defined by declaring horizontality to be with respect to A instead. Nevertheless, c is of course a perfectly good connection on the $\mathcal{A}/\mathcal{G}$ part of $\mathcal{Q}$, since it evidently satisfies the conditions (5.1), (5.2).

Summarizing the above discussion, we have now completed in detail the identification of the zero mode sector of Donaldson theory with the geometry of the universal bundle, and in particular therefore the identification of the observables of Donaldson theory with the Donaldson polynomials.

Having pinned down the theory to $\mathcal{A}/\mathcal{G}$, we can—by imposing further constraints on A (and thus ψ) compatible with gauge invariance—restrict the theory to some moduli (sub)space of $\mathcal{A}/\mathcal{G}$. Donaldson theory, for instance, follows from imposing $F_A^+ = 0$, which implies $(d_A \psi)^+ = 0$. Pairs (A, ψ) satisfying these conditions in addition to those encountered above will then represent a point in the moduli space of anti-instantons, and a cotangent vector to that point.

But we can clearly impose other conditions as well, in this way constructing topological gauge theories based on other moduli spaces. We will discuss this in the next section. First, however, we will turn to the issue of "triviality" of observables, since the problem as well as its resolutions are quite independent of the particular moduli space chosen.

5.3.3. Observables and triviality

Combining the description of the Donaldson polynomials we have given in section 5.2 with what we have learned above, we see that they are obtained from expansion of the second Chern class of $\mathcal{Q}$. The claim that these observables are in a sense trivial has been based on either of the following two observations:

(i) Locally $\text{tr}\,\hat{F}^2$ can be written as

$$\text{tr}\,\hat{F}^2 = (d + \delta)\,\text{tr}(\hat{A}\hat{F} - \tfrac{1}{3}\hat{A}^3) \ ;$$

this implies that all the W_k (where $\tfrac{1}{2}\text{tr}\,\hat{F}^2 = \sum_{k=0}^{4} W_k$) are δ-trivial modulo d, and thus that all the observables $W_k(\gamma) = \int_\gamma W_k$ are BRST-exact. Of course the emphasis here is on "locally"; but while this certainly means that

$$\text{tr}\,F_A^2 = d\,\text{tr}(AF_A - \tfrac{1}{3}A^3)$$

does not necessarily have any global implications, the situation concerning

$$\text{tr}\,\phi^2 = \delta\,\text{tr}(c\phi - \tfrac{1}{3}c^3) \tag{5.107}$$

is not so clear. From this point of view what needs to be explained is why c [and therefore (5.106)] does not make sense globally.

(ii) Perhaps more strikingly (but equivalently) the essence of the matter can be captured by the following argument [5.64]: By making the field redefinitions

$$\psi' = \psi - d_A c \ , \qquad \phi' = \phi - \tfrac{1}{2}[c,c]$$

the BRST algebra (5.98)–(5.101) can be brought into the form

$$\delta A = \psi' \ , \qquad \delta\psi' = 0 \ , \qquad \delta c = \phi' \ , \qquad \delta\phi' = 0 \ ,$$

which shows very clearly that any δ-closed functional of these fields will be δ-exact. From this point of view one has to explain why δ-cohomology is not the one relevant for Donaldson polynomials and their path integral evaluation.

We will now first give a very general topological explanation of what is going on here. Afterwards we will present a second (geometrical) and third (algebraic) argument pertinent to the point of view expressed in (i) and (ii), respectively.

Topological explanation. Characteristic classes of $\mathcal{Q}$ give rise to de Rham cohomology classes on $\mathcal{A}/\mathcal{G}$. The fact that these are δ-exact is—in view of the identification of δ with the exterior derivative on $\mathcal{A}$—nothing but the fact that any cohomology class of $\mathcal{A}/\mathcal{G}$ is trivial when regarded as a cohomology class on (the contractible space) $\mathcal{A}$. Thus clearly this whole issue of triviality is no issue from the mathematical point of view. What remains to be understood, though, is how physics in, say, the path integral formulation takes this into account.

Note that in Witten's (δ^H) formulation there is no problem. That only arises once one enlarges the space from $\mathcal{A}/\mathcal{G}$ (or a section of $\mathcal{A} \to \mathcal{A}/\mathcal{G}$) to all of $\mathcal{A}$, as manifested by the appearance of the gauge ghost c in, e.g., (5.107).

Geometrical explanation. The idea here is to explain why (5.106) is not valid globally and can therefore not be used to deduce the δ-triviality of observables derived from $\text{tr}\,\phi^2$ via the descent equations.

But this is easy to understand, since in general the bundle Q will be non-trivial over $\mathcal{A}/\mathcal{G}$. In that case there will be no global non-singular expression for the connection c of the curvature ϕ. But we have already seen this explicitly in (5.106)! Due to the Gribov ambiguity [5.76, 5.77], $K_A = (d_{A_0} * d_A)^{-1}$ will necessarily be singular somewhere under very general conditions, which shows that $c = K_A d_{A_0} * \psi$ can only be regarded as being defined locally on $\mathcal{A}/\mathcal{G}$. This point of view had been advocated in one way or another in refs. [5.49, 5.67, 5.66, 5.75] and was made precise by Kanno [5.78].

Algebraic explanation. In view of the topological argument given above, one way of phrasing the problem is: how does one detect the non-trivial cohomology of $\mathcal{A}/\mathcal{G}$ by computations on $\mathcal{A}$?, and the answer [5.64] is provided by basic cohomology [5.79], which is designed to do just this.

However, in order to make use of this in the present context, the quantization procedure adopted in section 5.2—based on the BRST operator δ, which combines the topological shift symmetry and the gauge symmetry into a single operator (cf. $\delta A = \psi - d_A c$)—is not the most suitable (fundamentally due to the fact of course that δ, as the exterior derivative on $\mathcal{A}$, cannot detect the non-trivial cohomology of $\mathcal{A}/\mathcal{G}$). An alternative method—advocated by Horne [5.80]—is to keep these symmetries separate and to augment Witten's action (5.44) by ordinary gauge fixing terms. If one proceeds in this manner one ends up with a theory which is—apart from a bigger field content—in every way equivalent to the one discussed in section 5.2 [5.67, 5.73, 5.75]. However, it is these extra fields which allow one to resolve the issue of triviality in the following way: although the observables are δ-exact, they cannot be written as the δ of something which is gauge invariant [5.64].

Thus we have seen that there are various ways of establishing the non-triviality of observables in Donaldson theory (and other topological gauge theories). Of course, ultimately all the arguments presented above are precisely equivalent; depending on the context, however, one or the other of these points of view may be more useful or to the point.

5.4. Construction of topological gauge theories

5.4.1. The classical action
As explained in section 2, part of the philosophy behind Witten type gauge theories is that they are cohomological theories associated with moduli problems, i.e. upon specifying the fields, the symmetries, and the field equations, there is a (perhaps non-unique) field theory whose correlation functions compute intersection numbers on the moduli space of solutions to the equations modulo the symmetries. The general procedure leading to these topological field theories given the above data was developed in refs. [5.49, 5.74].

Now that we have understood the geometry underlying Donaldson theory, the construction of topological gauge theories associated with arbitrary moduli spaces of connections (this also includes moduli spaces of Riemann surfaces, cf. section 5.4.3) is straightforward. In addition to the conditions $d_{A_0} * (A - A_0) = d_A * \psi = 0$, which project the theory down from $\mathcal{A}$ to $\mathcal{A}/\mathcal{G}$, we can now simply impose a further condition on A (and therefore, by "supersymmetry", on ψ),

$$\mathcal{F}(A) = 0 \tag{5.108}$$

(where $\mathcal{F}$ is some functional of A), as long as this condition is gauge invariant. Some obvious choices for $\mathcal{F}$ are

$$\mathcal{F}(A) = F_A^+ , \tag{5.109}$$

$$\mathcal{F}(A) = F_A ,$$ (5.110)

$$\mathcal{F}(A) = d_A * F_A .$$ (5.111)

These give rise to Donaldson theory, a topological gauge theory of flat connections, and a topological theory based on the moduli space of solutions to the Yang–Mills equations, respectively. Note that there is no problem of principle with the construction of topological gauge theories based on explicitly metric dependent field equations like (5.109) or (5.111). Note also that, while (5.109) only makes sense in four dimensions, (5.110) and (5.111) are not restricted in that way.

Combining (5.108) with the fundamental equation $\delta A = \psi - d_A c$, we learn that $\psi - d_A c$ has to satisfy the linearized equation

$$\frac{\delta \mathcal{F}}{\delta A}[\psi - d_A c] = 0 .$$ (5.112)

If A satisfies (5.108) then this is equivalent to

$$\frac{\delta \mathcal{F}}{\delta A}[\psi] = 0 ,$$ (5.113)

since gauge invariance of $\mathcal{F}$ implies

$$\frac{\delta \mathcal{F}}{\delta A}[d_A \Lambda] = 0$$ (5.114)

for all $(0,1)$-forms $\Lambda \in \Omega^0(M,\mathbf{g})$. Thus both ψ and $\psi - d_A c$ represent cotangent vectors to the moduli space $\mathcal{M}_{\mathcal{F}}$ determined by (5.108), albeit in different gauges,

$$d_A * \psi = 0 , \qquad d_{A_0} * (\psi - d_A c) = 0 .$$

In the examples (5.109)–(5.111) for $\mathcal{F}$ given above, (5.113) is, more explicitly,

$$(d_A \psi)^+ = 0 ,$$ (5.115)

$$d_A \psi = 0 ,$$ (5.116)

$$d_A * d_A \psi = [*F_A, \psi] ,$$ (5.117)

and (5.114) is easily verified directly in these cases.

It is now clear how to encode the geometry of $\mathcal{M}_{\mathcal{F}}$ and of the bundle $\mathcal{Q}$ over $M \times \mathcal{M}_{\mathcal{F}}$ (this is just the restriction of $\mathcal{Q}$ via the inclusion $M \times \mathcal{M}_{\mathcal{F}} \hookrightarrow M \times \mathcal{A}/\mathcal{G}$) into a supersymmetric action. To enforce (5.108) we introduce an anti-ghost (ghost number -1) field χ and a multiplier B with $\delta \chi = B, \delta B = 0$. The nature of χ and B will depend on the choice of $\mathcal{F}$: in the examples (5.109)-(5.111) they will be self-dual two-forms, $(n-2)$-forms and one-forms, respectively, in order to make $B\mathcal{F}(A)$ an n-form on M. Then we choose our *classical action* to be

$$S_c = \delta \int_M \chi \mathcal{F}(A) = \int_M B\mathcal{F}(A) \pm \chi \frac{\delta \mathcal{F}}{\delta A}[\psi] .$$ (5.118)

Evidently B imposes the desired classical equation of motion $\mathcal{F}(A) = 0$, whereas the χ equation of motion restricts ψ to be tangent to $\mathcal{F}(A) = 0$. For flat connections (5.118) reads more explicitly

$$S_c = \int_M BF_A + (-)^n \chi d_A \psi .$$ (5.119)

As in Donaldson theory (cf. section 5.2.2), we introduce anti-ghost–multiplier pairs $(\bar{\phi}, \eta)$ and $(\bar{c}, b)$ to impose the gauge fixing conditions on ψ and A. Then our (preliminary, cf. below) quantum action will be

$$S = \delta \int_M \chi \mathcal{F}(A) + \bar{\phi} d_A * \psi + \bar{c} d_{A_0} * (A - A_0) \; . \tag{5.120}$$

In the case $\mathcal{F}(A) = F_A^+$ this leads to the action for Donaldson theory given by Baulieu and Singer (in the $\alpha = 0$ gauge), whereas for flat connections we obtain

$$S = \int_M \{ BF_A + (-)^n \chi d_A \psi - (-)^n \chi [F_A, c] + \eta d_A * \psi + \bar{\phi} [\psi, *\psi]$$

$$+ \bar{\phi} d_A * d_A \phi - \bar{\phi} d_A * [c, \psi] + b d_{A_0} * (A - A_0) + \bar{c} d_{A_0} * \psi - \bar{c} d_{A_0} * d_A c \} \; . \tag{5.121}$$

The qualification "preliminary" refers to the fact that (5.120) will in general still possess residual local symmetries. In fact, in more than two dimensions, there will be a whole tower of on-shell reducible symmetries of (5.121), resulting from the obvious invariance $\delta B = -d_A \Sigma$ where Σ is an $(n - 3)$-form. One of the aims of the next section will be to show that the fully gauge fixed action can still be written as a BRST commutator, with an *off-shell* nilpotent BRST operator. This off-shell nilpotency (which is required for the arguments of section 2 on the topological nature of this model to go through directly) is not ensured by the Batalin–Vilkovisky algorithm (due to the on-shell reducibility), and we are therefore forced to construct the quantum action in a different way. The success of our method depends crucially on the fact that we work with a BRST operator which combines the topological, gauge, and the above p-form symmetries, instead of introducing a new BRST operator which accounts for the latter only (as would have resulted from the Batalin–Vilkovisky procedure).

Let us make some more comments concerning the action (5.121):

(i) As in Donaldson theory we could have equally well chosen an action based on $\chi (F_A - (\alpha/2)*B)$ with α non-zero. For $\alpha = 1$ the resulting action would then (upon having integrated out B) again have taken the form "Yang–Mills action + $\cdots$". And in order to construct this theory rigorously, $\alpha = 1$ may ultimately be a better (since less singular) choice, but for our present purposes $\alpha = 0$ is certainly more convenient.

(ii) The classical action (5.119) shows that this theory can be regarded as a supersymmetrization of BF systems [5.81, 5.82], Schwarz type topological gauge theories with classical action $\int BF_A$, which we will discuss in detail in section 6. Therefore we will whenceforth refer to theories described by the action (5.119), (5.121) as super-BF systems [5.49].

(iii) When B is a one-form (e.g. 3D super-BF) one can postulate a second nilpotent supersymmetry $\bar{\delta}$ in addition to δ, defined by $\bar{\delta}A = \chi, \bar{\delta}\psi = -B$. The super-BF action (5.119) can then be written as

$$S_c = \tfrac{1}{2}\delta\bar{\delta} \int A \, dA + \tfrac{2}{3}A^3 \; .$$

Grouping A, ψ, χ and B into an $N = 2$ superconnection shows (δ and $\bar{\delta}$ are, as usual, replaced by Berezin integrals) that (5.119) can alternatively be regarded as the Chern–Simons action of a certain supergroup (called super-IG by Witten [5.83]). It can be shown [5.49] that this second supersymmetry extends to an anti-BRST symmetry of the complete quantum action. Analogous remarks apply to the Yang–Mills action [based on (5.118) with $\mathcal{F}(A) = d_A * F_A$] in any dimension.

One also finds a counterpart of this observation in BF theories: in three dimensions these can be regarded as Chern–Simons theories for the group $IG = TG$ (the tangent bundle of G). We will explain this in more detail in section 6.2.3.

(iv) The third term in (5.121) can be eliminated by a shift of B or, equivalently, by redefining the transformations of χ and B to be

$$\delta\chi = -[c,\chi] + B , \qquad \delta B = -[c,B] + [\phi,\chi] . \tag{5.122}$$

Written this way the full quantum action, as well as the transformations for the required additional (ghost, anti-ghost, and multiplier) fields, take a more transparent form. We are therefore going to use this form of the transformations in the following section. Similar remarks apply to other couplings involving the gauge ghost c.

Before concluding this section let us make a short remark on the philosophy of the Baulieu–Singer approach to topological field theories. This approach is often referred to as one where one gauge fixes zero or a topological invariant (or $\sqrt{2}$ for that matter). Indeed, the form of the action (5.118) suggests an interpretation in which the classical action is zero and $\mathcal{F}(A) = 0$ is a "gauge fixing condition" for the shift symmetry $A \mapsto A + X$. However, here we see clearly that this cannot be quite right: firstly one is never just quantizing zero, but one has in mind a whole host of geometrical constructions; secondly the resulting theory clearly depends on the choice of the "gauge fixing condition" $\mathcal{F}$; and thirdly none of these conditions break the shift symmetry completely [indeed, the only one which does that is $\mathcal{F}(A) = A - A_0$, A_0 arbitrary but fixed, which leads to a trivial theory]. Thus, while this way of phrasing things provides some useful heuristics, it is certainly misleading if taken too literally.

5.4.2. The quantum action

In this section we will complete the gauge fixing of the super-BF theories introduced above, by taking into account the non-Abelian p-form symmetry

$$\delta B_{n-2} = d_A \Lambda_{n-3} , \tag{5.123}$$

and (for $n \geq 4$) its descendents $\delta \Lambda_{n-3} = d_A \Lambda_{n-4}, \dots$. This is discussed in general in ref. [5.49], where it is shown that, despite the on-shell reducibility of (5.123) for $n \geq 4$, the complete quantum action can be written as the commutator of an off-shell nilpotent BRST operator. Here we will only treat the three dimensional case in detail, where the quantization is straightforward, since that model is the most interesting from the mathematical point of view (being related to the Chern–Simons functional and the Casson invariant). In higher dimensions, the correct (minimal) field content is specified by the Batalin–Vilkovisky ghost triangles (for an explanation and the details see appendix A), and we will only quote the results in that case.

At this point we can already anticipate an interesting subtlety arising upon gauge fixing (5.123), namely that the required covariant gauge fixing condition on B, $d_A * B = 0$, will inevitably modify the B equation of motion from $F_A = 0$ to

$$F_A = -* d_A u , \tag{5.124}$$

where u is some ghost number zero (n-3)-form. In three dimensions this equation is known as the Bogomol'ny equation describing monopoles in the Prasad–Sommerfield limit, u acquiring the role of a Higgs field. We therefore appear to be dealing with (generalized) monopoles although we had set out to quantize a theory describing flat connections. Let us show that this is not the case and that in the situation at hand (5.124) implies $F_A = 0$ [5.49]. There are two cases to consider. If

the manifold M is closed (compact and no boundary), then an analog of the squaring argument of section 3 clearly leads to the conclusion that (5.124) implies that F_A and $d_A u$ are separately zero (alternatively, use the Bianchi identity to deduce $d_A * d_A u = 0$, multiply by u and integrate by parts to conclude that $d_A u$ is zero, which implies $F_A = 0$). If, on the other hand, M has a boundary or is non-compact, the above argument will remain valid provided that the boundary term

$$q = \int_{\partial M} F_A u = - \int_{\partial M} u * d_A u \qquad (5.125)$$

(the "monopole charge") is zero. The asymptotic behavior of u (as a secondary anti-ghost or multiplier field) can, however, not be specified arbitrarily, but is determined by that of the primary fields of the theory; if we are modeling flat connections the natural condition is that $F_A \to 0$ in a certain way at infinity, and this in turn fixes the asymptotic behavior of the multiplier B and consequently that of u. From either of the two above expressions it then follows that $q = 0$.

The origin of the Higgs field and its necessity at intermediate stages can also be understood in a different (but equivalent) way from the Langevin point of view. In order to describe a theory of flat connections one would naively start off with a classical action of the form $S = \int (G - F_A)^2$, where G is a two-form. But this action clearly does not have enough gauge freedom to set G to zero, a shift in A only permitting one to remove the exact part in a Hodge decomposition of G. In order to eliminate all of G one needs a complete Hodge decomposition on the right hand side of the Langevin equation $G = F_A + \cdots$. Thus (ignoring harmonic modes, which can trivially be incorporated along the lines of section 6.2.1) the correct Langevin equation to start off with is

$$G = F_A + *d_A u \ .$$

Setting $G = 0$ one recovers (5.124). The equivalence of the two arguments leading to (5.124) can be demonstrated by writing the classical action (the square of the Langevin equation) as

$$S = \frac{1}{2\alpha} \int (G - F_A - *d_A u)^2 = \int B(G - F_A - *d_A u) - \tfrac{1}{2}\alpha B^2 \ . \qquad (5.126)$$

Thus, having come to grips with this slight complication, let us complete our task of writing down the complete quantum action in the three dimensional case. Determining the required set of fields is straightforward: in addition to the fields already present in (5.121), the one-forms χ and B contribute one ghost–anti-ghost–multiplier triplet each, denoted by $(\rho_0, \bar\rho_0, \sigma_0)$ and $(\Sigma_0, \overline\Sigma_0, \pi_0)$, respectively. The required χ and B transformations (combining the previous shift and gauge symmetries with the p-form symmetry) are

$$\delta\chi = -[c,\chi] - d_A\rho_0 + B \ , \qquad (5.127)$$

$$\delta B = -[c,B] - d_A\Sigma_0 + [\phi,\chi] + [\psi,\rho_0] \ , \qquad (5.128)$$

and covariance and nilpotency then determine the action of δ on the remaining fields:

$$\begin{aligned}
\delta\rho_0 &= -[c,\rho_0] + \Sigma_0 \ , & \delta\Sigma_0 &= -[c,\Sigma_0] + [\phi,\rho_0] \ , \\
\delta\bar\rho_0 &= -[c,\bar\rho_0] + \sigma_0 \ , & \delta\overline\Sigma_0 &= -[c,\overline\Sigma_0] + \Pi_0 \ , \\
\delta\sigma_0 &= -[c,\sigma_0] + [\phi,\bar\rho_0] \ , & \delta\Pi_0 &= -[c,\Pi_0] + [\phi,\overline\Sigma_0] \ .
\end{aligned} \qquad (5.129)$$

Only two kinds of terms in the above transformations may require some explanation: Those of the form $\delta X = [\phi, X] + \cdots$ are required for nilpotency of the $[c, \ ]$ gauge rotations because of the shift by ϕ in δc, and the additional ψ term in δB is there to compensate the A variation in $d_A\Sigma_0$.

Choosing the action to be

$$S_q = \{Q, \int \chi F_A + \bar{p}_0 d_A * \chi + \Sigma_0 d_A * B + \bar{\phi} d_A * \psi + \bar{c} d * A\} \, , \tag{5.130}$$

we find that indeed all the invariances of the action are completely fixed. Integration over B enforces (as expected) the constraint $F_A = - * d_A(\bar{p}_0 - \Pi_0)$. As such $(\bar{p}_0 - \Pi_0)$ plays the role of the Higgs field we had previously called u. We have thus arrived at our goal of constructing an off-shell nilpotent BRST operator Q such that the full quantum action can be written as a BRST commutator, preserving explicitly the Witten type nature of this model, by combining the super- and gauge symmetries. Incidentally, we have on the way also achieved this for the super-Yang–Mills system [(5.118) with $\mathcal{F}(A) = d_A * F_A$], since the non-classical part of the above action coincides precisely with the one of the super-Yang–Mills system in any dimension, as B and χ will always be one-forms in this case.

After having discussed this three-dimensional example, the extension to higher dimensions turns out to be, somewhat surprisingly, fairly straightforward. Normally one would have expected considerable complications arising from the fact that in more than three dimensions the ghosts and ghosts for ghosts for χ and B will have their own (on-shell) gauge invariances. Looking at the B transformation (5.128) it is obvious that they will give rise to a term proportional to F_A in addition to those already present involving ψ and ϕ. As a consequence of the Bianchi identity for the universal curvature $F_A + \psi + \phi$ this is, however, basically the only modification required. But while the structure of the transformations in higher dimensions is similar to the above, there is the necessity of introducing extra ghosts and their corresponding anti-ghosts and Lagrange multipliers. In other words, one must ensure that the full field content, as specified by the Batalin–Vilkovisky triangles, is represented. Labeling, e.g., the fields in the B triangle by (Σ_i^j), $i, j = 0, 1, \ldots, n - 2$, where $\Sigma_{n-2}^0 = B$ and Σ_{n-2-i}^0 is the ith ghost of B (see appendix A), with analogous notation for the χ triangle (ρ_i^j), the BRST transformations are simply [cf. (5.129)]

$$\delta \Sigma_i^0 = -[c, \Sigma_i^0] - d_A \Sigma_{i-1}^0 + [\phi, \rho_i^0] + [\psi, \rho_{i-1}^0] + [F_A, \rho_{i-2}^0] \, ,$$
$$\delta \rho_i^0 = -[c, \rho_i^0] - d_A \rho_{i-1}^0 + \Sigma_i^0 \, , \tag{5.131}$$

and for $j = 1, \ldots, n - 2$,

$$\delta \Sigma_i^j = -[c, \Sigma_i^j] + \Pi_i^j \, , \qquad \delta \Pi_i^j = -[c, \Pi_i^j] + [\phi, \Sigma_i^j] \, ,$$
$$\delta \rho_i^j = -[c, \rho_i^j] + \sigma_i^j \, , \qquad \delta \sigma_i^j = -[c, \sigma_i^j] + [\phi, \rho_i^j] \, . \tag{5.132}$$

Here Π_i^j and σ_i^j are multiplier triangles, and δ is evidently nilpotent on all the fields. The complete quantum action can then be chosen to be

$$S_q = \left\{ Q, \int \chi F_A + \bar{\phi} d_A * \psi + \bar{c} d * A \sum_{j=1}^{n-2} \sum_{i=0}^{n-2-j} (\Sigma_i^j d_A * \Sigma_{i+1}^{j-1} + \rho_i^j d_A * \rho_{i+1}^{j-1}) \right\} \, , \tag{5.133}$$

and one verifies that all invariances of the action are gauge fixed and that all ghosts have (as required for the correct number of degrees of freedom to emerge) standard kinetic terms.

The situation here should be contrasted with that encountered in the quantization of the Schwarz type BF theories (section 6.2.4), where the construction of a complete quantum action is much more complicated and apparently only possible with a BRST operator which is not nilpotent off-shell.

5.4.3. Moduli spaces of flat connections I

As a preparation for our discussion of observables in super-BF theories in the following section we will now provide some information on moduli spaces of flat connections. This will also serve as a useful background for our subsequent discussion of Schwarz type topological gauge theories (section 6).

We begin by recalling the description of the moduli space $\mathcal{M} = \mathcal{M}(M, G)$ of flat G connections on a manifold M in terms of representations of the fundamental group $\pi_1(M) \equiv \pi$ of M, leading to the useful identification

$$\mathcal{M} = \mathrm{Hom}(\pi, G)/G \ , \tag{5.134}$$

where the quotient action of G is by conjugation: given a flat connection A and a point p in the fiber above some basepoint $x \in M$, the holonomy $h_p(\gamma)$ around a loop γ in M depends only on its homotopy class $[\gamma] \in \pi$, and this assignment of elements of G to homotopy classes behaves multiplicatively under composition of loops, determining an element $\rho_A \in \mathrm{Hom}(\pi, G)$; the latter is well defined modulo inner automorphisms of G, since changing the reference point p in the fiber to $pg, g \in G$, conjugates the holonomies by g, $h_{pg}(\gamma) = g^{-1}h_p(\gamma)g$; conversely every such homomorphism σ defines a natural flat connection $A(\sigma)$ with $\rho_{A(\sigma)} = \sigma$ in the principal G-bundle associated (via σ) to the principal π-bundle over M given by the universal covering of M; this establishes (5.134).

The spaces $\mathrm{Hom}(\pi, g)$ and $\mathcal{M}$ have a rich geometrical structure reflecting properties of both the manifold M and the group G, and for a wealth of information concerning these spaces the reader is referred to the work of Goldman [5.84] and Hitchin [5.85]. Here we just note that in general $\mathrm{Hom}(\pi, G)$ and $\mathcal{M}$ are not smooth manifolds (the latter may not even be Hausdorff if G is non-compact). If $M = \Sigma$ is a compact Riemann surface, and G is compact, $\mathcal{M}$ is a compact complex variety whose singular points are precisely the reducible representations. Under fairly general conditions, these singularities are not too bad (of quadratic type).

To get a feeling for the dimension of these spaces, let us take a look at some examples. If $M = \Sigma_g$ is a Riemann surface of genus g, the standard presentation of π is in terms of $2g$ generators $a_i, b_i, i = 1, \ldots, g$, satisfying the one relation $a_1 b_1 a_1^{-1} b_1^{-1} \cdots a_g b_g a_g^{-1} b_g^{-1} = 1$. For $g > 1$ and G simple, this implies that the dimension of $\mathcal{M}$ is $2g \dim G$ (for the assignment of group elements to the generators) minus $\dim G$ (one relation) minus $\dim G$ (identifying conjugacy classes), i.e.,

$$\dim \mathcal{M}(\Sigma_g, G) = (2g - 2)\dim G \ . \tag{5.135}$$

For $g = 0$, $\mathcal{M}$ is one point (the trivial representation), but for $g = 1$ the relation satisfied by a and b is $ab = ba$. This implies that (generically) a and b can be represented by elements lying in the maximal torus T of G. On $T \times T$, action by conjugation reduces to the action of the Weyl group W, and the moduli space of flat connections is the orbifold $(T \times T)/W$ of dimension

$$\dim \mathcal{M}(\Sigma_1, G) = 2 \dim T \ . \tag{5.136}$$

If $G = U(1)$ on the other hand, the one relation is automatically satisfied, since $U(1)$ is Abelian, and by the same token conjugation acts trivially, so that for all g one has $\mathcal{M}(\Sigma_g, U(1)) = U(1)^{2g}$ (the Jacobian variety of Σ_g once a complex structure on Σ_g has been chosen). Therefore $\dim \mathcal{M}(\Sigma_g, U(1)) = 2g$, corresponding to $2g$ "vacuum angles" or "Aharonov–Bohm phases".

As our final example, let us consider more concretely the case $\pi = \mathbb{Z}$ (i.e., M could be a circle or a solid torus), $G = SU(2)$. Every element of $SU(2)$ can be put into the maximal torus $T = U(1)$ by conjugation, but the fact that one is dealing with $SU(2)$ and not with $U(1)$ shows up through

the fact that there is a residual action of the Weyl group $W = \mathbb{Z}_2$. Thus if U(1) is described by a variable $\theta \in [0, 2\pi)$, the Weyl group identifies θ and $2\pi - \theta$. Therefore the moduli space of flat SU(2) connections is the closed interval (manifold with boundary) $[0, \pi]$.

We have already seen that (inner) automorphismse of G act on $\mathrm{Hom}(\pi, G)$, but so do automorphisms of π. In particular all the moduli spaces $\mathcal{M}(\Sigma_g, G)$ carry natural actions of the mapping class group $\pi_0(\mathrm{Diff}\,\Sigma_g)$, since—by a theorem of Nielsen [5.84]—$\pi_0(\mathrm{Diff}\,\Sigma_g) = \mathrm{Out}(\pi)$, the group of outer automorphisms of π. This remark acquires particular significance in the case $G = \mathrm{PSL}(2, \mathbb{R})$, where a component of $\mathcal{M}$ is Teichmüller space and $\mathcal{M}/\mathrm{Out}(\pi)$ is therefore the moduli space of Riemann surfaces of genus g. This observation underlies many of the constructions of topological gravity, and we will therefore take a closer look at the moduli space of flat $\mathrm{PSL}(2, \mathbb{R})$ connections in section 6.2.7.

The above discussion shows that moduli spaces of flat connections are reasonably nice and interesting spaces; it should therefore be possible to obtain topological information on M from the cohomology (and, in particular, intersection numbers) of $\mathcal{M}$. But on general grounds one should expect the topological invariants arising from these (obviously metric independent) moduli spaces to be less subtle than the Donaldson invariants constructed from the metric dependent instanton moduli.

What the above discussion does not tell us, however, is how to construct these invariants, and in particular, how the fermionic zero modes of the super-BF action are related to the dimension and singularities of $\mathcal{M}$. This information is, to a certain extent, provided by an index theorem approach to $\mathcal{M}$, based on the flat connection deformation complex which we will describe below. As we will see, this approach is in general not as powerful as in the case of instantons. For one, the index of the deformation complex, the formal dimension of $\mathcal{M}$, turns out to be identically zero in all odd dimensions; this implies that, unlike in the case of instantons, there is no relation between the net ghost number violation (zero) and the dimension of $\mathcal{M}$ in odd dimensions, and requires some rethinking concerning the construction of observables. Moreover, in the instanton case there is the added flexibility in a choice of metric, allowing one to prove vanishing theorems for particular [5.8] and generic [5.9] metrics, whereas here $\mathcal{M}$ and all the cohomology groups of the deformation complex are metric independent. Nevertheless, as we will show now, there are some things that can be learned from this approach.

The (Zariski) tangent space $\mathrm{T}_A\mathcal{M}$ to $\mathcal{M}$ at a flat connection A is the space of solutions to the linearized equations of motion modulo gauge equivalence. In other words, it is precisely the space of ψ zero modes ($d_A\psi = d_A * \psi = 0$) of the action (5.121). This is the first cohomology group $H_A^1(M, \mathbf{g})$ of the *twisted* de Rham (or flat connection deformation) complex

$$0 \to \Omega^0(M, \mathbf{g}) \xrightarrow{d_A} \Omega^1(M, \mathbf{g}) \xrightarrow{d_A} \cdots \xrightarrow{d_A} \Omega^n(M, \mathbf{g}) \to 0 \ . \tag{5.137}$$

Note that it was necessary to include all the higher rank forms to render the complex elliptic—a fact that is reflected in the necessity to gauge fix the non-Abelian p-form symmetry (5.123) by introducing the hierarchy of ghosts and multipliers into the action (5.133).

The Euler characteristic of this complex, or the index of the deformation operator

$$d_A + d_A^* : \Omega^{\mathrm{even}}(M, \mathbf{g}) \to \Omega^{\mathrm{odd}}(M, \mathbf{g}) \ , \tag{5.138}$$

is easy to compute: since it is independent of the local coefficient system, we may as well choose the trivial connection. This gives

$$\mathrm{ind}(d_A + d_A^*) = \chi(M)\dim G \ , \tag{5.139}$$

or, in terms of the dimensions h_A^i of the cohomology groups $H_A^i(M, \mathbf{g})$,

$$h_A^0 - h_A^1 + \cdots + (-)^n h_A^n = \chi(M) \dim G .$$ (5.140)

That (5.137) is the correct complex to consider can also be read off directly from the action. For instance in two dimensions the relevant part of the action is $\int \chi d_A \psi + \eta d_A * \psi$, in agreement with (5.138). In three dimensions χ also couples to the multiplier σ_0, and $\int -\chi d_A \psi + \eta d_A * \psi + \sigma_0 d_A * \chi$ again agrees with (5.138), $*\sigma_0$ representing $\Omega^3(M, \mathbf{g})$. From the above we see that the ghost number -1 χ zero modes are matched by the ghost number $+1$ ψ zero modes, and that there are an equal number of ghost number $+1$ σ_0 and ghost number -1 η scalar zero modes. This feature persists in all other odd dimensions, where, as announced above, both sides of (5.140) are identically zero by Hodge duality, giving no information on the dimension h_A^1 of $T_A \mathcal{M}$.

In two dimensions (5.140) reduces to

$$h_A^1 = (2g - 2) \dim G + 2h_A^0 ,$$ (5.141)

which is in precise agreement with the results obtained earlier in this section:
- for $g > 1$, $h_A^1 = (2g - 2) \dim G$ at an irreducible connection, which checks with (5.135);
- for $g = 1$ the discussion leading to (5.136) shows that all flat connections are reducible. The least reducible (simple points of $\mathcal{M}$) among these are the connections A with $h_A^0 = \dim T$; for these the prediction of (5.141), namely $h_A^1 = 2 \dim T$, coincides with the result (5.136);
- for $g = 0$, the only flat connection is the trivial connection, for which $h_A^0 = \dim G$, in agreement with $h_A^1 = 0$.

In four dimensions we conclude from (5.140) that

$$h_A^1 = h_A^0 + \tfrac{1}{2}[h_A^2 - \chi(M) \dim G] .$$ (5.142)

At an irreducible connection, h_A^2, measuring the cokernel of (5.138), will (as in section 5.1) be the obstruction to using the implicit function theorem to deduce smoothness of $\mathcal{M}$ near A. Contrary to the situation encountered in section 5.1.4, this obstruction is independent of the metric. If it happens to vanish, one obtains $-\tfrac{1}{2}\chi(M) \dim G$ as the dimension of the moduli space of irreducible flat connections. One thing to be learned from this is that (for $h_A^2 = 0$) $\chi(M) < 0$ is a necessary condition for non-trivial moduli spaces of flat connections to exist.

In higher even dimensions there are more and more unknown cohomology groups to worry about, and the index theorem provides less and less concrete information on h_A^1. The above result for the dimension of the moduli space of irreducible flat connections in four dimensions will, however, remain valid whenever $h_A^i = 0$ for $i \geq 2$.

One more important piece of information we will make use of in the following is that, under a very general condition on G, all the moduli spaces of flat connections on Riemann surfaces are symplectic manifolds. This condition on G—that its Lie algebra should admit a G-invariant scalar product—is very natural from the point of view of Chern–Simons theory (section 6.1): it ensures that a gauge invariant Chern–Simons action for G can be written down; $\mathcal{M}(\Sigma_g, G)$ is then the phase space of Chern–Simons theory on $\Sigma_g \times \mathbb{R}$ and should therefore be a symplectic manifold. Denoting this scalar product by tr, the symplectic two-form at a point $A \in \mathcal{M}$ is

$$\omega_A([X], [Y]) = \int_{\Sigma_g} \mathrm{tr}\, XY ,$$ (5.143)

where X, Y are arbitrary representatives of the cohomology classes $[X], [Y] \in H_A^1(M, \mathbf{g})$. This result in all its generality is due to Goldman [5.84], who has also shown that, for $G = \mathrm{PSL}(2, \mathbb{R})$,

ω_A agrees with the Weil–Petersson form on Teichmüller space. Note that, in two dimensions, the intersection pairing is anti-symmetric.

5.4.4. Observables and the Casson invariant

Let us now see what the previous section has taught us about observables. Again we have the characteristic classes of the universal bundle—this time restricted to the moduli space of flat connections—at our disposal to perform intersection theory. On $\mathcal{M}$, the non-vanishing components of $\frac{1}{2}(F_A + \psi + \phi)^2$ are

$$W_0 = \tfrac{1}{2}\operatorname{tr}\phi^2 \ , \qquad W_1 = \operatorname{tr}\psi\phi \ , \qquad W_2 = \operatorname{tr}\psi\psi \ , \tag{5.144}$$

giving rise to four-, three-, and two-forms on $\mathcal{M}$, respectively. For $G = \mathrm{U}(n)$ one has the additional possibility of constructing observables from $\operatorname{tr}\phi$ alone. This is of utmost importance in topological gravity in two dimensions, where the spin connection is a $\mathrm{U}(1)$ gauge field. In a careful and detailed analysis, Verlinde and Verlinde [5.86] have shown that the operators $\phi(x)$ of the Lorentz ghost for ghost are the basic building blocks of observables there. We postpone a further analysis of these matters to section 7 and continue here with the hierarchy of observables based on the second Chern class of the universal bundle.

We have already seen that in two dimensions the moduli spaces are even dimensional (in fact, symplectic). Since the rational cohomology of $\mathcal{A}/\mathcal{G}$ has a generator in two dimensions [5.87], we consider the observable $\int_{\Sigma_g} \psi\psi$. Ignoring reducible connections we have $(g > 1)$ $h_A^1 = (2g - 2)\dim G$, and we therefore expect the correlation function

$$V(\Sigma_g, G) = \left\langle \left(\int_{\Sigma_g} \psi\psi \right)^{(g-1)\dim G} \right\rangle \tag{5.145}$$

to be a non-zero topological invariant of Σ_g. In the context of topological gravity this observable appears in ref. [5.88]. But from the previous section we already know what (5.145) is! A glance at (5.143) shows that $\int_{\Sigma_g} \psi\psi$ is the symplectic form ω of $\mathcal{M}$ (this is another way of seeing why, among the candidates W_k, W_2 is the object of interest). Thus $V(\Sigma_g, G)$ is nothing other, than the symplectic volume of $\mathcal{M}$,

$$V(\Sigma_g, G) = \operatorname{Vol}(\mathcal{M}(\Sigma_g, G), \omega) \neq 0 \ . \tag{5.146}$$

Other correlation functions like

$$\left\langle \left(\int_{\Sigma_g} \psi\psi \right)^{(g-1)\dim G - 2k} \operatorname{tr}\phi^2(x_1)\operatorname{tr}\phi^2(x_2)\cdots\operatorname{tr}\phi^2(x_k) \right\rangle$$

(computing intersection numbers of the symplectic form with the curvature of the universal bundle on $\mathcal{M}$) will also be non-zero in general, and may lead to more refined invariants of Σ_g associated with $\mathcal{M}$.

In higher dimensions there may or may not be ψ zero modes. We will now take a look at the former case, and afterwards deal with the situation where $\mathcal{M}$ consists of isolated points.

The situation is then in *principle* the same as in two dimensions: the zero modes of the geometrical (A, ψ, ϕ, c) sector capture the geometry of the moduli space and the universal bundle; on these zero modes the BRST operator reduces to the exterior derivative; in particular, ψ zero modes still

represent (co)tangents to $\mathcal{M}$; observables constructed from these fields lead to cohomology classes of $\mathcal{M}$; correlation functions with the correct ghost number compute intersection numbers of $\mathcal{M}$; these are topological invariants.

In *practice*, however, one is confronted with the zero modes of the other fields of the theory, in particular those representing the higher cohomology groups of the deformation complex, which make correlation functions of operators from the geometrical sector ill-defined.

One possible strategy that suggests itself in that case is the following. In our discussion of zero modes in supersymmetric quantum mechanics we have already seen how the Faddeev–Popov procedure can be used to gauge away harmonic modes. This prescription is evidently not limited to quantum mechanics and has been explained in refs. [5.75, 5.81] within the context of topological gauge theories (cf. sections 6.2.1 and 6.2.5). The new (gauged) action is now BRST-invariant (in fact, BRST-exact) with respect to the combined BRST + Faddeev–Popov operator. This also ensures that no metric dependence is introduced into the theory by "dropping" the harmonic modes in that particular way. Since the fields in the non-geometrical sector were only introduced in the first place to reduce the theory to $\mathcal{M}$ in a well defined way, which is achieved without their zero modes, one should simply gauge all these zero modes away (this is essentially the procedure advocated in ref. [5.49]). In view of the above this is not only legitimate (compatible with BRST invariance and metric independence); it is also a reasonable thing to do, since one is then left with a situation which is essentially the same as in two dimensions or Donaldson theory, allowing one to evaluate intersection numbers of $\mathcal{M}$ in the standard way.

In three dimensions the partition function itself is well defined if the moduli space $\mathcal{M}$ consists of isolated flat connections which are—apart from the unavoidable trivial (product) connection—irreducible. As explained in our discussion of Floer homology in section 5.1.6 we are thus [if $G = \mathrm{SU}(2)$] interested in the case that M is a homology three-sphere.

The partition function will reduce to a sum of contributions from the points of $\mathcal{M}$, which—by supersymmetry—are plus or minus one, $Z(M) = \sum_{\mathcal{M}} \pm 1$. A look at the classical action $S = \int_M \mathrm{BF}_A - \chi d_A \psi$ reveals that the *relative* signs are determined by the (mod 2) spectral flow of the operator d_A, the same spectral flow that defines the relative Morse indices of Floer homology. Since d_A is the Hessian of the Chern–Simons functional whose first derivative defines a vector field on $\mathcal{A}/\mathcal{G}$, Z can be regarded as defining the Euler number $\chi(\mathcal{A}/\mathcal{G})$. From the Mathai–Quillen point of view this has also been established by Atiyah and Jeffrey [5.37] by considerations similar to those presented in section 5.2.6.

It is a result of Taubes [5.42] that this topological invariant agrees (possibly up to a sign) with the Casson invariant [5.89] $\lambda(M)$, or, more precisely,

$$Z(M) = 2\lambda(M) \ . \tag{5.147}$$

(Actually Taubes also fixes the absolute sign, but this requires considerations involving perturbations of the trivial connection, and we will not enter into this here.) $\lambda(M)$ is a very powerful invariant of three manifolds, which generalizes the classical Rohlin invariant (with which it agrees mod 2) and has already led to many interesting results in low-dimensional topology [5.89]. From the above it is apparent that the Casson invariant is closely related to Floer homology, the precise statement being that $\lambda(M)$ is the Euler characteristic of the Floer complex [5.11].

Casson's original definition of $\lambda(M)$ was somewhat different, involving Heegard splittings of M along a Riemann surface Σ_g, and intersection theory in $\mathcal{M}(\Sigma_g, \mathrm{SU}(2))$. We will now show how his definition can be recovered from the path integral point of view. Imagine splitting M along a Riemann surface Σ_g, i.e., $M = M_1 \#_{\Sigma_g} M_2$, where M_1 and M_2 are handlebodies (solid Riemann surfaces). Then—according to the general principles of quantum field theory—the path integral

over connections on the manifold M_1 with boundary $\partial M_1 = \Sigma_g$ will define a wave function Ψ_1 having support on those flat connections on Σ_g which extend to flat connections on M_1, i.e. on $\mathcal{M}(M_1, \mathrm{SU}(2)) \subset \mathcal{M}(\Sigma_g, \mathrm{SU}(2))$. Likewise the path integral over connections on M_2 will produce a wave function Ψ_2 having support on $\mathcal{M}(M_2, \mathrm{SU}(2)) \subset \mathcal{M}(\Sigma_g, \mathrm{SU}(2))$. The partition function $Z(M)$ can then be computed as the scalar product

$$Z(M) = \int\limits_{\mathcal{M}(\Sigma_g, \mathrm{SU}(2))} \Psi_1^* \Psi_2 \ , \tag{5.148}$$

and evidently only receives contributions from flat connections on Σ_g which extend to both M_1 and M_2 or—in other words—from flat connections on M. By our assumption that M is a homology three-sphere this implies that (5.148) is a sum over the points of $\mathcal{M}(M, \mathrm{SU}(2))$. The key point in Taubes' work is to show that the relative algebraic intersection numbers of $\mathcal{M}(M_1, \mathrm{SU}(2))$ and $\mathcal{M}(M_2, \mathrm{SU}(2))$ in $\mathcal{M}(\Sigma_g, \mathrm{SU}(2))$ can be determined from the spectral flow of d_A. Then one gets (denoting the total intersection number in $\mathcal{M}$ by $\#_{\mathcal{M}}$)

$$\lambda(M) = \frac{1}{2} \mathop{\#}_{\mathcal{M}(\Sigma_g, \mathrm{SU}(2))} (\mathcal{M}(M_1, \mathrm{SU}(2)), \mathcal{M}(M_2, \mathrm{SU}(2))) \ , \tag{5.149}$$

which is precisely Casson's original definition.

In the meantime, the invariant $\lambda(M)$ has been generalized to other classical groups G [5.90], where one has to come to terms with the stratification of $\mathcal{M}(M, G)$ determined by the degree of reducibility. Recovering this generalization (which involves equivariant Lagrangian perturbation theory) from the path integral point of view remains an interesting challenge, since it may teach one how to properly handle the scalar zero modes associated with reducible connections.

As can be seen from the above, super-BF theory in three dimensions bears a striking resemblance to supersymmetric quantum mechanics. This is brought out yet more clearly by the fact that there are an equal number of ψ and χ zero modes, with opposite ghost number, both representing tangents to $\mathcal{M}$. This is of course a feature also present in supersymmetric quantum mechanics, where the action (3.1) with $V = 0$ has an equal number of ψ and $\bar{\psi}$ zero modes, both representing tangents to M. In the latter case the partition function is non-zero in general, $Z(M) = \chi(M)$, despite the presence of fermionic zero modes: these zero modes appear in the action and can be soaked up by bringing down appropriate powers of the Riemann curvature tensor. An analogous situation occurs here in the 3D super-BF theory, provided that one works in a gauge with $\alpha \neq 0$ [cf. remark (i) after (5.121)]; again Z will in general be non-zero, giving $Z(M) = \chi(\mathcal{M}(M))$.

The underlying reason why three dimensions stand out in this respect is the presence of the second supersymmetry mentioned above [remark (iii) after (5.121)], which suggests that—as in supersymmetric quantum mechanics—one is really doing de Rham cohomology, this time on $\mathcal{M}$. Since a better understanding of this far-reaching analogy requires a more detailed explanation of the Atiyah–Jeffrey and Mathai–Quillen formalisms [5.37, 5.57] than we can present here, we will come back to these matters in ref. [5.61].

Further reading

A Morse theoretic interpretation of Witten type topological field theories (related to the Langevin/Nicolai map point of view) has been put forward in ref. [5.65]. The issue of metric dependence of the path integral measure of topological field theories has been addressed in refs. [5.91, 5.92].

Donaldson theory itself has naturally attracted much attention, and various aspects of its quantization and renormalization (which we shall discuss in detail in section 7) are treated in refs. [5.67, 5.80, 5.63, 5.93, 5.94] and refs. [5.47, 5.95, 5.49, 5.96], respectively, while supersymmetric extensions of Donaldson theory can be found in refs. [5.97, 5.98] as well as in ref. [5.73, 5.99]

Several other specific models in less than four dimensions have been constructed explicitly, in particular in the early days of topological field theories, when the generality of these constructions was not yet so clearly understood. Dimensionally reducing Donaldson theory to three dimensions [5.73], one obtains (depending on the manifold and boundary conditions) either a topological gauge theory of monopoles (which can alternatively be constructed directly from the Bogomol'ny equation along the lines of ref. [5.100] or section 5.4), or a theory equivalent to 3d super-BF theory. Reference [5.101] deals with supersymmetric topological theories in three dimensions. Two-dimensional theories have been discussed at length in the context of topological gravity (and here we refer to section 7 for references) as well as in refs. [5.102–5.104].

One important development we have not touched upon at all is the relation between stochastic quantization and topological field theories, discovered in ref. [5.105]. This has been further elaborated upon in refs. [5.106–5.109] with emphasis on the possibility of using stochastic quantization to establish relations among topological field theories in different dimensions. In a similar context, a relation between Yang–Mills theory in two dimensions and Donaldson theory has been proposed in ref. [5.63].

6. Schwarz type topological gauge theories

Having considered Witten type gauge theories in some detail, we now turn our attention to the Schwarz type counterparts. Recall (section 2) that these are defined as topological field theories with a non-trivial, but metric independent, classical action. We begin in section 6.1 with a review of the essential features of three-dimensional Chern–Simons gauge theory, and follow this in section 6.2 by introducing the arbitrary dimensional BF theories, which serve to model the moduli space of flat connections.

6.1. Chern–Simons theory

We shall now examine an extremely rich topological field theory, namely, the Chern–Simons model [6.1, 6.2]. Our discussion here will be, necessarily, brief; there is already a vast literature on this subject, dealing with both the three-dimensional computational point of view, as well as the two-dimensional conformal field theory aspects of the system. Our aim is simply to give a flavor of some of the special features which this model possesses, features not shared by the other models discussed in this report.

We first discuss the action and its symmetries, and present the candidate observables. The partition function, together with these observables, provide topological invariants for the three-manifold endowed with knot and link configurations. The beauty of Chern–Simons theory is that it presents an effective means of computing these invariants, for an arbitrary three-manifold, certainly in a large coupling expansion (recall that this model is of Schwarz type).

We go on to discuss some of the properties of these invariants, and follow this by making explicit contact with the 2D conformal field theory aspects of the model. Finally, in section 6.1.6, we briefly describe the Chern–Simons approach to 2 + 1 dimensional quantum gravity [6.3, 6.4].

6.1.1. Action, symmetries and observables

The Chern–Simons action is defined as follows:

$$S(A) = \frac{k}{4\pi} \int_M \text{tr}(A \wedge dA + \tfrac{2}{3} A \wedge A \wedge A) \ , \tag{6.1}$$

where M is an oriented three-dimensional manifold on which the theory is defined; $A = A^a T^a$ is a connection on a principal G-bundle over M, and T^a is a representation of the structure group G, which we take to be SU(n). The normalization of the tr will be specified presently. Since the integrand in (6.1) is a volume form on M, and is defined without any reference to the Hodge (star) duality operator, it is clearly metric independent and thus corresponds to a Schwarz type model. Given this action, our aim is to study (among other things) the partition function $Z = \int dA \, e^{iS}$. A minor point, which is perhaps worth a mention, is the fact that, since S is a first order system, a factor of i is required in the definition of the Euclidean path integral.

Let us now discuss the symmetries of the action. Clearly, the action is invariant under the usual infinitesimal Yang–Mills type gauge transformations, namely $\delta A_\alpha(x) = D_\alpha \epsilon(x)$, where $D_\alpha = \partial_\alpha + [A_\alpha, \ ]$ is the covariant derivative. However, given that gauge transformations in the present case are maps $g : M \to G$, together with the fact that $\pi_3(G) = \mathbb{Z}$ for any compact simple group G, one must exercise caution in studying the behavior of S under gauge transformations which are not connected to the identity. Indeed, the following situation arises [6.5–6.7]: defining a finite gauge transformation by

$$A_\alpha \to A_\alpha^g = g^{-1} D_\alpha g \ , \tag{6.2}$$

we find that

$$S(A^g) = S(A) + 2\pi k S(g) \ , \tag{6.3}$$

where

$$S(g) = \frac{1}{24\pi^2} \int_M \text{tr}(g^{-1} dg)^3 \tag{6.4}$$

is the winding number of the map g. Choosing a normalization for the tr to be $\text{tr}\, T^a T^b = -\tfrac{1}{2}\delta^{ab}$, in the fundamental representation, and demanding gauge invariance, fixes the normalization of k to be an integer. In other words, we now see that e^{iS} is indeed a single-valued functional which is invariant under all gauge transformations g.

A point worth noting here is that, even on $\mathbb{R}^3$, one finds the necessity of restricting to integer k values [6.8]. This arises due to the necessity of imposing boundary conditions, a suitable choice being, for example, to let $g \to 1$ at infinity. Since such a condition essentially compactifies $\mathbb{R}^3$ to S^3, the quantization condition again follows.

In addition to studying the partition function, one would also like to construct metric independent observables. In actual fact, it turns out that one can define suitable observables which are essentially computable on any three-manifold. We shall specify more precisely what is meant by "essentially computable", in section 6.1.4. Consider a closed oriented curve in M, and let R be an irreducible representation of G. We should point out that such a curve C corresponds to an embedding from the circle S^1 to the manifold M, and is called a knot. One then considers the Wilson loop operator defined by

$$W_R(C) = \text{tr}_R \, \text{P} \exp \int_C A \ , \tag{6.5}$$

where the trace is taken in the R representation, and P denotes path ordering. Since W is the holonomy of the one-form A around the loop C, it is clearly metric independent. The task then is the following: let M be an oriented three-manifold, and choose r oriented and non-intersecting knots $C_i, i = 1,...,r$ (the union of these knots is called a link). Assigning an irreducible representation R_i of G to each knot, we consider the observable [6.2]

$$Z(M; C_i, R_i) = \int \mathrm{d}A \, e^{iS} \prod_{i=1}^{r} W_{R_i}(C_i) \ . \tag{6.6}$$

Since the classical action and observables we have defined are metric independent, the hope is that the associated quantum partition function and correlation functions will enjoy similar properties. This is not, a priori, guaranteed; the aim is to make sense of these objects in a metric independent way, within the realms of quantum field theory. If this can be achieved, the Chern–Simons theory will indeed be a topological field theory.

Before treating the evaluation of the partition function and observables, we pause to discuss another important aspect of this theory, namely, its phase space.

6.1.2. Phase space

We first note that the equations of motion which follow from the action (6.1) are

$$F_A = \mathrm{d}A + \tfrac{1}{2}[A, A] = 0 \ . \tag{6.7}$$

Thus, the stationary points of the action are flat connections, and the reduced phase space is the moduli space of these connections. We have already discussed in section 5.4.3, at some length, the properties of this space. This allows us to be more brief here, and we will concentrate on the special features for the case in hand.

Canonical quantization. Our aim here is to give a more explicit description of the Hilbert space of Chern–Simons theory [6.9–6.14]. For this purpose it is most convenient to adopt the canonical approach, in which we take the three-manifold to be of the form $M = \Sigma \times \mathbb{R}$, for some genus g surface Σ. The coordinate along the real line $\mathbb{R}$ can be regarded as time. Quantization on this space will then produce the associated Hilbert space $\mathcal{H}_\Sigma$, in which the states of the system are, loosely speaking, functionals on the space $\mathcal{A}_\Sigma$ of gauge fields on Σ.

Writing $A_\alpha = (A_i, A_0)$, where A_0 is the time component of the gauge field, the action (6.1) takes the Gaussian form

$$S = -\frac{k}{4\pi} \int \mathrm{d}t \int_\Sigma \epsilon^{ij} \, \mathrm{tr} \left(A_i \frac{\mathrm{d}}{\mathrm{d}t} A_j - A_0 F_{ij} \right) \ , \tag{6.8}$$

allowing us to read off the Poisson brackets for the system,

$$\{A_i^a(x), A_j^b(y)\} = (4\pi/k)\epsilon_{ij}\delta^{ab}\delta(x - y) \ . \tag{6.9}$$

In order to set up the canonical formalism for a second order system, one first introduces canonical momenta for the fields, and then uses these to rewrite the original action in first order form. The Poisson brackets for the canonically conjugate variables can then be read off. In the present instance, we already have a first order system; this procedure is unnecessary, and the Poisson bracket structure is immediately evident.

We must now quantize these commutation relations, subject to the Gauss law constraint $(k/4\pi)F_{ij}^a = 0$, which is enforced by the multiplier field A_0. At this point we have two options: we can either first quantize the system, obtain the wave functions, and then construct the physical Hilbert space by imposing the constraints; or we can first impose the constraints, and then quantize the reduced phase space. We will discuss the second option here (for a comparison of the two approaches in the present context cf. ref. [6.10]). The constraint surface is the set of flat connections in $\mathcal{A}_\Sigma$. On this constraint surface the first class constraints $(k/4\pi)F_{ij}^a$ act by gauge transformations, and the reduced phase space is the moduli space $\mathcal{M} = \mathcal{M}(\Sigma, G)$ of flat connections on Σ.

We have already discussed many properties of $\mathcal{M}$ in section 5.4.3; the most important one for our present purposes is that $\mathcal{M}$ is a symplectic manifold. A direct way of seeing this is to note that $\mathcal{A}_\Sigma$ is an infinite dimensional symplectic manifold [with symplectic structure given by (6.9)], and that $\mathcal{M}$ is a symplectic quotient of $\mathcal{A}_\Sigma$ with moment map F_{ij}^a. As such $\mathcal{M}$ inherits a unique symplectic structure from $\mathcal{A}_\Sigma$ by the Marsden–Weinstein construction [6.15]. First reducing and then quantizing thus amounts to quantizing the symplectic manifold $\mathcal{M}(\Sigma, G)$. Let us digress briefly to understand what this means.

In general, when one wishes to quantize a classical system, the first step is to define the canonical coordinates q^i and momenta p_i. In a coordinate representation, for example, the quantum Hilbert space is the space of square integrable functions of the coordinates q^i. The operators $\hat{q}^i$ then act as multiplication operators on the wave functions, producing the c-number q^i, while the momentum operators $\hat{p}_i$ act by differentiation, $\hat{p}^i \sim \partial/\partial q^i$. This procedure works whenever the phase space N of the classical system is a cotangent bundle, $N = T^*Q$, because then there is a preferred separation of the space variables into coordinates and momenta.

Another representation which is often used is the so-called holomorphic (or coherent state) representation (in quantum mechanics this is also known as the Bargmann representation). One introduces the variable $a^i = p^i + iq^i$ and its complex conjugate $\bar{a}^i = p^i - iq^i$, i.e., one chooses an identification between $N = \mathbb{R}^{2n}$ and $\mathbb{C}^n$. Wave functions are taken to depend only on a^i. Then the operators $\hat{a}^i$ act as multiplication operators, producing the c-number a^i; this is in distinction to the coordinate representation, where the $\hat{a}^i$ act as a combination of multiplication and differentiation operators. For $N = \mathbb{R}^{2n}$ this representation is equivalent to the above coordinate (or Schrödinger) representation, the unitary equivalence being provided by the so-called Bargmann kernel. The advantage of the holomorphic representation is, however, that it is also available when N is not a cotangent bundle, provided that N admits a complex structure which is compatible with its phase space (symplectic) structure ω, or, in other words, if N is a Kähler manifold. In that case, the appropriate generalization of the above prescription is that the Hilbert space is the space of holomorphic sections of a complex line bundle L (the prequantum line bundle) over N whose curvature is the Kähler form ω. Since, as such, ω represents the first Chern class of L, such a bundle only exists if ω is an integral two-form on N.

In general—if neither of these additional structures (cotangent bundle, Kähler form) is available—there is no canonical way of cutting the phase space variables in half to construct a quantum Hilbert space, and although this splitting can certainly also be done in a non-canonical way (in the parlance of geometric quantization this is a choice of polarization), the resulting quantum theory will depend on such an (arbitrary) choice.

With this in mind let us now return to our problem of quantizing $\mathcal{M}(\Sigma, G)$. For compact G, $\mathcal{M}$ is compact as well, and can therefore not possibly be a cotangent bundle. Fortunately it turns out, however, that a choice of complex structure J on Σ equips $\mathcal{M}$ with a complex structure compatible with its symplectic structure, thus making it a Kähler manifold which we shall denote by $\mathcal{M}_J$. This

is a consequence of the Narashiman–Seshadri theorem [6.16], which identifies moduli spaces of flat vector bundles on Riemann surfaces with certain moduli spaces of holomorphic vector bundles. We thus have the holomorphic representation at our disposal to quantize $\mathcal{M}$. The above integrality condition on the symplectic form is equivalent to the requirement that k in (6.9) be an integer, and this is therefore another way of deriving the quantization condition on k as a consistency condition for the quantum theory. It now follows from either the general arguments of Quillen [6.17] or the more explicit construction of Pressley and Segal [6.18, 6.14] that the prequantum line bundle is a power (depending on G and the Chern–Simons coupling constant k) of the determinant line bundle associated with the family of operators $\bar{\partial}_A$ on $\Sigma = \Sigma_J$.

This then yields a Hilbert space $\mathcal{H}_\Sigma^J$, and it remains to investigate the residual quantization ambiguity inherent in the choice of complex structure J. From the three-dimensional point of view, no a priori specification of J was required, and one might therefore hope to be able to prove that there is a canonical identification of Hilbert spaces $\mathcal{H}_\Sigma^J$ and $\mathcal{H}_\Sigma^{J'}$, constructed from different complex structures J and J' on Σ. Actually this may be a little too much to ask for, since physical states correspond to rays in a Hilbert space, and the correct physical requirement should therefore be that at least the projective Hilbert spaces can be canonically identified.

To see what this amounts to geometrically, imagine smoothly (holomorphically) varying J. This gives a family of Hilbert spaces which can be regarded as forming a holomorphic vector bundle over Teichmüller space, the space of complex structures on Σ. The question is then if this vector bundle has a canonical *projectively flat* connection. Having initially had the status of a conjecture for some time, this question has in the meantime been answered in the affirmative [6.11, 6.12].

It is here that contact is made with the modular geometry approach to conformal field theory, initiated by Friedan and Shenker [6.19] and axiomatized by Segal [6.20]. From this point of view the Hilbert space $\mathcal{H}_\Sigma^J$ of Chern–Simons theory is nothing but the space of conformal blocks of a conformal field theory on Σ_J, the energy–momentum tensor of the latter governing the projectively flat parallel transport via the Knizhnik–Zamolodchikov equation [6.21]. It was this observation by Witten, that the Hilbert spaces of Chern–Simons theory provide a realization of Segal's axioms, which initially led to the discovery of the relation between Chern–Simons gauge theory and conformal field theory. In section 6.1.5 we will describe a more pedestrian way to understand this relation, which does not require a detailed knowledge of conformal field theory and modular geometry.

6.1.3. Evaluation of the partition function

We would now like to establish some formal properties of the Chern–Simons partition function. There will be a little overlap here with the material presented in section 8.4, although our emphasis in the latter is on the calculational aspects.

We wish to consider the partition function in the large-k (i.e. semi-classical) approximation. Such a limit corresponds to considering fluctuations about the stationary points of the action; these are the flat connections. Let us assume that the moduli space is zero dimensional, i.e., consisting of a finite number of isolated flat connections. In such a case, the partition function takes the form [6.2] $Z = \sum_i Z(A^i)$, where each $Z(A^i)$ is the one-loop contribution to Z, evaluated at the flat connection A^i, labelled by i. In order to give an explicit expression for $Z(A^i)$, we first need to quantize the action.

Since the symmetry here is of the standard Yang–Mills variety, no unnecessary gymnastics need to be performed in order to obtain the quantum action. For the purposes of calculation, we first decompose the A field as $A = A^i + A^q$, where A^i is the background flat connection, and A^q is the

quantum fluctuation. We also choose the background covariant gauge $D(A^i) \cdot A^q = 0$. The quantum action is then given by

$$S_q = \frac{k}{4\pi} \int d^3x \, \mathrm{tr}\, \epsilon^{\alpha\beta\gamma} (A_\alpha \partial_\beta A_\gamma + \tfrac{2}{3} A^\alpha A^\beta A^\gamma)$$

$$+ \int d^3x \, \mathrm{tr}[bD(A_i) \cdot (A - A_i) + \bar{c}D(A_i) \cdot D(A)c] \, , \tag{6.10}$$

where $\bar{c}, c$ are the ghosts and b is the multiplier field. Expanding (6.10) to second order in the fluctuations (bearing in mind that $\bar{c}, c, b$ are purely quantum fields) we obtain the quadratic action

$$S_q^2 = \int d^3x \, \mathrm{tr}(\epsilon^{\alpha\beta\gamma} A_\alpha^q D_\beta A_\gamma^q + bD \cdot A^q + \bar{c}D^2 c) \, , \tag{6.11}$$

where it is understood that the covariant derivative is with respect to the background flat connection A^i. This yields the following expression:

$$Z^i = e^{ikI(A^i)} \det[-D^2] \det^{-1/2}[H(A^i)] \, , \tag{6.12}$$

where D^2 is the ghost operator, $H(A^i)$ is the first order operator for the A^q, ϕ system, and I is defined as the Chern–Simons action S without the factor of k. Our task now is to study this ratio of determinants. The subtlety which arises (and which is described in more detail in section 8.4) is the fact that $\det H$ possesses a non-zero phase; other examples of this situation can be found in ref. [6.22]. We have

$$\det^{-1/2} H(A^i) = |\det^{-1/2} H(A^i)| e^{i\pi\eta_{H(A^i)}(0)/4} \, , \tag{6.13}$$

where

$$\eta_{H(A^i)}(s) \equiv \sum_n \mathrm{sign}(\lambda_n) |\lambda_n|^{-s} \, . \tag{6.14}$$

Here, $\eta_{H(A^i)}(s)$ is defined via the eigenvalues λ_n of the operator H at the flat connection A^i [6.23–6.25]. In section 8.4.3 we present an explicit computation of this function; here, we shall content ourselves with general arguments. The aim is to establish the topological nature of the partition function. At this point we have two quantities, the absolute value of the ratio of determinants, and the phase of $\det H$. The former corresponds to the Ray–Singer torsion [6.26, 6.27, 6.1] and is certainly metric independent; we shall denote it by T_i. The problem now is to understand the η phase. Appealing to the Atiyah–Patodi–Singer [6.23] index theorem we have the result

$$\tfrac{1}{4}[\eta_{H(A^i)}(0) - \eta_{H(0)}(0)] = (c_v/\pi)I(A^i) \, , \tag{6.15}$$

where c_v is the quadratic Casimir in the adjoint representation, and is an integer. Using this, we can write the partition function as

$$Z = e^{i\pi\eta_{H(0)}(0)/4} \sum_i e^{i(k+c_v)I(A^i)} T_i \, . \tag{6.16}$$

The first thing to notice is that the integer coupling constant k has been shifted by an amount c_v; we shall not dwell on this issue here, but will say more about it in section 8.4. This phenomenon was also noticed by explicit calculation in ref. [6.28]. Our main concern now is to understand the phase factor $\eta_{H(0)}(0)$. It is the η-function of the operator H coupled to the trivial connection

$A = 0$ and also coupled to the metric which was used in effecting the gauge fixing; it is thus metric dependent. Since our main goal is to maintain general covariance, we may wonder if it is possible to somehow modify this phase into a metric independent form. This is indeed the case, and again according to the Atiyah–Patodi–Singer theorem, we know that the combination $\frac{1}{4}\eta_{\text{grav}} + (1/24\pi)I(g)$ is a topological invariant, modulo a subtlety which we now discuss. In the above, $\eta_{H(0)}(0) = \dim(G)\eta_{\text{grav}}$, with η_{grav} being the purely gravitational η-function, and

$$I(g) = \frac{1}{4\pi} \int_M \text{tr}(\omega\, d\omega + \tfrac{2}{3}\omega^3) \,, \tag{6.17}$$

where ω is the Levi–Civita spin connection. The problem is that to define $I(g)$ as a number requires one to choose a particular trivialization of the tangent bundle of M. Although this can be done, different (i.e. homotopically inequivalent) trivializations will yield different values for $I(g)$. The rule, however, is that two trivializations which differ by a relative twist of s units, are related by $I(g) \rightarrow I(g) + 2\pi s$. This is similar to the occurrence of the winding number in the variation of the Chern–Simons action in (6.3). We can now consider the final form for the partition function,

$$Z = \exp\{i\pi \dim(G)[\tfrac{1}{4}\eta_{\text{grav}} + (1/24\pi)I(g)]\} \sum_i e^{i(k+c_v)I(A^i)} T_i \,. \tag{6.18}$$

To interpret this result, we define a framed manifold as follows [6.2, 6.29–6.36]: A framed manifold is one that is presented with a homotopy class of trivializations of the tangent bundle. Given this definition, we now see that the Chern–Simons partition function provides us with a topological invariant of framed three-manifolds, together with a prescription for how it behaves under a change in framing, viz., $Z \rightarrow Z\, e^{2\pi i s\, \dim G/24}$.

We remark here that we have implicitly assumed that $k > 0$ in the above analysis. Since the η-function is odd under a sign change in k, it is straightforward to establish the general k result, for details see section 8.4. We should also note that there will, in general, be situations where the above analysis is incomplete. We have assumed that the moduli space consists of isolated points; when the dimension of moduli space is non-zero, the discrete sum given above will be replaced by an integral over moduli space. Furthermore, one also needs to be aware that situations may arise where the ghost or multiplier fields possess zero modes, i.e., where reducible connections are present; more care is again required to deal properly with such cases.

We conclude this section with a discussion of an important feature of the partition function, that is, its behavior with respect to the connected sum of manifolds [6.2]. This is a concept already introduced in section 3.8.6.

Let M be the connected sum of M_1 and M_2, joined along a two-sphere S^2 (see section 5.4.4). Now, according to the tenets of quantum field theory, if we consider our action to be defined on a manifold with a single boundary component, then the resulting path integral must be performed with respect to field variables taking prescribed values on the boundary. Evaluating the path integral yields an object which is a functional of these boundary data; this object is a wave function or, in other words, a vector in the Hilbert space of the theory associated with the boundary. As an aside, if, for example, we choose a manifold with two disjoint boundary components, we obtain the propagation amplitude between the states defined on the two boundaries.

Resorting to a little visual gymnastics, one sees that the path integral over M_1 with boundary S^2 yields a vector in the S^2 Hilbert space; let us call this state ψ. Correspondingly, the integral over the other half of the manifold, M_2 with boundary S^2, yields a vector $\tilde{\psi}$ in the dual vector space. The entire path integral then gives the inner product between these two states, $Z(M) = (\psi, \tilde{\psi})$. At this

point, we know neither ψ nor $\tilde{\psi}$. However, let us repeat this performance by replacing M with S^3. In this way we find $Z(S^3) = (v, \tilde{v})$, where $v, \tilde{v}$ correspond to the vectors obtained by integrating over the two halves of S^3. Again, neither v nor $\tilde{v}$ is known. We can now recall our discussion in section 6.1.2 about the nature of the Hilbert space for Chern–Simons theory; and it is easy to show that for the case of S^2 this Hilbert space is one-dimensional. In the language of that section, the moduli space $\mathcal{A}_\Sigma$ associated with the Riemann sphere is a single point, and the line bundle L is the complex plane. As such, any two vectors enjoy the remarkable property that they are linearly dependent! We have $\psi = av$ and $\tilde{\psi} = b\tilde{v}$, for complex constants a and b. Using this information, we deduce that [6.2]

$$Z(M)Z(S^3) = (\psi, \tilde{\psi})(v, \tilde{v}) = (\psi, b\tilde{v})(v, b^{-1}\tilde{\psi}) = Z(M_1)Z(M_2) . \tag{6.19}$$

This property, that the partition function behaves multiplicatively with respect to connected sums, can also be used to establish a simple result when M contains a trivial link. Consider r unlinked and unknotted circles $C_i, i = 1, ..., r$ in S^3. By iterating the above argument, and avoiding to cut any of the knots, we find that

$$\frac{Z(S^3; C_1, ..., C_r)}{Z(S^3)} = \prod_i \frac{Z(S^3; C_i)}{Z(S^3)} . \tag{6.20}$$

The above two formulae are well known to knot theorists; however, it is pleasing to see how simply they arise from the path integral point of view.

6.1.4. Evaluation of the observables: knot invariants

The central idea in Chern–Simons theory is that it offers a means of computing invariants of knot and link configurations on an arbitrary three-manifold. Now, within knot theory an important role is played by the so-called skein relations [6.29–6.35]; the main property of such a relation is that it provides a way of relating a particular knot (or link) configuration to a simpler one. The first step is to picture the link projected to the plane; as such, there will be a finite number of "crossings", that is, points at which two projected lines meet. One must then distinguish between an over-crossing, an under-crossing, and a zero-crossing. The skein relation allows us to reduce the number of crossings; and by iteration, this relation essentially allows us to compute the invariant of a given link.

Confident of the visualization powers of the reader, we proceed to describe this construction in the simplest case; we consider S^3 with gauge group $SU(n)$, and consider a link L with all the component Wilson lines lying in the fundamental representation [6.2]. Let us focus our attention on a particular crossing where two Wilson lines meet. We imagine encompassing this crossing with a two-sphere, and removing it from S^3; the cut surface now consists of two pieces. One of these, denoted B_R, is a three-ball with boundary S^2 on which there are four marked points; these points are connected on the interior of B_R by the two Wilson lines. The remaining part of the surface is again a three-ball, denoted B_L, with an S^2 boundary containing four marked points; in this case the marked points are connected on the interior by the residual (complicated) part of the link. It may be helpful to bear in mind the two-dimensional analogue of this picture: consider the same crossing on a two-sphere; we then encompass this with a circle S^1, which is the boundary of a two-ball (equivalently, a two-disc).

Having conquered the visual problem, we can now proceed. As before, the path integrals on B_L, B_R determine vectors χ, ψ in the Hilbert space. We thus have $Z(S^3; L) = (\chi, \psi)$. As before, we know neither χ nor ψ; indeed, if these were directly computable, the present gyrations would

be superfluous. Now, in this case, the Hilbert space is that associated with the two-sphere containing four marked points in the fundamental representation. The dimension of this space can be determined from group properties to be two. Thus, any three vectors obey a relation of linear dependence; this is called the skein relation. We have

$$\alpha\psi + \beta\psi_1 + \gamma\psi_2 = 0 \ , \tag{6.21}$$

where ψ_1, ψ_2 are any two other vectors in the Hilbert space, and α, β, γ are complex numbers. Taking the inner product of (6.21) with χ, we find

$$\alpha(\chi,\psi) + \beta(\chi,\psi_1) + \gamma(\chi,\psi_2) = 0 \ . \tag{6.22}$$

We can rewrite the above equation more suggestively, as follows:

$$\alpha Z(S^3;L_+) + \beta Z(S^3;L_0) + \gamma Z(S^3;L_-) = 0 \ , \tag{6.23}$$

where we have introduced the notation L_+, L_0, L_- corresponding to an over-crossing, a zero-crossing, and an under-crossing. In order for such a relation to be of practical use, we need to determine the coefficients α, β, γ; we shall mention two methods for doing this. The first resorts to information available from conformal field theory [6.37–6.39], while the second maintains the three-dimensional point of view [6.40, 6.41].

In the former, we use the notion of the braiding (or half-monodromy) matrix [6.37]. Given a configuration of two Wilson lines with an over-crossing (L_+), we define an operator B which, acting on L_+, produces the configuration L_0. A repeated application then produces L_-. Since, in this instance, B acts in a two dimensional space, it has two eigenvalues, λ_1 and λ_2. We can now use the fact that every matrix obeys its own characteristic equation (by the Cayley–Hamilton theorem), to write the relation

$$(\chi, [(B - \lambda_1 I)(B - \lambda_2 I)]\psi) = (\chi, [B^2 - B(\lambda_1 + \lambda_2) + \lambda_1\lambda_2 I]\psi) = 0 \ . \tag{6.24}$$

Thus, with the convention that $B\psi = \psi_1$ and $B^2\psi = \psi_2$, we can now compute the coefficients in the skein relation from the eigenvalues of the braiding matrix, which are known [6.37–6.39].

It is perhaps useful to consider explicitly the case of SU(2), where the fundamental representation is the spin $j = \frac{1}{2}$ representation. Given the fact that

$$\tfrac{1}{2} \otimes \tfrac{1}{2} = 1 \oplus 0 \ , \tag{6.25}$$

we can define the eigenvalues of B via the conformal weights of the fields which transform as the $j = \frac{1}{2}, 0, 1$ representations. Noting that the conformal weight of a spin j field is $h_j = j(j+1)/(k+2)$ (see, e.g., ref. [6.42]) we find the eigenvalues

$$\lambda_1 = -e^{-i\pi(2h_{1/2}-h_1)} = -e^{i\pi/2(k+2)} \ , \qquad \lambda_0 = +e^{-i\pi(2h_{1/2}-h_0)} = e^{-3i\pi/2(k+2)} \ . \tag{6.26}$$

The relative plus and minus sign in the above eigenvalues arises due to the occurrence of the representation, either symmetrically or anti-symmetrically, in the decomposition (6.25). Equation (6.24) now takes the form (upon multiplication by $q^{1/4} = e^{i\pi/2(k+2)}$; we note for later use that q is called the monodromy parameter)

$$q^{1/4}Z(S^3;L_+) - q^{-1/4}Z(S^3;L_-) = (q^{1/2} - q^{-1/2})Z(S^3;L_0) \ . \tag{6.27}$$

It should be pointed out that the coefficients in the skein relation depend upon the "framing of knots", a concept which we shall define shortly. Suffice it to say at this point that the form given above corresponds to what is known as "vertical framing" [6.2, 6.43].

The alternative three-dimensional viewpoint allows one to compute the skein coefficients in a large-k approximation; nevertheless, geometrically it is quite elegant [6.40, 6.41].

The basic idea is to consider an arbitrary Wilson line with defining contour C; we then perform an infinitesimal deformation of C. It is well known, see, for example, ref. [6.44], that such a variation produces a factor of $F_{\mu\nu}(x)$, where x is the point on C at which the deformation is implemented. Thus, we have

$$U(x_1, x_2) \rightarrow U(x_1, x) \Sigma^{\mu\nu} F_{\mu\nu}(x) U(x, x_2) \ , \tag{6.28}$$

where $U(x, y)$ is a Wilson line operator, and is defined as in (6.5), with respect to an open contour C; $\Sigma^{\mu\nu}$ is the infinitesimal area element centered at x, and there is no summation here over μ, ν.

One now uses the fact that the equation of motion for the theory is proportional to the curvature tensor; this allows us to replace $F_{\mu\nu}$ in (6.28) by the derivative of the action with respect to the gauge potential. More care is required when working with the fully gauge fixed action; however, it can be shown that these additional terms do not affect the analysis [6.45]. It is straightforward to establish that

$$\langle F^a_{\mu\nu} O_1 O_2 \cdots \rangle = Z^{-1} \frac{4\pi i}{k} \int dA \ e^{iS} \epsilon_{\mu\nu\rho} \frac{\delta}{\delta A^a_\rho(x)} (O_1 O_2 \cdots) \ . \tag{6.29}$$

Thus, the presence of the curvature tensor has been replaced by a derivative, which now acts on the remaining observables. This is the key point, as such an effect is computable. For example, we can now evaluate the expectation value of the deformed Wilson line in (6.28). By a similar argument, one can relate an over-crossing to an under-crossing; that is, performing an infinitesimal deformation of an L_+ configuration will yield an L_- Wilson line, together with an infinitesimal correction. This correction is then computable with the above identity (6.29). In other words, such a procedure provides a geometrical means of obtaining the skein relation. We should remark that, since we are dealing with infinitesimal deformations, this computation yields the $O(1/k)$ term in the skein coefficients. In practice, the computation relies on the existence of a Fierz identity for the generators of the Lie algebra; such identities always exist, although in many cases their form may be quite unwieldy.

We should point out a subtlety in defining Wilson line expectation values; this is related to what is called the framing of knots, which we now define [6.2, 6.29–6.35]. Given a knot with defining contour C, one chooses a normal vector field along C; this produces a deformed contour C'. One can then consider an infinitesimal ribbon between the two paths C and C'; this is called the framing of the knot C. One can now define, for example, the self-linking number of a knot as the linking number between C and C'. However, such a definition clearly depends on the topological class of the normal vector field. In order to keep track of this situation, we require a rule for how a given knot expectation value changes under a change in framing. Indeed this is possible [6.2], and one should recall a similar feature emerging in our discussion of the framing of the three-manifold in section 6.1.3.

We conclude this section with our promise to specify more precisely the meaning of the phrase "effectively computable", which was introduced in section 6.1.1. In the above, we have established some basic properties of the partition function when $M = S^3$, and observables defined thereon. In order to generalize these results to an arbitrary three-manifold, one introduces the notion of surgery. The general idea is the following: we begin with an arbitrary three-manifold M, together with an embedded circle C. The circle is then thickened to a solid torus, which is then excised from M. As is now familiar, we have two manifolds, each with a boundary; we perform a diffeomorphism on the boundary of the solid torus, and then re-unite the two manifolds. This gives a new three-manifold

$\tilde{M}$. It can be shown that an arbitrary M can, in this way, be reduced to S^3; such a procedure is called surgery.

6.1.5. Connections with conformal field theory

We have already seen, in a somewhat abstract form, via the projectively flat bundle condition of section 6.1.2, the connection between Chern–Simons theory and two-dimensional conformal field theory. We shall now briefly describe how more explicit contact can be made. We treat the simplest of examples and take the manifold M to be $M = D \times \mathbb{R}$, where D is the two-dimensional disc; we are following here the treatment of ref. [6.9].

Recall (6.8), which is written for a manifold of the form $M = \Sigma \times \mathbb{R}$ as

$$S = -\frac{k}{4\pi} \int dt \int_D \epsilon^{ij} \operatorname{tr}\left(A_i \frac{d}{dt} A_j - A_0 F_{ij}\right) .$$ (6.30)

In the present case the boundary of M is a cylinder $\partial M = S^1 \times \mathbb{R}$, and we must be careful with boundary contributions to the action. It is easy to see that with the choice $A_0 = 0$ on the boundary, (6.8) is again the action of interest. Integrating over A_0 enforces the constraint $F_{ij} = 0$. This can be solved as follows: $A_i = -\partial_i g \cdot g^{-1}$, where g is a single-valued function $g : D \times \mathbb{R} \rightarrow G$. The single-valuedness of g is possible due to the fact that the disc is simply connected.

If we now change variables $A_i \rightarrow g$ (with unit Jacobian), and simply write the action (6.30) in terms of g we find

$$S = \frac{k}{4\pi} \int_{\partial M} \operatorname{tr}(g^{-1}\partial_\phi g g^{-1}\partial_t g) \, d\phi \, dt + \frac{k}{12\pi} \int_M \operatorname{tr}(g^{-1} dg)^3 .$$ (6.31)

where ϕ is the coordinate on S^1. One can immediately recognize this as being a WZW model, written in chiral coordinates [6.7]. In (6.31), the kinetic term appears in an off-diagonal form; this is the usual situation when ones uses complex coordinates on the plane, for example. However, here one obtains a chiral WZW model in terms of the real coordinates ϕ, t on $S^1 \times \mathbb{R}$. We thus see that, already at this level, we are making explicit contact with conformal field theory via the WZW model. The fact that the chiral form appears in terms of real coordinates ϕ, t, rather than the complex combinations $z = t + i\phi$, $\bar{z} = t - i\phi$, has an important consequence for the symmetry of the model, as we now discuss.

The symmetry of this action is given by [6.2, 6.9]

$$g(\phi, t) \rightarrow \tilde{h}(\phi) g h(t) .$$ (6.32)

It should be pointed out that the above symmetry is specified by the requirement that it preserves the chosen boundary conditions. The invariance under $h(t)$ corresponds to a local gauge symmetry which needs to be fixed; the remaining invariance, under $\tilde{h}(\phi)$, is a global symmetry, and thus the Hilbert space of the theory will carry representations of this group.

We can now proceed with the canonical quantization of the model; the solution to this problem is already well known [6.7], and it suffices to make a few remarks. Given an action of the form

$$S = \int dt \, L_i(\Phi) \frac{d\Phi^i}{dt} ,$$ (6.33)

where Φ^i are the field variables, one can consider its variation as follows:

$$\delta S = \int dt \, \omega_{ij} \delta\Phi^i \, d\Phi^j / dt .$$ (6.34)

Here $\omega_{ij} = \partial_i L_j - \partial_j L_i$. In fact ω can now be interpreted as the symplectic form, and the Poisson brackets of the system are defined via

$$[X, Y]_{\text{PB}} = \omega^{ij} \frac{\partial X}{\partial \Phi^i} \frac{\partial Y}{\partial \Phi^j} \, , \tag{6.35}$$

where ω^{ij} is the inverse matrix. In the present case, a similar construction can be set up, and one can check that the variation of the chiral WZW model is given by

$$\delta S = \frac{k}{4\pi} \int \mathrm{d}\phi \, \mathrm{d}t \, \mathrm{tr} \left(g^{-1} \delta g \frac{\mathrm{d}}{\mathrm{d}\phi} g^{-1} \frac{\mathrm{d}g}{\mathrm{d}t} \right) \, . \tag{6.36}$$

From here, one constructs the currents

$$J_\phi = \frac{k}{2\pi} \frac{\mathrm{d}g}{\mathrm{d}\phi} g^{-1} \, , \tag{6.37}$$

and a little work establishes that their Poisson bracket structure is precisely the defining relations for a Kač–Moody algebra, with a central extension proportional to the Chern–Simons coupling k.

An important remark is the following: The WZW model written in complex coordinates has an off-diagonal kinetic term of the form

$$\int \mathrm{tr}(g^{-1}\partial_z g g^{-1}\partial_{\bar{z}} g) \, \mathrm{d}z \, \mathrm{d}\bar{z} \, .$$

The corresponding symmetry is given by $g(z,\bar{z}) \to \Omega(z)g(z,\bar{z})\overline{\Omega}^{-1}(\bar{z})$, and this leads to the result that there are two commuting Kač–Moody algebras, given by the currents [6.7] $J_z \sim k(\partial_z g)g^{-1}$, $J_{\bar{z}} \sim k(\partial_{\bar{z}} g^{-1})g$. The fact that the Chern–Simons action leads to a chiral WZW model with action and symmetry given by (6.31), (6.32) means that there is a single Kač–Moody algebra associated with the model, corresponding to $\tilde{h}(\phi)$.

6.1.6. 2 + 1 gravity as a Chern–Simons theory

We will now describe yet another important application of Chern–Simons theory. It has oft been wondered if quantum gravity could be endowed with a gauge theory interpretation, see, for example, ref. [6.46] and references therein. The fields of interest here are the vierbein and the spin connection, and the basic idea is that these would combine to form a gauge field for the non-compact Poincaré group $\text{ISO}(d - 1, 1)$, where the symbol I refers to the addition of the translations to the Lorentz group $\text{SO}(d - 1, 1)$. The spin connection could play the role of the gauge field for the Lorentz group, while the vierbein takes care of the translations. However, in addition to a description in terms of a hypothetical gauge field, one also requires an action which will describe the corresponding dynamics of the system (and furthermore, these dynamics should be those of general relativity).

Rather than dwell on histories, let us proceed and describe how 2 + 1 dimensional gravity (with the usual Einstein–Hilbert action) can be re-interpreted as a gauge theory, in which the gauge field action is simply the Chern–Simons action for the Poincaré group [6.3, 6.4]. There have also been many previous studies of 2 + 1 dimensional gravity, as a more tractable alternative to the 3 + 1 dimensional theory [6.47].

Let M be 2 + 1 Lorentzian space-time; the initial field content is given by the dreibein $e_i{}^a$ and the spin connection $\omega_i{}^a{}_b$, where tangent space indices are i, j, k and Lorentz indices are a, b, c.

Consider now the Einstein–Hilbert action written in terms of these fields,

$$S = \frac{1}{2} \int_M \epsilon^{ijk} \epsilon_{abc} e_i^a (\partial_j \omega_k^{bc} - \partial_k \omega_j^{bc} + [\omega_j, \omega_k]^{bc}) \ . \tag{6.38}$$

The aim is to show how this action may be expressed in Chern–Simons form. To this effect, let us reconsider our original action (6.1). Writing $A = A^a T^a$ in a basis T^a of the Lie algebra, the quadratic part of (6.1) becomes

$$S_{\text{quad}} \sim \int_M d_{ab} (A^a \, \mathrm{d} A^b) \ , \tag{6.39}$$

where $d_{ab} = \mathrm{tr}(T^a T^b)$. Previously, we chose a specific normalization for this trace in the fundamental representation. However, more generally, d_{ab} is an ad-invariant quadratic form on the Lie algebra. We want this quadratic form to be non-degenerate, in the sense that all components of the gauge field possess a kinetic term. For semi-simple groups, a non-degenerate metric always exists, namely, the Cartan–Killing metric, which is positive definite for compact G. The novelty of $2 + 1$ dimensions is that a non-degenerate metric exists for the group $\mathrm{ISO}(2,1)$ (this is not the case in other dimensions), although $\mathrm{ISO}(2,1)$ is not semi-simple.

Let us denote the Lorentz generators by J_{ab}, and the translations by P_a. The metric of interest is then specified by

$$\langle J_a, P_b \rangle = \delta_{ab} \ , \qquad \langle J_a, J_b \rangle = \langle P_a, P_b \rangle = 0 \ , \tag{6.40}$$

where we have introduced, for convenience, the combination $J^a \equiv \frac{1}{2} \epsilon^{abc} J_{bc}$. The commutation relations of $\mathrm{ISO}(2,1)$ are now given by

$$[J_a, J_b] = \epsilon_{abc} J^c \ , \qquad [J_a, P_b] = \epsilon_{abc} P^c \ , \qquad [P_a, P_b] = 0 \ , \tag{6.41}$$

where it is important to realize that indices are raised and lowered with the Lorentz metric η_{ab}.

We introduce the following gauge field

$$A_i = e_i^a P_a + \omega_i^a J_a \ , \tag{6.42}$$

where $\omega_i^a = \frac{1}{2} \epsilon^{abc} \omega_{ibc}$. A simple exercise will establish that the Chern–Simons action (6.1), written in terms of this gauge field, and with this choice of Cartan–Killing metric, takes the form

$$S_{\text{CS}} = \int_M \epsilon^{ijk} e_{ia} (\partial_j \omega_k^a - \partial_k \omega_j^a + \epsilon_{abc} \omega_j^b \omega_k^c) \ . \tag{6.43}$$

Our achievements thus far lie at the level of the action; we must now examine the symmetries of the theory. As we already know, the Chern–Simons action is invariant under infinitesimal gauge transformations; in the case under study there are no large gauge transformations, due to the fact that $\pi_3(\mathrm{ISO}(2,1)) = 0$, hence the coupling constant is not necessarily quantized.

Let us consider an infinitesimal gauge transformation $\delta A_i = -D_i u$, where the gauge parameter is decomposed as $u = \rho^a P_a + \tau^a J_a$. This leads to the component transformations

$$\delta e_i^a = -\partial_i \rho^a - \epsilon^{abc} e_{ib} \tau_c - \epsilon^{abc} \omega_{ib} \rho_c \ , \qquad \delta \omega_i^a = -\partial_i \tau^a - \epsilon^{abc} \omega_{ib} \tau_c \ . \tag{6.44}$$

We require that these transformations are equivalent to the usual ones present in gravity theory, that is, local Lorentz transformations and diffeomorphisms. The parameter τ^a designates the local

Lorentz transformation, and a glance at (6.44) shows that the standard formulae are obtained here. The remaining question is the issue of diffeomorphism invariance; presumably this is related to the ρ^a parameter. However, another glance at (6.44) will establish the fact that the correspondence is not complete.

Consider a diffeomorphism generated by a vector field $-v^i$, then as usual the transformation of the dreibein and spin connection are obtained by taking the Lie derivative along this vector field. For the dreibein we have

$$
\begin{aligned}
\tilde{\delta} e_i{}^a &= -v^k (\partial_k e_i{}^a - \partial_i e_k{}^a) - \partial_i (v^k e_k{}^a) \ , \\
\tilde{\delta} \omega_i{}^a &= -v^k (\partial_k \omega_i{}^a - \partial_i \omega_k{}^a) - \partial_i (v^k \omega_k{}^a) \ ,
\end{aligned}
\tag{6.45}
$$

where D_i is the covariant derivative with respect to the spin connection ω. The aim is to study the difference between (6.44) and (6.45). Letting $\rho^a = v^k e_k{}^a$, we find that

$$
\tilde{\delta} e_i{}^a - \delta e_i{}^a = -v^k (D_k e_i{}^a - D_i e_k{}^a) + \epsilon^{abc} v^k \omega_{kb} e_{ic} \ .
\tag{6.46}
$$

However, we now see that the first term on the right hand side of (6.46) vanishes by the ω equation of motion, $D_i e_j{}^a - D_j e_i{}^a = 0$, while the second term corresponds to a local Lorentz transformation with parameter $\tau^a = v^k \omega_k{}^a$. Thus our final result is that the gauge transformations of the Chern–Simons theory (6.44) are equivalent to local Lorentz and diffeomorphism invariance (6.45), on-shell.

The above identification of $2 + 1$ gravity as a Chern–Simons gauge theory with gauge group $ISO(2, 1)$ has remarkable consequences for the solution of the model [6.48, 6.49]. We can, for example, proceed with the canonical quantization of the theory, along lines similar to 6.1.2. Furthermore, the conceptual aspects of a diffeomorphism invariant problem receive a new interpretation in gauge field terms; and practical calculations obtain a simplicity concomitant with this interpretation.

Further reading

Some of the early references in which the Chern–Simons term was considered in various other contexts are refs. [6.50–6.57]. Papers dealing with the formal aspects of the theory and its quantization are refs. [6.58–6.78] and explicit computations of observables can be found in refs. [6.79–6.96]. The conformal field theory aspects have been studied by the authors of refs. [6.97–6.113], and the quantum group structure inherent in Chern–Simons theory has been investigated in refs. [6.43, 6.80, 6.114]. The Chern–Simons interpretation of $2 + 1$ dimensional gravity, and extensions thereof, has received further attention in refs. [6.115–6.131].

6.2. BF theories

BF theories are Schwarz type topological gauge theories with classical action

$$
S_c(n,p) = \int_M B_p \, dA_{n-p-1}
\tag{6.47}
$$

in the Abelian case, and

$$
S_c(n) = \int_M \mathrm{tr}\, B_{n-2} F_A
\tag{6.48}
$$

in the non-Abelian case. M is a closed orientable n-dimensional manifold, and in (6.47) B and A are differential forms (possibly taking values in a flat vector bundle), the subscript indicating their rank, while in (6.48) F_A is the curvature of some flat principal G-bundle over M, with $B \in \Omega^{n-2}(M, \mathbf{g})$. These actions were introduced in refs. [6.132] and [6.122, 6.133], where they were analyzed from the canonical and covariant (path integral) point of view, respectively. They have also been suggested independently in refs. [6.134, 6.135].

At first sight the theories described by the above actions (the first linear and the second barely non-linear) may appear to be rather trivial and uninteresting, but, somewhat surprisingly, quite the opposite turns out to be the case, and we will in the following sections sketch some of the many things that can be done with, and learned from, BF theories.

Quantization of the Abelian models is quite straightforward, the quantum action taking the form

$$S_q(n,p) = S_c(n,p) + \{Q, \psi(n,p)\} , \tag{6.49}$$

where Q is a metric independent off-shell nilpotent BRST operator. From the general arguments of section 2 it then follows that the partition function $Z(n,p)$ is a topological invariant of M. Schwarz has shown a long time ago [6.1] that this invariant is related to the Ray–Singer torsion T_M [6.26], and we can use this observation in either of two ways. On the one hand, the BRST approach provides us with a formal proof of the metric independence of T_M. On the other hand we can, more profitably, use this known property of the Ray–Singer torsion to prove rigorously the topological nature of Abelian BF theories, if we define the determinants appearing in the evaluation of $Z(n,p)$, e.g., via ζ-function regularization. T_M has a number of other interesting properties, and it is tempting to try to establish these from the field theory point of view. As an illustration of how this can be done, we show that triviality of $T(M)$ in even dimensions, i.e., $T(M) = 1$, follows from a simple scale invariance of the quantum action (6.49) [6.122].

As in Chern–Simons theory, the partition function is not the only observable of interest, and we will show in section 6.2.2 that the analogs of the Chern–Simons Wilson loops, namely Wilson "surfaces" associated with A and B, determine linking and intersection numbers of manifolds in any dimension [6.122, 6.136, 6.137].

We then turn to the non-Abelian models described by the action (6.48). Quantization of these models is complicated by the fact that—in more than three dimensions—they have a string of on-shell reducible symmetries, the same non-Abelian p-form symmetries we have already encountered in the context of super-BF theories in section 5.4. In contrast to what we achieved there, here it will not be possible to construct a quantum action with an off-shell nilpotent operator. It is of course possible to construct a BRST invariant quantum action with an on-shell nilpotent BRST symmetry (this is guaranteed by the BV algorithm), but neither will the quantum action differ from the classical action by a BRST commutator, nor will the BRST operator be metric independent. This then casts serious doubts on the topological nature of these models, and more generally on the belief that "reasonable" metric independent classical actions lead to topological field theories. It is therefore gratifying to see that metric independence of the theory can nevertheless be established [6.138], although one needs to work a little bit harder in the presence of the above complications.

After some preliminary remarks on the classical action (6.48) (concerning for instance the relation between Chern–Simons and BF theory in three dimensions) we sketch the arguments leading to the above conclusion. But wishing not to burden this section with somewhat more technical issues, we refer to refs. [6.139] and [6.138] for the details of the construction of the quantum action and the proof of metric independence, respectively.

There are many things that can be done with non-Abelian BF theories, and in the following we sketch some of these. For instance, we show that (6.48) can be regarded as a zero coupling limit

of Yang–Mills theory, and how BF theories in turn provide us with a complete non-perturbative Nicolai map for Yang–Mills theory on any Riemann surface. Moreover, BF systems have a Nicolai map in any dimension (this somehow being the common link among all known topological field theories apart from Chern–Simons theory). This Nicolai map reduces the partition function to an integral over the moduli space of flat connections, with measure given by the Ray–Singer torsion [6.133]. We show this explicitly in two and three dimensions, and support the simple argument (somewhat cavalier in the handling of zero modes) by introducing a BRST method for keeping track of the zero mode integrals, i.e., the integrals over the collective coordinates of the moduli space. Returning to two dimensions, we take this opportunity to explain the relation between $PSL(2,\mathbb{R})$ BF theory (and its Witten type counterpart—the topological gravity model of section 7) and Hitchin's self-duality equations on a Riemann surface [6.140]. We begin with a review of the Hitchin equations, and summarize those results of ref. [6.140] which are relevant for us here (section 6.2.6). We then provide some more information (beyond that contained in section 5.4.3) on the moduli space of flat $PSL(2,\mathbb{R})$ connections (section 6.2.7), and, armed with that, investigate the corresponding BF theory (section 6.2.8). Finally, we draw these threads together and explain the relation between the two sets of equations, as well as some of its implications.

6.2.1. Quantization of Abelian BF theories

The action (6.47) has the (reducible) Abelian gauge symmetries

$$B_p \rightarrow B_p + d\Lambda_{p-1} \ , \qquad A_{n-p-1} \rightarrow A_{n-p-1} + d\Lambda'_{n-p-2} \ . \tag{6.50}$$

Gauge fixing these can be achieved via straightforward (repeated) application of the Faddeev–Popov trick, keeping in mind the extra ghosts that appear in the quantization of reducible theories. Additionally (6.47) is invariant under the shifts

$$B_p \rightarrow B_p + \Gamma_p \ , \qquad A_{n-p-1} \rightarrow A_{n-p-1} + \Gamma'_{n-p-1} \ , \tag{6.51}$$

where Γ and Γ' are *harmonic* forms. These zero mode symmetries can easily be dealt with (we will explain this below), the net effect being to gauge the harmonic pieces of all the fields in the theory (A, B, ghosts, multipliers, anti-ghosts) to zero. This reduces the partition function to an integral over the coexact pieces of the fields, the exact pieces having been taken care of by the gauge fixing of the symmetry (6.50), and the harmonic pieces having obediently dropped out upon gauge fixing (6.51).

The space of solutions to the equations of motion $dA_{n-p-1} = dB_p = 0$ modulo the gauge symmetries (6.50) is the finite dimensional vector space $\mathcal{N} = H^p(M) \oplus H^{n-p-1}(M)$. If M is of the form $M = \Sigma \times \mathbb{R}$, $\mathcal{N}$ is even dimensional and naturally a symplectic vector space, as behoves a phase space. If one mods out further by the harmonic shift symmetry (6.51), the reduced phase space is a point. The general covariance of the theory is reflected in the fact that, on shell, diffeomorphisms are equivalent to gauge transformations. The explicit formulae can be found—for the more general case of non-Abelian BF theories—in section 6.2.3, eq. (6.73).

Let us start by explaining how the harmonic modes can be eliminated, so that we will henceforth not have to worry about them. The approach we choose relies on a straightforward application of the Faddeev–Popov procedure, developed for this purpose in refs. [6.141–6.143]. Recent applications can be found in refs. [6.144, 6.145, 6.122].

The analogy with ordinary gauge invariance is of course, that in QED (say) the part of the vector potential A which lies in the gauge direction does not enter into the action and is the cause of

the problems associated with defining the partition function. This is precisely the situation we are confronted with in the presence of zero modes. Using the Hodge decomposition of a p-form,

$$B_p = \delta\alpha_{p+1} + d\beta_{p-1} + \gamma_p , \tag{6.52}$$

where γ_p is harmonic, we can read off what the appropriate gauge fixing should be. When the action is invariant under $B_p \rightarrow B_p + dA_{p-1}$ this means that β_{p-1} does not appear. Gauge fixing then amounts to projecting β_{p-1} out by means of a Lagrange multiplier enforcing a delta function constraint on B in the path integral. Thus, one adds a term to the action which does precisely this, $\int \pi_{p-1} d * B_p$, plus the corresponding ghost terms.

Now the zero mode problem is posed as the invariance of the action under $B_p \rightarrow B_p + \Gamma_p$, where Γ_p is harmonic, which means that γ_p of (6.52) does not enter into the action. Following the previous rationale we gauge fix by projecting out γ_p, i.e., by adding a term $\int \Sigma_p * B_p$ to the action, where Σ is an arbitrary harmonic form. Let $\{\gamma_j, \ j = 1, \ldots, b_p = \dim H^p(M)\}$ be an orthogonal basis of harmonic p-forms, i.e., $\int_M \gamma_j * \gamma_k = v\delta_{jk}$, where $v = \int_M *1$ is the volume of M, and expand $\Sigma = \epsilon^j \gamma_j$. Then ϵ^j are the multiplier fields, which come along with the ghosts e^j and their anti-ghosts $\bar{e}^j$, satisfying the *harmonic BRST algebra*

$$\{Q^h, B_p\} = \gamma_j e^j , \qquad \{Q^h, e^j\} = 0 , \qquad \{Q^h, \bar{e}^j\} = \epsilon^j , \qquad \{Q^h, \epsilon^j\} = 0 . \tag{6.53}$$

Note that $\epsilon^j, \bar{e}^j$ and e^j are constant real numbers, not functions of space–time, since $H^p(M)$ is a finite dimensional real vector space. The term to be added to the action is then

$$\{Q^h, \int \bar{e}^j \gamma_j * B_p\} = \epsilon^j \int \gamma_j * B \pm v\bar{e}^j e^k \delta_{jk} , \tag{6.54}$$

which shows that v is the Faddeev–Popov determinant in this case. Even without invoking BRST invariance, it is evident that the addition of (6.54) does not introduce any metric dependence into the partition function, since the v^{b_p} contribution from the second term cancels against the v^{-b_p} contribution arising from the integral over ϵ and the harmonic mode of B in the first term. This then gauge fixes the harmonic modes to zero, and in the following it is understood that the zero modes of all the other fields appearing upon gauge fixing (6.50) have been dealt with in a similar way. We will from now on ignore the harmonic modes and concentrate on the gauge symmetry (6.50).

In the three-dimensional model $S_c(3,1) = \int B_1 dA_1$ the gauge symmetry is irreducible, and the quantum action is simply

$$S_q(3,1) = \int B_1 dA_1 + \pi_0 d * B_1 + \bar{c}_0 d * dc_0 + \pi'_0 d * A_1 + \bar{c}'_0 d * dc'_0 , \tag{6.55}$$

with the obvious BRST symmetry. Integration over the ghost fields yields $\det^2 \Delta_0$ (Δ_p is the Laplacian on p-forms), while integration over the remaining (B, A, π_0, π'_0) system requires more care. To evaluate the determinant one squares the first order operator (which diagonalizes it), reads off the determinant, and takes the square root. In this way one finds the contribution to be $\det^{-1/2}\Delta_1 \det^{-1/2}\Delta_0$, giving for the partition function

$$Z(3,1) = \det^{-1/2}\Delta_1 \det^{3/2}\Delta_0 . \tag{6.56}$$

One may wonder at this point what has happened to the alleged metric independence of Z. After all, the Laplacians depend on the metric, so do their spectra and their determinants. But, as it turns out, this particular combination of determinants is indeed metric independent, equalling T_M^{-1}, the inverse of the Ray–Singer torsion.

Given a flat vector bundle E over a Riemannian manifold (M, g) $(\dim M = n)$, the Ray–Singer torsion is defined by

$$T_M(E, g) = \prod_{k=0}^{n} \det{}^{(-1)^k k/2} \Delta_k \ . \tag{6.57}$$

Here Δ_k is the Laplace operator on k-forms with values in E, depending on the metric g of M. $\det \Delta_k$ is its determinant, defined via ζ-function regularization, and possible zero modes of Δ_k are excluded by defining

$$\zeta_k(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} t^{s-1} \operatorname{tr}(\mathrm{e}^{-t\Delta_k} - P_k) \ , \tag{6.58}$$

where $P_k = \lim_{t \to \infty} \mathrm{e}^{-t\Delta_k}$ is the orthogonal projector onto the space $H^k(M)$ of harmonic modes.

The most remarkable property of T_M is that it is independent of the metric g. It was put forward by Ray and Singer as an analytic analog of the Reidemeister–Franz torsion τ_M (for a review cf. ref. [6.146]), defined in terms of the simplicial complex of a smooth triangulation of M, but independent of the latter. As such τ_M is, like T_M, an invariant of the manifold M. Ray and Singer showed that T_M has many more properties in common with τ_M, and conjectured that they are in fact equal, $T_M = \tau_M$. The proof of this was supplied some time later by Cheeger [6.147] and Müller [6.148].

In three dimensions the Ray–Singer torsion $T(3)$ [we write $T(n)$ for the torsion of some n-manifold whenever there is no need to emphasize which particular manifold we are talking about] is $T(3) = \det^{-1/2} \Delta_1 \det \Delta_2 \det^{-3/2} \Delta_3$. Now by Hodge duality, $*\Delta_k = \Delta_{n-k}*$, we have $\det \Delta_k = \det \Delta_{n-k}$, and therefore $T(3) = \det^{-3/2} \Delta_0 \det^{1/2} \Delta_1$, which shows that, as announced, $Z(3,1) = T(3)^{-1}$.

One more thing worth noting about the result (6.56) is that it allows us to read off directly that the theory has no degrees of freedom in the field theoretic sense: we regard the inverse square root of the scalar Laplacian, $\det^{-1/2} \Delta_0$, as representing one bosonic degree of freedom, and (solely for counting purposes) treat the Laplacian Δ_k as if it acts on $\dim \Omega^k(M)$ copies of $\Omega^0(M)$; then the degrees of freedom displayed by the partition function $Z(3,1)$ are $-3 + 3 = 0$. This is of course true quite generally, and we will come back to this below, after having obtained an expression for the partition function $Z(n,p)$.

The other action available in three dimensions, $S(3,0) = \int B_0 dA_2$, is our first example of a reducible theory. The BV ghost triangle (see appendix A) tells us that the additional fields we need are: a ghost–anti-ghost–multiplier triplet $(c_1, \bar{c}_1, \pi_1)$ for the gauge fixing condition on A_2, a ghost-for-ghost triplet $(c_0, \bar{c}_0, \pi_0)$ for the gauge fixing of c_1, and finally an anti-ghost–multiplier pair (c_0', π_0') for the gauge fixing of the anti-ghost $\bar{c}_1$. Here c_0' is the famous extra ghost, characteristic of reducible theories. The quantum action is then

$$S_q(3,0) = \int (B_0 dA_2 + \pi_1 d * A_2 - \bar{c}_1 d * dc_1 + \pi_0 d * c_1 - \bar{c}_0 d * dc_0 + \pi_0' d * \bar{c}_1 - c_0' d * \pi_1) \ , \tag{6.59}$$

which leads to the partition function $Z(3,0) = T(3)$.

As our last example we consider the four-dimensional theory $S(4,2)$. The ghost structure of B_2 is identical to that of A_2 above, and one finds the partition function to be $Z(4,2) =$

$\det^{-1/4} \Delta_2 \det^{1/2} \Delta_1 \det^{-1/2} \Delta_0$. As expected, the number of degrees of freedom is $3 - 4 + 1 = 0$. And, comparing with the Ray–Singer torsion, which in four dimensions is

$$T(4) = \det^{-1/2} \Delta_1 \det \Delta_2 \det^{-3/2} \Delta_3 \det^2 \Delta_4 = \det^2 \Delta_0 \det^{-2} \Delta_1 \det \Delta_2 \ ,$$

we find $Z(4,2) = T(4)^{-1/4}$, which again establishes the topological nature of the model.

It is worth mentioning at this point that there is a very elegant (and quicker) way of arriving at $Z(n,p)$, directly from the classical action, without having to determine the quantum action. This is Schwarz's method of resolvents [6.1], invented (prior to the discovery and understanding of the ghost-for-ghost mechanism in the BRST framework) to make sense of the partition function in what is now known as reducible theories. A comparison of this method with the BRST method, within the context of BF theories, can be found in ref. [6.122].

The general result, obtainable via either of the above methods, is that the contribution to $Z(n,p)$ from the ghost triangle of A_{n-p-1} is $Z_A = \prod_{k=0}^{n-p-1} \det^{\nu_k} \Delta_{k+p+1}$, while B_p contributes $Z_B = \prod_{k=0}^{p} \det^{\nu_k} \Delta_{p-k}$, where $\nu_k = (-1)^{k+1}(2k+1)/4$. Thus the partition function of the Abelian BF theory with classical action (6.47) is

$$Z(n,p) = \prod_{j=0}^{n-p-1} \det^{\nu_j} \Delta_{j+p+1} \prod_{k=0}^{p} \det^{\nu_k} \Delta_{p-k} \ . \tag{6.60}$$

Comparing (6.60) with the definition (6.57) of the Ray–Singer torsion, we obtain the general result

$$Z(n,p) = T(n)^{(-1)^p} \ , \quad \text{for } n \text{ odd } ; \qquad Z(n,p) = T(n)^{(-1)^p (n-2p-1)/n} \ , \quad \text{for } n \text{ even } .$$

Moreover, it can be read off from (6.60) that the number of degrees of freedom of $S(n,p)$, $n \geq 2$, is zero in general. In view of the above relations this can alternatively be deduced more directly from the fact that the number N of determinants of "bosonic" Laplacians in $T(n)$ is

$$N = \sum_{k=1}^{n} (-)^k k \binom{n}{k} = 0 \ .$$

The first equality follows from the definition (6.57), the second from the x-derivative of the binomial formula

$$(x+y)^n = \sum_{k=0}^{n} x^k y^{n-k} \binom{n}{k} \ .$$

We have already mentioned above that in even dimensions $T_M = 1$. This follows from the relations among the non-zero spectra of Δ_k implied by $d\Delta = \Delta d$, or, more explicitly, $d_k \Delta_k = \Delta_{k+1} d_k$. Now the path integral encodes a great detail of information about determinants and eigenvalues, and as an illustration of the fact that it is also aware of the above relation we will now show how to derive the triviality of T_M in even dimensions (this is theorem 2.3 of ref. [6.26]) from a simple scale invariance of the quantum action of Abelian BF theories.

For instance in two dimensions, it is easy to see that $\det f(\Delta_1) = \det^2 f(\Delta_0)$, where f is some function of the Laplacian [e.g., $f(\Delta) = \Delta$]. This is seen by considering the quantum action

$$S_q(2,0) = \int B_0 d A_1 + \pi_0 d * A_1 + \bar{c}_0 d * d c_0 \ .$$

The action is invariant under the following scaling of the fields: $A \to f(\Delta_1)A$, $B \to f^{-1}(\Delta_0)B$, and $\pi_0 \to f^{-1}(\Delta_0)\pi_0$. Since Z cannot be changed by this transformation, its Jacobian must be equal to one. This yields the desired result.

In odd dimensions, this procedure does not give us any information. Let us, for instance, consider the three-dimensional example (6.55). The transformation that leaves the action invariant is

$$A \to f(\Delta_1)A \ , \qquad B \to f^{-1}(\Delta_1)B \ , \qquad \pi_0' \to f^{-1}(\Delta_0)\pi_0' \ , \qquad \pi_0 \to f(\Delta_0)\pi_0 \ ,$$

whose total Jacobian is identically one. It is, however, precisely this extra information that we have at our disposal in even dimensions, which allows us to prove the triviality of the Ray–Singer torsion in that case. In two dimensions,

$$T(M_2) = \det{}^{-1/2}\Delta_1 \det \Delta_2 = \det{}^{-1/2}\Delta_1 \det \Delta_0 = 1 \ ,$$

as a consequence of duality and the above result. Generalizing these considerations, one finds that any function of the determinant of the Laplace operator on m-forms (where $\dim M = n = 2m$) can be expressed in terms of the determinants of the Laplace operator acting on lower rank forms as

$$\det f(\Delta_m) = \prod_{i=0}^{m-1} \det{}^{(-1)^i 2} f(\Delta_{m-i-1}) \ . \tag{6.61}$$

To prove this one scales B_p by a factor $f(\Delta_p)$ in the full BRST-extended quantum action $S_q(n,p)$; this may be compensated in the first term of the action (which is just the classical action) by scaling A_{n-p-1} by $f^{-1}(\Delta_{n-p-1})$. All other fields that appear can then also be scaled in such a way that one returns to the original action. The product of determinants obtained in this way must therefore be equal to one. This implies

$$\begin{aligned}
1 &= \det f(\Delta_p) \det{}^{-1} f(\Delta_{p-1}) \cdots \det{}^{(-1)^p} f(\Delta_0) \\
&\quad \times \det{}^{-1} f(\Delta_{n-p-1}) \det f(\Delta_{n-p-2}) \cdots \det{}^{(-1)^{n-p}} f(\Delta_0) \ ,
\end{aligned} \tag{6.62}$$

where we have collected the contributions from the fields coming from the B and A triangle on the first and second line, respectively. For n odd, (6.62) is identically satisfied because of Hodge duality, whereas for n even the two sets of terms do not cancel, but rather add up upon using duality. Collecting all the terms one then arrives at (6.61). Now in even dimensions $n = 2m$, the Ray–Singer torsion is

$$T(M) = \prod_{q=0}^{2m} \det{}^{(-1)^q q/2} \Delta_q = \det{}^{(-1)^m m/2} \Delta_m \prod_{q=0}^{m-1} \det{}^{m(-1)^q} \Delta_q \ ,$$

which is indeed equal to one by (6.61).

Before leaving these results and turning our attention to observables in the next section, we mention one more interesting property of the Ray–Singer torsion (which again has its counterpart for the Reidemeister–Franz torsion), namely that [6.26]

$$T_{M_1 \times M_2} = T_{M_1}^{\chi(M_2)}$$

if M_2 is simply connected [here $\chi(M_2)$ is the Euler number of M_2]. It should clearly be possible to derive this from the path integral point of view as well, but presently we do not know how to do that.

6.2.2. Observables in Abelian BF theories

So far we have restricted our attention exclusively to the partition function of Abelian BF theories. But, as in other topological field theories, more general observables are a rich source of topological invariants. We have seen that in Witten type theories observables are related to the ghost zero mode sector of the theory, and that correlation functions of these typically compute intersection numbers of moduli spaces associated with the "space-time" manifold M. In Chern–Simons gauge theory on the other hand, the fundamental observables (Wilson loops) have nothing to do with the ghost sector and compute topological information associated more directly with M itself (namely, invariants of knots embedded in M). This is a feature shared by all other Schwarz type topological gauge theories, and in particular BF theories.

We will now show, that in Abelian BF theories on $M = \mathbb{R}^n$ correlation functions of "Wilson surfaces" associated with A and B compute linking and intersection numbers of manifolds embedded in M. That this is possible is already suggested by Polyakov's observation [6.54] that expectation values of Wilson loops in Abelian Chern–Simons theory in three dimensions are related to the classical Gauss linking number of loops. BF theories not only allow us to generalize this to any dimension; there is also the added benefit in $n = 3$ that there is no necessity of framing the loops (as in ref. [6.2]) or regularizing the self-linking number in some other way [6.54]; since we have two fields (A and B) instead of just one, the question of framing simply does not appear in our calculation, which is therefore finite and unambiguous at all stages.

In order to generalize the linking number $L(\gamma, \gamma')$ of two loops in three dimensions to higher dimensions, we reinterpret it as the *intersection number* of a disc D bounded by γ with the loop γ', which is defined as follows (cf. ref. [6.149] for more information): since the dimension of D is equal to the codimension of γ', these will generically intersect transversally at isolated points x_k (see section 4.5.2). Having chosen orientations on $\mathbb{R}^3$, D, and γ', one assigns to each x_k the number $+1$ or -1, depending on whether the orientation of (D, γ') at x_k coincides with that of $\mathbb{R}^3$ or not. The intersection number of D and γ' is then defined as $I(D, \gamma') = \sum_{x_k} \pm 1$.

To have an integral representation for I, we introduce the the de Rham current $\Delta_{\gamma'}$ [6.149], Poincaré dual to the embedding of γ' into $M = \mathbb{R}^3$ (cf. section 4.5.2). $\Delta_{\gamma'}$ is a delta function two-form whose sole purpose is to restrict an integral over M to one over γ', i.e., which satisfies $\int_M \Delta_{\gamma'} \alpha_1 = \oint_{\gamma'} \alpha_1$ for all one-forms $\alpha_1 \in \Omega^1(M)$. This allows us to rewrite the intersection number as

$$I(D, \gamma') = \int_D \Delta_{\gamma'} = \sum_{x_k} \pm 1 \qquad (6.63)$$

(here the relative signs are taken care of by the functional properties of the delta function). We also accept this as the definition of the linking number of γ and γ'; this amounts to fixing an overall sign in the definition of the latter.

It is now clear how to generalize this to higher dimensions. We let $\partial\Sigma$ and $\partial\Sigma'$ be disjoint, compact, oriented p- and $(n - p - 1)$-dimensional boundaries of oriented submanifolds Σ and Σ' of $M = \mathbb{R}^n$. We also introduce the the de Rham currents $\Delta_{\partial\Sigma}$ and Δ_Σ, with the properties

$$\int_{\partial\Sigma} \alpha_p = \int_M \Delta_{\partial\Sigma} \alpha_p , \qquad \int_\Sigma \alpha_{p+1} = \int_M \Delta_\Sigma \alpha_{p+1} , \qquad (6.64)$$

with analogous definitions for Σ'. We now define the linking number of $\partial\Sigma$ and $\partial\Sigma'$ by

$$L(\partial\Sigma, \partial\Sigma') := \int_\Sigma \Delta_{\partial\Sigma'} . \qquad (6.65)$$

It is clear that this integral will only receive contributions (± 1) from points where Σ and $\partial \Sigma'$ intersect.

At the level of de Rham currents, the duality between homology and cohomology is expressed by

$$d\Delta_\Sigma = (-)^{n-p}\Delta_{\partial\Sigma} ,\tag{6.66}$$

as can easily be verified from (6.64). It follows that $L(\partial\Sigma,\partial\Sigma')$ and $L(\partial\Sigma',\partial\Sigma)$ are related by

$$L(\partial\Sigma',\partial\Sigma) = (-)^{\dim\partial\Sigma\,\dim\partial\Sigma'+1}L(\partial\Sigma,\partial\Sigma') .\tag{6.67}$$

Since this is a known property of the linking number, (6.67) is a useful check on the consistency of our sign conventions. L is a topological invariant in the sense that it is invariant under homotopies of the embeddings of $\partial\Sigma$ and $\partial\Sigma'$ into $M = \mathbb{R}^n$.

From the point of view of BF theory, the above set-up appears very naturally, because it allows us to define the metric independent and gauge invariant observables (Wilson surfaces) $\exp\int_{\partial\Sigma}B$ and $\exp\int_{\partial\Sigma'}A$. Since it is easily seen that the expectation value of either one of these [with respect to the action $S(n,p)$] is equal to 1, the simplest non-trivial correlator to consider is

$$\left\langle \exp\left(i\beta\int_{\partial\Sigma}B\right)\exp\left(i\alpha\int_{\partial\Sigma'}A\right)\right\rangle = \int d[A]\,d[B]\,\exp\left[i\left(\int_M BdA + \beta\int_{\partial\Sigma}B + \alpha\int_{\partial\Sigma'}A\right)\right] \tag{6.68}$$

(since we are computing a correlator of gauge invariant operators, we have ignored gauge fixing and ghost terms in the above). Our aim is now to show that this correlation function indeed computes $L(\partial\Sigma,\partial\Sigma')$, the precise relation being

$$\log\left\langle \exp\left(i\beta\int_{\partial\Sigma}B\right)\exp\left(i\alpha\int_{\partial\Sigma'}A\right)\right\rangle = (-)^p i\alpha\beta L(\partial\Sigma,\partial\Sigma') .\tag{6.69}$$

By now, various proofs of this have appeared in the literature [6.122, 6.136, 6.137], the simplest [6.122] being to compute directly the Gaussian integral (6.68). We use the de Rham currents introduced above to rewrite the "action" appearing in (6.68) as

$$S = \int_M (BdA + \beta\Delta_{\partial\Sigma}B + \alpha\Delta_{\partial\Sigma'}A) ,\tag{6.70}$$

which shows that $\Delta_{\partial\Sigma}$ and $\Delta_{\partial\Sigma'}$ play the role of sources coupled to the gauge fields B and A. The equations of motion following from (6.70) are

$$dA = (-)^{p(n-p)+1}\beta\Delta_{\partial\Sigma} , \qquad dB = (-)^p\Delta_{\partial\Sigma'} .$$

Plugging this back into (6.70) and making repeated use of the identities (6.64), (6.66) one finds that the first and third terms cancel, leaving $\alpha\beta(-)^p\int_\Sigma\Delta_{\partial\Sigma'}$, which establishes (6.69).

By using forms with values in a flat vector bundle, and an exterior derivative with respect to a non-trivial flat background connection, it is also possible to define generalized linking numbers [6.136, 6.137]. The extension to manifolds with boundary is treated in ref. [6.150].

6.2.3. Classical aspects of non-Abelian BF theories

In this and the following sections we shall study in some detail the non-Abelian BF action (6.48). In addition to the ordinary Yang–Mills gauge symmetry (with B transforming in the adjoint representation), the action (6.48) has (for $n \geq 3$) the p-form symmetry

$$B_{n-2} \to B_{n-2} + d_A\Lambda_{n-3} ,\tag{6.71}$$

which is on-shell reducible for $n \geq 4$, since the equations of motion are

$$F_A = 0 \; , \qquad d_A B_{n-2} = 0 \; . \tag{6.72}$$

Denoting by $\mathcal{L}_X$ the Lie derivative along the vector field X, and using the fact that, on differential forms, $\mathcal{L}_X = d i_X + i_X d$, where $i_{(.)}$ is the operation of interior multiplication (or contraction), one finds

$$\mathcal{L}_X A = i_X F_A + d_A \Lambda(X) \; , \qquad \mathcal{L}_X B = i_X d_A B + [B, \Lambda(X)] + d_A \Lambda'(X) \; , \tag{6.73}$$

where $\Lambda(X) = i_X A$ and $\Lambda'(X) = i_X B$. This shows that, on shell, diffeomorphisms are equivalent to the gauge and p-form symmetries of the action (6.48).

Due to the non-linearity of the action there is no analog of the harmonic shift symmetry (6.51) in the non-Abelian case. Another consequence of the non-linearity is that the space $\mathcal{N} = \mathcal{N}(M, G)$ of solutions modulo gauge symmetries is in general no longer a vector space. In two dimensions for instance, with $M = \Sigma_g$ a Riemann surface, $\mathcal{N}$ is (ignoring reducible connections) simply the moduli space of flat connections (cf. section 5.4.3), since there are no non-trivial solutions to the equation of motion $d_A B_0 = 0$, i.e., $\mathcal{N}(\Sigma_g, G) = \mathcal{M}(\Sigma_g, G)$. As in super-BF theories, three dimensions are particularly interesting in the present context. Here, the equation $d_A B_1 = 0$ determines the (co)tangent space to the moduli space $\mathcal{M}(M_3, G)$ of flat connections, so that

$$\mathcal{N}(M_3, G) = \mathrm{T}^* \mathcal{M}(M_3, G) \; . \tag{6.74}$$

One way of understanding this result is to note that the BF action $\int B_1 F_A$ for a group G is the same as a Chern–Simons action for the group $TG \approx G \times \mathfrak{g}$; we expand a connection C for the latter as $C = T_a A^a + P_a B^a$, where (T_a, P_a) are generators for the group TG with commutation relations $[T_a, T_b] = f_{ab}^c T_c$, $[T_a, P_b] = f_{ab}^c P_c$, and $[P_a, P_b] = 0$, and choose the invariant inner product to be

$$\langle T_a, P_b \rangle = \mathrm{tr}(T_a T_b) \; , \qquad \langle T_a, T_b \rangle = \langle P_a, P_b \rangle = 0 \; ;$$

then we find that indeed

$$\frac{1}{2} \int \langle C, dC + \tfrac{1}{3}[C, C] \rangle = \mathrm{tr} \int B F_A \; . \tag{6.75}$$

This is the non-supersymmetric counterpart of the observation made in ref. [6.48] and explained in remark (iii) of section 5.4.1, that the three-dimensional super-BF action is a super-Chern–Simons action. It also immediately implies (6.74), since the phase space of Chern–Simons theory with gauge group TG is $\mathcal{M}(M_3, TG) \approx \mathrm{T}\mathcal{M}(M_3, G)$.

Another consequence of (6.75) is that it provides us with a potentially interesting and non-trivial observable for 3D BF theory, namely the Wilson loop of the gauge field C. Note that, although B is a one-form, its Wilson loop is not a well defined observable, since it is not invariant under the p-form symmetry (6.71). For the same reason it is not clear if there are non-trivial B-dependent observables in more than three dimensions. Broda [6.151] has recently constructed non-Abelian "Wilson surfaces" for BF theories, depending on A and B, but it remains to be seen if these define good observables at the quantum and gauge fixed level.

Equation (6.75) also sheds some light on gravity in three dimensions: we have seen in section 6.1.6 that the Einstein–Hilbert action is equivalent to a Chern–Simons action with gauge group $\mathrm{ISO}(2,1)$; but $\mathrm{ISO}(2,1) = \mathrm{TSO}(2,1)$, so that we can alternatively write the action as $\int BF_A$, where A is now an $\mathrm{SO}(2,1)$ gauge field and B is the dreibein [this is just equation (6.38)]. As

such, A and B are the $(2 + 1)$-dimensional analogs [6.129, 6.152] of the Ashtekar variables for $(3 + 1)$-dimensional gravity [6.153], and it is the observation encoded in (6.75) which relates this formulation to Witten's.

In higher dimensions, there is less geometrical structure associated with BF theories, and all that we can say in that generality is that the tangent space to $\mathcal{N}(M_n, G)$ at a solution (A, B) is the vector space

$$T_{(A,B)}\mathcal{N}(M_n, G) = H_A^1(M_n, \mathbf{g}) \oplus H_A^{n-2}(M_n, \mathbf{g}) \ ,$$

which is naturally symplectic if $M_n = \Sigma_{n-1} \times \mathbb{R}$.

6.2.4. Quantization of non-Abelian BF theories

Quantization of the action (6.48) in two and three dimensions is completely straightforward, the quantum action $S_q(3)$, for instance, being the obvious covariant non-Abelian analog of the Abelian action (6.55). In addition, various studies have been made of the one-loop effective action in these cases [6.154, 6.155].

The on-shell reducibility of (6.71) in more than three dimensions complicates matters, and upon following the BV algorithm one ends up with a quantum action having the following unpleasant features [6.122, 6.139, 6.138]:

(i) the BRST operator is nilpotent only on-shell,

(ii) the BRST operator is metric dependent,

(iii) the quantum action does not differ from the classical action only by a BRST commutator,

(iv) the quantum action contains cubic ghost interaction terms (which generally are metric dependent).

This [and (ii), (iii) in particular] prevents us from using the standard arguments to establish metric independence of the partition function.

We will now explain (in the case $n = 4$) step by step, why these features arise and how they can in turn be eliminated again to establish the topological nature of BF theories. This proof applies equally well to $n \geq 5$, since the only non-generic property of the four-dimensional theory (the metric independence of the cubic ghost term) plays no role in our arguments.

(1) We start off with the "naive" quantum action $S_q'(4)$, the non-Abelian analog of the Abelian action $S_q(4, 2)$, namely [cf. (6.59)]

$$S_q'(4) = \int (B_2 F_A + \pi_1 d_A * B_2 - \bar{c}_1 d_A * d_A c_1 + \pi_0 d_A * c_1 - \bar{c}_0 d_A * d_A c_0$$
$$+ \pi_0' d_A * \bar{c}_1 - c_0' d_A * \pi_1) \tag{6.76}$$

(we are not concerned with the ordinary Yang–Mills symmetry here; since all the terms we introduce are covariant with respect to A, this field may be gauge fixed at the end in the usual way). This is not yet the correct quantum action, since—due to the reducibility of the symmetry (6.71), expressed by $Q c_1 = d_A c_0$—the Q variation of $S_q'(4)$ is non-zero,

$$Q S_q'(4) = \int [c_0, *d_A \bar{c}_1] F_A \ .$$

(2) This term can be canceled by modifying the B variation $QB = d_A c_1$ to $sB = (Q + R)B$, with $RB = -[c_0, *d_A \bar{c}_1]$. But now $s S_q'(4)$ picks up a term from $R(\pi_1 d_A * B_2) = c_0[d_A \bar{c}_1, d_A \pi_1]$ (modulo total derivatives).

(3) This term in turn is Q-exact and the complete quantum action invariant under $s = Q + R$ is

$$S_q(4) = S'_q(4) + \tfrac{1}{2}c_0[d_A\bar{c}_1, d_A\bar{c}_1] \ . \tag{6.77}$$

We see that a cubic ghost interaction term has appeared, and that the relevant BRST operator $s = Q + R$ is metric dependent; due to the B equation of motion $F_A + *d_A\pi_1 = 0$ and

$$s^2B = [c_0, F_A + *d_A\pi_1] \ , \tag{6.78}$$

s is on-shell nilpotent (as it should be), since s^2 is (like Q^2) identically zero on all the other fields. Moreover, the classical action is not s-invariant, and (6.78) shows that $S_q(4)$ cannot possibly be of the form $S_c(4) + \{s, \Psi\}$.

(4) Now that we have accumulated all these complications, let us try to get rid of them again—one by one. As a first step towards proving that BF theories are indeed topological, we now argue that the cubic ghost term of (6.77) contributes neither to the partition function nor to the expectation value of any operator depending only on A and B. This follows from the observation that one can assign charges to the fields in such a way that A and B and all the terms in $S_q(4)$ apart from the cubic term are charge singlets (a possible choice is $k + 1$ for c_k and $-(k + 1)$ for $\bar{c}_k$, giving the cubic term charge -1). This essentially eliminates obstacle (iv) from our list.

(5) It follows that—for the purposes of studying the partition function $Z(4)$ and suitable correlators—the relevant action is $S'_q(4)$, which *is* of the form

$$S'_q(4) = S_c(4) + \{Q, \Psi(4)\} \ . \tag{6.79}$$

Since $S_c(4)$ is Q-invariant and Q is metric independent, this clearly simplifies matters considerably, and is a satisfactory state of affairs, provided that we can show that Q is nilpotent (at least on-shell). Above, we have used the B equation of motion $F_A + *d_A\pi_1 = 0$ to show this for $s = Q + R$, eq. (6.78). But now we can once again make use of the squaring argument of sections 3.1 and 5.4.2 ($F_A + *d_A\pi_1 = 0$ implies $F_A = 0$) to conclude that, despite appearance, Q (with $Q^2B = [c_0, F_A]$) is on-shell nilpotent as well! This eliminates problems (ii) and (iii).

(6) It remains to overcome the problem that for an on-shell nilpotent BRST operator the BRST Ward identity is *not* $\langle\{Q, \Sigma\}\rangle = 0$ (here Σ is an arbitrary functional of the fields), which would imply directly the metric independence of $Z(4)$ upon setting $\Sigma = \delta_g\Psi(4)$. Rather, in the case at hand this Ward identity receives a correction from the Q-variation of the action and reads

$$\langle\{Q, \Sigma\}\rangle + \langle\{Q^2, \Psi\}\Sigma\rangle = 0 \ . \tag{6.80}$$

Now—provided that we can integrate over B (which enforces $Q^2 = 0$, as we have seen above) in the second term of (6.80)—this Ward identity reduces to the standard one. It can be checked (using the charge assignments of above) that, for $\Sigma = \delta_g\Psi(4)$, the terms in $\{Q^2, \Psi(4)\}\Sigma$ involving B do not contribute, which finally establishes that indeed $\delta_g Z(4) = 0$. Likewise, the above argument shows that expectation values of metric independent Q-invariant functionals of A are metric independent. In more than four dimensions, the complete quantum action $S_q(n)$ is also of the form "naive quantum action $S'_q(n)$ plus cubic ghost terms", and the charge assignments to the ghost fields can again be chosen in such a way that none of the cubic terms contribute. Then the same sequence of arguments as above establishes the topological nature of BF theories in general.

The preceding, somewhat technical, discussion shows that the BV algorithm can be over-sophisticated for certain purposes: it obscures the fact that it is really the naive BRST operator Q (and not $s = Q + R$), and the naive quantum action $S'_q(n)$ [instead of $S_q(n)$] which govern the

fundamental properties of the theory. That this should be the case, can also be understood from a different point of view. The guiding principle in the gauge fixing procedure should be that the expectation value of any gauge invariant operator is not affected by the introduction of the gauge fixing and ghost terms (up to a multiplicative group volume factor). But a classically gauge invariant functional (an observable) is not necessarily *s*-invariant, whereas it is certainly *Q*-invariant. The *R* term in *s* spoils this invariance, but it is the *R* term that is linked to the cubic ghost terms. For our purposes it is then correct to demand that good observables be *Q*-invariant, but this is only legitimate if the ghost interactions are ignorable. And this is indeed what we have shown above.

The above discussion raises some questions of a more general nature regarding the construction of quantum actions, such as: is it possible more generally to make sense of quantum actions which (like S'_q) are BRST invariant "in the path integral"? or, under what general conditions do the quantum equations of motion imply the classical equations of motion?

A supersymmetry of the four-dimensional quantum action—analogous to that discovered in Chern–Simons theory in the Landau gauge (cf. section 8.4.6)—has recently been discussed in ref. [6.156], where it is also shown that (in flat space) the complete quantum action can be written as a BRST + supersymmetry commutator. The consequences of this observation—suggesting a somewhat unexpected link between Witten and Schwarz type theories—remain to be worked out.

6.2.5. Nicolai maps and Yang–Mills theory

Two of the observations we have made in the previous section will be of interest to us now: that the partition function Z receives contributions only from flat connections, and that the cubic ghost terms do not contribute to Z. These observations taken together imply that Z is a bunch of background field determinants, or, in other words, that the one-loop approximation to Z is exact. While familiar from Witten type theories, this is a somewhat unexpected result for a Schwarz type theory—and one which is certainly not shared by Chern–Simons theory. As in Witten type theories, this result can alternatively be understood as a consequence of the existence of a Nicolai map: all BF theories have a complete non-perturbative Nicolai map! We will come back to this below.

Already at this stage, however, it is possible to be more precise about what the partition function $Z(n)$ will turn out to be. In section 6.2.1 we have seen that the partition function of the Abelian action $S(n, n - 2) = \int B_{n-2} dA_1$ is the inverse of the Ray–Singer torsion $T(n)$ of the de Rham complex. Likewise, the partition function of the action $\int B_{n-2} d_C A_1$ (here C is a flat background connection on some vector bundle) is the inverse of the Ray–Singer torsion $T(n, C)$ of the de Rham complex with coefficients in this vector bundle. This action is just of the form of the one-loop approximation to the non-Abelian action $S(n)$. Parametrizing the moduli space of flat connections by coordinates $\{\lambda^k\}$, $A = A(\lambda)$, we therefore expect $Z(n)$ to be of the form

$$Z(n) = \int_{\mathcal{M}} d\lambda \, T(n, A(\lambda))^{-1} . \tag{6.81}$$

The Ray–Singer torsion thus provides us with a measure on the moduli space of flat connections. In (6.81) we have suppressed other zero mode integrations, and we adopt the attitude that—for the purposes of calculating Z—these should be gauged away. In support of this point of view we mention that the B zero mode B_c does not appear in the path integral [6.133]: we expand A and B about classical solutions,

$$A = A_c + A_q , \quad F_{A_c} = 0 ; \qquad B = B_c + B_q , \quad d_{A_c} B_c = 0 ; \tag{6.82}$$

then the action is

$$S = \int B_\mathrm{c} d_{A_\mathrm{c}} A_\mathrm{q} + B_\mathrm{q}(d_{A_\mathrm{c}} A_\mathrm{q} + \tfrac{1}{2}[A_\mathrm{q}, A_\mathrm{q}]) + B_\mathrm{c}\tfrac{1}{2}[A_\mathrm{q}, A_\mathrm{q}] \; ; \tag{6.83}$$

by (6.82) the first term of (6.83) is zero; the path integral over B_q leads to a delta function constraint $d_{A_\mathrm{c}} A_\mathrm{q} + \tfrac{1}{2}[A_\mathrm{q}, A_\mathrm{q}] = 0$, so that, again by (6.82), the last term of (6.83) vanishes as well; thus B_c does not enter at all, and we will in the following set $B_\mathrm{c} = 0$. It is important to note that we arrived at this result by keeping all terms in (6.83) and not just those quadratic in the quantum fields. It should also be borne in mind that the zero mode integrals may still need to be performed when one computes expectation values of observables.

After these preparatory remarks we return to the subject of Nicolai maps. In two dimensions the complete quantum action is

$$S_\mathrm{q}(2) = \int BF_A + \pi_0 d_{A_\mathrm{c}} * A_\mathrm{q} + \bar{c}_0 d_{A_\mathrm{c}} * d_A c_0 \; ,$$

and the change of variables

$$\xi(A) = F_A \; , \qquad \eta(A) = d_{A_\mathrm{c}} * A_\mathrm{q} \tag{6.84}$$

trivializes the bosonic part of the action,

$$S_\mathrm{q}(2) = \int B\xi + \pi_0 \eta + \bar{c}_0 d_{A_\mathrm{c}} * d_A c_0 \; .$$

This Nicolai map is similar to that used in the calculation of the partition function of Donaldson theory in section 5.2.5. The determinants arising from the ghost integration and the Jacobian of this change of variables combine to give $Z(2) = T(2, A_\mathrm{c})^{-1} = 1$. The integral over the moduli space of flat connections arises because the zeros of (ξ, η) are in one to one correspondence with points of $\mathcal{M}$, and we can therefore use this above change of variables to trivialize the path integral over all but a finite dimensional space of fields, and the remaining integral over $\mathcal{M}$ is still to be performed. In the case of isolated flat connections, this again gives us the interpretation of Z as the winding number of the Nicolai map.

It is also possible to introduce the A zero modes into the path integral directly [6.145, 6.122]: implicit in the split (6.82) is the assumption that A_q contains no fluctuations tangent to $\mathcal{M}$, and this can be made more explicit via the harmonic BRST algebra of section 6.2.1. Associated with the coordinates λ^k of $\mathcal{M}$ we have their superpartners $\sigma^k = Q^\mathrm{h}\lambda^k$, as well as anti-ghosts $\bar{\sigma}^k$ and multipliers $\tau^k = Q^\mathrm{h}\bar{\sigma}^k$. The flat connection $A_\mathrm{c}(\lambda)$ then transforms as $Q^\mathrm{h}A_\mathrm{c}(\lambda) = \sigma^k\partial_k A_\mathrm{c}(\lambda)$, and the $\partial_k A_\mathrm{c}(\lambda)$ span the tangent space to $\mathcal{M}$ at $A_\mathrm{c}(\lambda)$. It is now straightforward to gauge fix A_q to be orthogonal to these fluctuations. One simply adds $\{Q^\mathrm{h}, \int \bar{\sigma}^k\partial_k A_\mathrm{c}(\lambda) * A_\mathrm{q}\}$ to the action [with $Q^\mathrm{h}A_\mathrm{q} = -\sigma^k\partial_k A_\mathrm{c}(\lambda)$, so that $Q^\mathrm{h}A = 0$]. Then everything runs as above, the B, π_0, and τ integrations setting A_q to zero, with the important difference that one is left with an explicit λ integration at the end, giving (6.81).

In three dimensions the quantum action is

$$S_\mathrm{q}(3) = \int B_1 F_A + \pi_0 d_A * B_1 + \bar{c}_0 d_A * d_A c_0 + \pi_0' d_{A_\mathrm{c}} * A_\mathrm{q} + \bar{c}_0' d_{A_\mathrm{c}} * d_A c_0' \; ,$$

and the slightly different change of variables

$$\xi(A, \pi_0) = F_A + *d_A\pi_0 \; , \qquad \eta(A, \pi_0) = d_{A_\mathrm{c}} * A_\mathrm{q} \tag{6.85}$$

trivializes the action in this case. Away from reducible connections zeros of this map are again in one to one correspondence with gauge equivalence classes of flat connections, and as in two dimensions the Jacobian and ghost determinants combine to give $T(3, A_c)^{-1}$, and thus

$$Z(3) = \int_{\mathcal{M}} \mathrm{d}\lambda \, T(3, A_c(\lambda))^{-1} \, . \tag{6.86}$$

However, it has been shown in refs. [6.154, 6.155] that, if the Jacobian is regulated in a gauge invariant manner, then a Chern–Simons term is induced at the one-loop level. In any case, one could add the Chern–Simons term (with arbitrary integer coefficient) to the 3D BF action, and retain all the symmetries. Certain generalizations are also possible [6.3, 6.154, 6.157].

It should now be clear that essentially the same procedure as above suffices to trivialize the action in any dimension n; since we can ignore the cubic terms, the relevant action is the naive quantum action $S'_q(n)$, for which a Nicolai map is $(k = 1, 2, \ldots)$ [6.133]

$$
\begin{aligned}
\xi(A, \pi) &= F_A + {}^*d_A \pi_{n-3} \, , \\
\xi_k(A, \pi) &= d_A * \pi_{n-2k-1} \pm {}^*d_A \pi_{n-2k-3} \, , \\
\eta(A, \pi) &= d_{A_c} * A_q
\end{aligned}
\tag{6.87}
$$

(with $\pi_i = 0$ for $i < 0$).

An interesting spin-off of the above considerations is the result that there is a complete Nicolai map for Yang–Mills theory on any two-dimensional surface! This comes about as follows. In any dimension the classical BF action can be regarded as the zero coupling limit of Yang–Mills theory since

$$\frac{1}{2g^2} \int F_A * F_A = \int B_{n-2} F_A - \tfrac{1}{2} g^2 B_{n-2} * B_{n-2} \to \int B_{n-2} F_A \, . \tag{6.88}$$

But, whereas for $n \geq 3$ the B^2 term breaks the p-form gauge invariance, this limit is non-singular in two dimensions where both theories have no degrees of freedom. Yang–Mills theory can thus be regarded as a kind of regularization of BF theory [6.158]. This relation extends to the complete quantum action, and evidently the change of variables (6.84) trivializes the Yang–Mills action (6.88) as well. In particular one sees that the partition function of Yang–Mills theory on a surface Σ receives contributions only from the moduli space $\mathcal{M}(\Sigma, G)$ of flat connections. These considerations [6.122], as well as the fact that the classical phase space of Yang–Mills theory on a Riemann surface is independent of the metric [6.159], have led to the suggestion [6.160, 6.138] that Yang–Mills theory is, in a certain sense, a topological field theory in its own right. There are also some indications that Yang–Mills theory is related to conformal field theory [6.160], but this has not yet been confirmed by other methods. The above Nicolai map has already proven useful (in conjunction with the non-Abelian Stokes theorem) in the calculation of correlators of Wilson loops in flat space [6.161].

6.2.6. The self-duality equations on a Riemann surface

In this section we will take a look at two, seemingly unrelated sets of equations in two dimensions: the dimensionally reduced self-duality equations for the group SO(3) (known as the Hitchin equations [6.140, 6.162]), and the equations of motion of a PSL(2, $\mathbb{R}$) BF theory. The latter can be thought of as a theory of topological gravity, since one component of the moduli space $\mathcal{M}(\Sigma_g, \mathrm{PSL}(2, \mathbb{R}))$ of flat connections (cf. sections 5.4.3 and 6.2.7 below) is Teichmüller space $\mathcal{T}_g$.

On the other hand, Hitchin has shown that the moduli space $\mathcal{M}_H = \mathcal{M}_H(\Sigma_g, \mathrm{SO}(3))$ of solutions to the Hitchin equations contains $\mathcal{M}(\Sigma_g, \mathrm{PSL}(2,\mathbb{R}))$ and T_g as (complex) submanifolds. It was therefore suggested in ref. [6.163] to study the dimensional reduction of Donaldson theory to two dimensions as a gauge theory of topological gravity. Although we will not pursue this approach directly, the relation between topological gravity and self-duality will be the underlying theme in the remainder of this chapter.

A priori this relation is far from obvious from the $\mathrm{SO}(3)$ self-duality point of view. One of the reasons why we have included a discussion of the Hitchin equations in the present context is that BF theory provides us with a rather simple way of understanding this result, and in particular Hitchin's construction of constant negative curvature metrics from solutions to the self-duality equations. In fact, we will see that in that sector of $\mathcal{M}_H$ the Higgs fields appearing in the dimensionally reduced self-duality equation can be interpreted as zweibeins on Σ_g, parametrized by Beltrami differentials. The BF or (equivalently) the self-duality equations then tell us directly that the corresponding metric has constant negative curvature.

Another reason for including the Hitchin equations is that this observation suggests a reformulation of the $\mathrm{PSL}(2,\mathbb{R})$ topological gravity theory (and its Witten type counterpart, to be discussed in section 7) as a $\mathrm{U}(1)$ gauge theory coupled to matter ($=$ Higgs fields), the zweibein. This has obvious implications for the cohomological aspects of the theory and, to a certain extent, explains why only the Lorentz ghosts for ghosts, and not those associated with translations or diffeomorphisms, are relevant for the construction of observables in the Witten type models.

Upon dimensional reduction from four to two dimensions, the self-duality equations on a principal $\mathrm{SO}(3)$-bundle on $\mathbb{R}^4$ may be written in a conformally invariant way to make sense on an arbitrary Riemann surface Σ_g, thus giving rise to Hitchin's [6.140] *self-duality equations on a Riemann surface*,

$$F_A = -[\Phi, \Phi^*] \,, \tag{6.89}$$

$$\overline{\partial}_A \Phi = 0 \,. \tag{6.90}$$

Here the notation is the following: F_A is the curvature of a connection A on a principal $\mathrm{SO}(3)$ bundle on Σ_g, $g \geq 2$, $\overline{\partial}_A = \overline{\partial} + A_{\bar{z}} \, \mathrm{d}\bar{z}$ is the anti-holomorphic part of d_A with respect to a given complex structure on Σ_g, and

$$\Phi = \Phi_z \, \mathrm{d}z \in \Omega^{1,0}(\Sigma_g, \mathrm{ad}\, P \otimes \mathbb{C}) \equiv \Omega \,, \qquad \Phi^* = \Phi_{\bar{z}}^* \, \mathrm{d}\bar{z} \in \Omega^{0,1}(\Sigma_g, \mathrm{ad}\, P \otimes \mathbb{C}) \equiv \overline{\Omega}$$

$[\Omega = \Omega^{1,0}(\Sigma_g, \mathbf{g} \otimes \mathbb{C})$ in the notation of section 5.1.2] are complex combinations of the three- and four-components of the original four-dimensional connection.

Equation (6.90) says that Φ is holomorphic with respect to the holomorphic structure on $\mathrm{ad}\, P \otimes_\mathbb{C} K$ [K is the canonical line bundle of $(1,0)$-forms] defined by the connection A on P (cf. ref. [6.159]) and by the complex structure of Σ_g on K, whereas (6.89) can be regarded as a unitarity condition. For our purposes there is no compelling reason for using $\mathrm{SO}(3)$ instead of $\mathrm{SU}(2)$, since the connections we will be interested in below come from principal $\mathrm{SO}(3)$ bundles whose structure group lifts to $\mathrm{SU}(2)$ [i.e. the second Stiefel–Whitney class $w_2(P) = 0$], but for simplicity we will stick to $\mathrm{SO}(3)$.

By an argument analogous to that sketched in section 5.1.4 for the moduli space of instantons (based on use of the Atiyah–Singer index theorem, combined with vanishing and implicit function theorems) Hitchin has shown that the moduli space $\mathcal{M}_H \subset (\mathcal{A} \times \Omega)/\mathcal{G}$ is a smooth $12(g - 1)$ dimensional manifold (we apologize to the reader for once again ignoring the problems caused by the presence of reducible connections—these are treated with great care in ref. [6.140]). It is

also known that $\mathcal{M}_H$ is a non-compact (in the "Φ directions") connected and simply connected *hyper-Kähler* manifold, i.e., there are three symplectic forms with compatible complex structures I, J, K satisfying the quaternion relations $IJ = K$ etc.

I comes from the natural complex structure on $\Omega \oplus \overline{\Omega}$, where we have identified $T_A\mathcal{A} = \Omega^1(\Sigma_g, \mathbf{g})$ (cf. section 5.1.2) with $\overline{\Omega}$, so that $T_{(A,\Phi)}(\mathcal{A} \times \Omega) = \overline{\Omega} \oplus \Omega$. The isomorphism

$$\alpha : \mathcal{A} \times \Omega \to \mathcal{A} \times \overline{\mathcal{A}} \ , \quad \alpha(A, \Phi) = (\overline{\partial}_A + \Phi^*, \partial_A + \Phi) \ ,$$

together with the natural complex structure on $\mathcal{A} \times \overline{\mathcal{A}}$, gives rise to the second complex structure, J, on $\mathcal{A} \times \Omega$, defined by $(T\alpha)J = i(T\alpha)$, where $T\alpha$ is the tangent map of α. The standard metric on $\mathcal{A} \times \Omega$ defines the corresponding symplectic (Kähler) forms ω_I and ω_J (as well as that of $K = IJ$), making $\mathcal{A} \times \Omega$ a hyper-Kähler manifold. As a consequence of a quaternionic version [6.164] of the Marsden–Weinstein theorem [6.15], these Kähler structures pass down to $\mathcal{M}_H$ [6.140] (a brief explanation of this can also be found in ref. [6.66]).

Two further observations will be of interest to us in the following. The first is that the Hitchin equations (6.89), (6.90) imply that under the isomorphism α the PSL(2, $\mathbb{C}$) connection $A + \Phi + \Phi^*$ is flat,

$$F_{A+\Phi+\Phi^*} = 0 \ . \tag{6.91}$$

For irreducible connections Donaldson [6.165] has established a converse to this result but, as a consequence of the presence of reducible connections, $(\mathcal{M}_H, J)$ is not the moduli space of flat PSL(2, $\mathbb{C}$) connections but rather a covering space thereof.

The second observation concerns the existence of a circle action $(A, \Phi) \to (A, e^{i\theta}\Phi)$ and, in particular, an involution $(A, \Phi) \to (A, -\Phi)$ on $\mathcal{A} \times \Omega$, which—since it maps solutions to solutions—passes down to a circle action (involution) on $\mathcal{M}_H$. As can be checked from the above definitions, this involution σ is anti-holomorphic with respect to J, i.e., $(T\sigma)J = -J(T\sigma)$, and thus equips $(\mathcal{M}_H, J)$ with a *real* structure. The fixed points of σ [the real points of $(\mathcal{M}_H, J)$] then satisfy an additional reality constraint.

A simple example may help to explain what is going on: consider a two-dimensional real vector space V, $(x, y) \in V$; the identification $V \sim \mathbb{C}$, $(x, y) \sim x + iy$, equips V with the complex structure $J(x, y) = (-y, x)$; the involution $\sigma(x, y) = (x, -y)$ satisfies $\sigma J = -J\sigma$, and the fixed points of σ are the points $(x, 0)$, corresponding to the standard real line $\mathbb{R} \subset \mathbb{C}$ under the above identification.

If the pair (A, Φ) itself is fixed by σ (and not only its gauge equivalence class), then obviously $\Phi = 0$, and (6.89) then tells us that we are dealing with flat SO(3) connections [and SO(3) is indeed a real form of PSL(2, $\mathbb{C}$)]. Otherwise we are dealing with flat PSL(2, $\mathbb{R}$) connections (the other real form of PSL(2, $\mathbb{C}$)). We will say more about the corresponding moduli space $\mathcal{M}(\Sigma_g, \text{PSL}(2, \mathbb{R}))$ below. Here we just note that [6.140, Prop. 10.2] all but one of the components of $\mathcal{M}(\Sigma_g, \text{PSL}(2, \mathbb{R}))$ are smooth submanifolds of $\mathcal{M}_H$, which are in fact complex submanifolds of $(\mathcal{M}_H, I)$, since σ is holomorphic with respect to I, $(T\sigma)I = +I(T\sigma)$.

Hitchin now goes on to show how to construct constant curvature metrics from solutions (A, Φ) in a particular component $(= \mathcal{T}_g)$ of the fixed point set of σ (as corollaries giving new proofs of the uniformization theorem and the isomorphism $\mathcal{T}_g \approx \mathbb{C}^{3g-3}$). For later reference we sketch Hitchin's argument here.

Consider the complex rank 2 vector bundle $V = K^{1/2} \oplus K^{-1/2}$. With respect to this decomposition, any Φ of the form

$$\Phi(q) = \begin{pmatrix} 0 & q \\ 1 & 0 \end{pmatrix} \in \Omega^{1,0}(\Sigma_g, \text{End}_0 V) \tag{6.92}$$

(here we have replaced $\mathrm{ad}\,P$ by $\mathrm{End}_0\,V$, the traceless endomorphisms of V) will satisfy (6.90) for reducible connections with

$$A_{\bar{z}} \sim \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{6.93}$$

provided that $q \in \Omega^{1,0}(\Sigma_g, \mathrm{Hom}(K^{-1/2}, K^{1/2}))$ is holomorphic, i.e., $q \in H^0(\Sigma_g, \mathrm{Hom}(K^{-1/2}, K^{1/2})) \otimes K) = H^0(\Sigma_g, K^2)$ is a (holomorphic) quadratic differential. An existence theorem [based on what Hitchin calls the stability of a pair (V, Φ)] provides a unique solution $A_{\mathrm{H}}(q)$ to (6.89). $A_{\mathrm{H}}(q)$ determines a compatible metric $g(q)$ on K and Σ_g, from which Hitchin constructs a new metric $\hat{g}(q)$ with constant negative curvature. Instead of explaining this in more detail, we will now turn our attention towards topological gravity, which will lead us to an alternative and less sophisticated way of understanding parts of the above construction.

6.2.7. Moduli spaces of flat connections II: $\mathrm{PSL}(2,\mathbb{R})$

As a preparation for our discussion of topological gravity we summarize some results on the moduli space $\mathcal{M}(\Sigma_g, \mathrm{PSL}(2,\mathbb{R}))$ [6.166]. Let us start by clarifying the relation between $\mathrm{PSL}(2,\mathbb{R})$ and the groups $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{SO}(2,1)$, which have alternatively been used as gauge groups for topological gravity. At the Lie algebra level these are indistinguishable, but when dealing with global objects like moduli spaces the differences are important. $\mathrm{SL}(2,\mathbb{R})$ is the non-trivial double covering of $\mathrm{PSL}(2,\mathbb{R})$. The Killing form of $\mathrm{PSL}(2,\mathbb{R})$ defines an ad-invariant indefinite quadratic form on its Lie algebra. The adjoint representation thus embeds $\mathrm{PSL}(2,\mathbb{R})$ in $\mathrm{SO}(2,1)$, and under this embedding $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SO}_0(2,1)$, the component of the identity of $\mathrm{SO}(2,1)$. This also gives rise to the identification $\mathrm{SL}(2,\mathbb{R}) = \mathrm{Spin}(2,1)$. Now that there is no possible confusion, we will in the remainder of this section let G denote $\mathrm{PSL}(2,\mathbb{R})$ and $\mathcal{M}$ its moduli space.

The components of $\mathrm{Hom}(\pi, G)$ (cf. section 5.4.3) are indexed by the Euler number $\chi(E)$ of the associated $\mathbb{RP}^1$-bundle, and for a flat bundle one has the strict bound $|\chi(E)| \leq |\chi(\Sigma_g)| = 2g - 2$, so that the number of components of $\mathcal{M}$ is $2(2g - 2) + 1 = 4g - 3$. Bundles E, with even $\chi(E)$, lift to real two-plane bundles $\tilde{E}$ associated to $\mathrm{SL}(2,\mathbb{R})$, and $\chi(E) = 2\chi(\tilde{E})$. In particular therefore, $|\chi| \leq g - 1$ for a flat $\mathrm{SL}(2,\mathbb{R})$-bundle (this fact, discovered by Milnor [6.167], provided the first examples of topologically non-trivial flat bundles). The components of the $\mathrm{SL}(2,\mathbb{R})$ moduli space are, in contrast to those of $\mathcal{M}$, not classified completely by their Euler number. In genus 2, for instance, $\mathrm{Hom}(\pi, G)$ has five components ($|\chi| \leq 2$), while $\mathrm{Hom}(\pi, \mathrm{SL}(2,\mathbb{R}))$ has 33 components, so that the situation tends to get out of hand.

For all $k \neq 0, |k| \leq 2g - 2$, the components $\mathcal{M}_k$ of $\mathcal{M}$ are real analytic Hausdorff manifolds, and since they contain only irreducible connections (a reducible connection would give a section of the associated bundle E and would therefore force $k = 0$), the index formula (5.140) of section 5.4.3 determines their dimension to be $(6g - 6)$.

We will be interested primarily in one particular component of $\mathcal{M}$, namely $\mathcal{M}_{2g-2}$. The reason for this is that uniformization of a Riemann surface, $\Sigma_g = H/\Gamma_g$ (H is the upper half plane and Γ_g is a discrete subgroup of G acting on H by isometries of the Poincaré metric), defines a representation ϕ of $\Gamma_g = \pi$ in $\mathrm{PSL}(2,\mathbb{R})$ which determines a flat $\mathbb{RP}^1$-bundle E with $\chi(E) = |\chi(\Sigma_g)| = 2g - 2$: $\mathcal{M}_{2g-2} = T_g$ is Teichmüller space. The condition that $\phi \in \mathrm{Hom}(\pi, G)$ be an isomorphism onto a *discrete* subgroup of G (which singles out $\mathcal{M}_{2g-2}$) will reappear in the next section in the form of an invertibility condition for the zweibein. Since $2g - 2$ is even, T_g can alternatively be regarded as one component of the moduli space of flat $\mathrm{SL}(2,\mathbb{R})$ connections, and this is the reason why we could just as well have worked with $\mathrm{SU}(2)$ instead of $\mathrm{SO}(3)$ in the previous section.

6.2.8. Topological gravity and self-duality

Given all this information, the construction of a topological gravity theory is now straightforward. We choose generators (J, P_1, P_2) of the Lie algebra of PSL$(2,\mathbb{R})$ [or SO$(2,1)$] with the commutation relations

$$[P_a, P_b] = -\rho\varepsilon_{ab}J \ , \qquad [J, P_a] = \varepsilon_a{}^b P_b \ . \tag{6.94}$$

Here ρ is a positive real parameter, $\varepsilon_{12} = 1$, and indices are raised and lowered with the metric δ_{ab}. By a rescaling of the generators P_a, ρ can be set to $+1$, but since in section 7 we will be interested in the $\rho \to 0$ contraction of (6.94), the Lie algebra of ISO(2), it is more convenient to keep the ρ dependence explicit.

This algebra [with the field assignments as in (6.96) below] describes Euclidean quantum gravity with a negative cosmological constant. Had we chosen ρ to be negative instead, the above algebra [now that of SO(3)] would have described the same theory with a positive cosmological constant (i.e. at genus zero). Lorentzian quantum gravity is obtained by replacing δ_{ab} by $\eta_{ab} = \mathrm{diag}(+1, -1)$.

In this basis for PSL$(2,\mathbb{R})$, the non-degenerate invariant Killing–Cartan metric is

$$\langle J, J \rangle = -1 \ , \qquad \langle P_a, P_b \rangle = \rho\delta_{ab} \ , \tag{6.95}$$

with signature $(- + +)$, and allows us to write down a non-degenerate gauge invariant BF action in the usual way. We will comment below on how that can also be achieved in the $\rho \to 0$ limit, where (6.95) obviously becomes degenerate. We expand the PSL$(2,\mathbb{R})$ connection A and the multiplier B as

$$A = J\omega + P_a e^a \ , \qquad B = JB^0 + P_a B^a \ , \tag{6.96}$$

where the coefficients ω and e^a are of course ultimately to be identified with the spin connection and zweibein, respectively. Under gauge transformations $\delta A = d_A\lambda$ they transform as

$$\delta\omega = d\lambda^0 - \rho\varepsilon_{ab}e^a\lambda^b \ , \qquad \delta e^a = d\lambda^a + \varepsilon^a{}_b(\lambda^0 e^b - \lambda^b\omega) \ . \tag{6.97}$$

The action

$$S = \int_{\Sigma_g} BF_A = \int_{\Sigma_g} -B^0(d\omega - \tfrac{1}{2}\rho\varepsilon_{ab}e^a e^b) + \rho\delta_{ab}B^a(de^b - \omega\varepsilon^b{}_c e^c) \tag{6.98}$$

leads to the equations of motion

$$d\omega = \rho\tfrac{1}{2}\varepsilon_{ab}e^a e^b \ , \tag{6.99}$$

$$de^a = \omega\varepsilon^a{}_b e^b \ , \tag{6.100}$$

$$d_A B = 0 \ . \tag{6.101}$$

In (6.100) we recognize the no-torsion equation for the spin connection $\omega^a{}_b = -\omega\varepsilon^a{}_b$. Provided that e is invertible, there is a unique solution $\omega(e)$ to (6.100). With $\omega = \omega(e)$, (6.99) is then the statement that the metric $g_{\mu\nu} = \delta_{ab}e^a_\mu e^b_\nu$ associated to e has constant negative scalar curvature, in our conventions

$$R(g) = -2\rho \ . \tag{6.102}$$

Conversely, of course, the flat $PSL(2,\mathbb{R})$ connection associated to g via uniformization is $A(g) = J\omega(e) + P_a e^a$. As mentioned above, the condition of invertibility singles out the component $T_g = \mathcal{M}_{2g-2}$ of the moduli space of flat connections (an explicit proof of this can be found in ref. [6.168]). The above then establishes very directly the relation between Teichmüller space [defined in terms of $PSL(2,\mathbb{R})$ connections] and the space of $\mathrm{Diff}_0(\Sigma_g)$ classes of constant negative curvature metrics, since, as in Chern–Simons gravity, diffeomorphisms are on shell equivalent to the gauge transformations (6.97).

It remains to analyze the third of the equations of motion, $d_A B = 0$. Since, as mentioned in the previous section, the $PSL(2,\mathbb{R})$ moduli spaces $\mathcal{M}_k$ for $k \neq 0$ contain only irreducible connections, there are no non-trivial solutions to (6.101) in these sectors of the theory. In the gravity sector $\mathcal{M}_{2g-2} = T_g$, where this assertion can, in fact, easily be verified directly, this implies that the space of solutions to the complete set of equations of motion (6.99)–(6.101) is simply Teichmüller space T_g itself. If one considers Lorentzian gravity and non-compact surfaces instead, non-trivial solutions to (6.101) will generally exist.

The action (6.98) appears to have been first written down in 1985 by Fukuyama and Kamimura [6.169] as a gauge theory description of the Jackiw–Teitelboim model [6.170] of 2D gravity. In the formulation of Jackiw, this gravity theory is governed by the action

$$S_{\mathrm{JT}} = \int \sqrt{-g}\, N(R(g) - \Lambda) \ . \tag{6.103}$$

Here N is an auxiliary field enforcing the field equation (6.102) of the BF theory, in this context also known as the Liouville equation. The $SO(2,1)$ invariance of this equation had long been recognized, and played an important role in early attempts at quantizing Liouville theory (see, e.g., ref. [6.171]). Note also that, upon substitution of ω by $\omega(e)$, (6.98) reduces to (6.103) (with $B^0 = N$).

The action (6.98) was subsequently rediscovered and studied, in the context of topological field theory, by various groups [6.121, 6.124, 6.122]. Of course, all our general considerations, concerning its quantization, the existence of a Nicolai map, and the relation to the Ray–Singer torsion, are equally valid in this particular case.

Let us comment briefly on the $\rho \to 0$ contraction of the above. In that case, the Lie algebra (6.94) reduces to that of $ISO(2)$, which has no non-degenerate invariant scalar product. In spite of this fact, it is possible to construct an invariant BF action for this group This is accomplished by adopting transformation rules for B which are not the conventional $ISO(2)$ transformations, but which nevertheless arise quite naturally from the $PSL(2,\mathbb{R})$ transformations via contraction. The latter are $\delta B = [B, \lambda]$, i.e.,

$$\delta B^0 = \rho \varepsilon_{ab} \lambda^a B^b \ , \qquad \delta B^a = \varepsilon^a{}_b (\lambda^0 B^b - \lambda^b B^0) \ . \tag{6.104}$$

Naively taking the limit $\rho \to 0$ in eqs. (6.95), (6.97), (6.98) and (6.104), one is led to the invariant, but degenerate and quite boring, action $S = -\int B^0 d\omega$, with its invariances $\delta\omega = d\lambda^0, \delta B^0 = 0$. But if we rescale B^a by ρ in (6.104) and then take the limit $\rho \to 0$, we arrive at

$$\delta B^0 = \varepsilon_{ab} \lambda^a B^b \ , \qquad \delta B^a = \varepsilon^a{}_b \lambda^0 B^b \ . \tag{6.105}$$

It can now be checked that, with these transformation rules [and the conventional $ISO(2)$ transformation rules for A, the contraction of (6.97)], the action

$$S = \int -B^0 d\omega + \delta_{ab} B^a (de^b - \omega \varepsilon^b{}_c e^c) \tag{6.106}$$

is indeed invariant. As it stands, (6.106) is not particularly useful, since the equation of motion $d\omega = 0$ obviously implies that we are working on the torus. But in section 7 we will show, following Verlinde and Verlinde [6.172], how a clever modification of this action (or, rather, its super-BF counterpart) can describe Witten type topological gravity on surfaces of any genus.

We now return to the self-duality equations. Equipped with all this information, the identification between the fields (A_H, Φ) appearing in the Hitchin equations, and the connection $A = J\omega + P_a e^a$, eq. (6.96), of our topological gravity theory, is now straightforward. Before proceeding, we should perhaps emphasize that, while this identification immediately implies that the Hitchin equation (6.89) is a constant curvature condition on a metric, the power of Hitchin's argument (sketched at the end of section 6.2.6) lies in the fact that it automatically provides *solutions* to this equation. This is something the more simple-minded BF approach cannot do for us.

Let us introduce the following basis (with $\rho = 1$) of $\mathrm{End}_0(K^{1/2} \oplus K^{-1/2})$:

$$J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \qquad P_1 = \begin{pmatrix} 0 & 0 \\ \partial_z & 0 \end{pmatrix} , \qquad P_2 = \begin{pmatrix} 0 & dz \\ 0 & 0 \end{pmatrix} . \tag{6.107}$$

With respect to this basis, the reducible connection $A_H = A_H(q)$, compatible with the metric $g(q)$, takes the form $A_H = J\omega$ [cf. eq. (6.93)]. Expanding the Higgs fields $\Phi(q)$, eq. (6.92), and $\Phi^*(q)$ as

$$\Phi_z(q)\, dz + \Phi_{\bar{z}}^*(q)\, d\bar{z} = P_a \Phi^a(q) \ ,$$

one finds [writing $q = q_{zz}(dz)^2$]

$$\Phi^1(q)dz + g^{z\bar{z}}\bar{q}_{\bar{z}\bar{z}}\, d\bar{z} \ , \qquad \Phi^2(q) = q_{zz}\, dz + g_{z\bar{z}}\, d\bar{z} \ . \tag{6.108}$$

We thus see that the Higgs fields $\Phi(q)$ and $\Phi^*(q)$ can be interpreted as zweibeins on Σ_g. In terms of Beltrami differentials $\mu^z{}_{\bar{z}}$, related to the quadratic differentials q_{zz} via $\mu^z{}_{\bar{z}} = g^{z\bar{z}}\bar{q}_{\bar{z}\bar{z}}$, these take the more familiar form

$$\Phi^1(q) \equiv e^1(\mu) = dz + \mu^z{}_{\bar{z}}\, d\bar{z} \ , \qquad \Phi^2(q) \equiv e^2(\mu) = g_{z\bar{z}}(\mu)(d\bar{z} + \bar{\mu}^{\bar{z}}{}_z\, dz) \ . \tag{6.109}$$

As $(A_H(q), \Phi(q))$ is a solution to the Hitchin equations,

$$A_H + \Phi(q) + \Phi^*(q) = J\omega + P_a e^a(\mu)$$

is flat, eq. (6.91), and therefore defines a solution of topological gravity in the T_g sector. In particular, the Hitchin equation $F_{A_H} = -[\Phi(q), \Phi^*(q)]$ is then nothing other than the statement that the metric

$$\hat{g}(\mu) = e^1(\mu) \otimes e^2(\mu) = g_{z\bar{z}}(\mu)(dz + \mu^z{}_{\bar{z}}\, d\bar{z})(d\bar{z} + \bar{\mu}^{\bar{z}}{}_z\, dz) \tag{6.110}$$

has constant negative curvature. This is precisely the metric constructed by Hitchin—in the shorthand notation of ref. [6.140, Theorem 11.2],

$$\hat{g}(q) = q + (g + q\bar{q}/g) + \bar{q} \ .$$

These observations on the relation between topological gravity and the Hitchin equations may have some interesting consequences. They suggest, for instance, that topological gravity can be formulated as a $U(1)$ gauge theory, coupled to matter (Higgs) fields which can be interpreted as zweibeins on a Riemann surface in a particular phase of that theory. In this phase the metric

282 D. Birmingham et al., Topological field theory

emerges as a composite matter field. Moreover, the above suggests that relevant cohomological information about the theory is encoded only in the $U(1)$ part of the connection (6.96), and that the translation or (in other formulations) diffeomorphism sectors do not contribute non-trivially to observables in the corresponding Witten type theory. This conclusion is supported by the fact that $\mathcal{M}_H$ contains the cotangent bundle of the moduli space of stable vector bundles as an open dense set, the topologically trivial fiber directions being spanned essentially by the Higgs fields Φ.

7. Topological gravity

7.1. Introduction

Quantum gravity in two dimensions is a subject that has received considerable attention recently. String theory, in its first quantized form, as formulated by Polyakov [7.1, 7.2], is the study of two-dimensional gravity coupled to d bosonic fields. There is a critical dimension, $d = 26$, in which this theory is easily analyzed, while for other values of d there have been, and still are, great conceptual and computational difficulties to surmount. In an attempt to move away from this critical dimension, Polyakov, Knizhnik and Zamolodchikov [7.3, 7.4] quantized the quantum gravity and matter action in a light cone gauge. In this gauge, it is possible to completely solve the theory when $d < 1$. The reason for this is that there is a residual $SL(2, \mathbb{R})$ symmetry that may be employed to determine the anomalous dimensions of all conformal field theory operators that have the correct conformal weight.

Distler and Kawai [7.5] and David [7.6] have reproduced these results in the conformal gauge. The significance of this approach is that it allows for an extension to higher genus surfaces, a possibility not directly available in the light cone gauge. The critical exponents were straightforwardly calculated and the partition function as a function of the area was determined. But this is not the complete story yet, since there are some assumptions in these derivations that, although natural, are non-trivial to check, and we refer to the recent work of D'Hoker [7.7] for a discussion and a resolution of some of the problems involved.

The approaches outlined above are based on continuum field theory on Riemann surfaces. An alternative that has been developed is to replace the two-dimensional surface with a triangulation of it. The dynamics of the geometry is then encoded into the sum over all triangulations, which replaces the path integral over the metric. The weightings assigned to the vertices and edges are determined by the requirements of a fixed area—to be integrated over at the end—and correct Euler number. Technically, this is achieved by considering the dual lattice, and treating it as being generated by a Φ^3 matrix theory of $N \times N$ Hermitian matrices. These discrete theories are, somewhat surprisingly, easier to deal with, and the results obtained agree with their continuum counterparts. The advantages of these methods are that they yield non-perturbative information. This has been made manifest in the remarkable exact solutions of Brézin and Kazakov, Douglas and Shenker, and Gross and Migdal [7.8–7.10]. These authors were able to turn the problem of determining the matrix model partition function into one of solving a particular differential equation. With these developments, one has the prospect of finding nonperturbative ground states in string theory (albeit still with $d < 1$).

In a seemingly different direction, Labastida, Pernici and Witten [7.11] constructed a topological field theory (of Witten type) for gravity in two dimensions. Theirs is a metric approach to topological gravity. A gauge theory version, having some advantages over the metric formulation, was put forward in ref. [7.12]. The natural observables in this context, suggested in ref. [7.11],

are related to the so called Mumford classes [7.13–7.15]. This identification was further pursued in refs. [7.12, 7.16]. Soon after, Witten [7.17] and Distler [7.18] established that particular observables in topological gravity correspond to the correlation functions of the $k = 1$ matrix models, the evaluation of the topological observables agreeing with the values obtained in ref. [7.10]. Distler arrived at this result by firstly identifying $d = -2$ matter coupled to quantum gravity as the topological gravity model of Labastida, Pernici and Witten. Using the properties of this topological field theory and repeated bosonization he was able to determine the correlation functions of the arbitrary k matrix models on the sphere. Witten's approach does not rely on any particular Lagrangian. Instead, correlation functions are related recursively, $n + 1$-point functions to n-point funtions, using general properties of topological field theory on Riemann surfaces, as well as additional information provided by the structure of the compactified moduli space of Riemann surfaces. In this way, all the correlation functions on the sphere could be determined, and in ref. [7.19] these considerations were extended to genus $g = 1$ as well as to the higher matrix models. The explicit formulae on higher genus surfaces are difficult to arrive at in this very general setting. Verlinde and Verlinde [7.20] were able, using a particular gauge theoretic action for topological gravity, to derive recursion relations on arbitrary genus surfaces. Together with the results of ref. [7.21], this established the equivalence of pure topological gravity and the $k = 1$ matrix model. Based on this and other evidence it is now generally believed that minimal conformal matter coupled to two-dimensional gravity is equivalent to topological gravity coupled to certain topological matter.

We cannot do justice to all of the above topics in a short space. Reviewing conformal field theory and matrix models would take us too far afield, so in this section we have settled for the more modest aim of establishing the connection between the various formulations of topological gravity in etwo dimensions mentioned above.

7.2. Two-dimensional gravity

The moduli space of interest in two-dimensional gravity is the moduli space $\mathcal{M}_g$ of Riemann surfaces of genus g. As we have already alluded to in section 2, several different, but equivalent, definitions of $\mathcal{M}_g$ are possible. Quite generally, we will take topological gravity to be a field theoretic realization of any of these classical descriptions.

Let us then formally define $\mathcal{M}_g$ to be the space of metrics on a (compact, oriented) surface Σ_g with g handles, quotiented by the action of the group $\text{Diff}(\Sigma_g)$ of orientation preserving diffeomorphisms of Σ_g, and the Weyl group $W(\Sigma_g)$ of conformal rescalings of the metric. The finite-dimensional space $\mathcal{M}_g$ is not quite a smooth manifold but if, instead of $\text{Diff}(\Sigma_g)$, one takes the group $\text{Diff}_0(\Sigma_g)$ of diffeomorphisms connected to the identity, one obtains Teichmüller space $\mathcal{T}_g$, which is smooth. $\mathcal{M}_g$ is then the orbifold quotient of $\mathcal{T}_g$ by the modular (or mapping class) group $\pi_0(\text{Diff}(\Sigma_g)) = \text{Diff}(\Sigma_g)/\text{Diff}_0(\Sigma_g)$. Since large diffeomorphisms (like large gauge transformations) are difficult to implement at the level of Lagrangians, the topological theories we will discuss below appear to give us only a description of $\mathcal{T}_g$ (which is topologically trivial). But, provided that the action and the observables are modular invariant, this is sufficient to describe the topologically non-trivial moduli space $\mathcal{M}_g$.

The first approach to topological gravity, suggested by this definition of $\mathcal{M}_g$, will then be based on the fact that two metrics on Σ_g can only differ from each other by diffeomorphisms, Weyl rescalings, and finite deformations. It is therefore possible, using the shift, diffeomorphism and scaling symmetries, to descend directly to $\mathcal{M}_g$ by demanding that the dynamical metric (i.e., the one to be integrated over in the path integral) be equal to a given fixed metric on Σ_g, perhaps up to terms that parametrize the moduli space (e.g., Beltrami or quadratic differentials). Even so,

there are still some choices to be made in the gauge fixing. In section 7.3 we will choose one gauge, studied by Labastida, Pernici and Witten [7.11], which retains conformal invariance of the action with respect to the metric that is singled out on the manifold. The BRST symmetry of this action is not the supersymmetry used by Distler [7.18] in his approach to topological gravity, and in order to make contact with his work (section 7.4), we explain the existence of this second supersymmetry and, directly related to that, the equivariant nature of the Labastida–Pernici–Witten action. Another "metric" construction of topological gravity could be based on the field equation $R = \Lambda = $ const., together with diffeomorphism invariance, but we will not pursue this here.

In sections 6.2.7 and 6.2.8 we have already seen that a gauge theoretic description of T_g is possible. In this case the spin connection and the zweibein are (initially) treated on an equal footing—as independent components of a gauge field. This theory is an example of the super-BF theories discussed in section 5.4, and was introduced in ref. [7.12]. Here too, there is some latitude in the possible gauge choices (and gauge groups), and we shall concentrate on an economical and efficient formulation due to Verlinde and Verlinde [7.20].

There is another description of T_g and $\mathcal{M}_g$ which we have not yet mentioned, in terms of complex structures on the surface Σ_g. In two dimensions, a complex structure J is equivalent to a conformal equivalence class of metrics. Teichmüller space may then be identified with the space of complex structures quotiented by $\mathrm{Diff}_0(\Sigma_g)$. In this way of dealing with the theory, the Weyl invariance is an irrelevant concept. One should then get a third description of topological gravity following these lines. Again we do not pursue this here. However, for an application to string theory see ref. [7.22].

The various possibilities of describing the moduli space of interest encourage one to believe that it is possible to dispense with the action altogether. Indeed, topological gravity should perhaps (like all other Witten type theories) most fundamentally be regarded as *being* intersection theory on moduli space. The relevance of a field theoretic realization should nevertheless not be underestimated. A judicious choice of starting action, combined with standard field theoretic manipulations, can lead to considerable conceptual clarity and power of computation.

7.3. The Labastida–Pernici–Witten action

In this section we will outline an alternative to the usual method of constructing a topological field theoretic description of moduli spaces. Our constructions so far have been based on the use of defining equations (field equations) for the moduli space. It is, however, also possible to constrain the fields purely algebraically to lie "in the moduli space". In the case of the moduli space of Riemann surfaces this amounts to requiring

$$g_{\alpha\beta} = g(\gamma)_{\alpha\beta} , \qquad (7.1)$$

where $g(\gamma)_{\alpha\beta}$ represents the metrics on Σ_g that parametrize the moduli space; the Teichmüller parameters γ can, for instance, be chosen to be Beltrami differentials. The usual shift invariance is, of course, large enough to guarantee that this is possible. In the context of Donaldson theory the analogous condition would be that the gauge field is set to one parametrized by the moduli of self-dual instantons.

Though in (7.1) we have indicated that the metric is fixed to lie on the moduli space, it is possible to impose the stronger condition that the metric is restricted to be a preferred (fixed) metric on Σ_g,

$$g_{\alpha\beta} = g^0_{\alpha\beta} , \qquad (7.2)$$

and we shall follow this alternative in the sequel.

It may seem unlikely at first that both gauge choices describe the moduli space; nevertheless, this is indeed the case. With the first condition, obviously describing the moduli space, the topological field theory action, in a simplified form, is that of the diffeomorphism ghosts and their BRST partners. In this action the ghosts have no zero modes and the moduli already appear in $g(\gamma)$. The second condition leads to a formally similar ghost action. In this instance, however, there are anti-ghost zero modes to parametrize the moduli, even though the metric g^0 is fixed. This will be exhibited below.

The field content and (nilpotent) transformation rules adopted by Labastida, Pernici and Witten are ($\mathcal{L}$ is the Lie derivative)

$$\{Q, g_{\alpha\beta}\} = \psi_{\alpha\beta} - \rho g_{\alpha\beta} + \mathcal{L}_c(g_{\alpha\beta}) \ , \qquad \{Q, \psi_{\alpha\beta}\} = -\mathcal{L}_\phi(g_{\alpha\beta}) - \mathcal{L}_c(\psi_{\alpha\beta}) - \rho\psi_{\alpha\beta} + \tau g_{\alpha\beta} \ ,$$

$$\{Q, c^\alpha\} = \tfrac{1}{2}\mathcal{L}_c(c^\alpha) + \phi^\alpha \ , \qquad \{Q, \phi^\alpha\} = \mathcal{L}_c(\phi^\alpha) - \mathcal{L}_\phi(c^\alpha) \ ,$$

$$\{Q, b^{\alpha\beta}\} = d^{\alpha\beta}, \qquad \{Q, d^{\alpha\beta}\} = 0 \ ,$$

$$\{Q, \rho\} = \mathcal{L}_c(\rho) + \tau \ , \qquad \{Q, \tau\} = \mathcal{L}_c(\tau) - \mathcal{L}_\phi(\rho) \ ,$$

$$\{Q, B^{\alpha\beta}\} = D^{\alpha\beta} \ , \qquad \{Q, D^{\alpha\beta}\} = 0 \ . \tag{7.3}$$

The transformations for the metric are decomposed into the shift symmetry, conformal transformations and diffeomorphisms. The ghost fields parametrizing these all have ghost number one. The other fields (ϕ^α, τ, $B^{\alpha\beta}$, $D^{\alpha\beta}$, $b^{\alpha\beta}$, $d^{\alpha\beta}$) have ghost number $(2, 2, -2, -1, -1, 0)$, respectively. The tensors B, D, ψ, b and d are symmetric.

The action is taken to be

$$S = \int_\Sigma \sqrt{g^0} \{Q, [b^{\alpha\beta}(g_{\alpha\beta} - g^0_{\alpha\beta}) + B^{\alpha\beta}\psi_{\alpha\beta}]\} \ . \tag{7.4}$$

Before expanding this out, let us note that there is a second symmetry in this theory. The integrand is clearly invariant under the transformations

$$s g_{\alpha\beta} = \psi_{\alpha\beta} \ , \qquad s\psi_{\alpha\beta} = 0 \ , \qquad sB^{\alpha\beta} = b^{\alpha\beta} \ , \qquad sb^{\alpha\beta} = 0 \ , \tag{7.5}$$

and the action may indeed be rewritten as

$$S = \int_\Sigma \sqrt{g^0} \{Q, s[B^{\alpha\beta}(g_{\alpha\beta} - g^0_{\alpha\beta})]\}. \tag{7.6}$$

The s symmetry is of course the usual shift symmetry and so differs from Q by the diffeomorphisms and conformal scalings. They satisfy the algebra

$$s^2 = \{Q, Q\} = \{Q, s\} = 0 \ , \tag{7.7}$$

which determines the transformation rules under s for the rest of the fields.

The importance of the observation that the action has two invariances lies in the fact that one may work "equivariantly". The two symmetries allow for a good description of the observables in topological gravity quite generally [7.20, 7.23]. However, for the moment, let us continue with our analysis of the action. Expanding, we find

$$S = \int_\Sigma \sqrt{g^0} \Big[d^{\alpha\beta}(g_{\alpha\beta} - g^0_{\alpha\beta}) + D^{\alpha\beta}\psi_{\alpha\beta} - b^{\alpha\beta}(\mathcal{L}_c(g_{\alpha\beta}) + \psi_{\alpha\beta} - \rho g_{\alpha\beta})$$

$$+ B^{\alpha\beta}(\mathcal{L}_c(\psi_{\alpha\beta}) - \rho\psi_{\alpha\beta}) - B^{\alpha\beta}(\mathcal{L}_\phi(g_{\alpha\beta}) - \tau g_{\alpha\beta}) \Big] \ . \tag{7.8}$$

Matters are greatly simplified by integrating out the multiplier fields d and D, since then the constraint (7.2) is imposed and ψ is set to zero. On performing the ρ and τ integrals, the trace parts, with respect to the metric g^0, of b and B are also set to zero. The action now has the particularly simple form

$$S = \int_\Sigma \sqrt{g^0} \left(b^{\alpha\beta} \nabla^0_\alpha c_\beta + B^{\alpha\beta} \nabla^0_\alpha \phi_\beta \right) , \qquad (7.9)$$

where the covariant derivative ∇^0 is with respect to the background metric and the labels on ϕ and c have been lowered with that metric. b and B satisfy the tracelessness conditions $g^0_{\alpha\beta} b^{\alpha\beta} = 0$ and $g^0_{\alpha\beta} B^{\alpha\beta} = 0$. In complex notation this action takes the more familiar form $S = \int_\Sigma b\bar\partial c + B\bar\partial\phi + \text{c.c.}$ of a supersymmetric b–c system [7.24].

After all these manipulations one may wonder which of the symmetries survive. In fact, since we are only using algebraic equations of motion to eliminate fields, we maintain all the symmetries (see section 3.3.1). We note that the s symmetry remains manifest. It is straightforward to see that the action of Q on the remaining fields equals that of s up to diffeomorphisms.

We now have the action and the symmetries, and it only remains to interpret the theory. The Riemann–Roch theorem gives us the number of zero modes of the traceless symmetric tensor B (b) minus the number of zero modes of ϕ (c). On the sphere there are no B (b) zero modes, while on the torus there is one. Let us concentrate on genus $g \geq 2$. We have $6g - 6$ B and b zero modes, whereas there are no zero modes of ϕ or c. Now, $\dim T_g = 6g - 6$, and the traceless symmetric B zero modes (equivalently, holomorphic quadratic differentials) parametrize Teichmüller space. Moreover, due to $sB = b$, the anti-commuting b zero modes may be considered as one-forms on T_g, with s acting as the exterior derivative. In accordance with the arguments of section 5.3.2 and the general structure of Witten type theories, the complete BRST operator Q also reduces to the exterior derivative on T_g and $\mathcal{M}_g$, because local diffeomorphisms are inoperative there. The identification of the B zero modes as local coordinates on Teichmüller space and the b zero modes as differentials is strengthened by the fact that, when the moduli are explicitly parametrized in the metric (7.1), there are precisely enough degrees of freedom generated so as to be able to enforce that the B and b zero modes are set to zero [7.11]. Such a cancellation prevents an over-representation of Teichmüller space.

7.4. Relationship with quantum gravity and matrix models

Distler [7.18] has derived the model of the previous section from a totally different point of view. Consider a theory of d bosons coupled in the normal way to quantum gravity in two dimensions. One may gauge fix the quantum metric to the conformal gauge[*]

$$g_{\alpha\beta} = e^\sigma g^0_{\alpha\beta} . \qquad (7.10)$$

Naive conformal invariance would seem to indicate that the theory is, in fact, independent of σ. However, as is well known, the conformal anomaly prevents this from being true. It can be shown

---

[*] We have chosen to be consistent in using the symbol ϕ to denote the ghost associated with the shift field ψ throughout this report; hence our use of σ to denote the Liouville field. Our notation therefore differs slightly from that of the papers we are describing.

that an effective theory of "induced gravity" is generated, with the action [7.5, 7.6]

$$S = \frac{1}{2\pi} \int\limits_{\Sigma} \mathrm{d}^2x \sqrt{g^0} \left(g^{0\alpha\beta} \partial_\alpha \sigma \partial_\beta \sigma - \tfrac{1}{4} q R^0 \sigma - 2\mu e^{a\sigma} \right) + \frac{1}{\pi} \int\limits_{\Sigma} \sqrt{g^0} b^{\alpha\beta} \nabla^0_\alpha c_\beta \ . \tag{7.11}$$

The constants that appear are

$$q = \sqrt{(25-d)/3} \ , \qquad a = -(1/\sqrt{12}) \left[\sqrt{(25-d)} - \sqrt{(1-d)} \right] \ , \tag{7.12}$$

and R^0 is the curvature scalar of the metric g^0. Notice that the ghosts appear in exactly the same way as in (7.9). This is because the condition (7.2) matches (7.10), when the σ field is scaled out (as it is).

The following first order action, with anticommuting fields, gives a $d = -2$ [7.24] contribution to the action (7.11)[*],

$$\frac{1}{\pi} \int\limits_{\Sigma} \mathrm{d}^2x \left(\xi_1 \epsilon^{\alpha\beta} \partial_\alpha \eta_\beta + \sqrt{g^0} \xi_2 g^{0\alpha\beta} \partial_\alpha \eta_\beta \right) \ . \tag{7.13}$$

By bosonization, (7.13) plus the first two terms on the right hand side of (7.11) (with $d = -2$), may be represented by a bosonic system of commuting fields,

$$\frac{1}{\pi} \int\limits_{\Sigma} \mathrm{d}^2x \sqrt{g^0} B^{\alpha\beta} \nabla^0_\alpha \phi_\beta \ . \tag{7.14}$$

Equation (7.11), without the cosmological term, combined with (7.13) may therefore be represented by the same action as derived previously (7.9),

$$S = \frac{1}{\pi} \int\limits_{\Sigma} \mathrm{d}^2x \sqrt{g^0} \left(b^{\alpha\beta} \nabla^0_\alpha c_\beta + B^{\alpha\beta} \nabla^0_\alpha \phi_\beta \right) \ . \tag{7.15}$$

Within this approach the cosmological constant term is treated as a perturbation.

The symmetry that Distler attributes to this action is the one that we have designated s. It is independently invariant under the usual BRST diffeomorphisms. It is possible to establish (when $\Sigma = S^2$) that correlation functions calculated with this action agree with those obtained for the k matrix models in ref. [7.10]. From the topological field theory point of view, the construction of observables, using the basic set of fields (7.3), is not obvious. On the reduced set, that is those appearing in (7.15), there are some, more or less obvious, candidates that arise from the conformal field theory point of view. One such operator is

$$\mathcal{O}_1 = \tfrac{1}{4} (B_{zz}\phi^z + b_{zz}c^z)(B_{\bar{z}\bar{z}}\phi^{\bar{z}} + b_{\bar{z}\bar{z}}c^{\bar{z}}) \ , \tag{7.16}$$

which is to be integrated over the sphere. Before evaluating this we need to deal with the problem of c and ϕ zero modes on the sphere. There are three of each, corresponding to the three conformal Killing vector fields on the sphere. Following the procedures outlined in sections 3.9.1 and 6.2.1

---

[*] By making use of the results of section 6.2 it is easy to see that this action has $d = -2$. The first term is gauge invariant and metric independent, gauge fixing represented by the second term spoils the metric independence. However, the ghosts that one would generate would restore the metric independence. In this case there are two commuting ghosts giving $d = 2$, overall $d = 0$, and the result follows.

we could parametrize these modes and, in fact, set them to zero. We do this directly by setting the value of the ghosts at three preferred points x_i to zero. This is done in a manifestly BRST invariant fashion, whence nothing depends on the particular points chosen. Let $\bar{\sigma}^i_\alpha$ and η^i_α be constant fields transforming as

$$\{Q, \bar{\sigma}^i_\alpha\} = \eta^i_\alpha \ , \qquad \{Q, \eta^i_\alpha\} = 0 \ . \tag{7.17}$$

Now add to the action (7.15) the following term:

$$\sum_{i=1}^{3} \{Q, \bar{\sigma}^i_\alpha c^\alpha(x_i)\} = \sum_{i=1}^{3} [\eta^i_\alpha c^\alpha(x_i) + \bar{\sigma}^i_\alpha(\phi^\alpha(x_i) + c^\beta(x_i)\partial_\beta c^\alpha(x_i)] \ , \tag{7.18}$$

which on integrating over the $\bar{\sigma}$ and η yields the following insertion in the path integral:

$$\prod_{i=1}^{3} c^z(x_i)c^{\bar{z}}(x_i)\delta(\phi^z(x_i))\delta(\phi^{\bar{z}}(x_i)) \equiv \prod_{i=1}^{3} c^z c^{\bar{z}}\delta(\phi^z)\delta(\phi^{\bar{z}})(x_i) \ . \tag{7.19}$$

These insertions will reappear as puncture operators in our discussion of observables in section 7.5. We would like to calculate

$$\left\langle \exp\left(\lambda \int_{S^2} \mathcal{O}_1\right) \prod_{i=1}^{3} c^z c^{\bar{z}}\delta(\phi^z)\delta(\phi^{\bar{z}})(x_i) \right\rangle \ , \tag{7.20}$$

where the expectation value is taken with respect to the action (7.15) and the zero mode insertions (7.19) are included. The quartic term in the action may be turned into a cubic term with the help of multiplier fields. Specifically one replaces $\lambda \int_{S^2} \mathcal{O}_1$ by

$$\int_{S^2} [\sqrt{\lambda}(B_{zz}\phi^z + b_{zz}c^z)\Sigma_{\bar{z}} + \sqrt{\lambda}(B_{\bar{z}\bar{z}}\phi^z + b_{\bar{z}\bar{z}}c^z)\Sigma_z - 4\Sigma_z\Sigma_z] \ . \tag{7.21}$$

With the introduction of this term into the action one recognizes the theory as a Thirring type model (the fields, however, having exotic statistics). Such a model has been analyzed before [7.25] and one may call on this work to complete the evaluation. One finds that (7.20) is equal to $1/(1 - \lambda)$, in agreement with the result of Gross and Migdal [7.10] for the $k = 1$ matrix model on the sphere.

7.5. A gauge theory of topological gravity

The problem of constructing and interpreting observables may be overcome in a Witten type gauge theory formulation of topological gravity. Such a description, for genus $g \geq 2$ as a $(P)SL(2,\mathbb{R})$ super-BF theory, has first been given by Montano and Sonnenschein [7.12]. At this point we remind the reader of section 6.2.8, where we showed that the condition $F_A = 0$, for $A = J\omega + P_a e^a$, is equivalent to the statement that the spin connection ω is torsion free and has constant negative curvature, provided that the zweibein is invertible. If one imposes these equations, one removes the Liouville field (the scale factor of the zweibein) from the dynamics; the theory is then described entirely in terms of the ghost action. With a conventional covariant gauge fixing condition, the latter is a second order action.

In the following, we shall describe topological gravity in a way which differs in three important aspects from the above scenario. Firstly, instead of a covariant gauge fixing of the Yang–Mills

symmetry, we will impose algebraic gauge conditions on the zweibein, somewhat in line with the conditions imposed on the metric in section 7.3. Upon elimination of the Liouville sector one is then left with a first order ghost action which, not too surprisingly, turns out to be precisely the Labastida–Pernici–Witten action discussed in section 7.3. This has been shown by Li [7.26] and will also follow from our considerations below. Secondly, we will not eliminate the Liouville sector, since it has been noticed [7.20] that keeping it considerably simplifies the subsequent analysis of the theory. Another important observation made in ref. [7.20] is that, instead of using a constant curvature constraint to describe the moduli space, one can profitably constrain the curvature to vanish everywhere except at certain isolated points. The resulting gauge fixed theory is then a free, conformally invariant field theory, consisting of a Liouville sector, the first order $(b-c)$ ghost system, and their superpartners. This modification essentially amounts to replacing $PSL(2,\mathbb{R})$ by $ISO(2)$, and we shall therefore, thirdly, choose this to be our gauge group.

In section 6.2.8 we already discussed the contraction of $PSL(2,\mathbb{R})$ to $ISO(2)$, and we briefly summarize the relevant equations here.

Under a gauge transformation, $\Lambda = \lambda^a P_a + \lambda^0 J$, the components of the connection transform as

$$\delta e^a = d\lambda^a + \varepsilon^a{}_b(\lambda^0 e^b - \lambda^b \omega) \ , \qquad \delta\omega = d\lambda^0 \ , \tag{7.22}$$

and the J and P_a components of the curvature are

$$F^0 = d\omega \ , \qquad F^a = de^a - \omega\varepsilon^a{}_b e^b \ . \tag{7.23}$$

The gauge transformations (7.22) are readily seen to be equivalent to Lorentz transformations and diffeomorphisms on-shell ($F^a = 0$), the argument being the same as in sections 6.1.6 and 6.2.3. This also implies that the field equations $F^0 = F^a = 0$ themselves are Lorentz and diffeomorphism invariant. The modified contracted B transformations, which allow us to write down a non-degenerate gauge invariant action, despite the fact that an invariant trace exists only on the $U(1)$ subalgebra of $ISO(2)$, are ($B_0 = -B^0$, $B_a = B^a$, with analogous conventions for the J and P_a components of other fields)

$$\delta B_0 = -\varepsilon_{ab}\lambda^a B^b \ , \qquad \delta B_a = \varepsilon_a{}^b \lambda^0 B_b \ . \tag{7.24}$$

As in any super-BF theory, we introduce ghosts c^a and c_0 for the gauge transformations, their ghost for ghosts ϕ^a and ϕ_0, superpartners ψ^a and ψ_0 of e^a and ω [transforming as the supervariation of (7.22)], as well as "anti-ghosts" χ_a and χ_0 (transforming as their superpartners B_a and B_0).

As they stand, eq. (7.23) and the putative action $S = \int BF_A + \chi d_A \psi$ are obviously unacceptable for a theory of topological gravity on surfaces of genus $g \neq 1$, since $F^0 = 0$ says neither that the curvature is constant nor that it is concentrated at isolated points, but that it vanishes. In ref. [7.20] this problem was overcome by noting that, in order to obtain non-vanishing amplitudes in this theory, it is necessary to insert operators $e^{-q_i B_0(x_i)}$, which screen the background charge produced by the curvature of the Riemann surface. Integrating over B_0 one sees that these insertions generate δ-function singularities in the curvature,

$$d\omega(z) = \sum_i q_i \delta(z - x_i) \tag{7.25}$$

(here the δ's are two-forms, cf. section 6.2.2). This leads to the correct equation

$$\frac{1}{2\pi}\int_{\Sigma_g} \sqrt{g}\, R(g) = 2 - 2g = \chi(\Sigma_g) \tag{7.26}$$

(and hence to non-vanishing amplitudes) provided that $(1/\pi)\sum_i q_i = 2g - 2$. Instead of using the screening operators, we can equivalently modify the flatness conditions (on which we build the super-BF action) directly to read

$$de^a - \omega \varepsilon^a{}_b e^b = 0 \ , \qquad d\omega(z) = \sum_i q_i \delta(z - x_i) \ . \tag{7.27}$$

The price one then has to pay in either approach for the introduction of the delta function singularities is that these equations are no longer gauge invariant at the point singularities themselves. The equations are invariant under the $U(1)$, local Lorentz, transformations but break the inhomogeneous part of the $ISO(2)$ algebra, which corresponds to diffeomorphisms. The gauge invariance of the flatness conditions must be given up for us to move away from the torus. However, this turns out to be a reasonable way to proceed. The shift symmetry allows one to have "overall" gauge invariance of the action.

Nevertheless, we should point out that if we had tried to write this theory as a modified (Schwarz type) BF model, then we would only have had gauge invariance for those gauge transformations which vanish at the point singularities. As these gauge transformations correspond on-shell to diffeomorphisms, then we would be considering diffeomorphisms that leave those points fixed. This leads us to the notion of a punctured Riemann surface. A punctured Riemann surface is a Riemann surface together with some preferred points (these points are not deleted but rather "marked"). The moduli space $\mathcal{M}_{(g,s)}$ of a compact Riemann surface of genus g and s punctures arises on considering the space of metrics on the genus g Riemann surface and factoring out by the diffeomorphisms that leave the s "punctures" fixed. The gauge theory we have constructed is thus seen to describe precisely $\mathcal{M}_{(g,s)}$. The punctured Riemann surface plays an important role also in the Witten type super-BF theory, to whose construction we return.

We see that, even within this restricted framework of an $ISO(2)$ gauge theory, we have several options available. We can use either the original or the modified flatness conditions, and we can declare our symmetries to be either gauge symmetries or Lorentz transformations and diffeomorphisms. At least for our purposes, all these approaches are equivalent. We will use the modified flatness condition and the diffeomorphism ghosts for the above symmetries, so that $\lambda^a \to c^a = e^a_\mu c^\mu \equiv e^a.c$, while the local Lorentz transformation becomes, $\lambda \to c_0$. The transformation rules we adopt for the spin connection are therefore the usual $ISO(2)$ transformation rules augmented with the shift

$$\{Q, \omega\} = dc_0 + \psi_0 \ , \qquad \{Q, \psi_0\} = d\phi_0 \ , \qquad \{Q, c_0\} = \phi_0 \ , \qquad \{Q, \phi_0\} = 0 \ . \tag{7.28}$$

On the other hand, the transformation rules that are adopted in the zweibein sector are

$$
\begin{aligned}
\{Q, e^a\} &= \psi^a - d(e^a.c) + \epsilon^a{}_b \omega e^b.c + \epsilon^a{}_b e^b c_0 \ , \\
\{Q, \psi^a\} &= -\epsilon^a{}_b \psi^b c_0 + \epsilon^a{}_b e^b \phi_0 + d(e^a.\phi) - \epsilon^a{}_b \omega e^b.\phi - d(\psi^a.c) + \epsilon^a{}_b \psi_0 e^b.c + \epsilon^a{}_b \omega \psi^b.c \ , \\
\{Q, c^\alpha\} &= \phi^\alpha + c^\beta \partial_\beta c^\alpha \ , \qquad \{Q, \phi^\alpha\} = c^\beta \partial_\beta \phi^\alpha - \phi^\beta \partial_\beta c^\alpha \ ,
\end{aligned}
\tag{7.29}
$$

and while these have the form of the $ISO(2)$ transformations, the BRST operator is nilpotent only when the no-torsion equation is used. The reason for the failure of the algebra to close off-shell lies in the fact that, even though the transformations can be made nilpotent off-shell in the gauge theory setting (after all, this is then just a two dimensional version of Donaldson theory), the equivalence with diffeomorphisms holds only *on shell*. We should therefore expect that the gauge algebra will close only on shell, when the diffeomorphism ghosts are adopted and we insist that the diffeomorphism ghost transformations agree with those given in eq. (7.3).

At this point it is convenient to switch to complex and (internal) light cone notation,

$$e^{\pm} = (1/\sqrt{2})(e^1 \pm ie^2) = e_z^{\pm} dz + e_{\bar z}^{\pm} d\bar z \;,$$

so that (with the substitution $\omega \to i\omega$) the no-torsion conditions and their super(shift)partners read

$$De^{\pm} \equiv de^{\pm} \mp \omega e^{\pm} = 0 \;, \qquad D\psi^{\pm} \equiv d\psi^{\pm} \pm \psi_0 e^{\pm} \mp \omega\psi^{\pm} = 0 \;.$$

As the algebraic gauge conditions on the zweibein and its superpartner we choose

$$e_{\bar z}^+ = e_z^- = \psi_{\bar z}^+ = \psi_z^- = 0 \tag{7.30}$$

to fix the λ^a symmetry, and

$$e_z^+ = e_{\bar z}^- \;, \qquad \psi_z^+ = \psi_{\bar z}^- \tag{7.31}$$

to fix the λ^0 symmetry. The remaining component $e_z^+ > 0$ of the zweibein may then be written as $e_z^+ = \mathrm{e}^{\sigma}$, so that $g_{z\bar z} = \mathrm{e}^{2\sigma}$ and $\psi_z^+ = \mathrm{e}^{\sigma}\psi$, where ψ is the superpartner of the Liouville field σ. This exhausts all the gauge symmetries, so that the spin connection and its superpartner are unconstrained. In order to impose these gauge conditions we introduce the anti-ghosts $\bar c_z$, $\bar c_{\bar z}$ and $\bar c_0$ for the zweibeins, and $\bar\phi_z$, $\bar\phi_{\bar z}$ and $\bar\phi_0$ for the ψ's, as well as their multipliers (the notation used here is precisely that of Donaldson theory and its offsprings, as in sections 5.2–5.4).

The action is

$$
\begin{aligned}
S = \int_{\Sigma_g} \mathrm{d}^2 z \,\{ & Q, \chi \,(d\omega - \textstyle\sum_i q_i \delta(z - x_i)) + \chi_+ De^+ + \chi_- De^- \\
&+ \bar c_z e_{\bar z}^+ + \bar c_{\bar z} e_z^- + \bar c_0 (e_z^+ - e_{\bar z}^-) + \bar\phi_z \psi_{\bar z}^+ + \bar\phi_{\bar z}\psi_z^- + \bar\phi_0(\psi_z^+ - \psi_{\bar z}^-)\} \;.
\end{aligned}
\tag{7.32}
$$

Rather than expanding this out immediately, we can simplify matters considerably by imposing the algebraic constraints directly, which amounts to integrating out the corresponding multipliers. The last simplification that we make at this point is to impose the no-torsion constraint and its Q variation

$$D\psi^{\pm} = \sum_i q_i * e^{\pm}.c\delta^2(z - x_i) \;.$$

These combined with the other constraints allow us to solve for the spin connection and its superpartner,

$$\omega = -*d\sigma \;, \qquad \psi_0 = *d\psi + *c\sum_i q_i \delta^2(z - x_i) \;. \tag{7.33}$$

Then the variation of the first equation of (7.31) yields

$$c_0 = (\partial_z c^z + c^z \partial_z \sigma - \bar\partial_{\bar z} c^{\bar z} - c^{\bar z}\bar\partial_{\bar z}\sigma) \;, \tag{7.34}$$

which arises on integrating out $\bar c_0$, and may be used to eliminate c_0. Likewise, the variation of the second equation of (7.31) gives an algebraic condition on ϕ_0 (and arises on integrating out $\bar\phi_0$), namely

$$\phi_0 = (\partial_z \phi^z + \phi^z \partial_z \sigma + c^z \partial_z \psi - \bar\partial_{\bar z}\phi^{\bar z} - \phi^{\bar z}\bar\partial_{\bar z}\sigma - \bar c^{\bar z}\bar\partial_{\bar z}\psi) + \sum_i q_i c_z c_{\bar z}\delta(z - x_i) \;. \tag{7.35}$$

We have been a bit sloppy in this discussion as the delta function constraints that arise are actually of the form $e^{\sigma} c_0 = \cdots$ and $e^{\sigma} \phi_0 = \cdots$. However, the determinants that one gets in factoring out the Liouville field cancel between the two delta functions.

Taking all of these identifications into account allows us to express (7.32) in the particularly simple form

$$
S = \int_{\Sigma} d^2 z \sqrt{g^0} \left[B \left(\Delta\sigma - \sum_i q_i \delta(z - x_i) \right) + \chi \Delta\psi \right.
$$
$$
\left. - \sum_i q_i \delta^2(z - x_i)(c^z \partial_z + c^{\bar{z}} \bar{\partial}_{\bar{z}})\chi + b^{\alpha\beta} \nabla^0_\alpha c_\beta + B^{\alpha\beta} \nabla^0_\alpha \phi_\beta \right] \ . \tag{7.36}
$$

The anti fields $\bar{c}_\mu$ and $\bar{\phi}_\mu$ have been replaced by the symmetric traceless tensors $b^{\mu\nu}$ and $B^{\mu\nu}$; for example, $b_{zz} = \bar{c}_z e_z^+$. We see that, upon elimination of the Liouville sector and its superpartner, this action agrees with the Labastida–Pernici–Witten action (7.15), confirming the observation by Li [7.26] mentioned above. Actually, one could have started with the formalism of section 7.3, dropped the conformal terms in the transformation rules (7.3), i.e., set $\rho = \tau = 0$, and imposed instead the conditions $g_{zz} = g_{\bar{z}\bar{z}} = 0$, $\psi_{zz} = \psi_{\bar{z}\bar{z}} = 0$ and $\sqrt{g} R(g) = \sum_i q_i \delta^2(z - x_i)$ (and its shift variation) to arrive precisely at (7.36)[*]. We seem to have gained nothing by considering the model as a gauge theory. The differences between the two formalisms lie in the fact that from the gauge theory point of view we naturally have extra fields, namely the spin connection and its associated ghosts. It is with these fields that we may create observables.

7.6. Gauge theory observables

The invariant polynomials of section 5.2.7 are the obvious candidates for observables. There are, however, some problems that we need to overcome. Firstly, the invariant polynomials are defined with the help of an invariant trace. We have seen that no such trace exists for ISO(2) in general, but have also observed that there is an invariant U(1) trace. This means that as long as we only consider the fields in the geometric sector of the theory, the invariants that one can form come only from the U(1) part of the algebra. The second difficulty is that, as we explained in section 7.3, we wish to work equivariantly, so that the "descent equations" are produced by acting with $d + s$ rather than $d + Q$, with the action of the shift symmetry on the U(1) geometric fields being [cf. (7.5)],

$$
s\omega = \psi_0 \ , \qquad s\psi_0 = 0 \ , \qquad sc_0 = \phi_0 \ , \qquad s\phi_0 = 0 \ . \tag{7.37}
$$

It is quite straightforward to determine the form of the relevant polynomials from the known ones of section 5.2.7. In the U(1) sector, the invariant polynomials associated with Q are $(d\omega + \psi_0 + \phi_0)^n$. If we substitute $\psi_0 \rightarrow \psi'_0 = \psi_0 + dc_0$ in (7.37) then s becomes Q on this set of fields. Then substituting ψ'_0 for ψ_0 in the polynomials makes them s invariant. In equations this means that we are interested in the operators

$$
(d\omega + \psi_0 + dc_0 + \phi_0)^n = \sigma_n^2 + \sigma_n^1 + \sigma_n^0 \ .
$$

---

[*] Incidentally this shows that it is indeed safe to use the group ISO(2) in the gauge theory setting to describe the moduli of higher genus Riemann surfaces.

In this expression the terms are ordered according to their form degree, which is indicated by the value of the superscript. Then the equivariant Bianchi identity

$$(d + s)(d\omega + \psi_0 + dc_0 + \phi_0)^n = 0$$

becomes the descent equation

$$0 = s\sigma_n^0 , \qquad d\sigma_n^0 = s\sigma_n^1 , \qquad d\sigma_n^1 = s\sigma_n^2 , \qquad d\sigma_n^2 = 0 . \tag{7.38}$$

Recall the discussion on the triviality of observables in section 5.3.3. The argument we have given here is essentially a running backwards of observation (ii) [7.27] made there.

The set of operators σ_n^i form the basic observables in topological gravity. The fields out of which they are constructed, however, were eliminated from the functional integral, by their algebraic equations of motion, and so do not appear in the reduced action (7.36). The basic set of observables in the reduced theory are then made up of the above polynomials, but with the fields ω, ψ_0, c_0 and ϕ_0 eliminated in favor of their equations of motion.

In sections 5.2 and 5.3, we identified the building blocks of the observables of Donaldson theory as the characteristic classes of the Atiyah–Singer universal bundle $\mathcal{Q}$ by showing that the zero mode sector of Donaldson theory describes precisely the geometry of $\mathcal{Q}$. This established the equivalence between the field theoretic and topological definition of the correlation functions of Donaldson theory. In ref. [7.17] Witten has given a purely topological definition of correlation functions in 2D gravity, in terms of certain line bundles L_i, $i = 1,\ldots,s$, on the moduli spaces $\mathcal{M}_{(g,s)}$ of punctured Riemann surfaces, and in the following we shall sketch how, similarly, $\sigma_1^0(x_i) = \phi_0(x_i)$ can be identified as a representative of $c_1(L_i)$.

Thus let $\mathcal{M}_{(g,s)}$ be the moduli space of Riemann surfaces of genus g with s punctures. The dimension of $\mathcal{M}_{(g,s)}$ is $6g - 6 + 2s$, i.e., $6g - 6$ for the Riemann surface and 2 for the location of the puncture x_i. The fibre $K_{x_i} = T_{x_i}^{*(1,0)}\Sigma_g$ is a complex one-dimensional vector space, and as one moves in $\mathcal{M}_{(g,s)}$ the K_{x_i} vary holomorphically to form a holomorphic line bundle L_i over $\mathcal{M}_{(g,s)}$. Such a line bundle has a first Chern class $c_1(L_i)$, which can be represented rationally by a two-form α_i on $\mathcal{M}_{(g,s)}$ and the topological definition of the amplitudes is

$$\langle \sigma_{n_1}(x_1)\cdots\sigma_{n_k}(x_k)\rangle = \int_{\mathcal{M}_{(g,s)}} \alpha_1^{n_1}\cdots\alpha_s^{n_s} , \tag{7.39}$$

where $\sum_i 2n_i = 6g - 6 + 2s$, i.e., $\sum_{i=1}^s (n_i - 1) = 3g - 3$.

For any holomorphic line bundle L with a smooth Hermitian norm $|\cdot|$, $c_1(L)$ can be represented by the $(1,1)$-form $\alpha = \bar{\partial}\partial \log|s|^2$, where s is a locally trivializing section of L. If $L = K$, the Chern (monopole) number of L is

$$\frac{\mathrm{i}}{2\pi} \int_{\Sigma_g} \alpha = \deg(K) = 2g - 2 . \tag{7.40}$$

In particular, choosing s to be a meromorphic section of K with only simple zeros or poles at points z_i, one recovers the definition of the degree of K as the degree $\deg D(s) = \sum_i p_i$ of the *divisor* $D(s) = \sum_i p_i z_i$ of s [$p_i = +1(-1)$ if z_i is a pole (zero) of s], since

$$\partial\bar{\partial} \log|s|^2(z) = 2\pi\mathrm{i}\sum_k p_k \delta(z - z_k) \tag{7.41}$$

in that case.

Comparing (7.41) with the equation of motion (or vacuum expectation value)

$$\Delta\sigma(z) = \sum_k q_k \delta(z - x_k) \tag{7.42}$$

following from the action (7.36), and recalling the discussion following (7.25) we see that we can identify σ (up to an irrelevant scaling—this will normalize the amplitudes by a factor depending only on the dimension of $\mathcal{M}_{(g,s)}$) with $\log|s|^2$ for some section s of K. Almost tautologically, any section s of K gives rise to a section of L_i which, by abuse of notation, we shall denote by $s(x_i)$, and a possible choice of α_i is then $\alpha_i = \partial_{\mathcal{M}}\partial_{\mathcal{M}}\log|s(x_i)|^2$; the problem of identifying ϕ_0 as a representative of $c_1(L_i)$ is then reduced to the more concrete task of showing that $\phi_0 = \partial_{\mathcal{M}}\bar{\partial}_{\mathcal{M}}\sigma$.

We now recall that on $\mathcal{M}_{(g,s)}$ both the shift operator s and the BRST operator Q reduce to the exterior derivative $d_{\mathcal{M}} = \partial_{\mathcal{M}} + \bar{\partial}_{\mathcal{M}}$. Since $s\sigma = \psi$ and, on shell, $\Delta\psi = 0$ [*] so that we can locally write $\psi(z, \bar{z}) = \psi(z) + \bar{\psi}(\bar{z})$, we identify $\partial_{\mathcal{M}}\sigma = \psi(z)$ and $\bar{\partial}_{\mathcal{M}}\sigma = \bar{\psi}(\bar{z})$. The last step, showing that $Q(\psi - \bar{\psi}) = \phi_0$, follows from the transformation laws (7.29) for ψ and the constraint (7.35) for ϕ_0. This finally completes the identification of $\phi_0(x_i)$ as a representative of $c_1(L_i)$, and whence the equivalence of the topological and field theoretic definitions of correlation functions in topological gravity—modulo one subtlety, which we will now address.

The observables we have discussed are defined on punctured Riemann surfaces. Recall that the act of "puncturing" is to quotient only by those diffeomorphisms that leave the marked points fixed. This would seem to imply that we should reconsider the construction of the BRST operator and organize for it to leave the chosen points fixed. Rather than tampering with Q, we can insert an operator at some preferred points whose only effect is that there is no action of the diffeomorphisms there. An operator that does just this was introduced in section 7.4 [see the discussion just before equation (7.17), and the equations following that one]. This *puncture operator* is

$$\mathcal{P}(x_i) = c_z c_{\bar{z}} \delta(\phi^z) \delta(\phi^{\bar{z}})(x_i) \ . \tag{7.43}$$

It does not spoil any of the symmetries, and the new composite observables

$$\sigma_n^0 \mathcal{P}(x_i) = (\partial_z \phi^z - \bar{\partial}_{\bar{z}} \phi^{\bar{z}})^n \mathcal{P}(x_i)$$

are BRST invariant. In the course of the above discussion we have already chosen the singular points x_i of the curvature [cf. (7.25)] to coincide with those points at which the observables are placed. As there are now also puncture operators at the singularities, the fact that the diffeomorphism and gauge symmetries were broken originally at those points causes no technical difficulties, as the diffeomorphisms are now restricted not to act at these points.

The evaluation of the observables requires special care on two points. The first, which is by now familiar, is a correct handling of the zero modes in the theory. The situation here is somewhat more involved due to the fact that the puncture operators introduce extra zero modes in the anti-field sector. This comes about as each puncture operator fixes one of the ghost modes to zero so that that mode does not appear in the action; the corresponding anti-ghost mode is then not matched and so also does not appear in the action. However, the anti-ghost mode still needs to be integrated over and it is not weighted. The second feature that requires special care is the treatment of products of observables. The observables now involve delta functions and when operators come into "contact"

---

[*] This equation obviously holds away from the delta function singularities. It also holds at the singular points when the ghost fields c_z and $c_{\bar{z}}$ are put to zero there, as they will be shortly.

a correct analysis of the resulting singularities must be made. The proper treatment of these two points is spelled out in detail in ref. [7.20].

Correlation functions may be determined via a sequence of Ward and Schwinger–Dyson equations; these relate n-point functions to $(n - 1)$-point functions. The two identities, called the "dilaton" and "puncture" equations, are

$$\left\langle \sigma_1 \prod_{i=1}^{s} \sigma_{n_i} \right\rangle_g = (2g - 2 + s) \left\langle \prod_{i=1}^{s} \sigma_{n_i} \right\rangle_g ,$$ (7.44)

and

$$\left\langle \mathcal{P} \prod_{i=1}^{s} \sigma_{n_i} \right\rangle_g = \sum_{i=1}^{s} \left\langle \sigma_{n_j-1} \prod_{i \neq j}^{s} \sigma_{n_i} \right\rangle_g ,$$ (7.45)

respectively. The subscript g indicates the genus of the surface Σ and s is the number of punctures. These and more general relations (relating amplitudes in different genera) are derived in refs. [7.20, 7.23] and are generalizations and extensions of the equations derived by Witten [7.17] and Dijkgraaf and Witten [7.19]. These equations have been employed by Horne [7.28] to determine the genus 3 and 4 intersection numbers of the stably compactified moduli space of Riemann surfaces. This is an impressive achievement as these numbers are very difficult to compute from the algebra-geometric point of view. In particular, the genus 4 intersection numbers had not been known before.

The actual derivation of these identities requires consideration of the contact terms alluded to above. For example, in the dilaton equation one would expect that, as σ_1 is essentially $\int_\Sigma d\omega$, the factor on the right hand side would be $(2g - 2)$. The extra s contribution comes from the contact terms. Geometrically this is indeed the correct answer as $(2g - 2 + s)$ is the Euler character of the punctured Riemann surface and $d\omega$ should be thought of as a two-form there. A complete exposition may be found in the literature [7.17, 7.19, 7.20, 7.23].

Further reading

The study of dynamically triangulated random surface models was initiated in refs. [7.29–7.31], the relation to matrix models following from the classical work [7.32] of Brézin, Itzykson, Parisi and Zuber. The link with 2D gravity was first suggested in refs. [7.33, 7.34], and subsequent analytical and numerical studies were performed, e.g., in refs. [7.35–7.41] (for a review see ref. [7.42]). The discovery of the so-called double scaling limit in random matrix models [7.8–7.10, 7.43], a careful continuum limit for surfaces of all genera at the same time, has led to a vast number of papers on 2D quantum gravity, non-critical strings and matrix models during the past year. Instead of giving a, necessarily incomplete, list of references on the subject, we refer the reader to the review talks [7.44–7.48] and the references therein.

Topological gravity as a subject began with the work of Witten [7.49], where a description of self-dual Weyl gravitational instantons in four dimensions was given. There were problems, however, with regards to the conformal symmetry in this description. Labastida and Pernici [7.50], using a Langevin approach, were able to get around this difficulty. In more than two dimensions, however, there is considerable freedom (and hence ambiguity) in the choice of moduli space. Moreover, the construction of observables had been, and still is somewhat, problematic. While this is not a difficulty, as we have seen, in two dimensions, the general construction of observables in other dimensions remains an open problem. Some progress in this direction has been made by Myers

and Periwal [7.51–7.53]. For other work on topological gravity in four dimensions see refs. [7.54, 7.55] and the interesting suggestions in ref. [7.56].

The topological investigation of the multi-matrix models [7.19] suggested its equivalence with topological gravity coupled to some topological matter theory. Li [7.57] proposed that the appropriate topological theories are the twisted $N = 2$ superconformal models of Eguchi and Yang [7.58] and provided substantial circumstantial evidence in favor of this proposal. The correctness of this suggestion has been confirmed by Dijkgraaf, Verlinde and Verlinde [7.59] by comparison with the results of Douglas [7.60] on the correlation functions of the multi-matrix models. The contact algebra and recursion relations in these and other models have been investigated in refs. [7.61–7.63]. A rather thorough account of the possible descriptions of the moduli space of Riemann surfaces and the relevant symmetry groups may be found in ref. [7.64]. These authors give yet another topological field theory description of the moduli space.

SL$(n,\mathbb{R})$ super-BF theories in two dimensions, the original candidates for the "missing matter" in multi-matrix models [7.65], have been shown by Li to be interesting in their own right, describing what may rightfully be called topological W_n gravity [7.26, 7.66], the topological counterpart of W_n gravity [7.67–7.70] (for an update on W geometry and W gravity see ref. [7.71]).

8. Renormalization

8.1. Introduction

We would now like to address the issue of renormalization in topological field theories. The most important question which arises is whether the topological nature of these models is preserved by the renormalization procedure. This is crucial from the mathematical point of view; however, in addition questions of how and why symmetries may be broken in topological field theories are of physical relevance if these models are to correspond to unbroken phases of physical systems.

Now, since a topological field theory is in essence a finite dimensional quantum mechanical system, one may wonder as to the relevance of a renormalization discussion. The main point to be stressed here, however, is that the quantum mechanical system of interest is simply encoded in a true local quantum field theory. From the field theory point of view divergences can certainly occur [8.1–8.3]; it is only when the theory is restricted to the appropriate moduli space that the finiteness of the model is manifest [8.4–8.6]. However, as we have seen for the case of Witten type theories, this restriction to a finite dimensional moduli space is simply a gauge choice, viz. the delta function gauge. Hence, these theories are indeed finite. In the case of Schwarz type theories (e.g. Chern–Simons theory [8.7, 8.8]), one can establish the finiteness simply from the fact that the space of solutions to the field equations, modulo the gauge symmetries, is finite dimensional.

As our first example, we shall examine Donaldson theory [8.9], in several different gauges. The gauges that we are referring to here correspond to the gauge fixing of the topological shift symmetry, as described in section 5.2. Choosing a Feynman type gauge, one finds both a one-loop divergence and a non-zero β-function [8.1, 8.2]. However, in a Landau type gauge both of these are absent [8.4, 8.5]. One thus sees that a non-zero β-function is really a gauge artifact, and that one's intuition regarding the finiteness of a topological field theory is indeed borne out. It is perhaps worthwhile recalling that the nomenclature "gauge dependence" is meant to refer to gauge choice dependence, in other words, the choice of the gauge fixing parameter. Following this, we study the same issue in topological sigma models [8.10], and obtain similar results [8.3, 8.6].

The renormalization issue in the Schwarz type theories is of a more subtle nature, and as an

example we study the pure Chern–Simons theory [8.2, 8.11–8.26]. Here the primary issue is the presence of a phase in a one-loop determinant, as we saw in section 6.1.3. We evaluate this phase using a momentum expansion [8.18, 8.19], and discuss its physical relevance in section 8.4.5. There is the related issue of a peculiar supersymmetry which is present when the theory is quantized in the Landau gauge [8.2, 8.27–8.29]. The Ward identities which ensue from this supersymmetry, as well as a potential anomaly (connected with the above phase), are studied.

8.2. Donaldson theory at one loop

Let us begin with the complete quantum action in the form

$$S_q = \int d^4x \, \mathrm{tr}\{Q, \chi^{\alpha\beta}(F^+_{\alpha\beta} - \tfrac{1}{2}\alpha B_{\alpha\beta}) + \bar{\phi} D_\alpha \psi^\alpha + \bar{c}[D_\alpha(A_c)(A^\alpha - A_c^\alpha) - \tfrac{1}{2}\alpha' b]\} \ . \tag{8.1}$$

Here, α and α' are the gauge fixing parameters, and we should note the distinction between the full covariant derivative $D_\alpha = \partial_\alpha + [A_\alpha, \ ]$, and that defined with respect to the background gauge field A_c, $D_\alpha(A_c) = \partial_\alpha + [A_{c\alpha}, \ ]$. This is because, for the purposes of a one-loop calculation, we will be decomposing the Yang–Mills field A into a background plus a quantum part,

$$A = A_c + A_q \ . \tag{8.2}$$

First, we expand the action (8.1)

$$S_q = \int d^4x \, \mathrm{tr}\{B^{\alpha\beta}(F^+_{\alpha\beta} - \tfrac{1}{2}\alpha B_{\alpha\beta}) - 2\chi^{\alpha\beta} D_\alpha \psi_\beta - \tfrac{1}{2}\alpha\phi\{\chi^{\alpha\beta}, \chi_{\alpha\beta}\} + \eta D \cdot \psi$$
$$- \bar{\phi}(D^2\phi - \{\psi^\alpha, \psi_\alpha\} + \{c, D \cdot \psi\})$$
$$+ b[D_\alpha(A_c)(A^\alpha - A_c^\alpha) - \tfrac{1}{2}\alpha' b] - \bar{c}D_\alpha(A_c)(D^\alpha c + \psi^\alpha)\} \ . \tag{8.3}$$

Our first calculation involves choosing the Feynman gauge $\alpha = \alpha' = 1$ [8.2]. Upon integrating out the multiplier fields $B_{\alpha\beta}$ and b, we find that the action which is second order in the quantum fields is given by

$$S_q^{(2)} = \int d^4x \, \mathrm{tr}\{-\tfrac{1}{2}A_\alpha^q[D^2(A_c)\delta^{\alpha\beta} + 2[F^{\alpha\beta}(A_c), \ ]]A_\beta^q - \bar{c}D^2(A_c)c$$
$$- 2\chi^{\alpha\beta}D_\alpha(A_c)\psi_\beta - \bar{\phi}D^2(A_c)\phi - (\bar{c} - \eta)D_\alpha(A_c)\psi^\alpha\} \ , \tag{8.4}$$

where only the Yang–Mills field is given a background component, all other fields being purely quantum. Upon making the simple field redefinition

$$2\eta' = \bar{c} - \eta \ , \tag{8.5}$$

(8.4) reduces to

$$S_q^{(2)} = \int d^4x \, \mathrm{tr}\{-\tfrac{1}{2}A^\alpha(D^2\delta_{\alpha\beta} + 2F_{\alpha\beta})A^\beta - \bar{c}D^2c - 2\chi^{\alpha\beta}D_\alpha\psi_\beta - 2\eta'D_\alpha\psi^\alpha - \bar{\phi}D^2\phi\} \ . \tag{8.6}$$

For simplicity of notation, we have now omitted the quantum label q on A, as well as the indication that all covariant derivatives are with respect to the background A_c.

The one-loop correction to the effective action can now be represented as a ratio of determinants. First we note that the determinants arising from the $\bar{c}c$ and $\bar{\phi}\phi$ systems cancel against each other, while the AA system yields $\det^{-1/2}(-D^2\delta_{\alpha\beta} - 2F_{\alpha\beta})$.

The χ-η-ψ system requires a little more care. We notice that this system defines a linear map

$$T : \Omega^1 \to \Omega^2_+ \oplus \Omega^0 \ . \tag{8.7}$$

Here Ω^1, Ω^2_+ and Ω^0 are the spaces of one-forms, self-dual two-forms and zero-forms, and are represented by ψ, χ and η, respectively. The difficulty here is that the operator T is not a map from a space into itself, and so the definition of its determinant requires a little care. However, following Schwarz [8.30], we consider the adjoint operator

$$T^* : \Omega^2_+ \oplus \Omega^0 \to \Omega^1 \ , \tag{8.8}$$

and then form the product

$$T^*T : \Omega^1 \to \Omega^1 \ . \tag{8.9}$$

In this case, the determinant of T^*T can be defined, and one takes

$$\det T \equiv \det^{1/2}(T^*T) \ . \tag{8.10}$$

One finds

$$(T^*T)_{\alpha\beta} = (-D^2\delta_{\alpha\beta} - 2F^-_{\alpha\beta}) \ . \tag{8.11}$$

Alternatively, this can be verified, for example, by writing T as

$$\begin{pmatrix} \eta' & \chi^{0i} \end{pmatrix} \begin{pmatrix} D_0 & D_j \\ -D_i & D_0\delta_{ij} - \epsilon_{ijk}D^k \end{pmatrix} \begin{pmatrix} \psi^0 \\ \psi^j \end{pmatrix} \ , \tag{8.12}$$

where we have used the self-duality of $\chi_{\alpha\beta}$ to eliminate the χ_{ij} components.

The effective action to one-loop order is now given by

$$\Gamma(A_c) = S(A_c) - \tfrac{1}{2}\log\left(\frac{\det(-D^2\delta_{\alpha\beta} - 2F^-_{\alpha\beta})}{\det(-D^2\delta_{\alpha\beta} - 2F_{\alpha\beta})}\right) \ . \tag{8.13}$$

We should note that we have been assuming that the map T has no zero modes, which is true if we restrict our attention to the case of isolated instantons (i.e. zero-dimensional moduli space).

We can regularize these determinants by using, for example, the proper time representation. The one-loop contribution to Γ can be written as

$$\tfrac{1}{2}\text{Tr} \int\limits_0^\infty \frac{dt}{t^{1+\epsilon}}\{\exp[+t(D^2\delta_{\alpha\beta} + 2F^-_{\alpha\beta})] - \exp[+t(D^2\delta_{\alpha\beta} + 2F_{\alpha\beta})]\} \ , \tag{8.14}$$

where ϵ is a regularization parameter, the limit $\epsilon \to 0$ being taken. In evaluating (8.14) we find a divergent term proportional to $1/\epsilon$, and we can compute it using the Schwinger–DeWitt expansion for the heat kernel. The coefficients in this expansion have been given by Gilkey [8.31] for general operators of the form $-D^2 + X$. The relevant coefficient here is the a_2 coefficient, which is given by

$$a_2 = \tfrac{1}{12}\text{tr}\,F^2 + \tfrac{1}{2}\text{tr}\,X^2 - \tfrac{1}{6}\text{tr}\,D^2X \ . \tag{8.15}$$

Combining the a_2 coefficients from the two operators in (8.14), we find the divergent term has the structure [8.2]

$$\epsilon^{-1}\text{tr}(F^+)^2 \ . \tag{8.16}$$

The coefficient of this divergent contribution is the β-function, and the reader can check that it has the usual $N = 2$ super-Yang–Mills value [8.1, 8.32]. The important point to note here is that it is $(F^+)^2$ which is renormalized, rather than F^2, as in the case of conventional super-Yang–Mills theory. This is essential in preserving the topological nature of the model, as it is the former quantity which appears in the tree level action. Thus the one-loop effective action remains a BRST commutator, guaranteeing metric independence of the partition function.

In ref. [8.1], the β-function was computed using dimensional regularization. Although the same numerical coefficient was obtained, it was found that only F^2 was renormalized. This discrepancy can be traced to the fact that dimensional regularization is a momentum space procedure, and thus local surface terms are discarded [F^2 and $(F^+)^2$ differ by such a term, namely $F\tilde{F}$].

A natural question to ask at this point is: since a topological field theory is a theory with no local excitations, i.e., its phase space is finite dimensional, why is there a divergence at one loop? The answer to this question is in fact quite simple and is as follows: we first note that the infinity mentioned above is an off-shell divergence; on-shell, however, i.e., when $F^+ = 0$, we find that the theory is finite. In fact there is no one-loop correction to the effective action, since the ratio of determinants in (8.13) cancel on-shell (when $F = F^-$). This is exactly as we would have expected, since on-shell in this case is defined to mean a restriction of the theory to the instanton moduli space, which is certainly finite dimensional.

A nice way to see this result immediately, without generating the off-shell divergence, is to choose the δ-function gauge. This is defined by choosing $\alpha = \alpha' = 0$ in (8.3). Since the action now contains a term of the form $B^{\alpha\beta}F^+_{\alpha\beta}$, we can integrate over the multiplier B to enforce a delta function constraint in the path integral. This ensures that only instanton configurations are counted, and hence no divergence will appear.

More explicitly, let us examine the theory to one-loop order within the background field method, in the Landau gauge $\alpha = \alpha' = 0$ [8.4, 8.5]. Proceeding from (8.3), rescaling b and using (8.5) we find the quadratic action

$$S_q^{(2)} = \int \mathrm{d}^4x \; \mathrm{tr}\{2B^{\alpha\beta}D_\alpha(A_\mathrm{c})A^\mathrm{q}_\beta + 2bD_\alpha(A_\mathrm{c})A^\alpha_\mathrm{q} - \bar{c}D^2(A_\mathrm{c})c$$
$$-2\chi^{\alpha\beta}D_\alpha(A_\mathrm{c})\psi_\beta - 2\eta'D_\alpha(A_\mathrm{c})\psi^\alpha - \bar{\phi}D^2(A_\mathrm{c})\phi\} \; . \tag{8.17}$$

As before, the $\bar{c}c$ and $\bar{\phi}\phi$ determinants cancel. However, we now note that in addition to the Grassmann odd χ–η–ψ system, we have an even B–b–A system. Both of these define the linear map T, given in (8.7). Thus, irrespective of how we choose to define the determinant, we see that the corresponding ratio of determinants is equal to 1. We thus obtain the result that the entire one-loop correction to the effective action vanishes, the theory is finite and the β-function is zero.

It is important to study for a moment the dependence of these results on the gauge chosen. We have just found that in the Landau gauge ($\alpha = \alpha' = 0$), there is no one-loop renormalization. This agrees with our arguments concerning the fact that the delta function gauge restricts us immediately to the appropriate finite dimensional moduli space, thereby ensuring the absence of divergences. However, if we choose the Feynman gauge ($\alpha = \alpha' = 1$), then we obtain the result given in (8.16). In this case there is a one-loop divergence, of the form $\epsilon^{-1}(F^+)^2$. If we now go back to (8.3), and integrate out the B field, we obtain a term in the action of the form $\alpha^{-1}(F^+)^2$. Choosing $\alpha = 1$ means that the kinetic term for the A field is the usual Yang–Mills action, F^2, plus the instanton number $F\tilde{F}$ (which does not affect the equations of motion). Thus when we use the background field method we will be expanding about a solution of the Yang–Mills equations. It is therefore not surprising that in this case we do generate a divergence, as it is only when the background is

further restricted to be an instanton that this divergence cancels.

We can interpret the above results as affecting the renormalization of the gauge parameter α, as follows. In the Landau gauge there is no renormalization of α. However, in an arbitrary $\alpha \neq 0$ gauge, we have seen that the $\alpha^{-1}(F^+)^2$ term gets renormalized by $\epsilon^{-1}(F^+)^2$; this corresponds to a renormalization of α. This situation is analogous to the one that arises in QCD for the conventional Yang–Mills gauge fixing parameter α'. In the latter case one finds that only in the Landau gauge ($\alpha' = 0$) does the parameter α' receive no renormalization, see, for example, ref. [8.33].

Furthermore, one should also note that we can identify the gauge fixing parameter α with the coupling constant of the theory g^2. In particular then, since the theory is independent of the choice of α, it is also coupling constant independent. One-loop results (and indeed classical results) are valid to all orders of perturbation theory, and a legitimate gauge choice is $\alpha = 0$, or equivalently $g^2 = 0$.

Given the above analysis, it is natural to ask: what is the unique value of the β-function within the Vilkovisky–DeWitt [8.34, 8.35] effective action program? For those readers unfamiliar with this subject we refer to the reviews [8.36]. We shall also illustrate this technique in section 8.4.4 in relation to Chern–Simons theory.

To implement a Vilkovisky–DeWitt construction, one must begin with the full set of classical fields and their local symmetry transformations. The difficulty here is that the symmetries of the classical action are reducible (i.e., they possess zero modes). The usual Vilkovisky–DeWitt construction is not suitable in this case. It is, however, reasonable to conjecture that a modified Vilkovisky–DeWitt procedure will yield a vanishing β-function as the gauge independent value.

We saw in section 5.2.4 that Donaldson theory can be obtained by twisting conventional $N = 2$ super-Yang–Mills theory. In the latter case, it was known [8.37] that the β-function was exact at one-loop order. Such a property can be seen to possibly have its origin in the fact that there exists a twisted (i.e. topological) version of the theory (see section also 4.4.6). Indeed, the question arises as to which properties are left invariant under such a twisting procedure. As we noted in section 3.6, the different interpretations of a theory lie in the physical state conditions imposed. For a study of some of these issues, we refer to ref. [8.38].

8.3. Topological sigma models at one loop

To further illustrate the features of the delta function gauge, let us examine the two dimensional sigma models of section 4 within this gauge [8.6]. The one-loop analysis in this case requires use of the background covariant coordinate expansions that were used in the study of the Nicolai map for these theories (see section 4.4.3).

Recall from (4.48) that the sigma model action in the delta function gauge is simply

$$S_q = \int d^2\sigma \, \{ B_{\alpha i} P^{\alpha i}_{+\,\beta j} \partial^\beta u^j + \overline{C}_{\alpha i} [D^\alpha \delta^i{}_j + \tfrac{1}{2} \epsilon^\alpha{}_\beta (D_k J^i{}_j) \partial^\beta u^j] C^k \} \ . \tag{8.18}$$

To perform a one-loop calculation we write $u^i \rightarrow u^i + \hat{u}^i$, where u^i is an isolated classical background, and $\hat{u}^i$ is the quantum fluctuation. We then expand the action to second order in the quantum fields, bearing in mind that the ghost, anti-ghost, and multiplier are purely quantum. Thus, we only need to expand $P^{\alpha i}_{+\,\beta j} \partial^\beta u^j$ to first order in ξ^i. From (4.57) we have

$$P^{\alpha i}_{+\,\beta j} \partial^\beta u^j \rightarrow P^{\alpha i}_{+\,\beta j} \partial^\beta u^j + P^{\alpha i}_{+\,\beta j} D^\beta \xi^j + \tfrac{1}{2} \epsilon^\alpha{}_\beta (D_j J^i{}_k)(\partial^\beta u^k)\xi^j \ . \tag{8.19}$$

The quadratic action is then

$$S_q^{(2)} = \int d^2\sigma \, \{ B_{\alpha i} [D^\alpha \delta^i{}_j + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k) \partial^\beta u^k] \xi^j + \overline{C}_{\alpha i} (D^\alpha \delta^i{}_j + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k) \partial^\beta u^k) C^j \} \ .$$

(8.20)

Integrating over the quantum fields, we see that we have a ratio of determinants which is equal to 1. Each system defines a map from the space of vectors ξ^i or C^i to the space of self-dual tensors $B_{\alpha i}$ or $\overline{C}_{\alpha i}$. Irrespective of how we choose to regularize these determinants, we have the result that the entire one-loop correction to the effective action vanishes. There are no divergences, the theory is finite, and the β-function vanishes.

The renormalization of the sigma model action in the $\alpha = 1$ gauge has been studied in ref. [8.3]. While the result in that case is quite tedious to obtain, it was found that the β-function for both the target space metric and complex structure were equal and non-zero. This guarantees the one-loop preservation of the topological properties of the theory, since, with these values for the β-functions, it is $P^\alpha_{+i}{}^\beta{}_j \partial_\alpha u^i \partial_\beta u^j$ which is renormalized, rather than simply $h^{\alpha\beta} g_{ij} \partial_\alpha u^i \partial_\beta u^j$ as in conventional supersymmetric sigma models.

We have illustrated that by using the freedom to choose a delta function gauge, one can prove that the entire one-loop correction to the effective action vanishes for both Donaldson theory and topological sigma models. This result can be seen to hold for all Witten type field theories, as it is in these cases that the gauge fixing of a topological shift symmetry is performed. Thus, for these models a non-zero β-function is really a gauge artifact.

8.4. Renormalization in Chern–Simons theory

In section 6.1, the pure Chern–Simons theory in three dimensions was discussed and the salient topological features of the model were explored. Of necessity, one-loop quantum corrections were included at this point as they enter a proper discussion of the framing issue, but the approach taken there was to appeal to general results in index theory. One need only recall that the phase in the one-loop partition function was obtained by an application of the Atiyah–Patodi–Singer index theorem.

Here, we would like to readdress the entire issue of quantum corrections in this model from a more pedestrian point of view. By regularizing determinants of operators, the one-loop results can be calculated in a simple and direct fashion. As in our discussion of Donaldson theory, we will then deal with off-shell corrections to the effective action, within the context of the Vilkovisky–DeWitt program. The physical relevance of these results will then be examined, and finally, we close with a section describing a peculiar type of supersymmetry which is present when the theory is quantized in the Landau gauge (on a flat three-manifold), and examine the potentially anomalous nature of this symmetry.

8.4.1. Regularization of determinants

The background field method is a powerful tool for computing correlation functions in quantum field theory, and we have already applied this technique in the previous subsection. In order to gain a better understanding of the subtle phase that arises in a one-loop analysis of the pure Chern–Simons theory, it is useful to first review how determinants of operators encode the lowest order quantum corrections.

In a path integral formulation of any quantum field theory, one is confronted with formal

expressions like

$$Z = \int d\phi \, e^{i \int \phi H \phi} \quad . \tag{8.21}$$

Here ϕ is a generic field, and the functional integral over all physically distinct field configurations needs to be made more precise. H is an elliptic Hermitian operator which may depend on the background fields and we will assume that the theory is formulated on some compact manifold, so that the spectrum of H will be real and discrete. Although a compact space–time may be unnatural in a truly physical theory, it is the case of greatest interest from the topological point of view and conveniently circumvents potential infrared problems. In the above expression, we could, for example, be considering an integral over the fermionic degrees of freedom in QCD, where $H = \not{D}$ (Dirac operator coupled to a Yang–Mills potential). In the present case H is a twisted Dirac operator [8.7].

One standard approach to defining Z [8.39] is to decompose the fields ϕ into eigenfunctions of H,

$$\phi(x) = \sum_n a_n \phi_n(x) \quad , \qquad H\phi_n(x) = \lambda_n \phi_n(x) \quad . \tag{8.22}$$

The measure $d\phi$ is taken to be $\prod_n da_n / \sqrt{\pi}$, and with the eigenfunctions appropriately normalized $(\phi_n, \phi_m) = \delta_{n,m}$, Z becomes

$$\prod_n \int_{-\infty}^{\infty} \frac{da_n}{\sqrt{\pi}} \, e^{ia_n^2 \lambda_n} = \prod_n \frac{1}{|\sqrt{\lambda_n}|} \, e^{(i\pi/4)\,\mathrm{sign}(\lambda_n)} \quad . \tag{8.23}$$

Such an expression needs to be regulated and the standard procedure is to first define the zeta and eta functions of a first order operator H [8.40, 8.41],

$$\zeta_H(s) \equiv \sum_n |\lambda_n|^{-s} \quad , \tag{8.24}$$

$$\eta_H(s) \equiv \sum_n \mathrm{sign}(\lambda_n)|\lambda_n|^{-s} \quad . \tag{8.25}$$

Many properties of these functions are well known for the type of operators we are considering [8.40, 8.42], and they have well defined analytic continuations in s. Other references to the η-function in the physics literature can be found in ref. [8.43]. A natural regulated definition of Z becomes

$$Z_{\mathrm{reg}} = e^{\zeta_H'(0)/2 + (i\pi/4)\eta_H(0)} = (\det H)^{-1/2} \quad . \tag{8.26}$$

It is the η-function which measures the spectral asymmetry of an operator—the possible mismatch of positive and negative eigenvalues—and is important for theories with first order operators. This analysis has assumed that the original fields ϕ are bosonic, but one can likewise treat the fermionic case.

For the purposes of calculation, it is most convenient to begin with integral representations of ζ and η; these are given by

$$\zeta_H(s) = \frac{1}{\Gamma(s)} \int_0^{\infty} dt \, t^{s-1} \mathrm{Tr} \, e^{-Ht} \quad , \tag{8.27}$$

$$\eta_H(s) = \frac{1}{\Gamma((s+1)/2)} \int\limits_0^\infty \mathrm{d}t \, t^{(s-1)/2} \mathrm{Tr}(H \, \mathrm{e}^{-H^2 t}) \ .$$ (8.28)

The above expression for $\zeta_H(s)$ is valid when H is a positive operator, i.e., all its eigenvalues are greater than zero. Although this is not the case for the first order operator we are considering, we can evaluate $\zeta'_H(0)$ in the form $|\zeta'_{H^2}(0)^{1/2}|$. One can easily check that the above representations reproduce the defining equations (8.24), (8.25) by computing the trace in a diagonal basis of H and using standard Γ-function relations. From (8.25) or (8.28) we see that, if two operators differ by a constant factor, their $\eta(0)$ values differ only by the sign of that factor. For the case of Chern–Simons theory this is the sign of k.

In cases where H has a zero mode, it is possible to make sense of the above expression by inserting an extra regulating factor $\mathrm{e}^{-\epsilon t}$ in the above integrals, taking the limit $\epsilon \to 0$ at the end of the calculation [8.42].

Before directly trying to attack the problem of evaluating the integral expressions for the ζ and η functions, it is very convenient to first make some simple observations. The first point is that we only need the behavior of these functions near $s = 0$, and indeed, this will make the calculations tractable. The standard trick [8.42] in evaluating $\eta_H(0)$ is to first introduce a one-parameter family of operators $H(\lambda)$ such that $H(1) = H$ and such that $H(0)$ is an operator whose trace kernel can easily be evaluated. By differentiating the above expression (8.28) one obtains

$$\frac{\mathrm{d}\eta_{H(\lambda)}(s)}{\mathrm{d}\lambda} = \frac{-s}{\Gamma((s+1)/2)} \int\limits_0^\infty \mathrm{d}t \, t^{(s-1)/2} \mathrm{Tr}\left(\frac{\mathrm{d}H(\lambda)}{\mathrm{d}\lambda} \, \mathrm{e}^{-tH^2(\lambda)}\right) \ ,$$ (8.29)

showing that the $\lim_{s\to 0} \eta_H(s)$ is given by integrating ($\int_0^\infty \mathrm{d}\lambda$) the residue of the $s = 0$ pole of the t-integral. As for computing $\zeta'_H(0)$, we just note that

$$\zeta'_H(0) = \lim_{s\to 0} \frac{1}{\Gamma(s+1)} \int\limits_0^\infty \mathrm{d}t \, t^{s-1} \mathrm{Tr} \, \mathrm{e}^{-Ht} \ ,$$ (8.30)

since $\Gamma(s+1) = s\Gamma(s)$.

There are at least two standard methods for proceeding with the calculation for a given operator. The most powerful technique is to employ the Schwinger–DeWitt expansion, which is an asymptotic expression for the kernel $\mathrm{Tr} \, \mathrm{e}^{-Ht}$ [8.42], and applies to operators on curved manifolds. We took this path in section 8.2 when computing the β-function in Donaldson theory. Another procedure is to compute the trace kernels in momentum space (we are now on $\mathbb{R}^n$ or T^n) and this corresponds to an operator regularization scheme of McKeon and Sherry [8.44]. We will illustrate this technique by computing $\eta_H(0)$ for the relevant first order operator in Chern–Simons field theory.

8.4.2. Chern–Simons theory at one loop

We have seen that the pure Chern–Simons field theory is generated by the classical action

$$S(A) = \frac{k}{4\pi} \int\limits_M \mathrm{tr}(A \wedge \mathrm{d}A + \tfrac{2}{3} A \wedge A \wedge A) \equiv kI(A) \ ,$$ (8.31)

where $A = A^a T^a$ is a connection form on some principal bundle over a closed three-manifold M and T^a is a representation of the structure group G. The field configurations which are extrema of

this action are flat connections,

$$F_A = dA + [A, A] = 0 .$$

(8.32)

The quantum theory of this classical system is constructed in the usual way, and one considers the partition function

$$Z = \int dA \, e^{iS(A)} ,$$

(8.33)

where the functional integral is over all gauge equivalence classes of connections.

A perturbative analysis of this theory begins by making a background field expansion of the connection $A_\alpha \to A_\alpha + B_\alpha$, into a classical background A and a quantum field B. BRST quantization proceeds as in any garden variety Yang–Mills theory, and it is convenient to choose the background field gauge $D_\alpha B^\alpha = 0$, where $D_\alpha^{ab} = \partial_\alpha \delta^{ab} + A_\alpha^c f^{acb}$ is the covariant derivative with respect to the background field A.

The one-loop contribution to the partition function is simply given by

$$Z[A] = e^{ikI(A)} \int dB \, db \, d\bar{c} \, dc \, e^{iS_q^{(2)}} ,$$

(8.34)

with

$$S_q^{(2)} = \int d^3x \, \text{tr}(\epsilon^{\alpha\beta\gamma} B_\alpha D_\beta B_\gamma - 2bD \cdot B + \bar{c}D^2 c) ,$$

(8.35)

where we have rescaled the quantum fields to obtain a more convenient normalization, and kept only terms quadratic in the quantum fields. The b field is the multiplier which enforces the gauge condition, and $\bar{c}, c$ are the usual ghosts. In our conventions, the structure constants of the semi-simple Lie algebra are real and completely anti-symmetric with $[T^a, T^b] = f^{abc}T^c$. For the fundamental representation of SU(n), the matrices T^a are skew-hermitian and we take $\text{tr}\, T^a T^b = -\frac{1}{2}\delta^{ab}$, while the quadratic Casimir is defined by $\sum_{c,d} f^{acd} f^{bcd} = c_v \delta^{ab}$.

Following the preceding discussion, the one-loop corrected partition function can be represented as a combination of determinants,

$$Z[A] = e^{ikI(A)} \det(-D^2)/\sqrt{\det H} ,$$

(8.36)

where H is the operator which appears sandwiched between the B and b fields in the action

$$\int d^3x \begin{pmatrix} B_\alpha^a & b^a \end{pmatrix} \begin{pmatrix} -\epsilon^{\alpha\gamma\beta} D_\gamma & -D^\alpha \\ D^\beta & 0 \end{pmatrix}^{ab} \begin{pmatrix} B_\beta^b \\ b^b \end{pmatrix} .$$

(8.37)

Since H is a first order operator, there is the possibility of the η-function phase, namely

$$(\det H)^{-1/2} = (\det H^2)^{-1/4} e^{(i\pi/4)\eta_H(0)} ,$$

(8.38)

and thus the partition function takes the form

$$Z[A] = e^{i[k+(\pi/4)\eta_H(0)]I(A)} \det(-D^2)/(\det H^2)^{1/4} .$$

(8.39)

8.4.3. Evaluation of the phase

As we have seen in the preceding section, the one-loop corrections to the effective action are encoded in three pieces: $\det(-D^2)$, $\det(H^2)$, and $\eta_H(0)$. Computing the first two determinants

is relatively straightforward, since the operators are positive definite—one need only evaluate $\zeta'(0)$—and the essential techniques are well described in the physics literature [8.45]. Here we shall concentrate on the computation of the η-function phase using a momentum space procedure [8.18], which is in the spirit of an "operator regularization" scheme of McKeon and Sherry [8.44]. In effect, it extends their regularization procedure to the case of first order operators, where the potential for generating a phase exists. This procedure has the advantage of being both simple and direct, but is limited to the case of flat space–times ($\mathbb{R}^n$ or T^n). Although more powerful techniques exist which apply without this restriction [8.42] there is no need to enter into this discussion here.

We wish to calculate $\eta_H(0)$ for the operator H given in (8.37), when the theory is defined on $\mathbb{R}^3$. The momentum space technique that we will employ applies equally well to the case of the three-torus; one need only write Fourier series in place of Fourier integrals. The definition of the η-function (8.28) instructs us to evaluate the trace of a certain operator, and it is this trace that we shall evaluate in momentum space. For a general operator $\mathcal{O}$, $\mathrm{Tr}(\mathcal{O})$ is defined by

$$\mathrm{Tr}(\mathcal{O}) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \langle p|\mathcal{O}|p \rangle = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \, \mathrm{d}^3 x \, \mathrm{d}^3 y \, \langle p|x\rangle \langle x|\mathcal{O}|y\rangle \langle y|p \rangle \ . \tag{8.40}$$

[Note that we take $\langle p|x\rangle = \mathrm{e}^{-ipx}$, $\hat{A}_\alpha(p) = \int \mathrm{d}^3 x \, \mathrm{e}^{ipx} A_\alpha(x)$, and $\int \mathrm{d}^3 x \, \delta(x) = \int (2\pi)^{-3} \mathrm{d}^3 p \, \hat{\delta}(p)$ $= 1$.] If $\mathcal{O}$ has, in addition, any discrete labels, then one must augment the above formula with an additional trace over these labels. Written as a matrix element in coordinate space, the operator H is defined by

$$\langle x, a, \alpha | H | y, b, \beta \rangle = \begin{pmatrix} -\epsilon^{\alpha\gamma\beta} D_\gamma & -D^\alpha \\ D^\beta & 0 \end{pmatrix}_x^{ab} \delta(x - y) \ . \tag{8.41}$$

For the purposes of our calculation, we decompose $H = H_0 + H_1$ into a "free part" H_0, which does not contain the background field $A_\alpha^a(x)$, and a part H_1 which is proportional to this field. In momentum space, we have

$$\langle p, a, \alpha | H_0 | q, b, \beta \rangle = i\delta^{ab} \hat{\delta}(p - q) \begin{pmatrix} -\epsilon^{\alpha\gamma\beta} p_\gamma & -p^\alpha \\ p^\beta & 0 \end{pmatrix} \ , \tag{8.42}$$

$$\langle p, a, \alpha | H_1 | q, b, \beta \rangle = f^{acb} \begin{pmatrix} -\epsilon^{\alpha\gamma\beta} \hat{A}_\gamma^c(p - q) & -\hat{A}^{c\alpha}(p - q) \\ \hat{A}^{c\beta}(p - q) & 0 \end{pmatrix} \ . \tag{8.43}$$

We are now in a position to evaluate the η-function in the form (8.29), by taking $H(\lambda) = H_0 + \lambda H_1$. If we explicitly integrate over λ, we have

$$\eta_H(s) - \eta_{H_0}(s) = \int_0^1 \mathrm{d}\lambda \, \frac{-s}{\Gamma(s + 1/2)} \int_0^\infty \mathrm{d}t \, t^{(s-1)/2} \, \mathrm{Tr}(H_1 \mathrm{e}^{-tH^2(\lambda)}) \ . \tag{8.44}$$

Now note that $\eta_{H_0}(s) = 0$, since there is always a one-to-one correspondence between positive and negative eigenvalues of H_0. We can now compute $\eta_H(s)$ order by order in the background field A_α^a by expanding the exponential term using the Schwinger–DeWitt expansion [8.46, 8.44]. For any operator $M = M_0 + M_1$, this expansion takes the form

$$
e^{-Mt} = e^{-M_0 t} - t \int\limits_0^1 \mathrm{d}u \; e^{-M_0(1-u)t} M_1 e^{-M_0 u t}
$$

$$
+ t^2 \int\limits_0^1 u \, \mathrm{d}u \int\limits_0^1 \mathrm{d}v \; e^{-M_0(1-u)t} M_1 e^{-M_0(1-v)ut} M_1 e^{-M_0 uvt} + \cdots \, , \tag{8.45}
$$

where we have explicitly written only those terms which are essential to the case at hand.

At this point, it may appear that there is no end to the number of terms we must calculate; (8.45) is an infinite series. This would be the case if we wished to calculate η_H for a generic value of s, but a one-loop analysis only requires knowledge at $s = 0$. A glance at (8.44) shows that this series can be nipped in the bud; we only need to find those terms in the above expansion which give a pole at $s = 0$, after the t-integral is performed, and these fortunately turn out to be finite in number. A general analysis has been given in ref. [8.19], where it was argued that only terms up to order A^3 have a chance of contributing to $\eta_H(0)$. Here, we will be content with illustrating the technique by evaluating the lowest order piece, which is order A^2.

Our task now is to compute $\eta_H(0)$, keeping only two powers of the interaction H_1. At this order, (8.44) reduces to

$$
\frac{1}{2} \frac{s}{\Gamma((s+1)/2)} \int\limits_0^\infty \mathrm{d}t \, t^{(s+1)/2} \operatorname{Tr} \left(H_1 \int\limits_0^1 \mathrm{d}u \; e^{-H_0^2(1-u)t} \{H_0, H_1\} e^{-H_0^2 u t} \right) , \tag{8.46}
$$

where we have already carried out the λ integral. The first step is to evaluate the trace, and this is most conveniently carried out in momentum space. Using the above expressions for H_0 and H_1, it is straightforward to show that $\operatorname{Tr}(H_1 e^{-H_0^2(1-u)t} \{H_0, H_1\} e^{-H_0^2 u t})$ has the momentum space representation

$$
4 \mathrm{i} f^{dac} f^{cbd} \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{\mathrm{d}^3 q}{(2\pi)^3} \; e^{-q^2(1-u)t - p^2 u t} \epsilon^{\alpha\beta\gamma} \hat{A}_\alpha^a(p - q)(q - p)_\beta \hat{A}_\gamma^b(q - p) . \tag{8.47}
$$

If we now shift the q variable to $q + p$ and then complete the square in the exponential, this trace reduces to

$$
4 \mathrm{i} (-c_v \delta^{ab}) \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \; e^{-p^2 t} \int \frac{\mathrm{d}^3 q}{(2\pi)^3} \; e^{-q^2 u(1-u)t} \epsilon^{\alpha\beta\gamma} \hat{A}_\alpha^a(-q) q_\beta \hat{A}_\gamma^b(q) . \tag{8.48}
$$

Notice that we have factored out an elementary integral over p, $\int (2\pi)^{-3} \mathrm{d}^3 p \; e^{-p^2 t} = (4\pi)^{-3/2} t^{-3/2}$, and have used our convention $f^{acd} f^{bcd} = c_v \delta^{ab}$ for the quadratic Casimir. The t-integral is now easy to evaluate, giving

$$
-\frac{4\mathrm{i}}{(4\pi)^{3/2}} c_v \frac{(s/2)\Gamma(s/2)}{\Gamma((s+1)/2)} \int\limits_0^1 \mathrm{d}u \; [u(1 - u)]^{-s/2} \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \; (p^2)^{-s/2} \epsilon^{\alpha\beta\gamma} \hat{A}_\alpha^a(-p) p_\beta \hat{A}_\gamma^a(p) , \tag{8.49}
$$

for the order A^2 contribution to $\eta_H(s)$. The remaining integral over the u-parameter can be evaluated with the general relation

$$
\int\limits_0^1 \mathrm{d}u \, u^{\alpha-1} (1 - u)^{\beta-1} = \Gamma(\alpha) \Gamma(\beta) / \Gamma(\alpha + \beta) , \tag{8.50}
$$

and we obtain the final expression in momentum space,

$$\frac{-ic_v}{2\pi^{3/2}} \frac{\Gamma(1+s/2)}{\Gamma((s+1)/2)} \frac{\Gamma(1-s/2)^2}{\Gamma(2-s)} \int \frac{d^3p}{(2\pi)^3} \, (p^2)^{-s/2} \epsilon^{\alpha\beta\gamma} \hat{A}_\alpha^a(-p) p_\beta \hat{A}_\gamma^a(p) \; . \tag{8.51}$$

Notice that this is a non-local expression when s is different from zero. At $s = 0$, it reduces to

$$\frac{-ic_v}{2\pi^2} \int \frac{d^3p}{(2\pi)^3} \, \epsilon^{\alpha\beta\gamma} \hat{A}_\alpha^a(-p) p_\beta \hat{A}_\gamma^a(p) \; , \tag{8.52}$$

or equivalently in coordinate space,

$$\frac{c_v}{\pi^2} \int d^3x \, \mathrm{tr}(\epsilon^{\alpha\beta\gamma} A_\alpha \partial_\beta A_\gamma) \; , \tag{8.53}$$

where we have restored the trace over the fundamental representation $(\mathrm{tr}\, T^a T^b = -\frac{1}{2}\delta^{ab})$. In this form, we easily recognize this structure as the first term in the Chern–Simons action. Had we carried out the entire calculation, where terms of order A^3 are retained, we would have found that the complete answer is [8.19]

$$\eta_H(0) = \frac{c_v}{\pi^2} \int d^3x \, \epsilon^{\alpha\beta\gamma} \, \mathrm{tr}(A_\alpha \partial_\beta A_\gamma + \tfrac{2}{3} A_\alpha A_\beta A_\gamma) \; . \tag{8.54}$$

The reader who wishes to carry out the exercise of computing this A^3 contribution should be warned that the calculation is only tractable at $s = 0$; i.e., one does not get a simple expression for $\eta_H(s)$ when s is different from zero.

As remarked after (8.28), the value of $\eta_H(0)$ depends on the sign of k; from (8.39) we thus have the result

$$Z[A] = e^{i[k+\mathrm{sign}(k)c_v]I(A)} \det(-D^2)/(\det H^2)^{1/4} \; . \tag{8.55}$$

To complete the one-loop analysis, we need only evaluate the two determinants $\det(-D^2)$ and $\det(H^2)$. As we have previously remarked, this is nothing more than a standard ζ-function calculation, and can be carried out in a manner identical to the one we have outlined here; one just evaluates a slightly simpler trace kernel. Before quoting the result, a few remarks are in order. First, if we look at the operator H^2,

$$H^2 = \begin{pmatrix} -\delta_{\alpha\beta} D^2 - F_{\alpha\beta} & \frac{1}{2}\epsilon_{\alpha\sigma\tau} F^{\sigma\tau} \\ -\frac{1}{2}\epsilon_{\beta\sigma\tau} F^{\sigma\tau} & -D^2 \end{pmatrix} \; , \tag{8.56}$$

we immediately see that it is proportional to D^2, when the background field is on-shell $F_{\alpha\beta} = 0$, and in fact the ratio $\det(-D^2)/(\det(H^2))^{1/4}$ is identically one in this limit. If we are only interested in the on-shell corrections, there is nothing further to do. Away from $F = 0$, on the other hand, these determinants no longer cancel, and one finds that

$$\ln \frac{\det(-D^2)}{[\det(H^2)]^{1/4}} = \frac{c_v}{32} \int \frac{d^3p}{(2\pi)^3} \, (p^2)^{-1/2} \hat{F}_{\alpha\beta}^a(-p) \hat{F}^{a\alpha\beta}(p) \; , \tag{8.57}$$

for the lowest order correction [8.19, 8.17]. There will in general be further higher order corrections proportional to powers of the curvature F, but these have not been computed. However, we note that the effective action is of the form $iCS + F^2$, where the imaginary unit is crucial. Thus, flat connections remain as the stationary points. It is worth making some clarifying remarks on this

point: The topological nature of Chern–Simons theory arises essentially because of the fact that the space of solutions to the equations of motion modulo the gauge symmetries is a finite dimensional space, in this case the moduli space of flat connections. From the path integral point of view, one can establish the topological nature of the theory by performing, for example, a one-loop analysis. This one-loop analysis is performed by expanding about flat connections. As shown in section 6.1.3, this leads to a ratio of determinants defined with respect to the flat background connection, called the Ray–Singer torsion. The Ray–Singer torsion is a metric independent object; and its definition is in terms of determinants of a flat connection. In (8.57) we have examined this ratio for an off-shell (non-flat) background; and we have seen that F^2-like terms appear. The fact that such terms are metric dependent is not a problem, since in the off-shell case this ratio of determinants is no longer the Ray–Singer torsion. A discussion of F^2 terms has also been given in ref. [8.20].

8.4.4. Gauge dependence of the effective action

A natural issue that arises in the study of any quantum field theory is the dependence of a particular calculation on the way in which the fields have been parametrized; and in the case of gauge theories, there is the further issue of gauge dependence. We have already seen that the β-function in Donaldson theory depends on the way in which the topological shift symmetry is gauge fixed; one gauge yielded a β-function equal to that of $N = 2$ super-Yang–Mills, while the β-function vanished in the delta function gauge. In the preceding section, we computed one-loop corrections to Chern–Simons theory—in particular the quantum mechanical phase shift—and it is natural to pose the same question here. The answer is that the η-function phase does in fact depend on the gauge choice, and explicit calculations [8.19] have been carried out to verify this. What then is the significance of a calculation performed in a specific gauge?

One way to address this issue of gauge and parametrization dependence was used in refs. [8.34, 8.35], where the authors began by observing that these issues were, in a sense, a failure in the original definition of the effective action. The usual definition of the effective action is manifestly not a natural geometrical object on the space of field configurations. One can see this easily at one loop, where the background field method instructs us to compute the determinant of the Hessian $S_{,ij}$, where S is the gauge fixed action and the condensed index notation means that we are taking derivatives with respect to the ith and jth field coordinate [if i labels the field $\Phi^i(x)$, then by $S_{,ij}$ we mean $\delta^2 S/\delta\Phi^j(y)\delta\Phi^i(x)$]. If we demand that S is a scalar function on the space of field configurations, then, although $S_{,i}$ is a vector, $S_{,ij}$ is not a rank-2 tensor. To construct a tensor, one needs a covariant derivative on this infinite dimensional space, and the simplest solution is to take the Christoffel connection of some metric. The general program has been formulated in ref. [8.35] to all loop orders; at one loop the result of this procedure is to replace $\det S_{,ij}$ by $\det(S_{,ij} - \Gamma^k_{ij}S_{,k})$. This latter quantity can be shown to be sensitive to the gauge and parametrization issues raised above. We will now review this construction within the context of Chern–Simons theory.

We have seen that the quantized theory can be described in terms of the fields $A^a_\alpha(x)$, $b^a(x)$, $\bar{c}^a(x)$, $c^a(x)$, where $A^a_\alpha(x)$ is a connection form, $b^a(x)$ is a gauge fixing multiplier field, and $(\bar{c}, c)$ are the usual ghosts. To avoid any issue involving the geometry of the multiplier space, we proceed in the standard way by introducing an αb^2 term in (8.35), and then integrating over b. This leads to the usual $-\alpha^{-1}(D \cdot B)^2$ gauge fixing term; the Landau gauge is then recovered in the $\alpha \to 0$ limit. Our first step is to select a metric on the space of physically distinct field configurations. For the $\bar{c}, c$ fields this is trivial, since we can take the constant unit metric

$$G_{\bar{c}^a(x)\,c^b(y)} = \delta^{ab}\delta(x - y) \ . \tag{8.58}$$

As this metric is field independent (i.e. constant), the Christoffel connection in this direction is

trivial. The ghost determinant that we used before, $\det(-D^2)$, remains unchanged in this program.

Defining a metric on the space $\mathcal{A}/\mathcal{G}$ is more subtle. We begin by selecting a metric on the full space of connections $\mathcal{A}$. Since this space is topologically trivial we might as well select the unit metric as above,

$$G_{A_\mu^a(x) A_\nu^b(y)} = \delta^{ab} g_{\mu\nu} \delta(x-y) \ , \tag{8.59}$$

where $g_{\mu\nu}$ is a background space–time metric. It is not difficult to define a metric on $\mathcal{A}$ which ignores tangent vectors along the group flow of $\mathcal{G}$. In condensed notation, the gauge transformation of a field Φ^i is given by $\delta\Phi^i = R_\alpha^i \epsilon^\alpha$. For example, when $\Phi^i = A_\mu^a(x)$, we have $R_\alpha^i = (D_\mu)_x^{ab} \delta(x-y)$ and the gauge parameter is $\epsilon^\alpha = \epsilon^b(y)$. We now consider the metric [see (5.90)]

$$\gamma_{ij} = G_{ij} - R_\alpha^k R_\beta^l N^{\alpha\beta} G_{ki} G_{lj} \ , \tag{8.60}$$

with $N^{\alpha\beta}$ defined as the inverse of $N_{\alpha\beta} = R_\alpha^k R_\beta^l G_{kl}$. It is now a simple matter to show that $\gamma_{ij} R_\alpha^j = 0$, i.e., vectors tangent to the group flow are ignored, and so γ_{ij} can be considered as a metric on the quotient space $\mathcal{A}/\mathcal{G}$. We remark that one could more generally consider other metrics on $\mathcal{A}$; the only constraint is that one requires the gauge transformations to be Killing vectors on $\mathcal{A}$, $D_i R_{j\alpha} + D_j R_{i\alpha} = 0$. Here D_i is the covariant derivative generated by the metric G_{ij}. In our case, this is automatically satisfied.

The Christoffel symbol of the metric on $\mathcal{A}/\mathcal{G}$ is now given in the usual way by

$$\gamma_{kl} \Gamma_{ij}^l = \tfrac{1}{2}(\gamma_{ik,j} + \gamma_{jk,i} - \gamma_{ij,k}) \ . \tag{8.61}$$

Solving for the coefficients Γ_{ij}^k is complicated by the fact that we are really only interested in directions tangent to $\mathcal{A}/\mathcal{G}$, that is, γ_{ij} is only invertible on this submanifold. There is a trick which greatly simplifies this task [8.36]. If we multiply both sides of this equation by G^{mk}, then we have

$$(\delta_l^m - G^{mk} R_{k\alpha} R_{l\beta} N^{\alpha\beta}) \Gamma_{ij}^l = \tfrac{1}{2} G^{mk} [G_{ik,j} + G_{jk,i} - G_{ij,k}$$
$$- (R_{i\alpha} R_{k\beta} N^{\alpha\beta})_{,j} - (R_{j\alpha} R_{k\beta} N^{\alpha\beta})_{,i} + (R_{i\alpha} R_{j\beta} N^{\alpha\beta})_{,k}] \ . \tag{8.62}$$

Now using the two identities

$$R_{i\beta,j} - R_{j\beta,i} = D_j R_{i\beta} - D_i R_{j\beta} \ ,$$
$$N_{,i}^{\alpha\beta} = -N^{\alpha\gamma} N^{\beta\delta} N_{\gamma\delta,i} = -2N^{\alpha(\gamma} N^{\delta)\beta} D_{(i} R_{j)\delta} R_\gamma^j \ , \tag{8.63}$$

one can rewrite (8.62) as

$$\Gamma_{jk}^i = \hat{\Gamma}_{jk}^i + T_{jk}^i + R_\alpha^i H_{jk}^\alpha \ , \tag{8.64}$$

where

$$T_{jk}^i = -2B_{(j}^\alpha D_{k)} R_\alpha^i + R_\alpha^i D_l R_\beta^i B_{(j}^\alpha B_{k)}^\beta \ ,$$
$$B_k^\alpha = N^{\alpha\beta} R_{k\beta} \ , \qquad R_{k\beta} = G_{kl} R_\beta^l \ ,$$
$$H_{ij}^\alpha = B_k^\alpha \Gamma_{ij}^k - N^{\alpha\beta} (D_{(i} R_{j)\beta} - N_{\beta\gamma,(i} B_{j)}^\gamma) \ . \tag{8.65}$$

We are using the notation that $\hat{\Gamma}_{jk}^i$ represents the Christoffel symbol of the metric G_{ij}, and that round brackets represent symmetrization over indices, so that $T_{(ab)} = \tfrac{1}{2}(T_{ab} + T_{ba})$. It is important

to note that these gyrations have not produced a solution for Γ^i_{jk}; a glance shows that Γ^i_{jk} is also embedded in the right hand side of (8.64). The key point, however, is that the term in Γ^i_{jk} proportional to R^i_α is irrelevant; we need to compute $\Gamma^k_{ij} S_{,k}$ in the effective action, but $R^k_\alpha S_{,k} = 0$ by virtue of gauge invariance. Furthermore, we have noted that $\hat{\Gamma}^i_{jk} = 0$ for the case at hand, so the entire construction of a connection on $\mathcal{A}/\mathcal{G}$ simply reduces to computing the T^i_{jk} symbol.

The point now is that the Vilkovisky–DeWitt correction term vanishes in precisely the gauge we have chosen, namely the Landau gauge. This can be seen as follows: we observe that the Vilkovisky–DeWitt correction term appears in the combination $\Phi^i_q T^k_{ij} S_{,k} \Phi^j_q$, where Φ^i_q is the set of quantum fluctuations around the classical Φ^i backgrounds. A glance at the structure of T^k_{ij} in (8.65) reveals that the correction term vanishes if

$$B^\alpha_i \Phi^i_q = 0 , \quad \text{all } \alpha. \tag{8.66}$$

Using the definitions given in (8.65), we see that an equivalent statement is

$$G_{ij} R^i_\alpha \Phi^j_q = 0 , \quad \text{all } \alpha. \tag{8.67}$$

Now in the present case $\Phi^i = A^a_\mu(x)$; the corresponding quantum field is thus $\Phi^i_q = B^a_\mu(x)$. In addition, there is a single gauge generator $\epsilon^\alpha = \epsilon^c(z)$. Using the metric (8.59) and the definition of the gauge generator R^i_α, it is easy to establish that (8.67) implies $D_\mu B^\mu = 0$. Thus, from this point of view we have the a priori knowledge that the Landau gauge provides the unique result within the Vilkovisky–DeWitt framework. However, one is free to choose other gauges, and it is an illuminating exercise to actually implement the Vilkovisky–DeWitt procedure. The reader is referred to ref. [8.19] for details.

8.4.5. Physical relevance of the results

At this point we wish to discuss the physical relevance of the results obtained in the previous sections. There has been some controversy in the literature regarding the existence and physical meaning of the shift in k derived in (8.55); it is thus useful to present some (hopefully) clarifying remarks. A recent review article [8.47] may be useful in this regard.

First of all, our calculations have dealt solely with the off-shell effective action at one loop on $\mathbb{R}^3$. We have used a specific gauge invariant regularization scheme, and have shown that the η-function of the Chern–Simons operator, defined with respect to this off-shell background, is non-zero. Furthermore, the η-function is proportional to the Chern–Simons action of the background connection, with a proportionality coefficient given by c_v sign(k). Such a calculation has a relevance in its own right, since it provides a means of checking for anomalies at one-loop order. As we have seen, the integer nature of the Chern–Simons coupling k is required for invariance under large gauge transformations. Despite the finiteness of the model, it remains to check whether, if a shift does occur, it is by an integer amount.

We have then addressed the question as to the meaning of the value of the coefficient, within this particular regularization scheme. The result, as we have shown, is that indeed the value depends on the gauge choice. Using a different gauge, but the same regulator, yields a different value. However, the Vilkovisky–DeWitt effective action program assigns a meaning to the notion of a unique off-shell effective action. Including the appropriate Vilkovisky–DeWitt correction shows that the unique value for the η-phase is provided by the Landau gauge result (which is the gauge originally used by Witten [8.7] in the calculation sketched in section 6.1.3).

The most important question, however, is: what is the significance of a calculation performed with a particular regularization scheme? The aim is to make regularization independent statements;

it is these which have true physical meaning. To perform a given calculation, one may select any desired regularization scheme. At the end of the day, however, we must all agree when discussing physical quantities. The controversy has arisen over whether or not such a shift has any physical significance. We shall briefly discuss the two points of view which have been advanced on this issue.

Recall that the classical Chern–Simons action comes equipped with an integer coupling k; let us call this the bare coupling and denote it by k_B. We assume, for convenience, that k_B is positive. This integer appears in two separate guises [8.7, 8.15]. Firstly, it appears in the central extension of the associated Kač–Moody algebra, as shown in section 6.1.5. Secondly, in the computation of Wilson line expectation values, a natural variable appears; this variable is called the monodromy parameter, which we denote by q. Knot and link invariants can be expressed as polynomials in q, and q depends on k_B (see section 6.1.4).

Now, it is well known from the conformal field theory point of view that, if the central extension of the current algebra is equal to l (so that we are dealing with, for example, a level l WZW model [8.48]), then the monodromy parameter q is given as an expansion in powers of $l + c_v$ by $q = e^{2\pi i/(l+c_v)}$. The question is: do we see this shift in Chern–Simons theory, and if so, how?

Let us first discuss the argument due to Alvarez-Gaumé et al. [8.15]. The claim is that it is physically meaningful to compute both the central extension and the monodromy parameter as functions of the bare coupling. Both are gauge invariant quantities and, furthermore, it is also meaningful to compute their difference. Using a gauge invariant regulator, these authors necessarily ensure that the current algebra relations are satisfied, with a central extension given by k_B. Computing one-loop corrections to the two-point and three-point functions $\langle AA \rangle$ and $\langle AAA \rangle$, yields the result that the bare coupling receives the famous shift. Further work [8.49] claims to establish that a computation of the monodromy parameter to two-loop order yields an expansion for q in powers of $k_B + c_v$. Therefore, the final result here is that, purely from the perturbative Chern–Simons point of view, one does see a relative shift.

On the other side of the coin, Guadagnini et al. [8.11, 8.50], using a different regularization scheme, have noticed that the $\langle AA \rangle$ and $\langle AAA \rangle$ functions remain bare at one loop; the bare coupling is not shifted in this scheme. Further computations to two-loop order show that the monodromy parameter q is an expansion in powers of k_B [8.50].

To make physical statements we should express all our results in terms of a renormalized coupling, which we denote by k_R; it is then possible to compare results, and hopefully we all agree. The question to ask at this point is: how do the central extension and monodromy parameters depend on k_R. In the scheme of refs. [8.12, 8.15], the renormalized coupling is defined by $k_R = k_B + c_v$. The monodromy parameter is then an expansion in k_R, while the central extension is $k_R - c_v$. This can, and should be, regarded as the true physical information extracted from their regularization scheme. In terms of a WZW model, for example, one can then identify the level as being $k_R - c_v$. From the results of [8.11, 8.50], we see that the renormalized coupling is $k_R = k_B$; thus, q is again an expansion in powers of k_R, in agreement with the results of refs. [8.12, 8.15]. We would therefore expect that the central extension is again $k_R - c_v$, which in terms of this regularization scheme is equivalently $k_B - c_v$. The point to note here is that, since the regulator is gauge non-invariant, the currents are not conserved; therefore, more work is required to establish the level of the algebra in terms of k_R. Such a calculation has not yet been performed, but clearly the expectation is that one will find a level $k_R - c_v$ current algebra. However, we should point out that this is not the viewpoint expressed in refs. [8.50, 8.51].

We conclude with a remark: working purely from the two-dimensional WZW point of view, one is given a current algebra at a certain level l, the monodromy parameter is then an expansion

in $l + c_v$. To see this result from the pure Chern–Simons point of view, one must define the level and the monodromy parameter solely in terms of a renormalized coupling. In this way, one establishes the fact that a relative shift does occur, and such a statement should be regularization independent. However, it could also be the case that different regularizations define inequivalent quantum theories.

8.4.6. Chern–Simons supersymmetry

We close with a discussion of a peculiar supersymmetry which is present when the Chern–Simons theory is quantized in the Landau gauge. The action under study is

$$S_q = \frac{k_B}{4\pi} \int d^3x \ \text{tr}[\epsilon^{\alpha\beta\gamma}(A_\alpha\partial_\beta A_\gamma + \tfrac{1}{3}A_\alpha[A_\beta, A_\gamma]) - 2b\partial \cdot A - 2\bar{c}\partial \cdot Dc] \ , \tag{8.68}$$

where D_α is the covariant derivative with respect to the connection A_α and the normalization of the action is chosen so as to make the transformation rules more pleasant. This action possesses the usual Yang–Mills gauge invariance, namely

$$\delta A_\alpha = -\epsilon D_\alpha c \ , \qquad \delta c = \tfrac{1}{2}\epsilon\{c,c\} \ , \qquad \delta\bar{c} = \epsilon b \ , \qquad \delta b = 0 \ . \tag{8.69}$$

Here, ϵ is a constant Grassmann odd parameter. However, in addition to this symmetry, it is also straightforward to verify that the following set of transformations leave the action invariant [8.2]:

$$\delta A_\alpha = \epsilon^\beta \epsilon_{\alpha\beta\gamma}\partial^\gamma c \ , \qquad \delta\bar{c} = \epsilon^\alpha A_\alpha \ , \qquad \delta c = 0 \ , \qquad \delta b = \epsilon^\alpha D_\alpha c \ , \tag{8.70}$$

where ϵ^α is a Grassmann odd vector parameter.

Following this, other studies of this supersymmetry were made [8.27–8.29]. This supersymmetry is reminiscent of the usual super-Yang–Mills transformations, only here it is the bosonic field A which has first order field equations, while the classical equations for the Grassmann odd fields $\bar{c}$ and c are of second order. Since there are no spinors in this theory, one is led to consider infinitesimal transformations with an odd vector parameter. In the present case, however, the anticommutator of two supersymmetry charges vanishes (i.e., the supersymmetry algebra is Abelian); indeed one can easily check that

$$[\delta(\epsilon_1^\alpha), \delta(\epsilon_2^\beta)](A_\mu, \bar{c}, c, b) = 0 \ . \tag{8.71}$$

The importance of this supersymmetry can be seen by examining the associated Ward identities. One proceeds in the standard fashion by supplementing the action with the following source terms:

$$S_q \to \tilde{S}_q = S_q + \int d^3x \ \text{tr}(J^\alpha A_\alpha + M\bar{c} + Nc + Qb - S^\alpha D_\alpha c + V\tfrac{1}{2}\{c,c\}) \ . \tag{8.72}$$

The notation here is the following: (J_α, M, N, Q) are the usual source terms for the fields $(A_\alpha, \bar{c}, c, b)$ with Grassmann parity and ghost number assignments given by $(+, -, -, +)$ and $(0, 1, -1, 0)$, respectively. S_α and V are the composite sources for the non-linear transformations δA and δc. They are (odd, even) with ghost numbers $(-1, -2)$, respectively. We can now define the unconnected and connected generating functionals Z and W by

$$Z[J_\alpha, ..., Q; S_\alpha, V] = \int d\Phi \ \exp(-\tilde{S}_q) = \exp(-W[J_\alpha, ..., Q; S_\alpha, V]) \ , \tag{8.73}$$

where we have denoted the collective field content by Φ. The 1-PI generating functional is then defined via the Legendre transform of W as

$$W[J_\alpha, ..., Q; S_\alpha, V] = \Gamma[A_\alpha, ..., b; S_\alpha, V] + \int d^3x \, \text{tr}(J^\alpha A_\alpha + \cdots + Qb) \, , \tag{8.74}$$

where we note that the composite sources do not undergo a Legendre transform.

We first present the BRST Ward identity, which takes the form

$$0 = \int d^3x \, \text{tr}\left(\frac{\delta\Gamma}{\delta A_\alpha}\frac{\delta\Gamma}{\delta S^\alpha} + \frac{\delta\Gamma}{\delta\bar{c}}b + \frac{\delta\Gamma}{\delta c}\frac{\delta\Gamma}{\delta V}\right) \, . \tag{8.75}$$

In a similar way, we find that the Ward identity encoding the supersymmetry is given by

$$0 = \int d^3x \, \text{tr}\left(\epsilon_{\alpha\beta\gamma}\frac{\delta\Gamma}{\delta A_\alpha}\partial^\gamma c + \frac{\delta\Gamma}{\delta\bar{c}}A_\beta - \frac{\delta\Gamma}{\delta b}\frac{\delta\Gamma}{\delta S^\beta} - \epsilon_{\alpha\beta\gamma}S^\alpha\partial^\gamma\frac{\delta\Gamma}{\delta V}\right) \, . \tag{8.76}$$

Upon differentiating (8.76) with respect to A_μ and c, setting the fields to zero and using the transversality of the A field inverse propagator and the b equation of motion, we obtain the following relation:

$$\epsilon_{\mu\nu\lambda}\partial_x^\lambda\frac{\delta^2\Gamma[0]}{\delta A_\nu^a(x)\delta A_\rho^b(y)} = \left(\delta_\mu^\rho - \frac{\partial_\mu^x\partial_x^\rho}{\partial_x^2}\right)\frac{\delta^2\Gamma[0]}{\delta c^a(x)\delta\bar{c}^b(y)} \, . \tag{8.77}$$

This can be rewritten as

$$\frac{\delta^2\Gamma[0]}{\delta A_\mu^a(x)\delta A_\nu^b(y)} = \epsilon^{\mu\nu\lambda}\frac{\partial_\lambda^x}{\partial_x^2}\frac{\delta^2\Gamma[0]}{\partial c^a(x)\delta\bar{c}^b(y)} \, . \tag{8.78}$$

We thus see that the supersymmetry Ward identity fixes the Lorentz tensor structure of the A field inverse propagator. However, it should be emphasized that the supersymmetry discussed here is valid only in flat space. For our considerations to be valid on a curved manifold, we need a parameter ϵ_α which is covariantly constant, i.e., $D_\alpha\epsilon_\beta = 0$. This leads to an integrability constraint which is satisfied when the Riemann curvature vanishes, or has zero modes.

Given such a peculiar symmetry, one is naturally led to ask if it is potentially anomalous. One way to check this is to look for violations of the Ward identity (8.76). However, since Chern–Simons theory is finite, one may wonder how it is possible to generate anomalies. The situation which arises is the following: let us first restrict our attention to $\mathbb{R}^3$, where the supersymmetry is valid. At the classical level, we have a system which is invariant under both gauge and supersymmetry transformations. The question is whether or not we have an effective action at the quantum level which is invariant under both these symmetries. If we can find a regulator which preserves both symmetries, then we can conclude that there are no anomalies. The question remains: does such a regulator exist? The most convenient Ward identity to study in this regard is (8.78), as it is a combined BRST and supersymmetry Ward identity.

It is certainly possible to find regulators which manifestly preserve gauge invariance, see, e.g., refs. [8.12, 8.16, 8.18, 8.20, 8.21]. In all of these cases, the result of a one-loop computation has led to the result that the two-point $\langle AA \rangle$ function receives a correction term which has the effect of shifting the bare coupling by $\text{sign}(k_B)c_v$. Further computations [8.12] have shown that the two-point ghost function remains bare at the one-loop level. Again one must define the renormalized coupling, within this regularization scheme, to be $k_R = k_B + \text{sign}(k_B)c_v$. However, since the original

bare coupling multiplies the entire gauge and ghost action [see (8.68)], there is now a discrepancy between the renormalized couplings multiplying the gauge and ghost parts. As a result, the Ward identity (8.78) is broken.

On the other hand, a different regularization scheme has been used in ref. [8.11], where it is found that all two- and three-point functions remain bare up to two-loop order. In this scheme, the renormalized coupling is the same as the bare coupling. As pointed out in ref. [8.12], this regulator is not gauge invariant; nevertheless, no gauge breaking terms are generated up to two-loop order for the two- and three-point functions. As such, within this regularization scheme both the gauge and supersymmetry invariances are maintained. This appears somewhat strange since, with this regulator, we can establish the fact that there are no gauge or supersymmetry anomalies. However, if this is really the case, then one should also be able to establish this fact with the regularization scheme of ref. [8.12]. But in the latter case it is clear that the supersymmetry is broken (that is, if one decides that the usual freedom to add finite local counterterms is forbidden in this topological situation). Indeed, as remarked above, the supersymmetry in only valid for flat spaces, and hence defining a quantum theory to satisfy both the gauge and supersymmetry Ward identities is not possible for an arbitrary three-manifold. At this point we shall leave the discussion, and invite the reader who has reached thus far in the report to resolve this issue!

Before concluding, however, let us remark on some work in refs. [8.14, 8.23, 8.24] on the usefulness of this particular supersymmetry. In ref. [8.14] the authors establish, with an essential use of the supersymmetry, the finiteness of Chern–Simons theory. It has been shown in ref. [8.23] that, assuming the absence of anomalies, one can use the supersymmetry to prove that the two-point gauge and ghost functions remain bare to all orders of perturbation theory. It is then shown [8.24] that including the anomaly term, as present in the scheme of ref. [8.12], simply adds an extra term to the Ward identity. Whether or not this symmetry is anomalous, its usefulness is clear.

Acknowledgements

Appendix A. The Batalin–Vilkovisky quantization procedure

Here we will briefly sketch the relevant conceptual and computational features of the Batalin–Vilkovisky prescription, for the case of first stage reducible systems and systems with open gauge algebras [A.1]. This will allow us to construct the complete quantum action for the models discussed in sections 3–5. In addition, we explain the construction of the Batalin–Vilkovisky triangles which are necessary to perform the quantization of the reducible (super-)BF systems of sections 5 and 6.

One is first presented with a classical action $S_c(\Phi^i)$, which depends on some fields, generically denoted by Φ^i, together with the local symmetry transformations

$$\delta\Phi^i = R^i_\alpha(\Phi)\epsilon^\alpha \ .$$

(A.1)

Here, ϵ^α denotes the local infinitesimal parameters. We should point out that we are using condensed notation here, in other words, Φ^i and ϵ^α could represent several different fields and transformation parameters; and in addition a repeated index indicates both a sum over discrete labels and an integration over continuous labels. If $\delta\Phi^i = 0$ for some non-zero ϵ^α, then the transformations (A.1) are said to be first stage reducible; one also says that the gauge algebra (A.1) contains zero modes. It is then clear that if one gauge fixed the theory according to Faddeev–Popov, the resulting determinant would have zero modes. This manifests itself as a residual gauge symmetry in the ghost action. The necessity for further gauge fixing is then clear, and this leads to the ghost-for-ghost phenomenon. To correctly incorporate all of these terms it is possible to resort to the Batalin–Vilkovisky machinery, which is guaranteed to produce a BRST invariant quantum action, together with an on-shell nilpotent BRST charge Q. It should be noted that in many cases the above mentioned zero modes are on-shell zero modes, i.e., $\delta\Phi^i = 0$ when the classical equations of motion are used. It is for this reason that the nilpotency of Q is achieved only on-shell.

It may also turn out that the residual gauge symmetry of the ghost action has a zero mode; if this is the case, the theory is said to be second stage reducible, and so on. Examples of this case are provided by the higher dimensional (super-)BF systems which were treated in sections 5 and 6.

The other complication which can arise, and which requires use of this procedure, is when the gauge algebra for (A.1) closes only on-shell. Again an on-shell nilpotent Q is then generated.

Before proceeding, it is useful to make some general remarks about this quantization prescription. Firstly, if the theory is non-Abelian and first stage reducible then one is guaranteed to generate cubic ghost coupling terms; if the theory has an on-shell closed gauge algebra, then quartic ghost interactions are generated. Furthermore, when the zero modes and closure are on-shell, $Q^2 = 0$ only upon using the quantum versions of these equations.

Given this knowledge, it appears somewhat obvious that the topological actions of Witten (which contain cubic and quartic ghost terms) could indeed be obtained as the BRST quantization of a simpler gauge theory.

We now describe the procedure. To each of the parameters ϵ^α we assign a ghost field C^α, of opposite Grassmann character. When the transformations are reducible, the R^i_α which are enumerated by the label α are not linearly independent; in other words there are zero mode eigenvectors, denoted by Z^α_a and enumerated by a, satisfying $R^i_\alpha Z^\alpha_a = 0$. More generally, one may find that $R^i_\alpha Z^\alpha_a = 0$ when the equations of motion are used [see (A.4) below]. Such a situation corresponds to an on-shell reducible theory.

One now introduces Grassmann even ghost fields, denoted by η^a, and defines the minimal set of fields to be $\Phi^A_{\min} = (\Phi^i, C^\alpha, \eta^a)$. The next step is to introduce a set of anti-fields $\Phi^*_{A\,\min} = (\Phi^*_i, C^*_\alpha, \eta^*_a)$, and look for a solution to the "master equation"

$$(S,S) = \frac{\partial_r S}{\partial \Phi^A} \frac{\partial_\ell S}{\partial \Phi^*_A} - \frac{\partial_r S}{\partial \Phi^*_A} \frac{\partial_\ell S}{\partial \Phi^A} = 0 \ .$$

(A.2)

Here ∂_r and ∂_ℓ denote right and left derivatives, respectively.

The solution S exists as an expansion in powers of anti-fields, and the minimal solution is given by

$$S(\varPhi_{\mathrm{min}}, \varPhi^*_{\mathrm{min}}) = S_{\mathrm{min}}$$
$$= S_{\mathrm{c}} + \varPhi^*_i R^i_\alpha C^\alpha + C^*_\alpha (Z^\alpha_a \eta^a + T^\alpha_{\beta\gamma} C^\gamma C^\beta)$$
$$+ \eta^*_a (A^a_{b\alpha} C^\alpha \eta^b) + \varPhi^*_i \varPhi^*_j (B^{ji}_a \eta^a + E^{ji}_{\alpha\beta} C^\beta C^\alpha) + \cdots \ , \tag{A.3}$$

where we have explicitly written only those terms which are needed for the cases of interest in sections 3–5. The coefficients in (A.3) are determined by solving the following equations:

$$R^i_\alpha Z^\alpha_a \eta^a - 2 \frac{\partial_r S_{\mathrm{c}}}{\partial \varPhi^j} B^{ji}_a \eta^a (-1)^{\epsilon_i} = 0 \ , \tag{A.4}$$

$$\frac{\partial_r R^i_\alpha C^\alpha}{\partial \varPhi^j} R^j_\beta C^\beta + R^i_\mu T^\mu_{\alpha\beta} C^\beta C^\alpha - 2 \frac{\partial_r S_{\mathrm{c}}}{\partial \varPhi^j} E^{ji}_{\alpha\beta} C^\beta C^\alpha (-1)^{\epsilon_i} = 0 \ , \tag{A.5}$$

$$\frac{\partial_r Z^\alpha_a \eta^a}{\partial \varPhi^j} R^j_\beta C^\beta + 2 T^\alpha_{\mu\beta} C^\beta Z^\mu_a \eta^a + Z^\alpha_b A^b_{a\beta} C^\beta \eta^a = 0 \ , \tag{A.6}$$

where $\epsilon_i = (0,1)$ denotes the (even, odd) Grassmann parity of the ith field.

In addition to the minimal set of fields, we introduce the anti-ghosts and multipliers $\overline{C}_\alpha, \overline{\eta}_a$ and $\varPi_\alpha, \pi_a$, and consider the solution

$$S = S_{\mathrm{min}} + \overline{C}^{*\alpha} \varPi_\alpha + \overline{\eta}^{*a} \pi_a \ , \tag{A.7}$$

where $\overline{C}^{*\alpha}$ and $\overline{\eta}^{*a}$ are the anti-fields for the anti-ghosts. We should stress that to each field there is a corresponding anti-field; thus, for example, the ghost C_α has an anti-field $C^{*\alpha}$, while the anti-ghost for C_α, denoted by $\overline{C}_\alpha$, is also assigned an anti-field, namely $\overline{C}^{*\alpha}$. The terms anti-ghost and anti-field should not be confused. The gauge fixing is performed by choosing a "gauge-fermion"

$$\varPsi = \overline{C}_\alpha F^\alpha (\varPhi) + \overline{\eta}_a \omega^a_\alpha C^\alpha \ , \tag{A.8}$$

and the complete quantum action then becomes

$$S_{\mathrm{q}} = S(\varPhi, \varPhi^* = \partial \varPsi / \partial \varPhi) \ . \tag{A.9}$$

Equation (A.8) enforces the gauge constraints $F^\alpha (\varPhi) = 0$ and $\omega^a_\alpha C^\alpha = 0$. For algebras with closure problems, we will only need the first constraint, the second being present for first stage reducible theories. One can now check that this action is invariant under the BRST transformations, which are given by

$$\delta \varPhi^A = \epsilon \, \partial_r S / \partial \varPhi^*_A |_{\varPhi^* = \partial \varPsi / \partial \varPhi} \ , \tag{A.10}$$

where ϵ is a constant Grassmann odd parameter. We should remark here that we are following the conventions of ref. [A.1] by introducing ϵ in the transformations (A.10). This ensures that the δ operator commutes with all fields. To translate to the notation used in the main body of the report, we simply write $\delta = -\epsilon \{Q, \ \}$, where $\{Q, \ \}$ now denotes the action of the graded BRST operator Q which (anti-)commutes with Grassmann (odd) even fields.

It is also important to make the following observation: As we have noted in section 3, certain factors of i are required in the definition of the quantum action. These factors are necessary in order to ensure, for example, that integration over a multiplier field will indeed yield a delta function constraint, viz., $\int \mathrm{d}x \, \mathrm{e}^{ipx} = \delta(p)$. It turns out that the Batalin–Vilkovisky algorithm is not sensitive to these factors; however, they are needed in order to properly define the quantum action. In this appendix, and in the other sections of the report, we have chosen to omit these i factors; the reader

should feel free to explicitly include them where necessary. In essence, one simply needs to rescale various multiplier and anti-ghost fields by i.

A.1. Donaldson theory

As our first example, and as a means of gaining some familiarity with the procedure, let us treat the quantization of Donaldson theory from the Labastida–Pernici point of view [A.2]. The classical action is

$$S_c = \frac{1}{2} \int d^4x \, \mathrm{tr}(G_{\alpha\beta} - F^+_{\alpha\beta})^2 \, , \tag{A.11}$$

with the local symmetry transformations

$$\delta A_\mu = D_\mu \epsilon + \epsilon_\mu \, , \qquad \delta G_{\mu\nu} = -[\epsilon, G_{\mu\nu}] + D_{[\mu}\epsilon_{\nu]} + \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}D^\alpha\epsilon^\beta \, . \tag{A.12}$$

Here $\Phi^i = (A^a_\mu, G^a_{\mu\nu})$ and $\epsilon^\alpha = (\epsilon^a, \epsilon^a_\mu)$, and $D_{[\mu}\epsilon_{\nu]} = \frac{1}{2}(D_\mu\epsilon_\nu - D_\nu\epsilon_\mu)$. We shall adopt for convenience the normalization $\mathrm{tr}(T^a T^b) = \delta^{ab}$. Having obtained the final quantum action (A.20) or (A.26), respectively, we are free to introduce a more conventional normalization, such as that employed in sections 6 and 8.

If we now set $\epsilon = -\Lambda$ and $\epsilon_\mu = D_\mu\Lambda$, we see that both variations (A.12) are zero, upon using the G equation of motion. In other words, the transformations (A.12) have an on-shell zero mode. To apply the Batalin–Vilkovisky algorithm, we only need to read off the R and Z coefficients, and solve eqs. (A.4)–(A.6) for T, A, B and E. We have

$$R(A^a_\alpha)_{\epsilon^b} = D^{ab}_\alpha \, , \qquad R(A^a_\alpha)_{\epsilon^b_\beta} = \delta^{ab}\delta_{\alpha\beta} \, , \qquad R(G^a_{\mu\nu})_{\epsilon^b} = -f^{abc}G^c_{\mu\nu} \, ,$$

$$R(G^a_{\mu\nu})_{\epsilon^b_\beta} = \tfrac{1}{2}D^{ab}_\mu\delta_{\beta\nu} - \tfrac{1}{2}D^{ab}_\nu\delta_{\beta\mu} + \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}D^{ab}_\alpha \, , \tag{A.13}$$

and

$$Z(\epsilon^a)_{\Lambda^b} = -\delta^{ab} \, , \qquad Z(\epsilon^a_\mu)_{\Lambda^b} = D^{ab}_\mu \, . \tag{A.14}$$

We point out that we should really write the gauge generators as, for example,

$$R(A^a_\alpha(x))_{\epsilon^b(y)} = D^{ab}_{\alpha x}\delta(x-y) \, , \qquad R(A^a_\alpha(x))_{\epsilon^b_\beta(y)} = \delta^{ab}\delta_{\alpha\beta}\delta(x-y) \, .$$

However, since all quantities are diagonal in the continuous indices, we take the generators as given. Choosing $E = 0$, we find the non-zero coefficients T, A, B,

$$B(G^a_{\rho\sigma}, G^b_{\mu\nu})_{\Lambda^c} = -\tfrac{1}{2}f^{acb}\delta_{\rho\mu}\delta_{\sigma\nu} \, , \qquad T^{\epsilon b}_{\beta\gamma}C^\gamma C^\beta = \tfrac{1}{2}\{c, c\}^b \, , \qquad T^{\epsilon^a_\alpha}_{\beta\gamma}C^\gamma C^\beta = \{\psi_\alpha, c\}^a \, ,$$

$$A^{\Lambda^b}_{\alpha\beta}C^\beta \eta^a = -[c, \phi]^b \, , \tag{A.15}$$

where we have introduced the ghost fields $C^\alpha = (c^a, \psi^a_\alpha)$ and $\eta^a = \phi^a$. On a point of notation: by $\{c, \psi_\alpha\}^a$ we mean $f^{abc}c^b\psi^c_\alpha$; alternatively, we can denote this by the graded (according to ghost number) commutator $[c, \psi_\alpha]^a$, the latter being used in eqs. (5.44) and (5.52) of section 5.2, for example.

The minimal solution now takes the form

$$S_{\min} = S_c + \int d^4x \, \mathrm{tr}\{A^*_\alpha(D^\alpha c + \psi^\alpha) + G^*_{\mu\nu}(-[c, G^{\mu\nu}] + D^{[\mu}\psi^{\nu]} + \tfrac{1}{2}\epsilon^{\mu\nu\alpha\beta}D_\alpha\psi_\beta)$$

$$+ \psi^*_\mu(D^\mu\phi + \{\psi^\mu, c\}) + c^*(-\phi + \tfrac{1}{2}\{c, c\}) - \phi^*[c, \phi] - \tfrac{1}{2}\{G^*_{\mu\nu}, G^{*\mu\nu}\}\phi\} \, . \tag{A.16}$$

We augment this solution with the anti-ghosts and multipliers,

$$S = S_{\min} + \int d^4x \ \mathrm{tr}(\chi^*_{\alpha\beta}B'^{\alpha\beta} - \bar{c}^*b + \bar{\phi}^*\eta) \ ,$$ (A.17)

and take the gauge fermion to be

$$\Psi = -\int d^4x \ \mathrm{tr}(\chi^{\alpha\beta}G_{\alpha\beta} + \bar{c}\partial_\alpha A^\alpha + \bar{\phi}D_\alpha\psi^\alpha) \ .$$ (A.18)

To obtain the quantum action we choose the anti-fields to lie on the gauge surface $\Phi^* = \partial\Psi/\partial\Phi$, and we find

$$A^{*a}_\mu = -\bar{c}^a\partial_\mu - \bar{\phi}^b f^{bac}\psi^c_\mu \ , \qquad G^{*\mu\nu a} = -\chi^{\mu\nu a} \ ,$$

$$\psi^{*\alpha a} = -\bar{\phi}^b D^{ba}_\alpha \ , \qquad \bar{c}^* = -\partial_\alpha A^\alpha \ , \qquad \bar{\phi}^* = -D_\alpha\psi^\alpha \ ,$$

$$\chi^{*\alpha\beta} = -G^{\alpha\beta} \ , \qquad B'^*_{\alpha\beta} = c^* = \eta^* = \phi^* = 0 \ .$$ (A.19)

The quantum action is given by

$$S_q = S_c + \int d^4x \ \mathrm{tr}\{-\bar{c}(\partial \cdot Dc + \partial \cdot \psi) + \bar{\phi}\{\psi^\alpha, D_\alpha c + \psi_\alpha\} - \chi^{\mu\nu}(-[c, G_{\mu\nu}] + 2D_\mu\psi_\nu)$$

$$-\bar{\phi}(D^2\phi + D_\alpha\{\psi^\alpha, c\}) - \tfrac{1}{2}\{\chi^{\mu\nu}, \chi_{\mu\nu}\}\phi - G^{\alpha\beta}B'_{\alpha\beta} + b\partial \cdot A + \eta D \cdot \psi\} \ .$$ (A.20)

The BRST transformations are

$$\delta A_\alpha = -\epsilon(D_\alpha c + \psi_\alpha) \ , \qquad \delta\psi_\alpha = \epsilon(D_\alpha\phi + \{c, \psi_\alpha\}) \ ,$$

$$\delta c = -\epsilon(-\tfrac{1}{2}\{c, c\} + \phi) \ , \qquad \delta\phi = \epsilon[c, \phi] \ , \qquad \delta\chi_{\alpha\beta} = \epsilon B'_{\alpha\beta} \ , \qquad \delta B'_{\alpha\beta} = 0 \ ,$$

$$\delta G_{\mu\nu} = \epsilon([c, G_{\mu\nu}] - D_{[\mu}\psi_{\nu]} - \tfrac{1}{2}\epsilon_{\mu\nu\alpha\beta}D^\alpha\psi^\beta + [\phi, \chi_{\mu\nu}]) \ ,$$

$$\delta\bar{c} = -\epsilon b \ , \qquad \delta b = 0 \ , \qquad \delta\bar{\phi} = -\epsilon\eta \ , \qquad \delta\eta = 0 \ .$$ (A.21)

At this point we have used the Batalin–Vilkovisky algorithm to determine the quantum action together with the BRST transformations. However, recall that these transformations are nilpotent only on-shell; one needs to use the quantum G equation of motion. As we described in section 2, Witten type theories are classified by a quantum action which is a BRST commutator, with respect to an off-shell nilpotent BRST charge. To show that the above system can be written in this way, we simply need to integrate out the G field. We proceed as follows: Consider the following terms in (A.20)

$$\tfrac{1}{2}(G_{\alpha\beta} - F^+_{\alpha\beta})^2 + \{\chi^{\alpha\beta}, c\}G_{\alpha\beta} - G_{\alpha\beta}B'^{\alpha\beta} \ .$$ (A.22)

Define

$$G'_{\alpha\beta} = G_{\alpha\beta} - F^+_{\alpha\beta} \ , \qquad B'_{\alpha\beta} = -B_{\alpha\beta} + \{\chi_{\alpha\beta}, c\} \ .$$ (A.23)

Equation (A.22) can now be written as

$$\tfrac{1}{2}(G'_{\alpha\beta})^2 + B_{\alpha\beta}(G'^{\alpha\beta} + F^{+\alpha\beta}) \ .$$ (A.24)

The G' equation of motion tells us that $G'_{\alpha\beta} = -B_{\alpha\beta}$, and hence (A.24) gives $-\tfrac{1}{2}(B_{\alpha\beta})^2 + B^{\alpha\beta}F^+_{\alpha\beta}$, with the modified BRST transformations

$$\delta\chi_{\alpha\beta} = -\epsilon(B_{\alpha\beta} - \{c, \chi_{\alpha\beta}\}) \ , \qquad \delta B_{\alpha\beta} = \epsilon([c, B_{\alpha\beta}] + [\chi_{\alpha\beta}, \phi]) \ ,$$ (A.25)

where we have used $\delta B = -\delta G' = \delta F^+ - \delta G$. The quantum action can now be written as a BRST commutator,

$$S_q = \int d^4x \ tr\{Q, \chi^{\alpha\beta}(F_{\alpha\beta}^+ - \tfrac{1}{2}\alpha B_{\alpha\beta}) + \bar{c}\partial_\alpha A^\alpha + \bar{\phi}D_\alpha\psi^\alpha\} \ , \tag{A.26}$$

where α is an arbitrary gauge fixing parameter, and Q is now an off-shell nilpotent BRST charge. We can recover the $(F^+)^2$ form of the action in (A.20) by choosing $\alpha = 1$, and integrating over B. Finally, we note that we can redefine the fields so that the transformations are again $\delta\chi = B$ and $\delta B = 0$. In (A.25) both fields transform covariantly with respect to the Yang–Mills gauge symmetry, the $[\chi_{\alpha\beta}, \phi]$ term in the B transformation is then present to achieve nilpotency. Upon writing $\delta = -\epsilon\{Q, \ \}$, we see that the Q obtained here coincides precisely with that used in section 5.

A.2. Supersymmetric quantum mechanics

The classical action is [A.3]

$$S_c = \frac{1}{2} \int d\tau \, g_{ij}(\phi)K^iK^j \ , \quad K^i = G^i - \dot{\phi}^i - g^{il}\partial_l V(\phi) \ . \tag{A.27}$$

The symmetries of the action are

$$\delta\phi^i = \epsilon^i \ , \qquad \delta G^i = \dot{\epsilon}^i + \partial_j(g^{il}\partial_l V)\epsilon^j - \Gamma_{jk}^i K^j \epsilon^k \ . \tag{A.28}$$

The subtlety which arises in the quantization of this system is the fact that the above gauge algebra only closes on-shell (upon using the G equation of motion). As such, we expect to generate quartic ghost coupling terms in the final quantum action, with a BRST charge which is nilpotent when the quantum version of the G equation of motion is used. The R coefficients are

$$R(u^i)_{\epsilon^j} = \delta_j^i \ , \qquad R(G^i)_{\epsilon^j} = (d/d\tau)\delta_j^i + \partial_j(g^{il}\partial_l V) - \Gamma_{jk}^i K^k \ . \tag{A.29}$$

We have only one other non-zero coefficient, namely

$$E(G^i, G^j)_{\epsilon^m\epsilon^l} = \tfrac{1}{2}g^{ir}(-\partial_m\Gamma_{lr}^j - \Gamma_{mn}^j\Gamma_{lr}^n) \ . \tag{A.30}$$

The solution to the master equation is

$$S = S_c + \int d\tau \, \phi_i^*\psi^i + G_i^*[\dot{\psi}^i + \partial_j(g^{ik}\partial_k V)\psi^j - \Gamma_{jk}^i K^j\psi^k]$$
$$+ \tfrac{1}{4}G_i^*G_j^* R^{ij}{}_{ml}\psi^m\psi^l + \bar{\psi}^{*i}B_i' \ , \tag{A.31}$$

and our conventions for the Christoffel connection and Riemann curvature are

$$\Gamma_{jk}^i = \tfrac{1}{2}g^{il}(\partial_j g_{lk} + \partial_k g_{lj} - \partial_l g_{jk}) \ , \qquad R^i{}_{ljk} = \partial_j\Gamma_{lk}^i - \partial_k\Gamma_{lj}^i + \Gamma_{jm}^i\Gamma_{lk}^m - \Gamma_{km}^i\Gamma_{lj}^m \ . \tag{A.32}$$

The gauge fermion is chosen to be

$$\Psi = -\int d\tau \, \bar{\psi}_i G^i \ . \tag{A.33}$$

This leads to the following values of the anti-fields on the constraint surface:

$$\phi_i^* = 0 \ , \qquad G_i^* = -\bar{\psi}_i \ , \qquad \psi_i^* = 0 \ , \qquad \bar{\psi}^{*i} = -G^i \ , \qquad B'^{*i} = 0 \ . \tag{A.34}$$

The quantum action is now

$$S_q = S_c - \int d\tau \{\overline{\psi}_i [\dot{\psi}^i + \partial_j (g^{ik}\partial_k V)\psi^j - \Gamma^i_{jk} K^j \psi^k] - \tfrac{1}{4}\overline{\psi}_i \overline{\psi}_j R^{ij}{}_{ml}\psi^m\psi^l + B'_i G^i\} \ . \tag{A.35}$$

The BRST transformations are

$$\delta\phi^i = -\epsilon\psi^i \ , \qquad \delta\psi^i = 0 \ ,$$
$$\delta G^i = -\epsilon[\dot{\psi}^i + \partial_j (g^{ik}\partial_k V)\psi^j - \Gamma^i_{jk} K^j \psi^k] + \tfrac{1}{2}\epsilon\overline{\psi}_j R^{ij}{}_{ml}\psi^m\psi^l \ ,$$
$$\delta\overline{\psi}_i = \epsilon B'_i \ , \qquad \delta B'_i = 0 \ . \tag{A.36}$$

It is straightforward to check that the corresponding BRST charge is on-shell nilpotent. The off-shell nilpotent BRST charge is constructed as before, and the modified transformations are

$$\delta\overline{\psi}_i = -\epsilon(B_i - \overline{\psi}_j \Gamma^j_{ik}\psi^k) \ ,$$
$$\delta B_i = -\epsilon(\Gamma^j_{ik} B_j \psi^k - \tfrac{1}{2}\overline{\psi}_j R^j{}_{ilk}\psi^l\psi^k) \ , \tag{A.37}$$

where $B_i = -B'_i + \overline{\psi}_j \Gamma^j_{ik}\psi^k$, and the quantum action can be expressed as

$$S_q = \int d\tau \{Q, \overline{\psi}_i (\dot{\phi}^i + g^{il}\partial_l V - \tfrac{1}{2}\alpha B^i)\} \ . \tag{A.38}$$

Writing $\delta = -\epsilon\{Q, \ \}$, we again recover the transformation rules as used in section 3. In order to recover the precise form of the action as used there, we simply need to rescale both B^i and $\overline{\psi}_i$ by a factor of i, as explained at the beginning of this appendix.

A.3. Topological sigma models

The classical action is [A.3]

$$S_c = \int d^2\sigma\, h_{\alpha\beta} g_{ij} K^{\alpha i} K^{\beta j} \ , \qquad K^{\alpha i} = G^{\alpha i} - \tfrac{1}{2}(\partial^\alpha u^i + \epsilon^\alpha{}_\beta J^i{}_j \partial^\beta u^j) \ . \tag{A.39}$$

The symmetries are

$$\delta u^i = \epsilon^i \ ,$$
$$\delta G^{\alpha i} = \tfrac{1}{2}\epsilon^\alpha{}_\beta \epsilon^j (D_j J^i{}_k) K^{\beta k} - \Gamma^i_{jk}\epsilon^j K^{\alpha k} + P^{\alpha i}_+{}_{\beta j}[D^\beta\epsilon^j + \tfrac{1}{2}\epsilon^\beta{}_\gamma \epsilon^l (D_l J^j{}_k)\partial^\gamma u^k]$$
$$+ \tfrac{1}{2}\epsilon^\alpha{}_\beta \epsilon^j (D_j J^i{}_k) P^{\beta k}_+{}_{\gamma l}\partial^\gamma u^l - \Gamma^i_{jk}\epsilon^j P^{\alpha k}_+{}_{\beta l}\partial^\beta u^l \ . \tag{A.40}$$

As for the case of quantum mechanics, the gauge algebra here suffers from closure problems. The R coefficients are

$$R(u^i)_{\epsilon^j} = \delta^i{}_j \ ,$$
$$R(G^{\alpha i})_{\epsilon^j} = \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k) K^{\beta k} - \Gamma^i_{jk} K^{\alpha k} + P^{\alpha i}_+{}_{\beta j} D^\beta + P^{\alpha i}_+{}_{\beta l} \tfrac{1}{2}\epsilon^\beta{}_\gamma (D_j J^l{}_k)\partial^\gamma u^k$$
$$+ \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k) P^{\beta k}_+{}_{\gamma l}\partial^\gamma u^l - \Gamma^i_{jk} P^{\alpha k}_+{}_{\beta l}\partial^\beta u^l \ . \tag{A.41}$$

The only non-zero E coefficient is

$$E(G^{\beta j}, G^{\alpha i})_{\epsilon^r \epsilon^s} = \tfrac{1}{4}g^{jk}[-h^{\alpha\beta}\partial_r \Gamma^i_{sk} + h^{\alpha\beta}\Gamma^i_{ls}\Gamma^l_{rk} - \tfrac{1}{4}h^{\alpha\beta}(D_s J^i{}_l)(D_r J^l{}_k)$$
$$+ \tfrac{1}{2}\epsilon^{\alpha\beta}(\partial_r D_s J^i{}_k) - \tfrac{1}{2}\epsilon^{\alpha\beta}(D_s J^i{}_l)\Gamma^l_{rk} - \tfrac{1}{2}\epsilon^{\alpha\beta}\Gamma^i_{ls}(D_r J^l{}_k)] \ . \tag{A.42}$$

The solution to the master equation is

$$S = S_c + \int d^2\sigma \, G_{\alpha i}^* R(G^{\alpha i})_j C^j + u_i^* R(u^i)_j C^j$$
$$+ G_{\alpha i}^* G_{\beta j}^* E(G^{\beta j}, G^{\alpha i})_{rs} C^s C^r + \overline{C}^{*\alpha i} B_{\alpha i}' \ . \tag{A.43}$$

The gauge fermion is taken to be

$$\Psi = \int d^2\sigma \, \overline{C}_{\alpha i} G^{\alpha i} \ . \tag{A.44}$$

The anti-fields on the gauge surface are then

$$G_{\alpha i}^* = \overline{C}_{\alpha i} \ , \qquad u_i^* = 0 \ , \qquad C_i^* = 0 \ , \qquad \overline{C}^{*\alpha i} = G^{\alpha i} \ , \qquad B'^{*\alpha i} = 0 \ . \tag{A.45}$$

The quantum action is now

$$S_q = S_c + \int d^2\sigma \, \overline{C}_{\alpha i} [D^\alpha C^i + \tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_k)(\partial^\beta u^k) C^j - \Gamma^i_{kj} G^{\alpha k} C^j] + G^{\alpha i} B_{\alpha i}'$$
$$+ \tfrac{1}{8}\overline{C}_\alpha^m \overline{C}^{\alpha k} R_{mkjr} C^j C^r + \tfrac{1}{16}\overline{C}_{\alpha i}\overline{C}^{\alpha k}(D_j J^{li})(D_r J_{lk}) C^j C^r \ . \tag{A.46}$$

A point worth noting here is the fact that the term $\tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^l{}_k) G^{\beta k} C^j$, which is present in (A.41), does not contribute to the action, since it multiplies the self-dual field $\overline{C}^{\alpha i}$, while it itself is anti-self-dual. The BRST transformations are

$$\delta u^i = -\epsilon C^i \ , \qquad \delta C^i = 0 \ ,$$
$$\delta \overline{C}_{\alpha i} = \epsilon [B_{\alpha i}' + \tfrac{1}{2}\epsilon_\alpha{}^\beta (\partial_k J_i{}^j)\overline{C}_{\beta j} C^k] \ , \qquad \delta B_{\alpha i}' = -\epsilon \tfrac{1}{2}\epsilon_\alpha{}^\beta (\partial_k J_i{}^j) B_{\beta j}' C^k \ ,$$
$$\delta G^{\alpha i} = -\epsilon [\tfrac{1}{2}\epsilon^\alpha{}_\beta (D_j J^i{}_l) G^{\beta l} C^j - \Gamma^i_{lj} G^{\alpha l} C^j]$$
$$\qquad\quad -\epsilon \tfrac{1}{2}(\delta^i{}_j \delta^\alpha{}_\beta + J^i{}_j \epsilon^\alpha{}_\beta)[D^\beta C^j + \tfrac{1}{2}\epsilon^\beta{}_\gamma (D_r J^j{}_k)(\partial^\gamma u^k) C^r$$
$$\qquad\quad -\tfrac{1}{4}\overline{C}_m^\beta R^{mj}{}_{sr} C^s C^r - \tfrac{1}{8}\overline{C}_m^\beta (D_s J^{lm})(D_r J_l{}^j) C^s C^r] \ . \tag{A.47}$$

The off-shell nilpotent BRST transformations are given by

$$\delta \overline{C}_{\alpha i} = \epsilon (B_{\alpha i} + \tfrac{1}{2}\epsilon_\alpha{}^\beta (D_k J_i{}^j)\overline{C}_{\beta j} C^k + \Gamma^k_{ij}\overline{C}_{\alpha k} C^j) \ ,$$
$$\delta B^{\alpha i} = \tfrac{1}{4}\epsilon C^k C^l (R_{kl}{}^i{}_t + R_{klrs} J^{ri} J^s{}_t)\overline{C}^{\alpha t} - \tfrac{1}{2}\epsilon\epsilon^\alpha{}_\beta (D_k J^i{}_j) C^k B^{\beta j}$$
$$\qquad\quad + \tfrac{1}{4}\epsilon (C^k D_k J^i{}_s)(C^l D_l J_t{}^s)\overline{C}^{\alpha t} + \epsilon \Gamma^i_{jk} C^j B^{\alpha k} \ , \tag{A.48}$$

where

$$B_{\alpha i} = B_{\alpha i}' - \tfrac{1}{2}(\delta_\alpha{}^\beta \delta_i{}^j + \epsilon^\beta{}_\alpha J^k{}_i)\overline{C}_{\beta l} \Gamma^l_{jk} C^j \ , \tag{A.49}$$
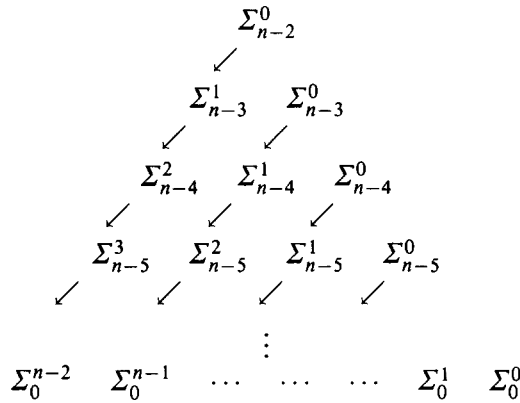
and the quantum action can be expressed as

$$S_q = -\int d^2\sigma \{Q, \overline{C}_{\alpha i}(\partial^\alpha u^i - \tfrac{1}{4}\alpha B^{\alpha i})\} \ . \tag{A.50}$$

A.4. Batalin–Vilkovisky triangles

As our final example of the Batalin–Vilkovisky procedure, we consider the super-BF systems (for $n > 3$) of section 5.4. The quantization of these reducible models is facilitated by the use of the so-called Batalin–Vilkovisky triangles. These serve as a useful bookkeeping device, and allow one to keep track of the proliferating plethora of ghost and multiplier fields required to effect the quantization. For the purposes of illustration, we shall deal solely with the triangle for the B part of the super-BF system. The application to the χ field, and also to similar fields in the BF models of section 6.2, follows by analogy.

Consider the super-BF system with $n > 3$. In n dimensions B is an $(n-2)$-form, and we denote by Σ_i ($i = 0, ..., n-2; \Sigma_{n-2} = B$) collectively B and its hierarchy of ghosts, and ghosts-for-ghosts. The Batalin–Vilkovisky triangle for the Σ system takes the form:

$$
\begin{array}{ccccccccc}
 & & & & \Sigma^0_{n-2} & & & & \\
 & & & \swarrow & & & & & \\
 & & & \Sigma^1_{n-3} & \Sigma^0_{n-3} & & & & \\
 & & \swarrow & & \swarrow & & & & \\
 & & \Sigma^2_{n-4} & \Sigma^1_{n-4} & \Sigma^0_{n-4} & & & & \\
 & \swarrow & & \swarrow & & \swarrow & & & \\
 & \Sigma^3_{n-5} & \Sigma^2_{n-5} & \Sigma^1_{n-5} & \Sigma^0_{n-5} & & & \\
\swarrow & & \swarrow & & \swarrow & & \swarrow & & \\
 & & & & \vdots & & & & \\
\Sigma^{n-2}_0 & \Sigma^{n-1}_0 & \cdots & \cdots & \cdots & & \Sigma^1_0 & \Sigma^0_0
\end{array}
$$

In order to simplify notation we have introduced the collective label Σ^j_i ($i, j = 0, ..., n-2$) to denote all fields in the triangle. The lower index indicates the form rank (with respect to the manifold M) of the field, while the upper index labels the various NW–SE diagonal ledges. An explanation of this structure is as follows:

(i) The horizontal lines contain all the ghosts which arise at each stage of reducibility of the system.

(ii) The right hand ledge ($j = 0$) contains the original reducible gauge field (Σ_{n-2} in this case), together with its ghost and ghost-for-ghosts.

(iii) Given the jth ledge, the range of i is $i = 0, ..., n-2-j$.

(iv) The ghost numbers of the $j = 0$ ledge are specified as ($\Sigma^0_i = (n-2-i)$); the ghost numbers of all the remaining fields (including the multipliers) are now fixed.

(v) The $j = 1$ ledge contains the anti-ghosts for the $j = 0$ ledge; to each arrow connecting the two ledges, there corresponds a gauge fixing condition.

(vi) The fields on the $j = 2$ ledge are called extra ghosts; they are simply the anti-ghosts for the $j = 1$ ledge, and so on.

(vii) To perform the quantization we must introduce a total of $\sum_{j=0}^{n-2}(n-2-j)$ gauge fixing conditions, one for each arrow in the triangle. In this way, one ensures that all the reducible symmetries have been correctly accounted for.

(viii) Each gauge fixing condition requires a multiplier field; one thus has a corresponding multiplier triangle Π^j_i, where $j = 1, ..., n-2$.

Appendix B. Conventions

In this appendix we state our conventions regarding graded forms and present the properties which are necessary for our calculations in sections 5 and 6. We introduce Lie algebra valued differential forms $[\in \Omega^*(M \times \mathcal{A}/\mathcal{G}, \text{ad}\,\mathcal{Q})]$ on $M \times \mathcal{A}/\mathcal{G}$, which carry a natural bigrading, a (p_1, p_2)-form referring to a p_1-form on M and a p_2-form on $\mathcal{A}/\mathcal{G}$. The p_1 label therefore refers to the usual exterior form degree, while the p_2 label is the ghost number of the form. A (p_1, p_2)-form is then called a graded form of degree $p_1 + p_2$ and is denoted by X_p, where $p = p_1 + p_2$. The following are the general properties of graded forms:

$$X_p Y_q = (-1)^{pq} Y_q X_p \ . \tag{B.1}$$

The usual graded commutator is defined as

$$[X_p, Y_q] = X_p Y_q - (-1)^{pq} Y_q X_p \ . \tag{B.2}$$

Thus if p or q is even we obtain the commutator, while if p and q are odd we get the anti-commutator. We remark that we have also used the notation $S_q = \{Q, V\}$ for the graded commutator of Q with V when expressing the quantum action as a BRST commutator. We next note that the usual exterior derivative d and the BRST operator δ are graded derivations, with bigradings $(1, 0)$ and $(0, 1)$, respectively. The standard result for the exterior derivative acting on a product of forms also holds in this case for both derivations d and δ, for example,

$$\delta X_p Y_q = (\delta X_p) Y_q + (-1)^p X_p \delta Y_q \ . \tag{B.3}$$

Given a pure ghost form X_p, i.e., where $p = (0, p)$, together with an arbitrary q-form Y_q, we have the following important result:

$$*(X_p Y_q) = X_p * Y_q \ . \tag{B.4}$$

From this we can derive the properties

$$*(Y_q X_p) = (-1)^{np} (*Y_q) X_p \ , \tag{B.5}$$

$$*[c, Y_q] = [c, *Y_q] \ , \tag{B.6}$$

where $\dim M = n$, c is the $(0, 1)$-ghost form, and $*$ is the Hodge star operator. (B.5) also tells us that the BRST operator commutes with the Hodge star operator,

$$*(\delta Y_q) = \delta * Y_q \ , \tag{B.7}$$

since δ is a pure ghost one-form. Other properties to note are the trace formulae and Jacobi identity,

$$\text{Tr}\, X_p Y_q = (-1)^{pq} \text{Tr}\, Y_q X_p \ , \tag{B.8}$$

$$\text{Tr}\, X_p [Y_q, Z_r] = \text{Tr}\,[X_p, Y_q] Z_r \ , \tag{B.9}$$

$$[X_p, [Y_q, Z_r]] = [[X_p, Y_q], Z_r] + (-1)^{pq} [Y_q, [X_p, Z_r]] \ , \tag{B.10}$$

where X, Y, Z are arbitrary degree forms. The integration by parts formula is

$$\int X_p \mathrm{d} Y_q = (-1)^{(p+1)(q+1)} \int Y_q \mathrm{d} X_p \ . \tag{B.11}$$

If either p or q is odd we get a $+$ sign, while if both p and q are even we get a $-$ sign. Our final result refers to the inner product rule between X_p and Y_q, where $p = (p_1, p_2)$ and $q = (p_1, -p_2)$, defined by

$$\langle X_p, Y_p \rangle = \int_M \mathrm{tr}(X_p * Y_p) \; , \tag{B.12}$$

which satisfies

$$\langle X_p, Y_p \rangle = (-1)^{p_2} \langle Y_p, X_p \rangle \; . \tag{B.13}$$

Given these rules, it is straightforward to translate, for example, the transformation rules of section 5.2.2 to bigraded notation.

References

Section 1

[1.1] A.S. Schwarz, The partition function of a degenerate quadratic functional and the Ray–Singer invariants (Lett. Math. Phys. 2 (1978) 247).

[1.2] E. Witten (Supersymmetry and Morse theory, J. Diff. Geom. 17 (1982) 661).

[1.3] E. Witten, Constraints on supersymmetry breaking, Nucl. Phys. B 202 (1982) 253.

[1.4] E. Witten, Holomorphic Morse inequalities, in: Algebraic and Differential Topology—Global Differential Geometry, ed. G. Rassias, Teubner-Texte zur Mathematik, Band 70 (Teubner, Stuttgart, 1984) p. 318.

[1.5] A. Floer, Witten's complex and infinite dimensional Morse theory, J. Diff. Geom. 30 (1989) 207; An instanton invariant for 3-manifolds, Commun. Math. Phys. 118 (1988) 215.

[1.6] S. Donaldson, An application of gauge theory to four dimensional topology, J. Diff. Geom. 18 (1983) 279; Polynomial invariants of smooth four-manifolds, Topology 29 (1990) 257.

[1.7] M.F. Atiyah, New invariants of three and four dimensional manifolds in: The Mathematical Heritage of Hermann Weyl, eds. R. Wells et al., Proc. Symp. Pure Math. 48 (Am. Math. Soc., Providence, 1988).

[1.8] E. Witten, Topological quantum field theory, Commun. Math. Phys. 117 (1988) 353.

[1.9] V. Jones, A polynomial invariant for links via von Neumann algebras, Bull. Am. Math. Soc. 12 (1985) 103.

[1.10] E. Witten, Quantum field theory and the Jones polynomial, Commun. Math. Phys. 121 (1989) 351.

[1.11] E. Witten, 2 + 1 dimensional gravity as an exactly soluble system, Nucl. Phys. B 311 (1988) 46.

[1.12] E. Witten, Topological sigma models, Commun. Math. Phys. 118 (1988) 411.

[1.13] E. Witten, Topological gravity, Phys. Lett. B 206 (1988) 601.

[1.14] A. Floer, Morse theory for fixed points of symplectic diffeomorphisms, Bull. Am. Math. Soc. 16 (1987) 279.

[1.15] J. Labastida, M. Pernici and E. Witten, Topological gravity in two dimensions, Nucl. Phys. B 310 (1988) 611.

[1.16] E. Verlinde and H. Verlinde, A solution of two dimensional topological quantum gravity, Nucl. Phys. B 348 (1991) 457.

[1.17] E. Witten, On the structure of the topological phase of two dimensional gravity, Nucl. Phys. B 340 (1990) 281.

[1.18] J. Distler, 2D quantum gravity, topological field theory and the multicritical matrix models, Nucl. Phys. B 342 (1990) 523.

[1.19] R. Dijkgraaf, E. Verlinde and H. Verlinde, Notes on topological string theory and 2D quantum gravity, Princeton Univ. preprint, PUPT-1217 (1990).

[1.20] L. Baulieu and I.M. Singer, Topological Yang–Mills symmetry, Nucl. Phys. B (Proc. Suppl.) 5B (1988) 12.

[1.21] J.M.F. Labastida and M. Pernici, A gauge invariant action in topological quantum field theory, Phys. Lett. B 212 (1988) 56.

[1.22] R. Brooks, D. Montano and J. Sonnenschein, Gauge fixing and renormalization in topological field theory, Phys. Lett. B 214 (1988) 91.

[1.23] D. Birmingham, M. Rakowski and G. Thompson, BRST quantization of topological field theories, Nucl. Phys. B 315 (1989) 577.

[1.24] D. Birmingham, M. Blau and G. Thompson, Geometry and quantization of topological gauge theories, Int. J. Mod. Phys. A 5 (1990) 4721.

[1.25] J. Sonnenschein, Topological quantum field theories, moduli spaces, and flat connections, Phys. Rev. D 42 (1990) 2080.

[1.26] S.B. Giddings, Cohomological field theory from field-space cohomology, Nucl. Phys. B 357 (1991) 271.

[1.27] M.F. Atiyah, Topological quantum field theories, Publ. Math. IHES 68 (1988) 175.

[1.28] G. Segal, Conformal field theory, in: Proc. IAMP Congress (Swansea, July 1988).

[1.29] S. Axelrod, Geometric quantization of Chern–Simons gauge theory, Ph. D. thesis, Princeton Univ. (1991), chs. 3 and 4.

Section 2

[2.1] C. Becchi, A. Rouet and R. Stora, Renormalization of the Abelian Higgs–Kibble model, Commun. Math. Phys. 42 (1975) 127; Renormalization of gauge theories, Ann. Phys. (NY) 98 (1976) 287.

[2.2] I.V. Tyupin, Gauge invariance in field theory and in statistical physics in the operator formulation, Lebedev preprint, FIAN No. 39, unpublished.

[2.3] T. Kugo and I. Ojima, Manifestly covariant canonical formulation of Yang–Mills theories physical state subsidiary conditions and physical S-matrix unitarity, Phys. Lett. B 73 (1978) 459.

[2.4] P. Ramond, Field Theory: A Modern Primer, Frontiers in Physics (Benjamin-Cummings, Menlo Park, 1981).

[2.5] J.C. Taylor, Gauge Theory of Weak Interactions (Cambridge Univ. Press, Cambridge, 1976).

[2.6] J. Thierry-Mieg, Geometrical interpretation of Faddeev–Popov ghost particles and BRS transformations, J. Math. Phys. 21 (1980) 2834.

[2.7] L. Bonora and P. Cotta-Ramusino, Some remarks on BRS transformations, anomalies and the cohomology of the Lie algebra of the group of gauge transformations, Commun. Math. Phys. 87 (1983) 589.

[2.8] L. Baulieu, Perturbative gauge theories, Phys. Rep. 129 (1985) 1.

[2.9] M. Daniel and C.M. Viallet, The geometrical setting of gauge theories of the Yang–Mills type, Rev. Mod. Phys. 52 (1980) 175.

[2.10] E. Witten, Topological quantum field theory, Commun. Math. Phys. 117 (1988) 353.

[2.11] E. Witten, Topological sigma models, Commun. Math. Phys. 118 (1988) 411.

[2.12] D. Birmingham, M. Rakowski and G. Thompson, Renormalization of topological field theory, Nucl. Phys. B 329 (1990) 83.

[2.13] A.S. Schwarz, The partition function of degenerate quadratic functional and the Ray–Singer invariants, Lett. Math. Phys. 2 (1978) 247.

[2.14] E. Witten, Quantum field theory and the Jones polynomial, Commun. Math. Phys. 121 (1989) 351.

[2.15] E. Witten, On the structure of the topological phase of two-dimensional gravity, Nucl. Phys. B 340 (1990) 281.

[2.16] E. Witten, Introduction to cohomological field theories, Lectures given at Conf. on Topological Methods in Quantum Field Theory (Trieste, June 1990), Int. J. Mod. Phys. A 6 (1991) 2775.

[2.17] D. Birmingham, M. Blau and G. Thompson, Geometry and quantization of topological gauge theories, Int. J. Mod. Phys. A 5 (1990) 4721.

[2.18] J. Sonnenschein, Topological quantum field theories, moduli spaces, and flat gauge connections, Phys. Rev. D 42 (1990) 2080.

Section 3

[3.1] E. Witten, Constraints on supersymmetry breaking, Nucl. Phys. B 202 (1982) 253.

[3.2] E. Witten, Supersymmetry and Morse theory, J. Diff. Geom. 17 (1982) 661.

[3.3] L. Alvarez-Gaumé, Supersymmetry and the Atiyah–Singer index theorem, Commun. Math. Phys. 90 (1983) 161.

[3.4] L. Alvarez-Gaumé, Supersymmetry and index theory, in: Supersymmetry, NATO Adv. Study Institute, eds. K. Dietz et al. (Plenum, New York, 1985).

[3.5] D. Birmingham, M. Rakowski and G. Thompson, Topological field theories, Nicolai maps and BRST quantization, Phys. Lett. B 214 (1988) 381.

[3.6] D. Birmingham, M. Rakowski and G. Thompson, BRST quantization of topological field theories, Nucl. Phys. B 315 (1989) 577.

[3.7] R. Delbourgo and P.D. Jarvis Extended BRS invariance and Osp(4/2) supersymmetry, J. Phys. 15A (1982) 611.

[3.8] H. Nicolai, On a new characterization of scalar supersymmetric theories, Phys. Lett. B 89 (1980) 341.

[3.9] H. Nicolai, Supersymmetry and functional integration measures, Nucl. Phys. B 176 (1980) 419.

[3.10] H. Nicolai, On the functional integration measure of supersymmetric Yang–Mills theories, Phys. Lett. B 117 (1982) 408.

[3.11] A. Floer, Witten's complex and infinite dimensional Morse theory, J. Diff. Geom. 30 (1989) 207.

[3.12] P. Mansfield, The first Donaldson invariant as the winding number of a Nicolai map, Mod. Phys. Lett. A 3 (1988) 1647.

[3.13] S. Cecotti and L. Girardello, Funtional measure, topology and dynamical supersymmetry breaking, Phys. Lett. B 110 (1982) 39.

[3.14] G. Parisi and Y. Wu, Perturbation theory without gauge fixing, Sci. Sin. 24 (1981) 483.

[3.15] G. Parisi and N. Sourlas, Supersymmetric field theories and stochastic differential equations, Nucl. Phys. B 206 (1982) 321.

[3.16] S. Cecotti and L. Girardello, Stochastic and parastochastic aspects of supersymmetric functional measures: a new non-perturbative approach to supersymmetry, Ann. Phys. (NY) 145 (1983) 81.

[3.17] S. Cecotti and L. Girardello, A supersymmetry anomaly and the fermionic string, Nucl. Phys. B 239 (1984) 573.

[3.18] E. Witten, Topological quantum field theory, Commun. Math. Phys. 117 (1988) 353.

[3.19] E. Witten, Topological sigma models, Commun. Math. Phys. 118 (1988) 411.

[3.20] L. Baulieu and I.M. Singer, Topological Yang–Mills symmetry, Nucl. Phys. B (Proc. Suppl.) 5B (1988) 12.

[3.21] R. Brooks, D. Montano and J. Sonnenschein, Gauge fixing and renormalization in topological quantum field theory, Phys. Lett. B 214 (1988) 91.

[3.22] L. Baulieu and I.M. Singer, The topological sigma model, Commun. Math. Phys. 125 (1989) 227.

[3.23] T. Kugo and I. Ojima, Manifestly covariant formulation of Yang–Mills theories, physical states, subsidiary conditions and physical *S*-matrix unitarity, Phys. Lett. B 73 (1978) 459.

[3.24] E. Witten, Dynamical breaking of supersymmetry, Nucl. Phys. B 188 (1981) 513.

[3.25] J.M.F. Labastida, Morse theory interpretation of topological quantum field theories, Commun. Math. Phys. 123 (1989) 641.

[3.26] D. Friedan and P. Windey, Supersymmetric derivation of the Atiyah–Singer index and the chiral anomaly, Nucl. Phys. B 235 [FS11] (1984) 395.

[3.27] P. Windey, Supersymmetric quantum mechanics and the Atiyah–Singer index theorem, 23rd Cracow School of Theoretical Physics (Zakopane, Poland, 1983).

[3.28] L. Girardello, C. Imbimbo and S. Mukhi, On constant configurations and the evaluation of the Witten index, Phys. Lett. B 132 (1983) 69.

[3.29] D. Quillen, Superconnections and the Chern character, Topology 24 (1985) 89.

[3.30] V. Mathai and D. Quillen, Superconnections, Thom classes and equivariant differential forms, Topology 25 (1986) 85.

[3.31] M.B. Green, J.H. Schwarz and E. Witten, Superstring theory vol.2 (Cambridge Univ. Press, Cambridge, 1987).

[3.32] S. Ouvry and G. Thompson, Topological field theory: zero modes and renormalisation, Nucl. Phys. B 344 (1990) 371.

[3.33] P. Salomonson and J.W. van Holten, Fermionic co-ordinates and supersymmetry in quantum mechanics, Nucl. Phys. B 196 (1982) 509.

[3.34] J.W. van Holten, Instantons in supersymmetric quantum mechanics, Proc. Erice Summer School on Unification of Fundamental Particle Interactions (1981).

[3.35] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Instantons in supersymmetric theories, Nucl. Phys. B 223 (1983) 445.

[3.36] J. Milnor, Morse Theory (Princeton Univ. Press, Princeton, 1973).

[3.37] R. Bott, Morse theory indomitable, Publ. Math. IHES, 68 (1988) 99.

[3.38] J.M.F. Labastida and M. Pernici, A gauge invariant action for topological quantum field theory, Phys. Lett. B 212 (1988) 56.

[3.39] L. Baulieu and B. Grossman, A topological interpretation of stochastic quantisation, Phys. Lett. B 212 (1988) 223.

[3.40] E. Sánchez-Velasco, The Gauss–Bonnet–Chern theorem and supersymmetry, Phys. Lett. B 141 (1984) 57.

[3.41] K. Fujikawa, Dynamical stability of the BRS supersymmetry and the Gribov problem, Nucl. Phys. B 223 (1983) 218.

[3.42] R. Brooks, The spectrum of topological quantum field theories on *N* manifolds, Nucl. Phys. B 325 (1989) 481.

[3.43] A.M. Polyakov, Quark confinement and topology of gauge theories, Nucl. Phys. B 120 (1977) 429.

[3.44] D. Amati and A. Rouet, Renormalisation of Yang–Mills theory developed around an instanton, Nuovo Cimento 50A (1979) 265.

[3.45] O. Babelon, Contributions of the zero frequeny modes in the quantization of solitons, Nucl. Phys. B 131 (1977) 519.

[3.46] L. Baulieu and M. Bellon, BRST symmetry for finite dimensional invariances: applications to global zero modes in string theory, Phys. Lett. B 202 (1988) 67.

[3.47] M. Blau and G. Thompson, Topological gauge theories of antisymmetric tensor fields, Ann. Phys. (NY) 205 (1991) 130.

[3.48] A.J. Niemi and L.C.R. Wijewardhana, Fractionization of the Witten index, Phys. Lett. B 138 (1984) 389.

[3.49] A.J. Niemi, Fermion number fractionisation and the Witten index—a new approach, Phys. Lett. B 146 (1984) 213.

[3.50] A. Iwazaki, Topological $CP^{(n-1)}$ model and topological quantum mechanics, Prog. Theor. Phys. 81 (1989) 1104.

[3.51] J.M. Guilarte, The supersymmetric Yang–Mills vacuum. Topology and physics, Phys. Lett. B 222 (1989) 241.

[3.52] P. van Baal, An introduction to topological Yang—Mills theory, Acta Phys. Pol. B 21 (1990) 73.

Section 4

[4.1] A. Floer, An instanton invariant for three manifolds, Commun. Math. Phys. 118 (1988) 215; Morse theory for fixed points of sympletic diffeomorphisms, Bull. Am. Math. Soc. 16 (1987) 279.

[4.2] M. Gromov, Pseudo holomorphic curves in sympletic manifolds, Invent. Math. 82 (1985) 307; in: Proc. Intern. Congress of Mathematicians (Berkeley, August 1986), p.81.

[4.3] E. Witten, Topological sigma models, Commun. Math. Phys. 118 (1988) 411.

[4.4] S. Kobayashi and K. Nomizu, Foundations of Differential Geometry (Wiley, New York, 1963).

[4.5] P. Griffiths and J. Harris, Principles of Algebraic Geometry (Wiley, New York, 1978).

[4.6] D. Birmingham, M. Rakowski and G. Thompson, Topological field theories, Nicolai maps, and BRST quantization, Phys. Lett. B 214 (1988) 381.

[4.7] D. Birmingham, M. Rakowski and G. Thompson, BRST quantization of topological field theory, Nucl. Phys. B 315 (1989) 577.

[4.8] L. Baulieu and I. Singer, The topological sigma model, Commun. Math. Phys. 125 (1989) 227.

[4.9] J. Honerkamp, Chiral multi-loops, Nucl. Phys. B 36 (1972) 130;
L. Alvarez-Gaumé, D.Z. Freedman and S. Mukhi, The background field method and the ultraviolet structure of the supersymmetric non-linear σ model, Ann. Phys. 134 (1981) 85;
S. Mukhi, The geometric background-field method, renormalization and the Wess–Zumino term in non-linear σ models, Nucl. Phys. B 264 (1986) 640;
D.H. Friedan, Nonlinear models in $2 + \epsilon$ dimensions, Ann. Phys. (NY) 163 (1985) 318.

[4.10] S. Cecotti and L. Girardello, Stochastic and parastochastic aspects of supersymmetric functional measures: a new non-perturbative approach to supersymmetry, Ann. Phys. (NY) 145 (1983) 81.

[4.11] P.G. Drazin and R.S. Johnson, Solitons: An Introduction (Cambridge Univ. Press, Cambridge, 1989).

[4.12] E. Witten, On the structure of the topological phase of two dimensional gravity, Nucl. Phys. B 340 (1990) 281.

[4.13] E. Witten, Supersymmetric form of the nonlinear σ model in two dimensions, Phys. Rev. D 16 (1977) 2991.

[4.14] P. Di Vecchia and S. Ferrara, Classical solutions in two-dimensional supersymmetric field theories, Nucl. Phys. B 130 (1977) 93.

[4.15] A.A. Belavin and A.M. Polyakov, Pis'ma Zh. Eksp. Teor. Fiz. 22 (1975) 503.

[4.16] V.A. Novikov, M.A. Shifman, A.I. Vainshtein and V.I. Zakharov, Instantons and the exact Gell-Mann–Low function of the supersymmetric $O(3)$ sigma model, Phys. Lett. B 199 (1984) 389; Two-dimensional sigma models: modelling non-perturbative effects in quantum chromodynamics, Phys. Rep. 116 (1984) 103.

[4.17] R. Bott and L. Tu, Differential Forms in Algebraic Topology (Springer, New York, 1982).

[4.18] F. Warner, Foundations of Differentiable Manifolds and Lie Groups (Springer, New York, 1983).

[4.19] V. Guillemin and A. Pollack, Differential Topology (Prentice-Hall, Englewood Cliffs, NJ, 1974).

[4.20] J.H. Horne, Superspace versions of topological theories, Nucl. Phys. B 318 (1989) 22.

[4.21] S.P. Sorella, Algebraic characterization of the topological σ model, Phys. Lett. B 228 (1989) 69.

[4.22] D. Montano and J. Sonnenschein, Topological strings, Nucl. Phys. B 313 (1989) 258.

[4.23] Y. Yu and H.-L. Hu, Topological sigma models and conformal invariance, Phys. Rev. D 40 (1989) 1207.

[4.24] R. Dijkgraaf, E. Verlinde and H. Verlinde, Notes on topological string theory and 2D quantum gravity, Princeton Univ. preprint, PUPT-1217 (1990).

[4.25] T. Killingback, Two-dimensional topological gravity, topological matter and intersection theory on moduli space, Phys. Lett. B 255 (1991) 513.

[4.26] T. Eguchi and S.-K. Yang, $N = 2$ superconformal models as topological field theories, Mod. Phys. Lett. A 5 (1990) 1693.

[4.27] K. Ito, Topological phase of $N = 2$ superconformal field theory and topological Landau–Ginzburg field theory, Phys. Lett. B 250 (1990) 91.

[4.28] C. Vafa, Topological Landau–Ginzburg models, Mod. Phys. Lett. A 6 (1991) 337.

Section 5

[5.1] E. Witten, Topological quantum field theory, Commun. Math. Phys. 117 (1988) 353.
[5.2] M. Daniel and C.M. Viallet, The geometrical setting of gauge theories of the Yang–Mills type, Rev. Mod. Phys. 52 (1980) 175.
[5.3] D. Bleecker, Gauge Theory and Variational Principles (Addison-Wesley, London, 1981).
[5.4] S.K. Donaldson, An application of gauge theory to four dimensional topology, J. Diff. Geom. 18 (1983) 279.
[5.5] S.K. Donaldson, Polynomial invariants for smooth four-manifolds, Topology 29 (1990) 257.
[5.6] A. Floer, Witten's complex and infinite dimensional Morse theory, J. Diff. Geom. 30 (1989) 207.
[5.7] A. Floer, An instanton-invariant for 3-manifolds, Commun. Math. Phys. 118 (1988) 215.
[5.8] M.F. Atiyah, N.J. Hitchin and I.M. Singer, Self-duality in four dimensional Riemannian geometry, Proc. R. Soc. London A 362 (1978) 425;
see also A.S. Schwarz, On regular solutions of Euclidean Yang–Mills equations, Phys. Lett. B 67 (1977) 172.
[5.9] D.S. Freed and K.K. Uhlenbeck, Instantons and Four-Manifolds, MSRI publications (Springer, New York, 1984).
[5.10] D.S. Freed, Yang–Mills theory and the topology of 4-dimensional manifolds, Lectures given at the ICTP College on Global Geometric and Topological Methods in Analysis' (ICTP, Trieste, 1988).
[5.11] M.F. Atiyah, New invariants of three and four dimensional manifolds, in: The Mathematical Heritage of Hermann Weyl, Proc. Symp. Pure Math. 48, eds. R. Wells et al. (Am. Math. Soc., Providence, 1988).
[5.12] C.H. Taubes, Self-dual connections on non-self-dual four-manifolds, J. Diff. Geom. 17 (1982) 139.
[5.13] S.K. Donaldson, The geometry of 4-manifolds, in: Proc. Intern. Congress of Mathematicians 1986, ed. A. Gleason (Am. Math. Soc., Berkeley, CA, 1987) p. 43.
[5.14] H.B. Lawson, The Theory of Gauge Fields in Four Dimensions, CBMS Regional Conf. Ser. in Math. (Am. Math. Soc., Providence, RI, 1985).
[5.15] M.H. Freedman, There is no room to spare in four dimensional space, AMS Notices 31 (1984) 3;
R. Stern, Instantons and the topology of 4-manifolds, The Mathematical Intelligencer 5:3 (1983) 39.
[5.16] M.H. Freedman and F. Quinn, Topology of 4-Manifolds (Princeton Univ. Press, Princeton, NJ, 1990).
[5.17] S.K. Donaldson and P.B. Kronheimer, The Geometry of Four-Manifolds, Oxford Mathematical Monographs (Clarendon, Oxford, 1990).
[5.18] R. Kirby and L. Siebenmann, Foundational Essays on Topological Manifolds, Smoothings, and Triangulations, Ann. Math. Studies 88 (Princeton Univ. Press, Princeton, NJ, 1977).
[5.19] M. Kervaire and J. Milnor, Groups of homotopy spheres I, Ann. Math. 77 (1963) 504.
[5.20] E. Witten, Global gravitational anomalies, Commun. Math. Phys. 100 (1985) 197.
[5.21] A.T. Fomenko, Differential Geometry and Topology, Contemporary Soviet Mathematics (Plenum, New York, 1987).
[5.22] D. Husemoller and J. Milnor, Symmetric Bilinear Forms (Springer, New York, 1973).
[5.23] J. Milnor, On simply connected 4-manifolds, in: Proc. Symp. Internacional Topologia Algebraica (Mexico, 1958), p. 122;
J.H.C. Whitehead, On simply connected 4-dimensional polyhedra, Comment. Math. Helv. 22 (1949) 48.
[5.24] M.H. Freedman, The topology of four-dimensional manifolds, J. Diff. Geom. 17 (1982) 357.
[5.25] S.K. Donaldson, Irrationality and the h-cobordism conjecture, J. Diff. Geom. 26 (1987) 141.
[5.26] S. Friedman and J.W. Morgan, On the diffeomorphism type of certain elliptic surfaces I, J. Diff. Geom. 27 (1988) 297.
[5.27] C.H. Taubes, Gauge theory on asymptotically periodic 4-manifolds, J. Diff. Geom. 25 (1987) 363.
[5.28] R.E. Gompf, An infinite set of exotic R^4's, J. Diff. Geom. 21 (1985) 283.
[5.29] F. Hirzebruch, Topological Methods in Algebraic Geometry (Springer, New York, 1966).
[5.30] S.K. Donaldson, The orientation of Yang–Mills moduli space and 4-manifold topology, J. Diff. Geom. 26 (1987) 397.
[5.31] R. Fintushel and R. Stern, SO(3) connections and the topology of four-manifolds, J. Diff. Geom. 20 (1984) 523.
[5.32] M. Furuta, Perturbations of moduli spaces of self-dual connections, J. Fac. Sci. Univ. Tokyo Sect. IA Math. 34 (1987) 275.
[5.33] S.K. Donaldson, Connections, cohomology and the intersection form of 4-manifolds, J. Diff. Geom. 24 (1986) 257.
[5.34] D. Quillen, Determinants of Cauchy–Riemann operators over a Riemann surface, Funct. Anal. Appl. 19 (1986) 31.

[5.35] M.F. Atiyah and I.M. Singer, Dirac operators coupled to vector potentials, Proc. Natl. Acad. Sci. USA 81 (1984) 2597.

[5.36] M.F. Atiyah and I.M. Singer, The index of elliptic operators IV, Ann. Math. 93 (1971) 119.

[5.37] M.F. Atiyah and L. Jeffrey, Topological Lagrangians and cohomology, J. Geom. Phys. 7 (1990) 120.

[5.38] A. Floer, A relative Morse index for the symplectic action, Commun. Pure Appl. Math. XLI (1988) 393.

[5.39] E. Witten, Supersymmetry and Morse theory, J. Diff. Geom. 17 (1982) 661.

[5.40] I.B. Frenkel, H. Garland and G.J. Zuckermann, String theory and semi-infinite cohomology , Proc. Natl. Acad. Sci. USA 83 (1986) 8442.

[5.41] M.F.Atiyah, V.K. Patodi and I.M. Singer, Spectral asymmetry and Riemannian geometry III, Math. Proc. Camb. Philos. Soc. 79 (1976) 71.

[5.42] C.H. Taubes, Casson's invariant and Riemannian geometry, J. Diff. Geom. 31 (1990) 547.

[5.43] P.J. Braam, Floer homology groups for homology three-spheres, Utrecht preprint (1988).

[5.44] P. van Baal, An introduction to topological Yang–Mills theory, Acta Phys. Pol. B 21 (1990) 73.

[5.45] W. Ebeling and C. Okonek, Donaldson invariants, monodromy groups, and singularities, Int. J. Math. 1 (1990) 233.

[5.46] L. Baulieu and I.M. Singer, Topological Yang–Mills symmetry, Nucl. Phys. B (Proc. Suppl.) 5B (1988) 12.

[5.47] R. Brooks, D. Montano and J. Sonnenschein, Gauge fixing and renormalization in topological quantum field theory, Phys. Lett. B 214 (1988) 91.

[5.48] J.M.F. Labastida and M. Pernici, A gauge invariant action in topological quantum field theory, Phys. Lett. B 212 (1988) 56.

[5.49] D. Birmingham, M. Blau and G. Thompson, Geometry and quantization of topological gauge theories, Int. J. Mod. Phys. A 5 (1990) 4721.

[5.50] I.A. Batalin and G.A. Vilkovisky, Quantization of gauge theories with linearly dependent generators, Phys. Rev. D 28 (1983) 2567.

[5.51] S. Ferrara and B. Zumino, Supergauge invariant Yang–Mills theories, Nucl. Phys. B 79 (1974) 413;
A. Salam and J. Strathdee, Super-symmetry and non-Abelian gauges, Phys. Lett. B 51 (1974) 353.

[5.52] J.P. Yamron, Topological actions from twisted supersymmetric theories, Phys. Lett. B 213 (1988) 325.

[5.53] A. Karlhede and M. Roček, Topological quantum field theory and $N = 2$ conformal supergravity, Phys. Lett. B 212 (1988) 51.

[5.54] E. Witten, On the structure of the topological phase of two-dimensional gravity, Nucl. Phys. B 340 (1990) 281.

[5.55] T. Eguchi and S. Yang, $N = 2$ superconformal models as topological field theories, Mod. Phys. Lett. A 5 (1990) 1693.

[5.56] R. Brooks and D. Kastor, Morphisms between supersymmetric and topological quantum field theories, Phys. Lett. B 246 (1990) 99.

[5.57] V. Mathai and D. Quillen, Superconnections, Thom classes and equivariant differential forms, Topology 25 (1986) 85.

[5.58] D. Birmingham, M. Rakowski and G. Thompson, Topological field theories, Nicolai maps and BRST quantization, Phys. Lett. B 212 (1988) 187.

[5.59] P. Mansfield, The first Donaldson invariant as the winding number of a Nicolai map, Mod. Phys. Lett. A 3 (1988) 1647.

[5.60] R. Bott and L.W. Tu, Differential Forms in Algebraic Topology (Springer, New York, 1986).

[5.61] M. Blau and G. Thompson, $N = 2$ topological gauge theory, the Euler characteristic of moduli spaces, and the Casson invariant, NIKHEF preprint (October 1991), to appear.

[5.62] G. 't Hooft, Computation of the quantum effects due to a four dimensional pseudoparticle, Phys. Rev. D 14 (1976) 3432.

[5.63] A. Hietamaki, A. Niemi and O. Tirkkonen, Four-dimensional topological Yang–Mills theory: symmetries, local observables, Wilson loops and confinement, Phys. Lett. B 238 (1990) 291.

[5.64] S. Ouvry, R. Stora and P. van Baal, On the algebraic characterization of Witten's topological Yang–Mills theory, Phys. Lett. B 220 (1989) 159.

[5.65] J.M.F. Labastida, Morse theory interpretation of topological quantum field theories, Commun. Math. Phys. 123 (1989) 641.

[5.66] J.M. Maillet and A. Niemi, Topological Yang–Mills theory is Abelian chiral anomaly in superspace, Phys. Lett. B 223 (1989) 195.

[5.67] R. Myers, Gauge fixing topological Yang–Mills, Int. J. Mod. Phys. A 5 (1990) 1369.

[5.68] J.M. Guilarte, The supersymmetric Yang–Mills vacuum. Topology and physics, Phys. Lett. B 222 (1989) 241.

[5.69] A.N. Kuznetsov, Topological field theory and local supersymmetry, Phys. Lett. B 234 (1990) 65.

[5.70] Y. Igarashi, H. Imai, S. Kitakado and H. So, Hamiltonian BRST quantization of topological field theories, Phys. Lett. B 227 (1989) 239.

[5.71] E. Witten, Introduction to cohomological field theories, Lectures given at Conf. on Topological Methods in Quantum Field Theory (Trieste, 1990), Int. J. Mod. Phys. A 6 (1991) 2775.

[5.72] H. Kanno, Weyl algebra structure and geometrical meaning of BRST transformation in topological quantum field theory, Z. Phys. C 43 (1989) 477.

[5.73] D. Birmingham, M. Rakowski and G. Thompson. BRST quantization of topological field theories, Nucl. Phys. B 315 (1989) 577.

[5.74] J. Sonnenschein, Topological quantum field theories, moduli spaces and flat gauge connections, Phys. Rev. D 42 (1990) 2080.

[5.75] S. Ouvry and G. Thompson, Topological field theory: zero modes and renormalization, Nucl. Phys. B 344 (1990) 371.

[5.76] V.N. Gribov, Quantization of non-Abelian gauge theories, Nucl. Phys. B 139 (1978) 1.

[5.77] I.M. Singer, Some remarks on the Gribov ambiguity, Commun. Math. Phys. 60 (1978) 7.

[5.78] H. Kanno, Observables in topological Yang–Mills theory and the Gribov problem, Lett. Math. Phys. 19 (1990) 249.

[5.79] W. Greub, S. Halperin and R. Vanstone, Connections, Curvature and Cohomology I,II,III (Academic Press, New York, 1972, 1973, 1976).

[5.80] J. Horne, Superspace versions of topological theories, Nucl. Phys. B 318 (1989) 22.

[5.81] M. Blau and G. Thompson, Topological gauge theories of antisymmetric tensor fields, Ann. Phys. (NY) 205 (1991) 130.

[5.82] G.T. Horowitz, Exactly soluble diffeomorphism invariant theories, Commun. Math. Phys. 125 (1989) 417.

[5.83] E. Witten, Topology changing amplitudes in 2 + 1 dimensional gravity, Nucl. Phys. B 323 (1989) 113.

[5.84] W. Goldman, The symplectic nature of fundamental groups of surfaces, Adv. Math. 54 (1984) 200; Topological components of spaces of representations, Invent. Math. 93 (1988) 557.

[5.85] N.J. Hitchin, The self-duality equation on a Riemann surface, Proc. London Math. Soc. (3) 55 (1987) 59; Stable bundles and integrable systems, Duke Math. J. 54 (1987) 91.

[5.86] E. Verlinde and H. Verlinde, A solution of two dimensional topological quantum gravity, Nucl. Phys. B 348 (1991) 457.

[5.87] M.F. Atiyah and R. Bott, The Yang–Mills equations over Riemann surfaces, Philos. Trans. R. Soc. (London) A 308 (1982) 523.

[5.88] D. Montano and J. Sonnenschein, The topology of moduli space and quantum field theory, Nucl. Phys. B 324 (1989) 348.

[5.89] S. Akbulut and J.D. McCarthy, Casson's Invariant for Oriented Homology 3-Spheres—An Exposition, Mathematical Notes 36 (Princeton Univ. Press, Princeton, 1990).

[5.90] S.E. Cappell, R. Lee and E.Y. Miller, A symplectic geometry approach to generalized Casson's invariants of 3-manifolds, Bull. Am. Math. Soc. 22 (1990) 261.

[5.91] L.F. Cugliandolo, G. Lozano and F.A. Schaposnik, Path integral measure for topological field theories, Phys. Lett. B 244 (1990) 249.

[5.92] R.K. Kaul and R. Rajaraman, Path integral quantization of topological field theories, Phys. Lett. B 249 (1990) 433.

[5.93] H.D. Dahmen, S. Marculescu and L. Szymanowski, Quantization of the topological Yang–Mills theory, Phys. Lett. B 252 (1990) 591.

[5.94] S. Deguchi, Superfield formalism of topological Yang–Mills theory, Mod. Phys. Lett. A 4 (1989) 2625.

[5.95] D. Birmingham, M. Rakowski and G. Thompson, Renormalization of topological field theory, Nucl. Phys. B 329 (1990) 83.

[5.96] J.O. Madsen, The beta-function in four-dimensional topological Yang–Mills theory, Mod. Phys. Lett. A 5 (1990) 1013.

[5.97] J.P. Ader, F. Gieres and J.C. Wallet, Topological super Yang–Mills theories, Phys. Lett. B 226 (1989) 302.

[5.98] J.P. Ader and F. Gieres, Topological super Yang–Mills and super Donaldson polynomials, Z. Phys. C 46 (1990) 285.

[5.99] D. Birmingham, M. Rakowski and G. Thompson, Supersymmetric instantons and topological quantum field theory, Phys. Lett. B 212 (1988) 187.

[5.100] L. Baulieu and B. Grossman, Monopoles and topological field theory, Phys. Lett. B 214 (1988) 223.

[5.101] R. Brooks, Topological invariants and a gauge theory of the super-Poincaré algebra in three dimensions, Nucl. Phys. B 320 (1989) 440.

[5.102] G. Chapline and B. Grossman, Conformal field theory and a topological quantum theory of vortices and knots, Phys. Lett. B 223 (1989) 336.
[5.103] F.A. Schaposnik and G. Thompson, Topological field theory and two-dimensional Abelian-Higgs instantons, Phys. Lett. B 224 (1989) 379.
[5.104] L. Cugliandolo, G. Lozano and F.A. Schaposnik, A topological field theory for nonabelian vortices, Phys. Lett. B 234 (1990) 52.
[5.105] L. Baulieu and B. Grossman, A topological interpretation of stochastic quantization, Phys. Lett. B 212 (1988) 351.
[5.106] O.F. Dayi, A general action for topological quantum field theories, Nucl. Phys. B 329 (1990) 667.
[5.107] Y. Yu, Topological Yang-Mills field theory as stochastic quantization of Chern-Simons gauge theory, Phys. Rev. D 41 (1990) 661.
[5.108] L. Baulieu, Stochastic and topological gauge theories, Phys. Lett. B 232 (1989) 479.
[5.109] L. Baulieu, Chern-Simons three-dimensional and Yang-Mills-Higgs two-dimensional systems as four-dimensional topological quantum field theories, Phys. Lett. B 232 (1989) 473.

Section 6

[6.1] A.S. Schwarz, The partition function of degenerate quadratic functional and Ray-Singer invariants, Lett. Math. Phys. 2 (1978) 247; The partition function of a degenerate functional, Commun. Math. Phys. 67 (1979) 1; New topological invariants arising in the theory of quantized fields, Talk at the Baku Intern. Topological Conf., Abstracts (Part II) (1987).
[6.2] E. Witten, Quantum field theory and the Jones polynomial, Commun. Math. Phys. 121 (1989) 351.
[6.3] E. Witten, 2+1 dimensional gravity as an exactly solvable system, Nucl. Phys. B 311 (1988/89) 46.
[6.4] M. Roček and P. van Nieuwenhuizen, $N \geq 2$ supersymmetric Chern-Simons terms as $d = 3$ extended conformal supergravity, Class. Quantum Grav. 3 (1986) 43;
A. Achúcarro and P.K Townsend, A Chern-Simons action for three-dimensional anti-de Sitter supergravity theories, Phys. Lett. B 180 (1986) 89.
[6.5] E. Witten, Global aspects of current algebra, Nucl. Phys. B 223 (1983) 422.
[6.6] S.P. Novikov, The Hamiltonian formalism and a many-valued analogue of Morse theory, Russian Math. Surveys 37:5 (1982) 1 [Usp. Mat. Nauk 37:5 (1982) 3].
[6.7] E. Witten, Non Abelian bosonization in two dimensions, Commun. Math. Phys. 92 (1984) 455.
[6.8] S. Deser, R. Jackiw and S. Templeton, Three dimensional massive gauge theories, Phys. Rev. Lett. 48 (1983) 975; Topologically massive gauge theory, Ann. Phys. (NY) 140 (1984) 372.
[6.9] S. Elitzur, G. Moore, A. Schwimmer and N. Seiberg, Remarks on the canonical quantization of the Chern-Simons-Witten theory, Nucl. Phys. B 326 (1989) 108.
[6.10] G.V. Dunne, R. Jackiw and C.A. Trugenberger, Chern-Simons theory in the Schrödinger representation, Ann. Phys. (NY) 194 (1989) 197.
[6.11] S. Axelrod, S. Della Pietra and E. Witten, Geometric quantization of Chern-Simons gauge theory, J. Diff. Geom. 33 (1991) 787.
[6.12] N.J. Hitchin, Flat connections and geometric quantization, Commun. Math. Phys. 131 (1990) 347.
[6.13] T.R. Ramadas, Chern-Simons gauge theory and projectively flat vector bundles on $\mathcal{M}_g$, Commun. Math. Phys. 128 (1990) 421.
[6.14] T.P. Killingback, Quantum Chern-Simons theory, Phys. Lett. B 219 (1989) 448.
[6.15] J. Marsden and A. Weinstein, Reduction of symplectic manifolds with symmetry, Rep. Math. Phys. 5 (1974) 121.
[6.16] M.S. Narashiman and C.S. Seshadri, Stable and unitary vector bundles on a compact Riemann surface, Ann. Math. 82 (1965) 540.
[6.17] D. Quillen, Determinants of Cauchy-Riemann operators over a Riemann surface, Funct. Anal. Appl. 19 (1986) 31.
[6.18] A.N. Pressley and G. Segal, Loop Groups (Oxford Univ. Press, Oxford, 1986).
[6.19] D. Friedan and S. Shenker, The analytic geometry of two-dimensional conformal field theory, Nucl. Phys. B 281 (1987) 509.
[6.20] G. Segal, Conformal field theory, in: Proc. IAMP Congress (Swansea, 1988).
[6.21] V.G. Knizhnik and A.B. Zamolodchikov, Current algebra and Wess-Zumino model in two dimensions, Nucl. Phys. B 247 (1984) 83.
[6.22] A.N. Redlich, Gauge invariance and parity conservation of three dimensional fermions, Phys. Rev. Lett. 52 (1984) 18; Parity violation and gauge noninvariance of the effective gauge field action in three dimensions, Phys. Rev. D 29 (1984) 2366;

L. Alvarez-Gaumé, S. Della Pietra and G. Moore, Anomalies and odd dimensions, Ann. Phys. 163 (1985) 288;
L. Alvarez-Gaumé and E. Witten, Gravitational anomalies, Nucl. Phys. B 234 (1983) 269;
A.J. Niemi and G.W. Semenoff, Axial-anomaly-induced fermion fractionization and effective gauge theory actions
in odd-dimensional space–times, Phys. Rev. Lett. 51 (1983) 2077.

[6.23] M.F. Atiyah, V. Patodi and I.M. Singer, Spectral asymmetry and Riemannian geometry, Math. Proc. Camb. Philos.
Soc. 77 (1975) 43; 78 (1975) 405; 79 (1976) 71.

[6.24] P.B. Gilkey, Invariance Theory, the Heat Equation and the Atiyah–Singer Index Theorem (Publish or Perish,
Berkeley, 1984).

[6.25] I.M. Singer, Families of Dirac operators with applications to physics, Astérisque (1985) 323.

[6.26] D. Ray and I.M. Singer, R torsion and the Laplacian on Riemannian manifolds, Adv. Math. 7 (1971) 145.

[6.27] D. Ray and I.M. Singer, Analytic torsion for complex manifolds, Ann. Math. 98 (1973) 154.

[6.28] U. Danielsson, Properties of the three-dimensional Chern–Simons partition function, Phys. Lett. B 220 (1989)
137.

[6.29] V.F.R. Jones, Index for subfactors, Invent. Math. 71 (1983) 1; A polynomial invariant for links via von Neumann
algebras, Bull. Am. Math. Soc. 12 (1985) 103; Hecke algebra representations of braid group and link polynomials,
Ann. Math. 126 (1987) 335.

[6.30] P. Freyd, D. Yetter, J. Hoste, W.B.R. Lickorish, K. Millett and A. Ocneanu, A new polynomial invariant of knots
and links, Bull. Am. Math. Soc. 12 (1985) 183.

[6.31] L. Kauffman, State models and the Jones polynomial, Topology 26 (1987) 395; Statistical mechanics and the
Jones polynomial, in: Proc. 1986 Conf. on Artin's Braid Group (Santa Cruz, CA).

[6.32] V.G. Turaev, The Yang–Baxter equation and invariants of links, Invent. Math. 92 (1988) 527.

[6.33] J.H. Przytycki and P. Traczyk, Invariants of links of Conway type, Kobe J. Math. 4 (1988) 115.

[6.34] J. Birman, On the Jones polynomial of closed 3-braids, Invent. Math. 81 (1985) 287.

[6.35] A. Tsuchiya and Y. Kanie, Vertex operators in conformal field theory on P^1 and monodromy representation of
braid groups, in: Conformal Field Theory and Solvable Lattice Models, Adv. Stud. Pure Math. 16 (1988) 297;
Erratum ibid. 19 (1989) 675.

[6.36] M.F. Atiyah, On framing of 3 manifolds, Topology 29 (1990) 1; Topological quantum field theories, Publ. Math.
IHES 68 (1988) 175.

[6.37] G. Moore and N. Seiberg, Polynomial equations for rational conformal field theories, Phys. Lett. B 212 (1988)
454.

[6.38] T. Kohno, Monodromy representations of braid groups and Yang–Baxter equations, Ann. Inst. Fourier, Grenoble
37 (4) (1987) 139.

[6.39] N. Yu. Reshetikhin, Quantized universal enveloping algebras, the Yang–Baxter equation and invariants of links, I
and II, LOMI preprints, E-4-87, E-17-87, Leningrad.

[6.40] L. Smolin, Invariants of links and critical points of the Chern–Simons path integral, Mod. Phys. Lett. A 4 (1989)
1091.

[6.41] P. Cotta-Ramusino, E. Guadagnini, M. Martellini and M. Mintchev, Quantum field theory and link invariants,
Nucl. Phys. B 330 (1990) 557.

[6.42] D. Gepner and E. Witten, String theory on group manifolds, Nucl. Phys. B 278 (1986) 493.

[6.43] E. Witten, Gauge theories and integrable lattice models, Nucl. Phys. B 322 (1989) 629; Gauge theories, vertex
models, and quantum groups, Nucl. Phys. B 330 (1990) 285.

[6.44] A.M. Polyakov, Gauge fields and strings (Harwood, New York, 1987).

[6.45] E. Guadagnini, M. Martellini and M. Mintchev, Perturbative aspects of Chern–Simons field theory, Phys. Lett. B
227 (1989) 111.

[6.46] R. Utiyama, Invariant theoretical interpretation of interaction, Phys. Rev. 101 (1956) 1597;
T.W.B. Kibble, Lorentz invariance and the gravitational field, J. Math. Phys. 2 (1961) 212;
Y.M. Cho, Einstein Lagrangian as the translational Yang–Mills Lagrangian, Phys. Rev. D 14 (1976) 2521;
F.W. Hehl, P. von der Heyde and G.D. Kerlick, General relativity with spin and torsion: foundations and prospects,
Rev. Mod. Phys. 48 (1976) 393;
R.J. Petti, Some aspects of the geometry of first-quantized theories, Gen. Rel. Grav. 7 (1976) 869;
R. Giachetti, R. Ricci and E. Sorace, Tetrad fields and metric tensor in the gauge theory of gravitation, Lett. Math.
Phys. 5 (1981) 85;
K.A. Pilch, Geometrical meaning of the Poincaré gauge group theory, Lett. Math. Phys. 4 (1980) 49;
J. Hennig and J. Nitsch, Gravity as an internal Yang–Mills gauge field theory of the Poincaré group, Gen. Rel.
Grav. 13 (1981) 947;
D.W. Sciama, in: Recent Developments in General Relativity (Festschrift for Infeld) Pergamon, Oxford, 1962;
A. Trautman, The geometry of gauge fields, Czech. J. Phys. B 29 (1979) 107;

D. Ivanenko and G. Sardanashvily, The gauge treatment of gravity, Phys. Rep. 94 (1983) 1;
A. Trautman, Fibre bundles, gauge fields, and gravitation, in: General Relativity and Gravitation, ed. A. Held (Plenum, New York, 1980).

[6.47] S. Deser, R. Jackiw and G. 't Hooft, Three-dimensional Einstein gravity: dynamics of flat space, Ann. Phys. (NY) 152 (1984) 220;
S. Deser and R. Jackiw, Three-dimensional cosmological gravity: dynamics of constant curvature, Ann. Phys. (NY) 153 (1984) 405;
J.R. Gott and M. Alpert, General relativity in a (2+1)-dimensional space–time, Gen. Rel. Grav. 16 (1984) 243;
J. Abbott, S. Giddings and K. Kuchar, Einstein's theory in a three-dimensional space–time, Gen. Rel. Grav. 16 (1984) 751;
T. Banks, W. Fischler and L. Susskind, Quantum cosmology in 2+1 and 3+1 dimensions, Nucl. Phys. B 262 (1985) 159;
E. Martinec, Soluble systems in quantum gravity, Phys. Rev. D 30 (1984) 1198;
G. 't Hooft, Non-perturbative 2 particle scattering amplitudes in 2+1 dimensional quantum gravity, Commun. Math. Phys. 117 (1988) 685;
S. Deser and R. Jackiw, Classical and quantum scattering on a cone, Commun. Math. Phys. 118 (1988) 495;
J.D. Brown, Lower Dimensional Gravity (World Scientific, Singapore, 1988).

[6.48] E. Witten, Topology-changing amplitudes in 2+1 dimensional gravity, Nucl. Phys. B 323 (1989) 113.

[6.49] S. Carlip, Exact quantum scattering in 2+1 dimensional gravity, Nucl. Phys. B 324 (1989) 106;
S.P. Martin, Observables in 2+1 dimensional gravity, Nucl. Phys. B 327 (1989) 178;
S. Carlip and S.P. deAlwis, Wormholes in 2+1 dimensions, Nucl. Phys. B 337 (1990) 681.

[6.50] W. Siegel, Unextended superfields in extended supersymmetry, Nucl. Phys. B 156 (1979) 135.

[6.51] J. Schonfeld, A mass term for three dimensional gauge fields, Nucl. Phys. B 185 (1981) 157.

[6.52] R. Jackiw and S. Templeton, How superrenormalizable theories cure their infrared divergences, Phys. Rev. D 23 (1981) 2291.

[6.53] G. Zuckerman, Action principles and global geometry, in: Proc. 1986 San Diego Summer Workshop, ed. S.-T. Yau.

[6.54] A.M. Polyakov, Fermi–Bose transmutation induced by gauge fields, Mod. Phys. Lett. A 3 (1988) 325.

[6.55] J. Fröhlich, Statistics of fields, the Yang–Baxter equation, and the theory of links and knots, 1987 Cargese Lectures, in: Nonperturbative Quantum Field Theory (Plenum, New York).

[6.56] C.R. Hagen, A new gauge theory without an elementary photon, Ann. Phys. 157 (1984) 342.

[6.57] D. Arovas, R. Schrieffer, F. Wiczek and A. Zee, Statistical mechanics of anyons, Nucl. Phys. B 251 [FS 13] (1985) 117.

[6.58] E. Witten, Quantization of Chern–Simons gauge theory with complex gauge group, Commun. Math. Phys. 137 (1991) 29.

[6.59] I.R. Ramadas, I.M. Singer and J. Weitsman, Some comments on Chern–Simons theory, Commun. Math. Phys. 126 (1989) 409.

[6.60] Y. Igarashi, H. Imai. S. Kitakado, J. Kubo and H. So, BRST quantization of Chern–Simons gauge theory, Mod. Phys. Lett. A 5 (1990) 1663.

[6.61] W. Ogura, Path integral quantization of Chern–Simons gauge theory, Phys. Lett. B 229 (1989) 61.

[6.62] G.V. Dunne and C.A. Trugenberger, Schrödinger representation for Abelian Chern–Simons theories on nontrivial spacetimes, Mod. Phys. Lett. A 4 (1989) 1635.

[6.63] M. Bos and V.P. Nair, Coherent state quantization of Chern–Simons theory, Mod. Phys. Lett. A 5 (1990) 959.

[6.64] H. Murayama, Explicit quantization of the Chern–Simons action, Z. Phys. C 48 (1990) 79.

[6.65] A.P. Polychronakos, Abelian Chern–Simons theories in (2+1) dimensions, Ann. Phys. (NY) 203 (1990) 231.

[6.66] T.P. Killingback, Quantization of SL(2,R) Chern–Simons theory, CERN preprint, CERN-TH-5698-90 (1990).

[6.67] T.P. Killingback, Chern–Simons theory for compact non-semisimple Lie groups, Class. Quantum Grav. 7 (1990) 2179.

[6.68] R. Dijkgraaf and E. Witten, Topological gauge theories and group cohomology, Commun. Math. Phys. 129 (1990) 393.

[6.69] B. Grossman, Topological quantum field theories: relations between knot theory and four manifold theory, talk at Intern. Conf. on Differential Geometric Methods in Theoretical Physics (Tahoe City, CA, 1989).

[6.70] A.P. Polychronakos, On the quantization of the coefficient of the Abelian Chern–Simons term, Phys. Lett. B 241 (1990) 37.

[6.71] L. Baulieu, Chern–Simons three-dimensional and Yang-Mills–Higgs two-dimensional systems as four-dimensional topological quantum field theories, Phys. Lett. B 232 (1989) 473.

[6.72] M. Crescimanno and S.A. Hotes, Monopoles, modular invariance and Chern–Simons field theory, Berkeley preprint, UCB-PTH-90-38 (1990).

[6.73] V.P. Nair and J. Schiff, A Kähler–Chern–Simons theory and quantization of instanton moduli space, Phys. Lett. B 246 (1990) 423.

[6.74] Y. Yu, Topological Yang–Mills field theory as stochastic quantization of Chern–Simons gauge theory, Phys. Rev. D 40 (1989) 1301.

[6.75] C. Nash, Singularly perturbed Chern–Simons theory, J. Math. Phys. 9 (1990) 2258.

[6.76] R. Floreanini, R. Percacci and R. Rajaraman, Four-dimensional current algebra from Chern–Simons theory, Phys. Lett. B 231 (1989) 119.

[6.77] G.V. Dunne and C.A. Trugenberger, Current algebra from Chern–Simons theories, Phys. Lett. B 248 (1990) 311.

[6.78] R. Floreanini and R. Percacci, Higher dimensional Abelian Chern–Simons theory, Phys. Lett. B 224 (1989) 291.

[6.79] E. Guadagnini, M. Martellini and M. Mintchev, Wilson lines in Chern–Simons theory and link invariants, Nucl. Phys. B 330 (1990) 575; Chern–Simons model and new relations between the HOMFLY coefficients, Phys. Lett. B 228 (1989) 489.

[6.80] E. Guadagnini, M. Martellini and M. Mintchev, Chern–Simons holonomies and the appearence of quantum groups, Phys. Lett. B 235 (1990) 275; Braids and quantum group symmetry in Chern–Simons theory, Nucl. Phys. B 336 (1990) 581.

[6.81] E. Guadagnini, The link polynomials of the Chern–Simons theory, Phys. Lett. B 251 (1990) 115; The universal link polynomial, CERN preprint, CERN-TH-5826/90 (1990).

[6.82] K. Yamagishi, M.-L. Ge and Y.-S. Wu, New hierarchies of knot polynomials from topological Chern–Simons gauge theory, Lett. Math. Phys. 19 (1990) 13.

[6.83] M.A. Awada, Quantum geometry of loops and the exact solubility of nonabelian gauge Chern–Simons theory, Commun. Math. Phys. 129 (1990) 329; The exact equivalence of Chern–Simons theory with fermionic string theory, Phys. Lett. B 221 (1989) 21.

[6.84] J. Fröhlich and C. King, The Chern–Simons theory and knot polynomials, Commun. Math. Phys. 126 (1989) 167.

[6.85] S.U. Park and B.H. Cho, The Chern–Simons theory, WZW models and invariant polynomials for links, Korea preprint, PRINT-89-0811 (1989).

[6.86] S.P. Martin, Observables, skein relations and tetrahedra in Chern–Simons gauge theory, Nucl. Phys. B 338 (1990) 244.

[6.87] J.H. Horne, Skein relations and Wilson loops in Chern–Simons gauge theory, Nucl. Phys. B 334 (1990) 669.

[6.88] Y.-S. Wu and K. Yamagishi, Chern–Simons theory and Kauffman polynomials, Int. J. Mod. Phys. A 5 (1990) 1165.

[6.89] S. Mukhi, Skein relations and braiding in topological gauge theory, Tata Inst. preprint, TIFR/TH/89-39 (1989).

[6.90] D. Birmingham and S. Sen, Generalized skein relations from Chern–Simons field theory, J. Phys. A 24 (1991) 1229.

[6.91] B. Broda, A three dimensional covariant approach to monodromy (skein relations) in Chern–Simons theory, Mod. Phys. Lett. A 5 (1990) 2747.

[6.92] T.H. Hansson, A. Karlhede and M. Roček, On Wilson loops in Abelian Chern–Simons theories, Phys. Lett. B 225 (1989) 92.

[6.93] G. Zemba, An interpretation of the Abelian Chern–Simons vacuum holonomy, Int. J. Mod. Phys. A 5 (1990) 559.

[6.94] J.M.F. Labastida, P.M. Llatas and A.V. Ramallo, Knot operators in Chern–Simons gauge theory, Nucl. Phys. B 348 (1991) 651.

[6.95] K. Gawedzki and A. Kupiainen, $SU(2)$ Chern–Simons theory at genus zero, Commun. Math. Phys. 135 (1991) 531.

[6.96] M. Camperi, F. Levstein and G. Zemba, The large N limit of Chern–Simons gauge theory, Phys. Lett. B 247 (1990) 549.

[6.97] J.M.F. Labastida and A.V. Ramallo, Chern–Simons theory and conformal blocks, Phys. Lett. B 228 (1989) 214.

[6.98] A.P. Polychronakos, Abelian Chern–Simons theories and conformal blocks, Florida preprint, UFIFT-89-9 (1989).

[6.99] J.M.F. Labastida and A.V. Ramallo, Operator formalism for Chern–Simons theories, Phys. Lett. B 227 (1989) 92.

[6.100] M. Bos and V. P. Nair, $U(1)$ Chern–Simons theory and $c = 1$ conformal blocks, Phys. Lett. B 223 (1989) 61.

[6.101] P. Murphy and S. Sen, Constraints on two-dimensional conformal field theories from knot invariants, Lett. Math. Phys. 18 (1989) 287; Rational conformal theories from Witten's knot approach: constraints and their analysis, Ann. Phys. (NY) 205 (1991) 173.

[6.102] J. Fröhlich and C. King, Two dimensional conformal field theory and three dimensional topology, J. Mod. Phys. A 4 (1989) 5321.

[6.103] M. Li and M. Yu, Braiding matrices, modular transformation and toplogical field theories in $(2+1)$ dimensions, Commun. Math. Phys. 127 (1990) 195.

[6.104] N. Chair and C.-J. Zhu, Polynomial equations and modular transformations in topological field theory, Phys. Lett. B 247 (1990) 323.

[6.105] P. Degiovanni, Topological field theory and rational conformal field theory, Paris preprint, LPTENS-89/25 (1989).

[6.106] T. Hollowood, A connection between connections: Toda conformal field theories and topological gauge theories, Phys. Lett. B 234 (1990) 57.

[6.107] G. Siopsis, On the quantum group structure in WZW models and Chern–Simons theories, Mod. Phys. Lett. A 6 (1991) 1515.

[6.108] N. Sakai and Y. Tanii, Super Wess–Zumino–Witten models from super Chern–Simons theories, Prog. Theor. Phys. 83 (1990) 968.

[6.109] A. Blasi and R. Collina, Chern–Simons model in the Landau gauge and its connection to the Kač–Moody algebra, Genoa preprint, GEF-TH-14-1990 (1990).

[6.110] E.J. Mlawer, S.G. Naculich, H.A. Riggs and H.J. Schnitzer, Group level duality of WZW fusion coefficients and Chern–Simons link observables, Phys. Lett. B 244 (1990) 235.

[6.111] A. Blasi and R. Collina, Chern–Simons and two-dimensional conformal field theories in a covariant Lagrangian approach, Phys. Lett. B 243 (1990) 99.

[6.112] V.V. Fock and Y.I. Kogan, The generating function for the 2d WZW model correlators and the Sugawara construction from the $(2+1)$ Chern–Simons theory, Mod. Phys. Lett. A 5 (1990) 210.

[6.113] S.G. Naculich, H.A. Riggs and H.J. Schnitzer, Group level duality in WZW models and Chern–Simons theory, Phys. Lett. B 246 (1990) 417.

[6.114] L. Alvarez-Gaumé, C. Gomez and G. Sierra, Duality and quantum groups, Nucl. Phys. B 330 (1990) 347.

[6.115] M. Li, $N = 1$, $N = 2$ supergravities in $(2+1)$- dimensions as Chern–Simons theories, ICTP preprint, IC/88/406 (1988).

[6.116] R. Brooks, Topological invariants and a gauge theory of the super-Poincaré algebra in three dimensions, Nucl. Phys. B 320 (1989) 440.

[6.117] M.P. Blencowe, A consistent interacting massless higher spin theory in $D = 2 + 1$, Class. Quantum Grav. 6 (1989) 443.

[6.118] V. Moncrief, Reduction of the Einstein equations in $2+1$ dimensions to a Hamiltonian system over Teichmüller space, J. Math. Phys. 30 (1989) 2907.

[6.119] F. Yu, $D = 2 + 1$ super ISO(2,1) gauge theory and its equivalence to $D = 2 + 1$ supergravity, Phys. Lett. B 222 (1989) 237.

[6.120] J.H. Horne and E. Witten, Conformal gravity in three dimensions as a gauge theory, Phys. Rev. Lett. 62 (1989) 501.

[6.121] K. Isler and C.A. Trugenberger, Gauge theory of two dimensional gravity, Phys. Rev. Lett. 63 (1989) 834.

[6.122] M. Blau and G. Thompson, Topological gauge theories of antisymmetric tensor fields, Ann. Phys. (NY) 205 (1991) 130.

[6.123] A.H. Chamseddine, Topological gauge theory of gravity in five dimensions and all odd dimensions, Phys. Lett. B 233 (1989) 291; Topological gravity and supergravity in various dimensions, Nucl. Phys. B 346 (1990) 213.

[6.124] A.H. Chamseddine and D. Wyler, Gauge theory of topological gravity in $(1+1)$ dimensions, Phys. Lett. B 228 (1989) 75; Topological gravity in $(1+1)$ dimensions, Nucl. Phys. B 340 (1990) 595.

[6.125] J. Soda, Topological aspects of classical and quantum $(2+1)$-dimensional gravity, Hiroshima preprint, RRK-90-4 (1990).

[6.126] P. de Sousa Gerbert, On spin and (quantum) gravity in $2+1$ dimensions, Nucl. Phys. B 346 (1990) 440.

[6.127] O. Dayi, $d = 3$ Chern–Simons action, supergravity and quantization, ICTP preprint, IC/89/6 (1989).

[6.128] A. Achúcarro and P.K. Townsend, Extended supergravities in $d = (2 + 1)$ as Chern–Simons theories, Phys. Lett. B 229 (1989) 383.

[6.129] A. Ashtekar and J.D. Romano, Chern–Simons and Palatani actions and $(2+1)$ gravity, Phys. Lett. B 229 (1989) 56.

[6.130] J. Soda and Y. Yamanaka, Moduli space of Chern–Simons gravity, Mod. Phys. Lett. A 6 (1991) 303.

[6.131] K. Koehker, F. Mansouri, C. Vaz and L. Witten, Wilson loop observables in $(2+1)$-dimensional Chern–Simons supergravity, Nucl. Phys. B 341 (1990) 167; De Sitter gravity and supergravity in three-dimensions as Chern–Simons theories, Cincinnati preprint, UCTP-105/89 (1990); Some anti De Sitter and De Sitter supergravity Chern–Simons theories in three dimensions, Cincinnati preprint, UCTP-108/89 (1990); Two particle scattering in the Chern–Simons–Witten theory of gravity in $(2+1)$ dimensions, Cincinnati preprint, UCTP-105/90 (1990); Extended supergravity Chern–Simons theories in $(2+1)$ dimensions, Cincinnati preprint, UCTP-315-90 (1990).

[6.132] G.T. Horowitz, Exactly soluble diffeomorphism invariant theories, Commun. Math. Phys. 125 (1989) 417.

[6.133] M. Blau and G. Thompson, A new class of topological field theories and the Ray–Singer torsion, Phys. Lett. B 228 (1989) 64.

[6.134] R. Myers and V. Periwal, New symmetries in topological field theories, Phys. Lett. B 225 (1989) 352.

[6.135] A. Karlhede and M. Roček, Topological quantum field theories in arbitrary dimensions, Phys. Lett. B 224 (1989) 58.

[6.136] G.T. Horowitz and M. Srednicki, A quantum field theoretic description of linking numbers and their generalization, Commun. Math. Phys. 130 (1990) 83.

[6.137] I. Oda and S. Yahikozawa, Linking numbers and variational method, Phys. Lett. B 238 (1990) 272.

[6.138] M. Blau and G. Thompson, Do metric independent classical actions lead to topological field theories ?, Phys. Lett. B 255 (1991) 535.

[6.139] J.C. Wallet, Algebraic set-up for the gauge fixing of BF and super BF systems, Phys. Lett. B 235 (1990) 71.

[6.140] N.J. Hitchin, The self-duality equations on a Riemann surface, Proc. London Math. Soc. 55 (1987) 59.

[6.141] A.M. Polyakov, Quark confinement and topology of gauge theories, Nucl. Phys. B 120 (1977) 429.

[6.142] D. Amati and A. Rouet, Renormalisation of Yang–Mills theory developed around an instanton, Nuovo Cimento A 50 (1979) 265.

[6.143] O. Babelon, Contribution of the zero frequency mode in the quantization of solitons, Nucl. Phys. B 131 (1977) 519.

[6.144] L. Baulieu and M. Bellon, BRST symmetry for finite dimensional invariances: applications to global zero modes in string theory, Phys. Lett. B 202 (1988) 67.

[6.145] S. Ouvry and G. Thompson, Topological field theory: zero modes and renormalisation, Nucl. Phys. B 344 (1990) 371.

[6.146] J. Milnor, Whitehead torsion, Bull. Am. Math. Soc. 72 (1966) 358.

[6.147] J. Cheeger, Analytic torsion and the heat equation, Ann. Math. 109 (1979) 259.

[6.148] W. Müller, Analytic torsion and the *R*-torsion of Riemannian manifolds, Adv. Math. 28 (1978) 233.

[6.149] R. Bott and L.W. Tu, Differential Forms in Algebraic Topology (Springer, New York, 1986).

[6.150] S. Wu, Topological quantum field theories on manifolds with a boundary, Commun. Math. Phys. 136 (1991) 157.

[6.151] B. Broda, Invariants for higher-dimensional links via non-Abelian topological gauge theory, Phys. Lett. B 254 (1991) 111.

[6.152] I. Bengtsson, Yang–Mills theory and general relativity in three and four dimensions, Phys. Lett. B 220 (1989) 51.

[6.153] A. Ashtekar (with invited contributions), New Perspectives in Canonical Gravity (Bibliopolis, Naples, 1988).

[6.154] I. Oda and S. Yahikozawa, Effective actions of 2 + 1 dimensional gravity and BF theory, ICTP preprint, IC/90/44, Trieste (1990).

[6.155] D. Birmingham, H.T. Cho, R. Kantowski and M. Rakowski, Operator phases in BF-type topological field theories, Phys. Lett. B 264 (1991) 324; Operator phases in BF-type topological field theories II, CERN preprint, CERN-TH. 6160/91 (1991).

[6.156] E. Guadagnini, N. Maggiore and S.P. Sorella, Supersymmetric structure of four-dimensional antisymmetric tensor fields, Phys. Lett. B 255 (1991) 65.

[6.157] D. Birmingham, R. Gibbs and S. Mokhtari, The effective action in (2 + 1)-dimensional gravity and generalized BF topological field theory, Phys. Lett. B 263 (1991) 176.

[6.158] E. Witten, Introduction to cohomological field theories, Lectures given at Conf. on Topological Methods in Quantum Field Theory (Trieste, 1990), Int. J. Mod. Phys. A 6 (1991) 2775.

[6.159] M.F. Atiyah and R. Bott, The Yang–Mills equations over Riemann surfaces, Philos. Trans. R. Soc. London A 308 (1982) 523.

[6.160] T. P. Killingback, Quantum Yang–Mills theory on Riemann surfaces and conformal field theory, Phys. Lett. B 223 (1989) 357.

[6.161] B. Broda, Two-dimensional topological Yang–Mills theory, Phys. Lett. B 244 (1990) 444.

[6.162] N.J. Hitchin, Stable bundles and integrable systems, Duke Math. J. 54 (1987) 91.

[6.163] D. Birmingham, M. Rakowski and G. Thompson, BRST quantization of topological field theory, Nucl. Phys. B 315 (1989) 577.

[6.164] N.J. Hitchin, A. Karlhede, U. Lindström and M. Roček, Hyperkähler metrics and supersymmetry, Commun. Math. Phys. 108 (1987) 535.

[6.165] S.K. Donaldson, Twisted harmonic maps and the self-duality equations, Proc. London Math. Soc. 55 (1987) 127.

[6.166] W.M. Goldman, The symplectic nature of fundamental groups of surfaces, Adv. Math. 54 (1984) 200; Topological components of spaces of representations, Invent. Math. 93 (1988) 557; Representations of fundamental groups of surfaces, in: Geometry and Topology, eds. J. Alexander and J. Harer, Lecture Notes in Mathematics 1167 (Springer, Berlin, 1985) p. 95.

[6.167] J. Milnor, On the existence of a connection with curvature zero, Comments Math. Helv. 32 (1958) 215.

[6.168] H. Verlinde, Conformal field theory, two-dimensional quantum gravity and quantization of Teichmüller space, Nucl. Phys. B 337 (1990) 652.

[6.169] T. Fukuyama and K. Kamimura, Gauge theory of two-dimensional gravity, Phys. Lett. B 160 (1985) 259.

[6.170] R. Jackiw, Liouville field theory: a two-dimensional model for gravity?, in: Quantum Theory of Gravity, ed. S. Christensen (Hilger, Bristol, 1984) p. 403;
C. Teitelboim, The Hamiltonian structure of two-dimensional space–time and its relation with the conformal anomaly, in: Quantum Theory of Gravity, ed. S. Christensen (Hilger, Bristol, 1984) p. 327.

[6.171] E. D'Hoker, D.Z. Freedman and R. Jackiw, SO(2,1)-invariant quantization of the Liouville theory, Phys. Rev. D 28 (1983) 2583.

[6.172] E. Verlinde and H. Verlinde. A solution of two-dimensional topological quantum gravity, Nucl. Phys. B 348 (1991) 457.

Section 7

[7.1] A.M. Polyakov, Quantum geometry of bosonic strings, Phys. Lett. B 103 (1981) 207.

[7.2] A.M. Polyakov, Quantum geometry of fermionic strings, Phys. Lett. B 103 (1981) 211.

[7.3] A.M. Polyakov, Quantum gravity in two dimensions, Mod. Phys. Lett. A 2 (1987) 893.

[7.4] V.G. Knizhnik, A.M. Polyakov and A.B. Zamalodchikov, Fractal structure of 2d-quantum gravity, Mod. Phys. Lett. A 3 (1988) 819.

[7.5] J. Distler and H. Kawai, Conformal field theory and 2d quantum gravity, Nucl. Phys. B 321 (1989) 509.

[7.6] F. David, Conformal field theories coupled to 2-d gravity in the conformal gauge, Mod. Phys. Lett. A 3 (1988) 1651.

[7.7] E. D'Hoker, Equivalence of Liouville theory and 2-d quantum gravity, Mod. Phys. Lett. A 6 (1991) 745.

[7.8] E. Brézin and V.A. Kazakov, Exactly solvable field theories of closed strings, Phys. Lett. B 236 (1990) 144.

[7.9] M. Douglas and S. Shenker, Strings in less than one dimension, Nucl. Phys. B 335 (1990) 635.

[7.10] D.J. Gross and A.A. Migdal, Nonperturbative two-dimensional quantum gravity, Phys. Rev. Lett. 64 (1990) 127.

[7.11] J.M.F. Labastida, M. Pernici and E. Witten, Topological gravity in two dimensions, Nucl. Phys. B 310 (1988) 611.

[7.12] D. Montano and J. Sonnenschein, The topology of moduli space and quantum field theory, Nucl. Phys. B 324 (1989) 348.

[7.13] D. Mumford, Towards an enumerative geometry of the moduli space of curves, in: Arithmetic and Geometry, Vol. II (Birkhäuser, Boston, 1983) p. 271.

[7.14] S. Morita, Characteristic classes of surface bundles, Invent. Math. 90 (1987) 551.

[7.15] E. Miller, The homology of the mapping class group, J. Diff. Geom. 24 (1986) 551.

[7.16] R. Myers and V. Periwal, Topological gravity and moduli space, Nucl. Phys. B 333 (1990) 536.

[7.17] E. Witten, On the structure of the topological phase of two dimensional gravity, Nucl. Phys. B 340 (1990) 281.

[7.18] J. Distler, 2-d quantum gravity, topological field theory and the multicritical matrix models, Nucl. Phys. B 342 (1990) 523.

[7.19] R. Dijkgraaf and E. Witten, Mean field theory, topological field theory, and multi-matrix models, Nucl. Phys. B 342 (1990) 486.

[7.20] E. Verlinde and H. Verlinde, A solution of two dimensional topological quantum gravity, Nucl. Phys. B 348 (1991) 457.

[7.21] R. Dijkgraaf, E. Verlinde and H. Verlinde, Loop equations and Virasoro constraints in non-perturbative two-dimensional quantum gravity, Nucl. Phys. B 348 (1991) 435.

[7.22] L. Bonora, M. Martellini and C.M. Viallet, Bosonic strings and complex structures, Phys. Lett. B 197 (1987) 335.

[7.23] R.Dijkgraaf, E. Verlinde and H. Verlinde, Notes on topological string theory and 2d quantum gravity, Institute for Advanced Study preprint, IASSNS-HEP-90/80 (1990).

[7.24] D. Friedan, E. Martinec and S. Shenker, Conformal invariance, supersymmetry and string theory, Nucl. Phys. B 271 (1986) 93.

[7.25] D. Freedman and K. Pilch, Partition functions of the superconformal ghost Thirring model, Int. J. Mod. Phys. A 4 (1989) 5553.

[7.26] K. Li, Action for topological W gravity, Nucl. Phys. B 346 (1990) 329.

[7.27] S. Ouvry, R. Stora and P. van Baal, On the algebraic characterization of Witten's topological Yang–Mills theory, Phys. Lett. B 220 (1989) 159.

[7.28] J.H. Horne, Intersection theory and two dimensional gravity at genus 3 and 4, Mod. Phys. Lett. A 5 (1990) 2127.

[7.29] J. Ambjorn, B. Durhuus and J. Fröhlich, Diseases of triangulated random surface models, and possible cures, Nucl. Phys. B 257 (1985) 433.

[7.30] V. Kasakov, I. Kostov and A. Migdal, Critical properties of randomly triangulated planar random surfaces, Phys. Lett. B 159 (1985) 295.

[7.31] F. David, A model of random surfaces with non-trivial critical behaviour, Nucl. Phys. B 257 (1985) 543.

[7.32] E. Brézin, C. Itzykson, G. Parisi and J.B. Zuber, Planar diagrams, Commun. Math. Phys. 59 (1978) 35.

[7.33] V. Kazakov, Bilocal regularization of models of random surfaces, Phys. Lett. B 150 (1985) 282.

[7.34] F. David, Planar diagrams, two-dimensional lattice gravity and surface models, Nucl. Phys. B 257 [FS14] (1985) 45.

[7.35] F. David, Randomly triangulated surfaces in −2 dimensions, Phys. Lett. B 159 (1985) 303.

[7.36] J. Ambjorn, B. Durhuus, J. Fröhlich and P. Orland, The appearance of critical dimensions in regulated string theories, Nucl. Phys. B 270 [FS16] (1986) 457.

[7.37] D. Boulatov, V. Kazakov, I. Kostov and A. Migdal, Analytical and numerical study of a model of dynamically triangulated random surfaces, Nucl. Phys. B 275 [FS17] (1986) 641.

[7.38] J. Jurkievic, A. Krzywicki and B. Peterson, A numerical study of discrete euclidean Polyakov surfaces, Phys. Lett. B 168 (1986) 273.

[7.39] A. Billoire and F. David, Microcanonical simulations of randomly triangulated planar random surfaces, Phys. Lett 168B (1986) 279.

[7.40] D. Boulatov and V. Kazakov, Universality, critical exponents and mass gap in the $D \to 0$ limit of the model of dynamically triangulated random surfaces, Phys. Lett. B 184 (1987) 247.

[7.41] I. Kostov and M.L. Mehta, Random surfaces of arbitrary genus: exact results for $D = 0$ and -2 dimensions, Phys. Lett. B 189 (1987) 118.

[7.42] V.A. Kazakov, Exactly solvable models of 2d-quantum gravity on the lattice, in: Fields, Strings and Critical Phenomena, eds. E. Brézin and J. Zinn-Justin, Proc. Les Houches Summer School, Vol. 49 (North-Holland, Amsterdam, 1990).

[7.43] D. Gross and A.A. Migdal, A nonperturbative treatment of two-dimensional quantum gravity, Nucl. Phys. B 340 (1990) 333.

[7.44] A. Bilal, 2d gravity from matrix models, CERN preprint, CERN-TH.5867/90 (1990).

[7.45] J. Ambjorn, 2d gravity, random surfaces and all that, Niels Bohr Institute preprint, NBI-HE-90-76 (1990).

[7.46] N. Seiberg, Notes on quantum Liouville theory and quantum gravity, Rutgers preprint, RU-90-29 (1990).

[7.47] P. Di Francesco and D. Kutasov, Integrable models of two-dimensional quantum gravity, Princeton preprint, PUPT-1206 (1990).

[7.48] F. David, Matrix models and 2-d gravity, Saclay preprint, SACLAY-SPHT-90-127 (1990).

[7.49] E. Witten, Topological Gravity, Phys. Lett. B 206 (1988) 601.

[7.50] J.M.F. Labastida and M. Pernici, A Lagrangian for topological gravity and its quantization, Phys. Lett. B 213 (1988) 319.

[7.51] R. Myers and V. Periwal, Invariants of smooth 4-manifolds from topological gravity, Santa Barbara preprint, NSF-ITP-90-4.

[7.52] R. Myers, On alternate formulations of topological gravity, Phys. Lett. B 252 (1990) 365.

[7.53] R. Myers, New observables for topological gravity, Nucl. Phys. B 343 (1990) 705.

[7.54] C.G. Torre, A topological theory of gravitational instantons, Phys. Lett. B 252 (1990) 242.

[7.55] K. Yamagishi and G.F. Chapline, Induced 4D self-dual quantum gravity: $\hat{W}_\infty$ algebraic approach, Class. Quantum Grav. 8 (1991) 427.

[7.56] O. Pekonen, On the variational characterization on conformally flat 3-manifolds, J. Geom. Phys. 7 (1990) 109.

[7.57] K. Li, Topological gravity with minimal matter, Nucl. Phys. B 354 (1991) 711.

[7.58] T. Eguchi and S. Yang, $N = 2$ superconformal models as topological field theories, Mod. Phys. Lett. A 5 (1990) 1693.

[7.59] R. Dijkgraaf, E. Verlinde and H. Verlinde, Topological strings in $d < 1$, Nucl. Phys. B 352 (1991) 59.

[7.60] M. Douglas, Strings in less than one dimension and the generalized KdV hierarchies, Phys. Lett. B 238 (1990) 176.

[7.61] K. Li, Recursion relations in topological gravity with minimal matter, Nucl. Phys. B 354 (1991) 725.

[7.62] K. Aoki, D. Montano and J. Sonnenschein, The role of the contact algebra in multi-matrix models, Caltech preprint, CALT-DOE-68-1676.

[7.63] J. Hughes and D. Montano, Topological gravity on surfaces with boundaries, Phys. Lett. B 250 (1990) 33.

[7.64] L. Baulieu and I.M. Singer, Conformally invariant gauge fixed actions for 2-d topological gravity, Commun. Math. Phys. 135 (1991) 253.

[7.65] E. Witten, Surprises with topological field theories, Institute for Advanced Study preprint, IASSNS-HEP-90/37 (1990).

[7.66] K. Li, Construction of topological W_3 gravity, Phys. Lett. B 251 (1990) 54.

[7.67] C.M. Hull, Gauging the Zamolodchikov W-algebra, Phys. Lett. B 240 (1989) 110.

[7.68] K. Schoutens, A. Sevrin and P. van Nieuwenhuizen, A new gauge theory for W-type algebras, Phys. Lett. B 243 (1990) 245; Covariant formulation of classical W-gravity, Nucl. Phys. B 349 (1991) 791; Covariant w_∞ gravity and its relation to W_N-gravity, Phys. Lett. B 251 (1990) 355.

[7.69] E. Bergshoeff, C.N. Pope, L.J. Romans, E. Sezgin, X. Shen and K.S. Stelle, w_∞ gravity, Phys. Lett. B 243 (1990) 350.

[7.70] E. Bergshoeff, C.N. Pope and K.S. Stelle, Covariant w_∞ gravity, Phys. Lett. B 249 (1990) 208.

[7.71] E. Bergshoeff, A. Bilal and K.S. Stelle, W-symmetries: gauging and geometry, CERN preprint, CERN-TH.5924/90 (1990).

Section 8

[8.1] R. Brooks, D. Montano and J. Sonnenschein, Gauge fixing and renormalization in topological field theory, Phys. Lett. B 214 (1989) 91.

[8.2] D. Birmingham, M. Rakowski and G. Thompson, Renormalization of topological field theory, Nucl. Phys. B 329 (1990) 83.

[8.3] J.H. Horne, Superspace versions of topological theories, Nucl. Phys. B 318 (1989) 22.

[8.4] D. Birmingham, M. Blau and G. Thompson, Geometry and quantization of topological gauge theories, Int. J. Mod. Phys. A 5 (1990) 4721.

[8.5] S. Ouvry and G. Thompson, Topological field theory: zero modes and renormalisation, Nucl. Phys. B 344 (1990) 371.

[8.6] D. Birmingham and M. Rakowski, The β-function in topological sigma models, Mod. Phys. Lett. A 6 (1991) 129.

[8.7] E. Witten, Quantum field theory and the Jones polynomial, Commun. Math. Phys. 121 (1989) 351.

[8.8] A.S. Schwarz, The partiion function of degenerate quadratic functional and the Ray–Singer invariants, Lett. Math. Phys. 2 (1978) 247.

[8.9] E. Witten, Topological quantum field theory, Commun. Math. Phys. 117 (1988) 353.

[8.10] E. Witten, Topological sigma models, Commun. Math. Phys. 118 (1988) 411.

[8.11] E. Guadagnini, M. Martellini and M. Mintchev, Perturbative aspects of the Chern–Simons field theory, Phys. Lett. B 227 (1989) 111.

[8.12] L. Alvarez-Gaumé, J.M.F. Labastida and A.V. Ramallo, A note on perturbative Chern–Simons theory, Nucl. Phys. B 334 (1990) 103.

[8.13] A. Blasi and R. Collina, Finiteness of the Chern–Simons model in perturbation theory, Nucl. Phys. B 345 (1990) 472.

[8.14] F. Delduc, C. Lucchesi, O. Piguet and S.P. Sorella, Exact scale invariance of the Chern–Simons theory in the Landau gauge, Nucl. Phys. B 346 (1990) 313.

[8.15] L. Alvarez-Gaumé, J.M.F. Labastida and A.V. Ramallo, Remarks on topological Chern–Simons theory, Talk presented at 4th Annecy Meeting on Theoretical Physics Commemorating R. Stora's 60th Birthday (Annecy, France, 1990), CERN preprint, CERN-TH-5799/90 (1990).

[8.16] R.D. Pisarski and S. Rao, Topologically massive chromodynamics in the perturbative regime, Phys. Rev. D 22 (1985) 2081.

[8.17] G.P. Korchemsky, Parity anomaly in Chern–Simons theory, Mod. Phys. Lett. A 6 (1991) 727.

[8.18] D. Birmingham, R. Kantowski and M. Rakowski, The eta function in Chern–Simons field theory, Phys. Lett. B 251 (1990) 121.

[8.19] D. Birmingham, H.T. Cho, R. Kantowski and M. Rakowski, Gauge dependence of the eta-function in Chern–Simons field theory and the Vilkovisky-DeWitt correction, Phys. Rev. D 42 (1990) 3476.

[8.20] W. Chen, G.W. Semenoff and Y.-S. Wu, Probing topological features in perturbative Chern–Simons gauge theory, Mod. Phys. Lett. A 5 (1990) 1833.

[8.21] C.P. Martin, Dimensional regularization of Chern–Simons field theory, Phys. Lett. B 241 (1990) 513.

[8.22] M. Asorey and F. Falceto, Chern–Simons theory and geometric regularization, Phys. Lett. B 241 (1990) 31.

[8.23] D. Daniel and N. Dorey, The Schwinger–Dyson equations and non-renormalization in Chern–Simons theory, Phys. Lett. B 246 (1990) 82.

[8.24] N. Dorey, Ward identities for broken supersymmetry on Chern–Simons theory, Phys. Lett. B 246 (1990) 87.

[8.25] Y.-C. Kao, Loop corrections to the Chern–Simons term, Taiwan preprint, PRINT-90-0520 (1990).

[8.26] M.A. Shifman, Four-dimensional aspect of the perturbative renormalization in three-dimensional Chern–Simons theory, BERN preprint, BUTP-90-23-BERN (1990).

[8.27] D. Birmingham and M. Rakowski, Superfield formulation of Chern–Simons supersymmetry, Mod. Phys. Lett. A 4 (1989) 1753.

[8.28] F. Delduc, F. Gieres and S.P. Sorella, Supersymmetry of the $d = 3$ Chern–Simons action in the Landau gauge, Phys. Lett. B 225 (1989) 367.

[8.29] P.H. Damgaard and V.O. Rivelles, Symmetries of Chern–Simons theory in Landau gauge, Phys. Lett. B 245 (1990) 48.

[8.30] A.S. Schwarz, Instantons and fermions in the field of instanton, Commun. Math. Phys. 64 (1979) 233.

[8.31] P.B. Gilkey, The spectral geometry of a Riemannian manifold, J. Diff. Geom. 10 (1975) 601.

[8.32] J.O. Madsen, The beta function in four-dimensional topological Yang–Mills theory, Mod. Phys. Lett. A 5 (1990) 1013.

[8.33] R. Ramond, Field Theory: A Modern Primer, Frontiers in Physics (Benjamin-Cummings, Menlo Park, 1981).

[8.34] G.A. Vilkovisky, The unique effective action in quantum field theory, Nucl. Phys. B 234 (1984) 125; in: Quantum Theory of Gravity: The Gospel According to DeWitt, ed. S.M. Christensen (Hilger, Bristol).

[8.35] B. DeWitt, in: Architecture of Fundamental Interactions at Short Distances, eds. P.Ramond and R. Stora, Proc. Les Houches Summer School (1985), Vol. 44 (North-Holland, Amsterdam, 1987).

[8.36] G. Kunstatter, in: NATO Advanced Research Workshop on Super Field Theories (Simon Fraser Univ., 1986), eds. H.C. Lee, V. Elias, G. Kunstatter, R.B. Mann and K.S. Viswanthan (Plenum, New York, 1987).

[8.37] P.S. Howe, K.S. Stelle and P.C. West, A class of finite four-dimensional supersymmetric field theories, Phys. Lett. B 124 (1983) 55;
M. Grisaru and W. Siegel, Supergravity II: manifestly covariant rules and higher-loop finiteness, Nucl. Phys. B 201 (1982) 292.

[8.38] R. Brooks and D. Kastor, Morphisms between supersymmetric and topological quantum field theories, Phys. Lett. B 246 (1990) 99.

[8.39] K. Fujikawa, Path integral for gauge theories with fermions, Phys. Rev. D 21 (1980) 2848.

[8.40] M.F. Atiyah, V. Patodi and I.M. Singer, Spectral asymmetry and Riemannian geometry, Math. Proc. Camb. Philos. Soc. 77 (1975) 43.

[8.41] I.M. Singer, Families of Dirac operators with applications to physcis, Astérisque (1985) 323.

[8.42] P.B. Gilkey, Invariance Theory, the Heat Equation and the Atiyah–Singer Index Theorem (Publish or Perish, Berkeley, 1984).

[8.43] L. Alvarez-Gaumé, S. Della Pietra and G. Moore, Anomalies and odd dimensions, Ann. Phys. (NY) 163 (1985) 288.

[8.44] D. McKeon and T. Sherry, Operator regularization and one-loop Green's functions, Phys. Rev. D 35 (1987) 3854.

[8.45] J.S. Dowker and R. Critchley, Effective Lagrangian and energy–momentum tensor in de Sitter space, Phys. Rev. D 13 (1976) 3224;
S.W. Hawking, Zeta function regularization of path integrals in curved spacetime, Commun. Math. Phys. 55 (1977) 133.

[8.46] J. Schwinger, On gauge invariance and vacuum polarization, Phys. Rev. 82 (1951) 664.

[8.47] G. Bonneau, Some fundamental but elementary facts on renormalization and regularization: a critical review of the eighties, Int. J. Mod. Phys. A 5 (1990) 3831.

[8.48] E. Witten, Non Abelian bosonization in two dimensions, Commun. Math. Phys. 92 (1984) 455.

[8.49] J.M.F. Labastida and A.V. Ramallo, Santiago de Compostelo Univ. preprint, in preparation.

[8.50] E. Guadagnini, M. Martellini and M. Mintchev, Braids and quantum group symmetry in Chern–Simons theory, Nucl. Phys. B 336 (1990) 581.

[8.51] E. Guadagnini, The link polynomials of the Chern–Simons theory, Phys. Lett. B 251 (1990) 115; The universal link polynomial, CERN preprint, CERN-TH-5826/90 (1990).

Appendix A

[A.1] I.A. Batalin and G.A. Vilkovisky, Quantization of gauge theories with linearly dependent generators, Phys. Rev. D 28 (1983) 2567.

[A.2] J.M.F. Labastida and M. Pernici, A gauge invariant action in topological quantum field theory, Phys. Lett. B 212 (1988) 56.

[A.3] D. Birmingham, M. Rakowski and G. Thompson, BRST quantization of topological field theory, Nucl. Phys. B 315 (1989) 577.