THE
INTERNATIONAL SERIES
OF
MONOGRAPHS ON PHYSICS

SERIES EDITORS

J. BIRMAN  CITY UNIVERSITY OF NEW YORK
S. F. EDWARDS  UNIVERSITY OF CAMBRIDGE
R. FRIEND  UNIVERSITY OF CAMBRIDGE
M. REES  UNIVERSITY OF CAMBRIDGE
D. SHERINGTON  UNIVERSITY OF OXFORD
G. VENEZIANO  CERN, GENEVA
International Series of Monographs on Physics

134. G. Fredrickson: The equilibrium theory of inhomogeneous polymers
133. H. Suhl: Relaxation processes in micromagnetics
132. J. Terning: Modern supersymmetry
131. M. Mariño: Chern–Simons theory, matrix models, and topological strings
130. V. Gantmakher: Electrons and disorder in solids
129. W. Barford: Electronic and optical properties of conjugated polymers
128. R.E. Raab, O.L. de Lange: Multiple theory in electromagnetism
127. A. Larkin, A. Varlamov: Theory of fluctuations in superconductors
126. P. Goldbart, N. Goldenfeld, D. Sherrington: Stealing the gold
125. S. Atzeni, J. Meyer-ter-Vehn: The physics of inertial fusion
124. C. Kiefer: Quantum Gravity
123. T. Fujimoto: Plasma Spectroscopy
122. K. Fujikawa, H. Suzuki: Path integrals and quantum anomalies
121. T. Giamarchi: Quantum physics in one dimension
120. M. Warner, E. Terentjev: Liquid crystal elastomers
119. L. Jacak, P. Sitko, K. Wieczorek, A. Wójs: Quantum Hall systems
118. J. Wesson: Tokamaks Third edition
117. G. Volovik: The universe in a helium droplet
116. L. Pitaevskii, S. Stringari: Bose–Einstein condensation
115. G. Dissertori, I. G. Knowles, M. Schmelling: Quantum chromodynamics
114. B. DeWitt: The global approach to quantum field theory
112. R. M. Mazo: Brownian motion—fluctuations, dynamics, and applications
111. H. Nishimori: Statistical physics of spin glasses and information processing—an introduction
110. N. B. Kopnin: Theory of nonequilibrium superconductivity
108. R. Dobbs: Helium three
107. R. Wigmanns: Calorimetry
106. J. Kübler: Theory of itinerant electron magnetism
105. Y. Kuranoto, Y. Kitaoa: Dynamics of heavy electrons
104. D. Bardin, G. Passarino: The Standard Model in the making
103. G. C. Branco, L. Lavoura, J. P. Silva: CP violation
102. T. C. Choy: Effective medium theory
101. H. Araki: Mathematical theory of quantum fields
100. L. M. Pismen: Vortices in nonlinear fields
99. L. Mestel: Stellar magnetism
98. K. H. Bennemann: Nonlinear optics in metals
97. D. Salzmann: Atomic physics in hot plasmas
96. M. Brambilla: Kinetic theory of plasma waves
95. M. Wakatani: Stellarator and heliotron devices
94. S. Chikazumi: Physics of ferromagnetism
93. R. A. Bertlmann: Anomalies in quantum field theory
92. P. K. Gosh: Ion traps
91. E. Simánek: Inhomogeneous superconductors
90. S. L. Adler: Quaternionic quantum mechanics and quantum fields
89. P. S. Joshi: Global aspects in gravitation and cosmology
88. E. R. Pike, S. Sarkar: The quantum theory of radiation
87. V. Z. Kresin, H. Morawitz, S. A. Wolf: Mechanisms of conventional and high $T_c$ superconductivity
86. V. G. de Gennes, J. Prost: The physics of liquid crystals
85. B. H. Bransden, M. R. C. McDowell: Charge exchange and the theory of ion–atom collision
84. J. Jensen, A. R. Mackintosh: Rare earth magnetism
83. R. Gastmans, T. T. Wu: The ubiquitous photon
82. P. Luchini, H. Motz: Undulators and free-electron lasers
81. P. Weinberger: Electron scattering theory
80. H. Aoki, H. Kamimura: The physics of interacting electrons in disordered systems
79. J. D. Lawson: The physics of charged particle beams
Chern–Simons theory, matrix models and topological strings

MARCOS MARIÑO

Department of Physics, Theory Division, CERN, Geneva 23, CH-1211 Switzerland and Departamento de Matemática, Instituto Superior Técnico, Lisboa, Portugal

CLARENDON PRESS • OXFORD
2005
Para Ara
This page intentionally left blank
Even though string theory has not yet found a clear place in our understanding of Nature, it has already established itself as a source of fascinating results and research directions in mathematics. In recent years, string theory and some of its close cousins (like conformal field theory and topological field theory) have had an enormous impact in representation theory, differential geometry, low-dimensional topology, and algebraic geometry.

One mathematical area that has been deeply influenced by conformal field theory and topological field theory is knot theory. In a groundbreaking work, Witten (1989) found that many topological invariants of knots and links discovered in the 1980s (like the Jones and the HOMFLY polynomials) could be reinterpreted as correlation functions of Wilson loop operators in Chern–Simons theory, a gauge theory in three dimensions with topological invariance. Witten also showed that the partition function of this theory provided a new topological invariant of three-manifolds, and by working out the exact solution of Chern–Simons gauge theory he made a connection between these knot and three-manifold invariants and conformal field theory in two dimensions (in particular, the Wess–Zumino–Witten model).

In a seemingly unrelated development, it was found that the study of string theory on Calabi–Yau manifolds (which was triggered by the phenomenological interest of the resulting four-dimensional models) provided new insights into the geometry of these spaces. Some correlation functions of string theory on Calabi–Yau manifolds turn out to compute numbers of holomorphic maps from the string worldsheet to the target, and therefore they contain information about the enumerative geometry of the Calabi–Yau spaces. This led to the introduction of Gromov–Witten invariants in mathematics as a way to capture this information. Moreover, the existence of a powerful duality symmetry of string theory in Calabi–Yau spaces – mirror symmetry – allowed the computation of generating functions for these invariants, and made it possible to solve with physical techniques difficult enumerative problems (see Hori et al., 2003, for a review of these developments). The existence of a topological sector in string theory that captured the enumerative geometry of the target space also led to the construction of simplified models of string theory that kept only the topological information of the more complicated, physical theory. These models are called topological string theories and turn out to provide in many cases exactly solvable models of string dynamics.

The key idea that allowed the building of a bridge between topological string theory and Chern–Simons theory was the gauge theory/string theory correspondence. It is an old idea, going back to ’t Hooft (1974), that gauge theories can be described in the 1/N expansion by string theories. This idea has been difficult
to implement, but in recent years some spectacular progress was made thanks to the work of Maldacena (1998), who found a duality between type IIB string theory on $\text{AdS}_5 \times \text{S}^5$ and $\mathcal{N} = 4$ super Yang–Mills with gauge group $U(N)$. It is then natural to ask if gauge theories that are simpler than $\mathcal{N} = 4$ Yang–Mills – like for example Chern–Simons theory – also admit a string theory description. It was shown by Gopakumar and Vafa (1999) that Chern–Simons gauge theory on the three-sphere in fact has a closed string description in terms of topological string theory propagating on a particular Calabi–Yau target, the so-called resolved conifold.

The result of Gopakumar and Vafa has three important consequences. First, it provides a toy model of the gauge theory/string theory correspondence that makes it possible to test in detail general ideas about this duality. Secondly, it gives a stringy interpretation of invariants of knots in the three-sphere. More precisely, it establishes a relation between invariants of knots based on quantum groups and Gromov–Witten invariants of open strings propagating on the resolved conifold. These are a priori two very different mathematical objects, and in this way the physical idea of a correspondence between gauge theories and strings gives new and fascinating results in mathematics that we are only starting to unveil. Finally, one can use the results of Gopakumar and Vafa to completely solve topological string theory on certain Calabi–Yau threefolds in a closed form. As we will see, this gives the all-genus answer for certain string amplitudes, and it is in fact one of the few examples in string theory where such an answer is available. The all-genus solution to the amplitudes also encodes the information about all the Gromov–Witten invariants for those threefolds. Since the solution involves building blocks from Chern–Simons theory, it suggests yet another bridge between knot invariants and Gromov–Witten theory.

In this book, we will describe examples of string theory/gauge theory dualities involving topological strings. We will explain in detail the implications of the correspondence between Chern–Simons theory and topological strings, and we will also discuss a correspondence found by Dijkgraaf and Vafa that relates topological strings to matrix models. As we will see, the underlying logic in both correspondences is very similar: first, one shows that the gauge theory in question – Chern–Simons theory or the matrix model – is equivalent to an open string theory. This open string theory turns out to be related to a closed string theory after a geometric transition that changes the geometric background. In the end, one finds a closed string theory description of the gauge theories in the spirit of 't Hooft.

The organization of this book is as follows. The first part is devoted to a presentation of the gauge theories that we will consider and their $1/N$ expansions. The first chapter is on matrix models, while the second chapter is on Chern–Simons theory. In the second part of the book we turn our attention to topological string theory. In Chapter 3 we introduce topological sigma models in some detail, and then we discuss topological string theory in Chapter 4. Chapter 5 studies non-compact Calabi–Yau manifolds, which are the geometric backgrounds that
we will consider in this book. The third and final part presents the relations
between the gauge theories discussed in the first part and the topological string
theories discussed in the second part. In Chapter 6 we outline the general strategy
to establish such a relation. In Chapter 7 we show, using string field theory,
that Chern–Simons theory and matrix models can be realized as topological
open string theories. Chapter 8 studies the relevant geometric transitions relating
open and closed string backgrounds, and presents the closed string duals to the
gauge theories of the first part. In Chapter 9 we define and study the topological
vertex, an object that allows one to solve topological string theory on a wide
class of non-compact Calabi–Yau threefolds by purely combinatorial methods.
Finally, in Chapter 10 we present some applications of these developments to
knot theory and to $\mathcal{N} = 2$ supersymmetric gauge theory. A short Appendix
contains some elementary facts about the theory of symmetric polynomials that
are used throughout the book.

This book grew out of graduate level courses on topological strings and
Chern–Simons theory that I gave in 2003 and 2004, and that are partially writ-
ten down in Mariño, 2004$b$ and 2004$c$. I have tried to keep the spirit of these
courses by including some exercises, and I have avoided the temptation to give
an encyclopedic presentation of all the topics. Unfortunately, there are many
important and related issues that are not analysed in detail in this book. For
example, I do not discuss mirror symmetry, and I do not address the relation
between matrix models and $\mathcal{N} = 1$ gauge theories. These topics are covered in
detail in the book by Hori et al. (2003) and in the review paper by Argurio et
al. (2004), respectively. Finally, some of the material covered in this book is also
presented in the review articles by Grassi and Rossi (2003) and Neitzke and Vafa
(2004).
ACKNOWLEDGEMENTS

In the long process that led to this book, I have benefitted from the insights of many people. I would like to thank Bobby Acharya, Luis Álvarez-Gaumé, Emanuel Diaconescu, Jaume Gomis, Rajesh Gopakumar, Antonella Grassi, Amer Iqbal, Volodya Kazakov, Ivan Kostov, Gregory Moore, Andrew Neitzke, Nikita Nekrasov, Boris Pioline, Pichai Ramadevi, Justin Sawon, Miguel Tierz, George Thompson and Niclas Wyllard for conversations, comments, and inspiration. Some colleagues and friends read preliminary versions of this book and made valuable comments. Thanks are especially due to Arthur Greenspoon, whose suggestions were crucial to improve the text.

I also want to thank my collaborators on the topics discussed here: Mina Aganagic, Vincent Bouchard, Robbert Dijkgraaf, Bodgan Florea, Albrecht Klemm and Stefan Theisen. I learnt most of these topics thanks to our common work and their insights.

Finally, I want to give special thanks to José Labastida and Cumrun Vafa. José Labastida first introduced me to many of the topics covered in this book, when I was a graduate student, and he has been a constant source of knowledge and support since then. Cumrun Vafa has been at the origin of the deep ideas that I discuss here, and I was fortunate enough to work with him and learn from him. I hope this book will be able to capture some of his insights on topological string theory.
# CONTENTS

I MATRIX MODELS, CHERN–SIMONS THEORY, AND THE LARGE N EXPANSION

1 Matrix models 3
  1.1 Basics of matrix models 3
  1.2 Matrix model technology I: saddle-point analysis 11
  1.3 Matrix model technology II: orthogonal polynomials 18

2 Chern–Simons theory and knot invariants 25
  2.1 Chern–Simons theory: basic ingredients 25
  2.2 Perturbative approach 29
    2.2.1 Perturbative structure of the partition function 29
    2.2.2 Group factors 33
  2.3 Canonical quantization and surgery 36
  2.4 Framing dependence 44
  2.5 Results on Wilson loops and knot invariants 46
    2.5.1 General properties 46
    2.5.2 Torus knots 49
  2.6 $U(\infty)$ representation theory 51
  2.7 The $1/N$ expansion in Chern–Simons theory 59
  2.8 Chern–Simons theory as a matrix model 62

II TOPOLOGICAL STRINGS

3 Topological sigma models 71
  3.1 The $\mathcal{N} = 2$ supersymmetric sigma model 71
  3.2 Topological twist 73
  3.3 The topological type-A model 78
  3.4 The topological type-B model 83

4 Topological strings 87
  4.1 Coupling to gravity 87
  4.2 Relation to compactifications of type II string theory 88
  4.3 The type-A topological string 89
    4.3.1 Mathematical description 89
    4.3.2 Integrality properties and Gopakumar–Vafa invariants 92
  4.4 Open topological strings 97
    4.4.1 Type A model 97
    4.4.2 Type-B model 105
5 Calabi–Yau geometries 107
  5.1 Non-compact Calabi–Yau geometries: an introduction 107
  5.2 Constructing toric Calabi–Yau manifolds 109
    5.2.1 \( \mathbb{C}^3 \) 109
    5.2.2 The general case 110
  5.3 The conifold transition 115
  5.4 Examples of closed string amplitudes 118

III THE TOPOLOGICAL STRING/GAUGE THEORY CORRESPONDENCE

6 String theory and gauge theory 123

7 String field theory and gauge theories 127
  7.1 Open string field theory 127
  7.2 Chern–Simons theory as an open string theory 128
    7.2.1 Open topological strings on \( T^*M \) 128
    7.2.2 More general Calabi–Yau manifolds 132
  7.3 Matrix model as an open string theory 133
    7.3.1 Holomorphic Chern–Simons theory 133
    7.3.2 Type B topological strings and matrix models 134
    7.3.3 Open string amplitudes and multicut solutions 138

8 Geometric transitions 143
  8.1 The conifold transition and the large \( N \) duality 143
  8.2 Incorporating Wilson loops 144
  8.3 Geometric transitions for more general toric manifolds 148
  8.4 Matrix models and geometric transitions 153

9 The topological vertex 157
  9.1 Framing of topological open string amplitudes 157
  9.2 Definition of the topological vertex 158
  9.3 Gluing rules 160
  9.4 Derivation of the topological vertex 162
  9.5 Useful formulae for the vertex 166
  9.6 Some applications 168
  9.7 Further properties of the topological vertex 173

10 Applications of the topological string/gauge theory correspondence 177
  10.1 Applications to knot invariants 177
  10.2 Applications to \( \mathcal{N} = 2 \) supersymmetric gauge theory 180

A Symmetric polynomials 185

References 187

Index 197
Part I

Matrix models, Chern–Simons theory, and the large $N$ expansion
In this chapter we develop some basic aspects and techniques of matrix models. A more complete treatment can be found, for example, in Di Francesco et al. (1995) or Di Francesco (2001).

1.1 Basics of matrix models
Matrix models are the simplest examples of quantum gauge theories, namely, they are quantum gauge theories in zero dimensions. The basic field is a Hermitian $N \times N$ matrix $M$. We will consider an action for $M$ of the form

$$\frac{1}{g_s} W(M) = \frac{1}{2g_s} \text{Tr} M^2 + \frac{1}{g_s} \sum_{p \geq 3} \frac{g_p}{p} \text{Tr} M^p, \quad (1.1)$$

where $g_s$ and $g_p$ are coupling constants. This action has the gauge symmetry

$$M \rightarrow UMU^\dagger, \quad (1.2)$$

where $U$ is a $U(N)$ matrix. The partition function of the theory is given by

$$Z = \frac{1}{\text{vol}(U(N))} \int dM \, e^{-\frac{1}{g_s} W(M)} \quad (1.3)$$

where $\text{vol}(U(N))$ is the usual volume factor of the gauge group that arises after fixing the gauge. In other words, we are considering here a gauged matrix model. The measure in the ‘path integral’ is the Haar measure

$$dM = 2^{\frac{N(N-1)}{2}} \prod_{i=1}^{N} dM_{ii} \prod_{1 \leq i < j \leq N} d\text{Re} M_{ij} d\text{Im} M_{ij}. \quad (1.4)$$

The numerical factor in (1.4) is introduced to obtain a convenient normalization.

![Fig. 1.1. The propagator in the double-line notation.](image)
A particularly simple example is the Gaussian matrix model, defined by the partition function

$$Z_G = \frac{1}{\text{vol}(U(N))} \int dM \, e^{-\frac{1}{2g_s} \text{Tr} M^2}. \tag{1.5}$$

We will denote by

$$\langle f(M) \rangle_G = \frac{\int dM \, f(M) \, e^{-\frac{1}{2} \text{Tr} M^2 / 2g_s}}{\int dM \, e^{-\frac{1}{2} \text{Tr} M^2 / 2g_s}} \tag{1.6}$$

the normalized vacuum expectation values of a functional $f(M)$ in the Gaussian matrix model. This model is of course exactly solvable, and the expectation value (1.6) can be computed systematically by using Wick’s theorem. Let us recall how this goes. Consider first a Gaussian integral of the form

$$\langle x_{\mu_1} \cdots x_{\mu_n} \rangle = \frac{\int d^p x \, x_{\mu_1} \cdots x_{\mu_n} \exp \left( -\frac{1}{2} \sum_{\mu,\nu} x_{\mu} A_{\mu\nu} x_{\nu} \right)}{\int d^p x \, \exp \left( -\frac{1}{2} \sum_{\mu,\nu} x_{\mu} A_{\mu\nu} x_{\nu} \right)}, \tag{1.7}$$

where $A$ is a real, invertible matrix, which is taken to be definite-positive in order for the integral to be well defined. If $n$ is odd, (1.7) vanishes. For $n = 2$, the answer is simply

$$\langle x_{\mu_1} x_{\mu_2} \rangle = A^{-1}_{\mu_1\mu_2}, \tag{1.8}$$

while for even $n > 2$ one has

$$\langle x_{\mu_1} \cdots x_{\mu_{2n}} \rangle = \sum_{\text{pairings}} \prod_{i=1}^{n} \langle x_{\sigma_{2i-1}} x_{\sigma_{2i}} \rangle. \tag{1.9}$$

Here, the sum is over all distinct pairings of the indices, and we have denoted by $\sigma_i$, $i = 1, \cdots, 2n$ the indices associated to such a pairing. The results (1.8) and (1.9) can be easily proved with Gaussian integration, and the result (1.9) is Wick’s theorem. We can now come back to the original purpose of computing the integral (1.6). This is again a Gaussian integral, since

$$\frac{1}{2} \text{Tr} M^2 = \frac{1}{2} \left\{ \sum_i M_{ii}^2 + 2 \sum_{i<j} (\text{Re} M_{ij})^2 + (\text{Im} M_{ij})^2 \right\}. \tag{1.10}$$

Using the above results on Gaussian integration it is easy to see that

$$\langle M_{ij} M_{lk} \rangle_G = g_s \delta_{ik} \delta_{jl}. \tag{1.11}$$

According to Wick’s theorem, this is the basic ingredient in computing averages in the Gaussian matrix model. If we regard the matrix model as a quantum zero-dimensional theory, (1.11) corresponds to the propagator. Since the adjoint
representation of $U(N)$ is the tensor product of the fundamental representation $N$ and the anti-fundamental representation $\overline{N}$, we can look at the index $i$ (resp. $j$) of $M_{ij}$ as an index of the fundamental (resp. antifundamental) representation. It is then useful to represent this index structure through a double line with opposite directions. In this double-line notation, the propagator (1.11) can be represented as in Fig. 1.1. Different Wick contractions can now be represented by different double-line diagrams. Consider, for example, the Gaussian average

$$\langle (\text{Tr} M^3)^2 \rangle_G. \quad (1.12)$$

After writing

$$\text{Tr} M^3 = \sum_{i,j,k} M_{ij} M_{jk} M_{ki}$$

we see that the computation of (1.12) amounts to making all possible Wick contractions between two cubic vertices. The cubic vertex can be represented in the double-line notation as in Fig. 1.2. The different Wick contractions give three topologically inequivalent diagrams. We have first the diagram shown in Fig. 1.3, which contributes

$$3 \sum_{ijkmnp} \langle M_{ij} M_{mn} \rangle \langle M_{jk} M_{pm} \rangle \langle M_{ki} M_{np} \rangle = 3 g_s^3 N^3. \quad (1.13)$$

The second diagram is given by two circles joined by a single propagator. Its contribution is

$$9 \sum_{ijkmnp} \langle M_{ij} M_{ki} \rangle \langle M_{jk} M_{pm} \rangle \langle M_{mn} M_{np} \rangle = 9 g_s^3 N^3. \quad (1.14)$$

The remaining diagram is represented in Fig. 1.4 and gives a factor

$$3 \sum_{ijkmnp} \langle M_{ij} M_{mn} \rangle \langle M_{jk} M_{np} \rangle \langle M_{ki} M_{pm} \rangle = 3 g_s^3 N. \quad (1.15)$$

Putting everything together, we find

$$\langle (\text{Tr} M^3)^2 \rangle_G = g_s^3 (12 N^3 + 3 N). \quad (1.16)$$

One can regard the double-line diagrams or ‘fatgraphs’ as a refinement of the
Fig. 1.3. A planar diagram obtained by contracting two cubic vertices.

standard Feynman diagrams with single lines. For example, the diagram

\[
\begin{array}{c}
\includegraphics[width=0.1\textwidth]{figure1_3.png}
\end{array}
\]

in the standard single-line notation 'splits' in the double line notation into the diagram in Fig. 1.3 and the diagram in Fig. 1.4, which are topologically inequivalent. Fatgraphs are characterized topologically by the number of propagators or edges \( E \), the number of vertices with \( p \) legs \( V_p \), and the number of closed loops \( h \). The total number of vertices is \( V = \sum_p V_p \). Each propagator gives a power of \( g_s \), while each interaction vertex with \( p \) legs gives a power of \( g_p/g_s \). The fatgraph will then give a factor

\[
g_s^{E-V} N^h \prod_p g_p^{V_p}. \tag{1.17}
\]

We can also regard the fatgraph as a Riemann surface with holes, in which each closed loop represents the boundary of a hole. The genus \( g \) of such a surface is determined by the elementary topological relation

\[
2g - 2 = E - V - h, \tag{1.18}
\]

therefore we can write (1.17) as

\[
g_s^{2g-2+h} N^h \prod_p g_p^{V_p} = g_s^{2g-2+h} \prod_p g_p^{V_p}, \tag{1.19}
\]

where we have introduced the 't Hooft parameter

\[
t = N g_s. \tag{1.20}
\]

The fatgraphs with \( g = 0 \) are called planar, while the ones with \( g > 0 \) are called non-planar. The graph in Fig. 1.3 is planar: it has \( E = 3, V_3 = 2 \) and \( h = 3 \), therefore \( g = 0 \), and topologically it corresponds to a sphere with three holes. The graph in Fig. 1.4 is non-planar: it has \( E = 3, V_3 = 2 \) and \( h = 1 \), therefore \( g = 1 \), and represents a torus with one hole (it is easy to see this by drawing the diagram on the surface of a torus).

We then see that the computation of any vacuum expectation value (1.6) can be done in a systematic way by using fatgraphs. Notice that, if \( f(M) \) is a
BASICS OF MATRIX MODELS

Fig. 1.4. A non-planar diagram obtained by contracting two cubic vertices.

gauge-invariant function, it can be written as a linear combination of traces of $M$ in arbitrary representations $R$ of $U(N)$. Using Wick’s theorem one can in fact obtain a closed formula for the averages $\langle \text{Tr}_R M \rangle_G$ (Itzykson and Zuber, 1990; Di Francesco and Itzykson, 1993). We present this formula here for completeness, omitting the proof. Let us represent $R$ by a Young tableau with rows of lengths $l_i$, with $l_1 \geq l_2 \geq \cdots$, and with $\ell(R)$ boxes in total. We define the set of $\ell(R)$ integers $f_i$ as follows

$$f_i = \lambda_i + \ell(R) - i, \quad i = 1, \ldots, \ell(R). \quad (1.21)$$

Following Di Francesco and Itzykson (1993) we will say that the Young tableau associated to $R$ is even if the number of odd $f_i$s is the same as the number of even $f_i$s. Otherwise, we will say that it is odd. If $R$ is even, one has

$$\langle \text{Tr}_R M \rangle_G = c(R) \dim R, \quad (1.22)$$

where

$$c(R) = (-1)^{\frac{A(A-1)}{2}} \frac{\prod_{f \text{ odd}} f!! \prod_{f' \text{ even}} f'!!}{\prod_{f \text{ even}, f' \text{ odd}} (f - f')} \quad (1.23)$$

and $A = \ell(R)/2$ (notice that $\ell(R)$ has to be even in order to have a non-vanishing result). Here $\dim R$ is the dimension of the irreducible representation of $SU(N)$ associated to $R$, and can be computed for example by using the hook formula. On the other hand, if $R$ is odd, the above vacuum expectation value vanishes.

The partition function $Z$ of more general matrix models with action (1.1) can be evaluated by doing perturbation theory around the Gaussian point: one expands the exponential of $\sum_{p \geq 3} (g_p/g_s) \text{Tr} M^p/p$ in (1.3), and computes the partition function as a power series in the coupling constants $g_p$. The evaluation of each term of the series involves the computation of vacuum expectation values like (1.6). Of course, this computation can be interpreted in terms of Feynman diagrams, and as usual the perturbative expansion of the free energy

$$F = \log Z$$

will only involve connected vacuum bubbles. Consider, for example, the cubic matrix model, where $g_p = 0$ for $p > 3$. Using (1.16) we find that the free energy reads, at leading order,
\[ F - F_G = \frac{2}{3} g_s g_3^2 N^3 + \frac{1}{6} g_s g_3^2 N + \cdots. \]  

(1.24)

As we have seen, we can re-express the perturbative expansion of \( F \) in terms of fatgraphs, which are labelled by the genus \( g \) and the number of holes \( h \). Therefore, we can write

\[ F = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} F_{g,h} g_s^{2g-2} t^h, \]  

(1.25)

where the coefficients \( F_{g,h} \) (which depend on the coupling constants of the model \( g_p \)) take into account the symmetry factors of the different fatgraphs. We can now formally define the free energy at genus \( g \), \( F_g(t) \), by keeping \( g \) fixed and summing over all closed loops \( h \)

\[ F_g(t) = \sum_{h=1}^{\infty} F_{g,h} t^h, \]  

(1.26)

so that the total free energy can be written as

\[ F = \sum_{g=0}^{\infty} F_g(t) g_s^{2g-2}. \]  

(1.27)

This is the genus expansion of the free energy of the matrix model. In (1.27) we have written the diagrammatic series as an expansion in \( g_s \) around \( g_s = 0 \), keeping the ’t Hooft parameter \( t = g_s N \) fixed. Equivalently, we can regard it as an expansion in \( 1/N \), keeping \( t \) fixed, and then the \( N \) dependence appears as \( N^{2-2g} \). Therefore, for \( t \) fixed and \( N \) large, the leading contribution comes from planar diagrams with \( g = 0 \), which go like \( O(N^2) \). The non-planar diagrams give subleading corrections. Notice that \( F_g(t) \), which is the contribution to \( F \) to a given order in \( g_s \), is given by an infinite series where we sum over all possible numbers of holes \( h \), weighted by \( t^h \). This reformulation of the perturbation expansion of \( U(N) \) gauge theories was proposed by ’t Hooft (1974).

There is an alternative way of writing the matrix model partition function that is very useful. The original matrix model variable is a Hermitian matrix \( M \) that has \( N^2 \) independent real parameters, but after modding out by gauge transformations there are only \( N \) independent parameters left in \( M \). We can for example take advantage of our gauge freedom to diagonalize the matrix \( M \):

\[ M \rightarrow U M U^\dagger = D, \]  

(1.28)

with \( D = \text{diag}(\lambda_1, \cdots, \lambda_N) \). In other words, we can take (1.28) as a choice of gauge, and then use standard Faddeev–Popov techniques to compute the gauge-fixed integral (see for example Bessis et al., 1980). The gauge fixing (1.28) leads to the delta-function constraint

\[ \delta^U(M) = \prod_{i<j} \delta^{(2)}(U M_{ij}), \]  

(1.29)
where $UM = UMU^\dagger$. We introduce
\[ \Delta^{-2}(M) = \int dU \, \delta(UM). \] (1.30)

It then follows that the integral of any gauge-invariant function $f(M)$ can be written as
\[
\int dM \, f(M) = \int dM \, f(M) \Delta^2(M) \int dU \, \delta(UM) = \Omega_N \int \prod_{i=1}^N d\lambda_i \Delta^2(\lambda)f(\lambda),
\] (1.31)
where we have used the gauge invariance of $\Delta^2(M)$, and
\[ \Omega_N = \int dU \] (1.32)
is proportional to the volume of the gauge group $U(N)$, as we will see shortly. We have to evaluate the factor $\Delta(\lambda)$, which can be obtained from (1.30) by choosing $M$ to be diagonal. If $F(M) = 0$ is the gauge-fixing condition, the standard Faddeev–Popov formula gives
\[ \Delta^2(M) = \det \left( \frac{\delta F(U M)}{\delta A} \right)_{F=0}, \] (1.33)
where we write $U = e^A$, and $A$ is an anti-Hermitian matrix. Since
\[ F_{ij}(U D) = (UDU^\dagger)_{ij} = A_{ij}(\lambda_i - \lambda_j) + \cdots \] (1.34)
(1.33) leads immediately to
\[ \Delta^2(\lambda) = \prod_{i<j}(\lambda_i - \lambda_j)^2, \] (1.35)
which is the square of the Vandermonde determinant. Finally, we fix the factor $\Omega_N$ as follows. The Gaussian matrix integral can be computed explicitly by using the Haar measure (1.4), and is simply
\[ \int dM \, e^{-\frac{1}{2g_s} \text{Tr} M^2} = (2\pi g_s)^{N^2/2}. \] (1.36)
On the other hand, by (1.31) this should equal
\[ \Omega_N \int \prod_{i=1}^N d\lambda_i \Delta^2(\lambda)e^{-\frac{1}{2g_s} \sum_{i=1}^N \lambda_i^2}. \] (1.37)
The integral over eigenvalues can be evaluated in various ways, using for example the Selberg function (Mehta, 2004) or the technique of orthogonal polynomials that we describe in the next section. Its value is

\[ g_s^{N^2/2}(2\pi)^{N/2}G_2(N + 2), \]  

(1.38)

where \( G_2(z) \) is the Barnes function, defined by

\[ G_2(z + 1) = \Gamma(z)G_2(z), \quad G_2(1) = 1. \]  

(1.39)

Comparing these results, we find that

\[ \Omega_N = \frac{(2\pi)^{N(N-1)/2}}{G_2(N + 2)}. \]  

(1.40)

Using now (see, for example, Ooguri and Vafa, 2002):

\[ \text{vol}(U(N)) = \frac{(2\pi)^{N(N+1)/2}}{G_2(N + 1)}, \]  

(1.41)

we see that

\[ \frac{1}{\text{vol}(U(N))} \int dM f(M) = \frac{1}{N! (2\pi)^N} \int \prod_{i=1}^{N} d\lambda_i \Delta^2(\lambda)f(\lambda). \]  

(1.42)

The factor \( N! \) in the r.h.s. of (1.42) has an obvious interpretation: after fixing the gauge symmetry of the matrix integral by choosing the diagonal gauge, there is still a residual symmetry given by the Weyl symmetry of \( U(N) \). This is the symmetric group \( S_N \) acting as permutation of the eigenvalues. The ‘volume’ of this discrete gauge group is just its order, \( |S_N| = N! \), and since we are considering gauged matrix models we have to divide by it as shown in (1.42). As a particular case of the above formula, it follows that one can write the partition function (1.3) as

\[ Z = \frac{1}{N! (2\pi)^N} \int \prod_{i=1}^{N} d\lambda_i \Delta^2(\lambda)e^{-\frac{1}{2\sigma_N} \sum_{i=1}^{N} W(\lambda_i)}. \]  

(1.43)

The partition function of the gauged Gaussian matrix model (1.5) is given essentially by the inverse of the volume factor. Its free energy to all orders can be computed by using the asymptotic expansion of the Barnes function

\[ \log G_2(N + 1) = \frac{N^2}{2} \log N - \frac{1}{12} \log N - \frac{3}{4} N^2 + \frac{1}{2} N \log 2\pi + \zeta’(-1) \]

\[ + \sum_{g=2}^{\infty} \frac{B_{2g}}{2g(2g - 2)} N^{2-2g}, \]  

(1.44)

where \( B_{2g} \) are the Bernoulli numbers. Therefore, we find the following expression for the total free energy:
\begin{equation}
F_G = \frac{N^2}{2} \left( \log(N g_s) - \frac{3}{2} \right) - \frac{1}{12} \log N + \zeta'(-1) + \sum_{g=2}^{\infty} \frac{B_{2g}}{2g(2g-2)} N^{2-2g}, \quad (1.45)
\end{equation}

If we now put \( N = t/g_s \), we obtain the following expressions for \( F_g(t) \) in the Gaussian case:

\begin{align*}
F_0(t) &= \frac{1}{2} t^2 \left( \log t - \frac{3}{2} \right), \\
F_1(t) &= -\frac{1}{12} \log t, \\
F_g(t) &= \frac{B_{2g}}{2g(2g-2)} t^{2-2g}, \quad g > 1.
\end{align*}

1.2 Matrix model technology I: saddle-point analysis

The computation of the functions \( F_g(t) \) in closed form seems a difficult task, since in perturbation theory it involves summing up an infinite number of fatgraphs (with different numbers of holes \( h \)). However, in the classic paper of Brézin et al. (1978) it was shown that, remarkably, \( F_0(t) \) can be obtained by solving a Riemann–Hilbert problem. In this section we will review this procedure.

Let us consider a general matrix model with action \( W(M) \), and let us write the partition function after reduction to eigenvalues (1.43) as follows:

\begin{equation}
Z = \frac{1}{N!} \int \prod_{i=1}^{N} d\lambda_i \frac{2\pi}{e^{N^2 S_{\text{eff}}(\lambda)}}, \quad (1.46)
\end{equation}

where the effective action is given by

\begin{equation}
S_{\text{eff}}(\lambda) = -\frac{1}{tN} \sum_{i=1}^{N} W(\lambda_i) + \frac{2}{N^2} \sum_{i<j} \log |\lambda_i - \lambda_j|. \quad (1.47)
\end{equation}

Notice that, since a sum over \( N \) eigenvalues is roughly of order \( N \), the effective action is of order \( O(1) \). We can now regard \( N^2 \) as a sort of \( \hbar^{-1} \) parameter, in such a way that \( N \to \infty \) corresponds to a semi-classical approximation. In this limit the integral (1.46) will be dominated by a saddle-point configuration that extremizes the effective action. Varying \( S_{\text{eff}}(\lambda) \) w.r.t. the eigenvalue \( \lambda_i \), we obtain the equation

\begin{equation}
\frac{1}{2t} W''(\lambda_i) = \frac{1}{N} \sum_{j \neq i} \frac{1}{\lambda_i - \lambda_j}, \quad i = 1, \ldots, N. \quad (1.48)
\end{equation}

The eigenvalue distribution is formally defined for finite \( N \) as

\begin{equation}
\rho(\lambda) = \frac{1}{N} \sum_{i=1}^{N} \delta(\lambda - \lambda_i), \quad (1.49)
\end{equation}

where the \( \lambda_i \) solve (1.48). In the large \( N \) limit, it is reasonable to expect that this distribution becomes a continuous function with compact support. In this
section, we will make the stronger assumption that $\rho(\lambda)$ vanishes outside an interval $\mathcal{C}$. This is the so-called one-cut solution.

Qualitatively, what is going on is the following. Assume for simplicity that $W(x)$, the potential, has only one minimum $x_*$. We can regard the eigenvalues as co-ordinates of a system of $N$ classical particles moving on the real line. Equation (1.48) says that these particles are subject to an effective potential

$$W_{\text{eff}}(\lambda_i) = W(\lambda_i) - \frac{2t}{N} \sum_{j \neq i} \log |\lambda_i - \lambda_j|,$$

(1.50)

which involves a logarithmic Coulomb repulsion between eigenvalues. For small 't Hooft parameter, the potential term dominates over the Coulomb repulsion, and the particles tend to be in the minimum $x_*$ of the potential $W'(x_*) = 0$. This means that, for $t = 0$, the interval $\mathcal{C}$ collapses to the point $x_*$. As $t$ grows, the Coulomb repulsion will force the eigenvalues to be apart from each other and to spread out over an interval $\mathcal{C}$.

We can now write the saddle-point equation in terms of continuum quantities, by using the standard rule

$$\frac{1}{N} \sum_{i=1}^{N} f(\lambda_i) \to \int_{\mathcal{C}} f(\lambda) \rho(\lambda) d\lambda.$$

(1.51)

Notice that the distribution of eigenvalues $\rho(\lambda)$ satisfies the normalization condition

$$\int_{\mathcal{C}} \rho(\lambda) d\lambda = 1.$$  

(1.52)

Equation (1.48) then becomes

$$\frac{1}{2t} W'(\lambda) = P \int_{\mathcal{C}} \frac{\rho(\lambda') d\lambda'}{\lambda - \lambda'},$$

(1.53)

where $P$ denotes the principal value of the integral. This is an integral equation that, given the potential $W(\lambda)$, makes it possible in principle to compute $\rho(\lambda)$ as a function of the 't Hooft parameter $t$ and the coupling constants. Once $\rho(\lambda)$ is known, one can easily compute $F_0(t)$: in the saddle-point approximation, the free energy is given by

$$\frac{1}{N^2} F = S_{\text{eff}}(\rho) + \mathcal{O}(N^{-2}),$$

(1.54)

where the effective action in the continuum limit is a functional of $\rho$: 

$$S_{\text{eff}}(\rho) = -\frac{1}{t} \int_{\mathcal{C}} d\lambda \rho(\lambda) W(\lambda) + \int_{\mathcal{C} \times \mathcal{C}} d\lambda d\lambda' \rho(\lambda) \rho(\lambda') \log |\lambda - \lambda'|.$$

(1.55)

Therefore, the planar free energy is given by
\[ F_0(t) = t^2 S_{\text{eff}}(\rho). \] (1.56)

By integrating (1.53) with respect to \( \lambda \) one finds
\[ \frac{1}{2t} W(\lambda) = \int d\lambda' \rho(\lambda') \log |\lambda - \lambda'| + \xi(t), \] (1.57)
where \( \xi(t) \) is an integration constant that only depends on the coupling constants. It can be computed by evaluating (1.57) at a convenient value of \( \lambda \) (say, \( \lambda = 0 \) if \( W(\lambda) \) is a polynomial). Since the effective action is evaluated on the distribution of eigenvalues that solves (1.53), one can simplify the expression to
\[ F_0(t) = -\frac{t}{2} \int_C d\lambda \rho(\lambda) W(\lambda) + t^2 \xi(t), \] (1.58)

Similarly, averages in the matrix model can be computed in the planar limit as
\[ \frac{1}{N} \langle \text{Tr} M^\ell \rangle = \int_C d\lambda \lambda^\ell \rho(\lambda). \] (1.59)

We then see that the planar limit is characterized by a \textit{classical} density of states \( \rho(\lambda) \), and the planar part of a quantum average like (1.59) can be computed as a moment of this density. The fact that the planar approximation to a quantum field theory can be regarded as a classical field configuration was pointed out by Witten (1980) (see Coleman, 1988, for a beautiful exposition). This classical configuration is often called the \textit{master field}. In the case of matrix models, the master field configuration is given by the density of eigenvalues \( \rho(\lambda) \), and as we will see later it can be encoded in a complex algebraic curve with a deep geometric meaning.

We have shown that the density of eigenvalues is obtained as a solution to the saddle-point equation (1.53). This is a singular integral equation that has been studied in detail in other contexts of physics (see, for example, Muskhelishvili, 1953). The way to solve it is to introduce an auxiliary function called the \textit{resolvent}. The resolvent is defined as a correlator in the matrix model:
\[ \omega(p) = \frac{1}{N} \langle \text{Tr} \frac{1}{p - M} \rangle, \] (1.60)
which is, in fact, a generating functional of the correlation functions (1.59):
\[ \omega(p) = \frac{1}{N} \sum_{k=0}^{\infty} \langle \text{Tr} M^k \rangle p^{-k-1}. \] (1.61)

Being a generating functional of vacuum expectation values, it admits an expansion of the form (Coleman, 1988):
\[ \omega(p) = \sum_{g=0}^{\infty} g_s^{2g} \omega_g(p), \] (1.62)
and the genus zero part can be written in terms of the eigenvalue density as

$$\omega_0(p) = \int \frac{d\lambda \rho(\lambda)}{p - \lambda}.$$  \hfill (1.63)

The genus zero resolvent (1.63) has three important properties. First, as a function of $p$ it is an analytic function on the whole complex plane except on the interval $C$, since if $\lambda \in C$ one has a singularity at $\lambda = p$. Secondly, due to the normalization property of the eigenvalue distribution (1.52), it has the asymptotic behaviour

$$\omega_0(p) \sim \frac{1}{p}, \quad p \to \infty. \hfill (1.64)$$

Finally, one can compute the discontinuity of $\omega_0(p)$ as one crosses the interval $C$. This is just the residue at $\lambda = p$, and one then finds the key equation

$$\rho(\lambda) = -\frac{1}{2\pi i} \left( \omega_0(\lambda + i\epsilon) - \omega_0(\lambda - i\epsilon) \right). \hfill (1.65)$$

Therefore, if the resolvent at genus zero is known, the eigenvalue distribution follows from (1.65), and one can compute the planar free energy. On the other hand, by looking again at the resolvent as we approach the discontinuity, we see that the r.h.s. of (1.53) is given by $-(\omega_0(p + i\epsilon) + \omega_0(p - i\epsilon))/2$, and we then find the equation

$$\omega_0(p + i\epsilon) + \omega_0(p - i\epsilon) = -\frac{1}{t} W'(p), \hfill (1.66)$$

which determines the resolvent in terms of the potential. In this way, we have reduced the original problem of computing $F_0(t)$ to the Riemann–Hilbert problem of computing $\omega_0(p)$. There is a closed expression for the resolvent in terms of a contour integral (Migdal, 1983) that is very useful. Let $C$ be given by the interval $b \leq \lambda \leq a$. Then, one has

$$\omega_0(p) = \frac{1}{2t} \int_C \frac{dz}{2\pi i} \frac{W'(z)}{p - z} \left( \frac{(p - a)(p - b)}{(z - a)(z - b)} \right)^\frac{1}{2}. \hfill (1.67)$$

This equation is easily proved by converting (1.66) into a discontinuity equation:

$$\hat{\omega}_0(p + i\epsilon) - \hat{\omega}_0(p - i\epsilon) = -\frac{1}{t} \frac{W'(p)}{\sqrt{(p - a)(p - b)}}, \hfill (1.68)$$

where $\hat{\omega}_0(p) = \omega_0(p)/\sqrt{(p - a)(p - b)}$. This equation determines $\omega_0(p)$ to be given by (1.67) up to regular terms, but because of the asymptotics (1.64), these regular terms are absent, and (1.67) follows. The asymptotics of $\omega_0(p)$ also gives two more conditions. By taking $p \to \infty$, one finds that the r.h.s. of (1.67) behaves
like $c + d/p + O(1/p^2)$. Requiring the asymptotic behavior (1.64) imposes $c = 0$ and $d = 1$, and this leads to

$$
\oint_C \frac{dz}{2\pi i} \frac{W'(z)}{\sqrt{(z-a)(z-b)}} = 0,
$$

$$
\oint_C \frac{dz}{2\pi i} \frac{zW'(z)}{\sqrt{(z-a)(z-b)}} = 2t.
$$

(1.69)

These equations are enough to determine the endpoints of the cuts, $a$ and $b$, as functions of the 't Hooft coupling $t$ and the coupling constants of the model.

The above expressions are, in fact, valid for very general potentials (we will apply them to logarithmic potentials in Chapter 2), but when $W(z)$ is a polynomial, one can find a very convenient expression for the resolvent. If we deform the contour in (1.67) we pick up a pole at $z = p$, and another one at infinity, and we get

$$
\omega_0(p) = \frac{1}{2t} W'(p) - \frac{1}{2t} \sqrt{(p-a)(p-b)} M(p),
$$

(1.70)

where

$$
M(p) = \oint_0 \frac{dz}{2\pi i} \frac{W'(1/z)}{1-pz} \frac{1}{\sqrt{(1-az)(1-bz)}}.
$$

(1.71)

Here, the contour is around $z = 0$. These formulae, together with the expressions (1.69) for the endpoints of the cut, completely solve the one-matrix model with one cut in the planar limit, for polynomial potentials.

Another way to find the resolvent is to start with (1.48), multiply it by $1/((\lambda_i - p)$, and sum over $i$. One finds, in the limit of large $N$,

$$
(\omega_0(p))^2 - \frac{1}{t} W'(p) \omega_0(p) + \frac{1}{4t^2} R(p) = 0,
$$

(1.72)

where

$$
R(p) = 4t \int d\lambda \rho(\lambda) \frac{W'(p) - W'(\lambda)}{p - \lambda}.
$$

(1.73)

Notice that (1.72) is a quadratic equation for $\omega_0(p)$ and has the solution

$$
\omega_0(p) = \frac{1}{2t} \left( W'(p) - \sqrt{(W'(p))^2 - R(p)} \right),
$$

(1.74)

which is of course equivalent to (1.70).

A useful way to encode the solution to the matrix model is to define

$$
y(p) = W'(p) - 2t \omega_0(p).
$$

(1.75)

Notice that the force on an eigenvalue is given by

$$
f(p) = -W_{\text{eff}}'(p) = -\frac{1}{2} (y(p + i\epsilon) + y(p - i\epsilon)).
$$

(1.76)
In terms of $y(p)$, the quadratic equation (1.72) determining the resolvent can be written as

$$y^2 = W'(p)^2 - R(p). \quad (1.77)$$

This is the equation of a hyperelliptic curve given by a certain deformation (measured by $R(p)$) of the equation $y^2 = W'(p)^2$ typical of singularity theory. We will see in Chapter 8 that this result has a beautiful interpretation in terms of topological string theory on certain Calabi–Yau manifolds.

**Example 1.1** The Gaussian matrix model. Let us apply this technology to the simplest case, the Gaussian model with $W(M) = M^2/2$. First, we find the position of the endpoints using (1.69). Deforming the contour to infinity and introducing the variable $z = 1/p$, we find that the first equation in (1.69) becomes

$$\oint_0 \frac{dz}{2\pi i} \frac{1}{z^2} \frac{1}{\sqrt{(1-az)(1-bz)}} = 0, \quad (1.78)$$

where the contour is now around $z = 0$. Therefore, $a + b = 0$, in accord with the symmetry of the potential. Taking this into account, the second equation becomes:

$$\oint_0 \frac{dz}{2\pi i} \frac{1}{z^3} \frac{1}{\sqrt{1-a^2z^2}} = 2t, \quad (1.79)$$

and gives

$$a = 2\sqrt{t}. \quad (1.80)$$

We see that the interval $C = [-a, a] = [-2\sqrt{t}, 2\sqrt{t}]$ opens up as the 't Hooft parameter grows, and as $t \to 0$ it collapses to the minimum of the potential at the origin, as expected. We immediately find from (1.70)

$$\omega_0(p) = \frac{1}{2t} \left( p - \sqrt{p^2 - 4t} \right), \quad (1.81)$$

and from the discontinuity equation we derive the density of eigenvalues

$$\rho(\lambda) = \frac{1}{2\pi t} \sqrt{4t - \lambda^2}. \quad (1.82)$$

The graph of this function is a semicircle of radius $2\sqrt{t}$, and the above eigenvalue distribution is the famous Wigner–Dyson semicircle law. Notice also that (1.77) is, in this case,

$$y^2 = p^2 - 4t. \quad (1.83)$$

This is the equation for a complex curve of genus zero, which resolves the singularity $y^2 = p^2$. We then see that the opening of the cut as we turn on the 't Hooft parameter can be interpreted as a deformation of a geometric singularity.

**Exercise 1.1** Resolvent for the cubic matrix model. Consider the cubic matrix model with potential $W(M) = M^2/2 + g_3M^3/3$. Derive an expression for the endpoints of the one-cut solution as a function of $t$, $g_3$, and find the resolvent and the planar free energy. The solution is worked out in Brézin et al. (1978).
Although we will not need it in this book, we would like to point out that there are well-developed techniques to obtain the higher-genus $F_g(t)$ as systematic corrections to the saddle-point result $F_0(t)$ (Ambjørn et al., 1993; Eynard, 2004). These corrections can be computed in terms of integrals of differentials defined on the hyperelliptic curve (1.77).

We have so far considered the so-called one-cut solution to the one-matrix model. This is not, however, the most general solution, and we now will consider the multicut solution in the saddle-point approximation. Recall from our previous discussion that the cut appearing in the one-matrix model was centred around a minimum of the potential. If the potential has $n$ minima, one can have a solution with various cuts, centred around the different minima. The most general solution then has $s$ cuts, with $s \leq n$, and the support of the eigenvalue distribution is a disjoint union of $s \leq n$ intervals

\[ C = \bigcup_{i=1}^{s} C_i, \tag{1.84} \]

where

\[ C_i = [x_{2i}, x_{2i-1}], \tag{1.85} \]

and $x_{2s} < \cdots < x_1$. Equation (1.74) still gives the solution for the resolvent, and it is easy to see that the way to have multiple cuts is to require $\omega_0(p)$ to have 2$s$ branch points corresponding to the roots of the polynomial $W'(z)^2 - R(z)$. Therefore,

\[ \omega_0(p) = \frac{1}{2t} W'(p) - \frac{1}{2t} \sqrt{\prod_{k=1}^{2s} (p - x_k) M(p)}, \tag{1.86} \]

which can be solved in a compact way by

\[ \omega_0(p) = \frac{1}{2t} \oint_C \frac{dz}{2\pi i} \frac{W'(z)}{p - z} \left( \prod_{k=1}^{2s} \frac{p - x_k}{z - x_k} \right)^{\frac{1}{2}}. \tag{1.87} \]

In order to satisfy the asymptotics (1.64) the following conditions must hold:

\[ \delta_{\ell,s} = \frac{1}{2t} \oint_C \frac{dz}{2\pi i} \frac{z^{\ell} W'(z)}{\prod_{k=1}^{2s} (z - x_k)^{\frac{1}{2}}}, \quad \ell = 0, 1, \ldots, s. \tag{1.88} \]

In contrast to the one-cut case, these are only $s + 1$ conditions for the 2$s$ variables $x_k$ representing the endpoints of the cut. For $s > 1$, there are not enough conditions to determine the solution of the model, and we need extra input to determine the positions of the endpoints $x_k$. Usually, the extra condition that is imposed is that the different cuts are at equipotential lines (see for example Ake- mann, 1996). It is easy to see that, in general, the effective potential is constant on each cut,

\[ W_{\text{eff}}(p) = \Gamma_i, \quad p \in C_i, \tag{1.89} \]

but the values of $\Gamma_i$ will be, in general, different for the different cuts. This means that there can be eigenvalue tunneling from one cut to the other. The way to
guarantee equilibrium is to choose the endpoints of the cuts in such a way that
\[ \Gamma_i = \Gamma \text{ for all } i = 1, \cdots, s. \]
This gives the \( s - 1 \) conditions:
\[ W_{\text{eff}}(x_{2i+1}) = W_{\text{eff}}(x_{2i}), \quad i = 1, \cdots, s - 1, \]
which, together with the \( s + 1 \) conditions (1.88) provide \( 2s \) constraints that allow one to find the positions of the \( 2s \) endpoints \( x_k \). We can also write (1.90) as
\[ \int_{x_{2i+1}}^{x_{2i}} dz M(z) \prod_{k=1}^{2s} (z - x_k)^{\frac{1}{2}} = 0, \quad i = 1, \cdots, s - 1. \]
In Chapter 7 we will consider a different set of conditions determining the endpoints of the cuts in the multicut case.

1.3 Matrix model technology II: orthogonal polynomials

Another useful technique to solve matrix models involves orthogonal polynomials. This technique was developed by Bessis (1979) and Bessis et al. (1980) (which we follow quite closely), and provides explicit expressions for \( F_g(t) \) at least for low genus. We will use this technique to study Chern–Simons matrix models in Chapter 2.

The starting point of the technique of orthogonal polynomials is the eigenvalue representation of the partition function
\[ Z = \frac{1}{N!} \int \prod_{i=1}^{N} \frac{d\lambda_i}{2\pi} \Delta^2(\lambda)e^{-\frac{1}{g_s} \sum_{i=1}^{N} W(\lambda_i)}, \]
where \( W(\lambda) \) is an arbitrary potential. If we regard
\[ d\mu = e^{-\frac{1}{g_s} W(\lambda)} \frac{d\lambda}{2\pi}, \]
as a measure in \( \mathbb{R} \), one can introduce orthogonal polynomials \( p_n(\lambda) \) defined by
\[ \int d\mu p_n(\lambda)p_m(\lambda) = h_n\delta_{nm}, \quad n \geq 0, \]
where \( p_n(\lambda) \) are normalized by requiring the behaviour \( p_n(\lambda) = \lambda^n + \cdots \). One can now compute \( Z \) by noting that
\[ \Delta(\lambda) = \det p_{j-1}(\lambda_i). \]
By expanding the determinant as
\[ \sum_{\sigma \in S_N} (-1)^{\ell(\sigma)} \prod_k p_{\sigma(k)-1}(\lambda_k), \]
where the sum is over permutations $\sigma$ of $N$ indices and $\epsilon(\sigma)$ is the signature of the permutation, we find

$$Z = \prod_{i=0}^{N-1} h_i = h_0^N \prod_{i=1}^{N} \epsilon^N_{-i}, \quad (1.97)$$

where we have introduced the coefficients

$$r_k = \frac{h_k}{h_{k-1}}, \quad k \geq 1. \quad (1.98)$$

One of the most important properties of orthogonal polynomials is that they satisfy recursion relations of the form

$$(\lambda + s_n)p_n(\lambda) = p_{n+1}(\lambda) + r_n p_{n-1}(\lambda). \quad (1.99)$$

It is easy to see that the coefficients $r_n$ involved in this relation are indeed given by (1.98). This follows from the equality

$$h_{n+1} = \int d\lambda p_{n+1}(\lambda) \lambda p_n(\lambda), \quad (1.100)$$

together with the use of the recursion relation for $\lambda p_{n+1}(\lambda)$. For even potentials, $s_n = 0$.

**Example 1.2** *The Gaussian matrix model and the Hermite polynomials.* As an example of this technique, we can consider again the simple case of the Gaussian matrix model. The orthogonal polynomials of the Gaussian model are well known: they are essentially the Hermite polynomials $H_n(x)$, which are defined by

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} e^{-x^2}. \quad (1.101)$$

More precisely, one has

$$p_n(x) = \left( \frac{g_s}{2} \right)^{n/2} H_n(x/\sqrt{2g_s}), \quad (1.102)$$

and one can then check that

$$h_n^G = \left( \frac{g_s}{2\pi} \right)^{1/2} n! g_s^n, \quad r_n^G = n g_s. \quad (1.103)$$

Using now (1.97) we can confirm the result (1.38) that we stated before.

It is clear that a detailed knowledge of the orthogonal polynomials allows the computation of the partition function of the matrix model. It is also easy to see that the computation of correlation functions also reduces to an evaluation in
terms of the coefficients in the recursion relation. To understand this point, it is useful to introduce the orthonormal polynomials

$$P_n(\lambda) = \frac{1}{\sqrt{h_n}} p_n(\lambda), \quad (1.104)$$

which satisfy the recursion relation

$$\lambda P_n(\lambda) = -s_n P_n(\lambda) + \sqrt{r_{n+1}} P_{n+1}(\lambda) + \sqrt{r_n} P_{n-1}(\lambda). \quad (1.105)$$

Let us now consider the normalized vacuum expectation value $\langle \text{Tr} M^\ell \rangle$, which in terms of eigenvalues is given by the integral

$$\langle \text{Tr} M^\ell \rangle = \frac{1}{N!Z} \int \prod_{i=1}^N e^{-\frac{1}{g_s} W(\lambda_i)} \frac{d\lambda_i}{2\pi} \Delta^2(\lambda) \left(\sum_{i=1}^N \lambda_i^\ell\right). \quad (1.106)$$

By using (1.95) it is easy to see that this equals

$$\sum_{j=0}^{N-1} \int d\mu \lambda^\ell P_j^2(\lambda). \quad (1.107)$$

This integral can be computed in terms of the coefficients in (1.104). For example, for $\ell = 2$ we find

$$\langle \text{Tr} M^2 \rangle = \sum_{j=0}^{N-1} (s_j^2 + r_{j+1} + r_j), \quad (1.108)$$

where we put $r_0 = 0$. A convenient way to encode this result is by introducing the Jacobi matrix

$$J = \begin{pmatrix} 0 & r_1^{1/2} & 0 & 0 & \cdots \\ r_1^{1/2} & 0 & r_2^{1/2} & 0 & \cdots \\ 0 & r_2^{1/2} & 0 & r_3^{1/2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (1.109)$$

as well as the diagonal matrix

$$S = \begin{pmatrix} s_0 & 0 & 0 & 0 & \cdots \\ 0 & s_1 & 0 & 0 & \cdots \\ 0 & 0 & s_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (1.110)$$

It then follows that

$$\langle \text{Tr} M^\ell \rangle = \text{Tr} (J - S)^\ell. \quad (1.111)$$

The results we have presented so far give the exact answer for the correlators and the partition function, at all orders in $1/N$. As we have seen, we are particularly interested in computing the functions $F_g(t)$ that are obtained by resumming
the perturbative expansion at fixed genus. As shown in Bessis (1978) and Bessis et al. (1980), one can in fact use the orthogonal polynomials to provide closed expressions for \( F_g(t) \) in the one-cut case. We will now explain how to do this in some detail.

The object we want to compute is

\[
\mathcal{F} = F - F_G = \log Z - \log Z_G. \tag{1.112}
\]

If we write the usual series \( \mathcal{F} = \sum_{g \geq 0} \mathcal{F}_g g_s^{2g - 2} \), we have

\[
g_s^2 \mathcal{F} = \frac{t^2}{N^2} (\log Z - \log Z_G) = \frac{t^2}{N} \log \frac{h_0}{h_G^G} + \frac{t^2}{N} \sum_{k=1}^{N} (1 - \frac{k}{N}) \log \frac{r_k(N)}{k g_s}. \tag{1.113}
\]

The planar contribution to the free energy \( \mathcal{F}_0(t) \) is obtained from (1.113) by taking \( N \to \infty \). In this limit, the variable

\[
\xi = \frac{k}{N}
\]

becomes a continuous variable, \( 0 \leq \xi \leq 1 \), in such a way that

\[
\frac{1}{N} \sum_{k=1}^{N} f(k/N) \to \int_0^1 d\xi f(\xi)
\]

as \( N \) goes to infinity. Let us assume as well that \( r_k(N) \) has the following asymptotic expansion as \( N \to \infty \):

\[
r_k(N) = \sum_{s=0}^{\infty} N^{-2s} R_{2s}(\xi). \tag{1.114}
\]

We then find

\[
\mathcal{F}_0(t) = -\frac{1}{2} t^2 \log t + t^2 \int_0^1 d\xi (1 - \xi) \log \frac{R_0(\xi)}{\xi}. \tag{1.115}
\]

This provides a closed expression for the planar free energy in terms of the large \( N \) limit of the recursion coefficients \( r_k \).

One can recover the density of states \( \rho(\lambda) \) in the saddle-point approximation from orthogonal polynomials. Let us first try to evaluate (1.111) in the planar approximation, following Bessis et al. (1980). A simple argument based on the recursion relations indicates that, at large \( N \),

\[
(J^\ell)_{nn} \sim \frac{\ell!}{(\ell/2)!^2} r_n^{\ell/2}. \tag{1.116}
\]

Using now the integral representation
we find
\[ \frac{\ell!}{(\ell/2)!^2} = \int_{-1}^{1} \frac{dy}{\pi} \frac{(2y)^\ell}{\sqrt{1 - y^2}}. \]

\[ \frac{1}{N} \langle \text{Tr } M^\ell \rangle = \int_{0}^{1} d\xi \int_{-1}^{1} \frac{dy}{\pi} \frac{1}{\sqrt{1 - y^2}} (2y R_{0}^{1/2}(\xi) - s(\xi))^\ell, \]

where we have denoted by \( s(\xi) \) the limit as \( N \to \infty \) of the recursion coefficients \( s_k(N) \) that appear in (1.99). Since the above average can also be computed by (1.59), by comparing we find
\[ \rho(\lambda) = \int_{0}^{1} d\xi \int_{-1}^{1} \frac{dy}{\pi} \frac{1}{\sqrt{1 - y^2}} \delta(\lambda - (2y R_{0}^{1/2}(\xi) - s(\xi))), \]
or, more explicitly,
\[ \rho(\lambda) = \int_{0}^{1} d\xi \frac{\theta[4R_{0}(\xi) - (\lambda + s(\xi))^2]}{\sqrt{4R_{0}(\xi) - (\lambda + s(\xi))^2}}, \]  
(1.117)

where \( \theta(x) \) denotes here the step function. It also follows from this equation that \( \rho(\lambda) \) is supported on the interval \([b(t), a(t)]\), where
\[ b(t) = -2\sqrt{R_{0}(1)} - s(1), \quad a(t) = 2\sqrt{R_{0}(1)} - s(1). \]  
(1.118)

**Example 1.3** In the Gaussian matrix model \( R_{0}(\xi) = t\xi \), and \( s(\xi) = 0 \). We then find that the density of eigenvalues is supported in the interval \([-2\sqrt{t}, 2\sqrt{t}]\) and is given by
\[ \rho(\lambda) = \frac{1}{\pi} \int_{0}^{1} d\xi \frac{\theta[4\xi t - \lambda^2]}{\sqrt{4\xi t - \lambda^2}} = \frac{1}{2\pi t} \sqrt{4t - \lambda^2} \]
which reproduces, of course, Wigner’s semicircle law.

As shown by Bessis (1979) and Bessis et al. (1980), orthogonal polynomials can also be used to obtain the higher-genus free energies \( F_g \). The key ingredient to do this is simply the Euler–MacLaurin formula, which reads
\[ \frac{1}{N} \sum_{k=1}^{N} f \left( \frac{k}{N} \right) = \int_{0}^{1} f(\xi) d\xi + \frac{1}{2N} [f(1) - f(0)] \]
\[ + \sum_{p=1}^{\infty} \frac{1}{N^{2p}} \frac{B_{2p}}{(2p)!} [f^{(2p-1)}(1) - f^{(2p-1)}(0)], \]  
(1.119)

and should be regarded as an asymptotic expansion for \( N \) large that gives a way to compute systematically \( 1/N \) corrections. We can then use it to calculate (1.113) at all orders in \( 1/N \), where
\[ f(k/N) = \left( 1 - \frac{k}{N} \right) \log \frac{N r_k(N)}{k}, \]  
(1.120)
and we use the fact that \( r_k \) has an expansion of the form (1.114). In this way, we find, for example, that

\[
F_1(t) = t^2 \int_0^1 d\xi (1 - \xi) \frac{R_2(\xi)}{R_0(\xi)} + \frac{t^2}{12} \frac{d}{d\xi} \left[ (1 - \xi) \log \frac{R_0(\xi)}{\xi} \right]_0^1
\]

and one can also find explicit expressions for \( F_g(t) \), \( g \geq 2 \).

It is clear from the above analysis that matrix models can be solved with the method of orthogonal polynomials, in the sense that once we know the precise form of the coefficients in the recursion relation we can compute all quantities in a \( 1/N \) expansion. Since the recursion relation is only known exactly in a few cases, we need methods to determine its coefficients for general potentials \( W(M) \).

In the case of polynomial potentials, of the form

\[
W(M) = \sum_{p \geq 0} g_p \text{Tr} M^p,
\]

there are well-known techniques to obtain explicit results (Bessis et al. 1980), see Di Francesco (2001) and Di Francesco et al. (1995) for reviews. We start by rewriting the recursion relation (1.99) as

\[
\lambda p_n(\lambda) = \sum_{m=0}^{n+1} B_{nm} p_m,
\]

where \( B \) is a matrix. The identities

\[
\begin{align*}
& r_n \int d\lambda e^{-\frac{1}{g_s} W(\lambda)} W'(\lambda) p_n(\lambda) p_{n-1}(\lambda) = nh_n g_s, \\
& \int d\lambda \frac{d}{d\lambda} (p_n e^{-\frac{1}{g_s} W(\lambda)} p_n) = 0
\end{align*}
\]

lead to the matrix equations

\[
\begin{align*}
(W'(B))_{nn-1} &= ng_s, \\
(W'(B))_{nn} &= 0.
\end{align*}
\]

These equations are enough to determine the recursion coefficients. Consider, for example, a quartic potential

\[
W(\lambda) = \frac{g_2}{2} \lambda^2 + \frac{g_4}{4} \lambda^4.
\]

Since this potential is even, it is easy to see that the first equation in (1.122) is automatically satisfied, while the second equation leads to

\[
r_n \left\{ g_2 + g_4 \left( r_n + r_{n-1} + r_{n+1} \right) \right\} = ng_s,
\]

which at large \( N \) reads
\[ R_0(g_2 + 3g_4R_0) = \xi t. \]

In general, for an even potential of the form
\[ W(\lambda) = \sum_{p \geq 0} g_{2p+2} \lambda^{2p+2}, \quad (1.123) \]
one finds
\[ \xi t = \sum_{p \geq 0} g_{2p+2} \binom{2p+1}{p} R_0^{p+1}(\xi), \quad (1.124) \]
which determines \( R_0 \) as a function of \( \xi \). The above equation is sometimes called the \textit{string equation}, and by setting \( \xi = 1 \) we find an explicit equation for the endpoints of the cut in the density of eigenvalues as a function of the coupling constants and \( t \).

\textbf{Exercise 1.2} Verify, using saddle-point techniques, that the string equation correctly determines the endpoints of the cut. Compute \( R_0(\xi) \) for the quartic and the cubic matrix model, and use it to obtain \( F_0(t) \) (for the quartic potential, the solution is worked out in detail in Bessis \textit{et al.} 1980).
CHERN–SIMONS THEORY AND KNOT INVARIANTS

In this chapter we present various aspects of Chern–Simons theory. The essential reference for this is the paper by Witten (1989). An overview of Chern–Simons theory from the physical point of view can be found in the book by Guadagnini (1994).

2.1 Chern–Simons theory: basic ingredients

In a groundbreaking paper, Witten (1989) showed that Chern–Simons gauge theory, which is a quantum field theory in three dimensions, provides a physical description of a wide class of invariants of three-manifolds and of knots and links in three-manifolds.¹ The Chern–Simons action with gauge group \( G \) on a generic three-manifold \( M \) is defined by

\[
S = \frac{k}{4\pi} \int_M \text{Tr} \left( A \wedge dA + \frac{2}{3} A \wedge A \wedge A \right).
\]

Here, \( k \) is the coupling constant, and \( A \) is a \( G \)-gauge connection on the trivial bundle over \( M \). In the following, we will mostly consider Chern–Simons theory with gauge group \( G = U(N) \).

Chern–Simons theory is an example of a topological field theory, and since topological field theories play an important role in what follows, we discuss here some general properties of these theories. Let us consider a quantum field theory defined on a differentiable, compact manifold \( M \). We say that this theory is topological if it contains a set of operators (called observables) whose correlation functions do not depend on the metric that is chosen on \( M \). Formally,

\[
\frac{\delta}{\delta g_{\mu\nu}} \langle O_{i_1} \cdots O_{i_n} \rangle = 0
\]

for a set of operators \( O_{i_1}, \cdots, O_{i_n} \) labelled by indices \( i_1, \cdots, i_n \). In (2.2), \( g_{\mu\nu} \) denotes the metric of \( M \). The set of observables usually includes the identity operator, and in this case the partition function of the theory is independent of the metric and therefore defines a topological invariant of \( M \) (more precisely, a diffeomorphism invariant of \( M \)). There are two types of topological field theories, according to the terminology of Birmingham et al. (1991): topological field theories of the Schwarz type, and topological field theories of the Witten or cohomological type (also called cohomological field theories). In theories of the

¹This was also conjectured by Schwarz (1987).
Schwarz type, the classical Lagrangian does not contain any explicit metric dependence. In other words, the classical action has a symmetry given by invariance under change of background metric. If this symmetry of the classical theory is preserved after quantization, one obtains a topological field theory, since the partition function of the theory is metric independent as well. The set of observables in a theory of the Schwarz type is given by operators with no metric dependence. Cohomological field theories have a very different flavour, and will be discussed in Chapter 3.

Since the Chern–Simons theory action does not involve the metric of $M$ in order to be defined, it leads to a topological field theory of the Schwarz type. In particular, the partition function

$$Z(M) = \int [DA] e^{iS}$$

should define a topological invariant of the manifold $M$. The fact, however, that the classical Lagrangian is metric independent is not, in general, sufficient to guarantee that the quantum theory will preserve this invariance, since there could be anomalies that spoil the classical symmetry. A detailed analysis due to Witten (1989) shows that, in the case of Chern–Simons theory, topological invariance is preserved quantum mechanically, but with an extra subtlety: the invariant depends not only on the three-manifold but also on a choice of framing, i.e. a choice of trivialization of the bundle $TM \oplus TM$. The choice of framing changes the value of the partition function in a very precise way: if the framing is changed by $n$ units, the partition function $Z(M)$ changes as follows:

$$Z(M) \rightarrow \exp \left[ \frac{\pi i nc}{12} \right] Z(M),$$

where

$$c = \frac{kd}{k + y}.$$ (2.5)

In this equation, $d$ and $y$ are, respectively, the dimension and the dual Coxeter number of the group $G$ (for $G = U(N)$, $y = N$). Notice that $c$ is the central charge of the Wess-Zumino-Witten (WZW) model with group $G$ (see, for example, Di Francesco et al., 1997, for an exposition of the WZW model). As explained by Atiyah (1990), for every three-manifold there is in fact a canonical choice of framing, and the different choices are labelled by an integer $s \in \mathbb{Z}$ in such a way that $s = 0$ corresponds to the canonical framing. In the following, unless otherwise stated, all the results for the partition functions of Chern–Simons theory will be presented in the canonical framing.

Besides providing invariants of three-manifolds, Chern–Simons theory also provides invariants of knots and links inside three-manifolds (for a survey of modern knot theory, see Lickorish, 1998, and Prasolov and Sossinsky, 1997). Some examples of knots and links are depicted in Fig. 2.1. Given an oriented knot $K$ in $S^3$, we can consider the trace of the holonomy of the gauge connection
Fig. 2.1. Some knots and links. In the notation $x_n^L$, $x$ indicates the number of crossings, $L$ the number of components (when it is a link with $L > 1$) and $n$ is a number used to enumerate knots and links in a given set characterized by $x$ and $L$. The knot $3_1$ is also known as the trefoil knot, while $4_1$ is known as the figure-eight knot. The link $2_1^2$ is called the Hopf link.

around $\mathcal{K}$ in a given irreducible representation $R$ of $U(N)$. This gives the Wilson loop operator:

$$W^K_R(A) = \text{Tr}_R U_{\mathcal{K}},$$

(2.6)

where

$$U_{\mathcal{K}} = \text{P exp} \oint_{\mathcal{K}} A$$

(2.7)

is the holonomy around the knot. The operator in equation (2.6) is a gauge-invariant operator whose definition does not involve the metric on the three-manifold, therefore it is an observable of Chern–Simons theory regarded as a topological field theory. The irreducible representations of $U(N)$ will be labelled by highest weights or equivalently by the lengths of rows in a Young tableau,
When computing the linking number of two knots, the crossings are assigned a sign $\pm 1$ as indicated in the figure.

If we now consider a link $\mathcal{L}$ with components $\mathcal{K}_\alpha$, $\alpha = 1, \cdots, L$, we can in principle compute the normalized correlation function,

$$W_{R_1 \cdots R_L} (\mathcal{L}) = \langle W_{R_1}^{\mathcal{K}_1} \cdots W_{R_L}^{\mathcal{K}_L} \rangle = \frac{1}{Z(M)} \int [DA] \left( \prod_{\alpha=1}^L W_{R_\alpha}^{\mathcal{K}_\alpha} \right) e^{iS}. \quad (2.8)$$

The unnormalized correlation function will be denoted by $Z_{R_1 \cdots R_L} (\mathcal{L})$. The topological character of the action, and the fact that the Wilson loop operators can be defined without using any metric on the three-manifold, indicate that (2.8) is a topological invariant of the link $\mathcal{L}$. As we will see in Section 2.4, and similarly to what happens with the partition function, in order to define the invariant of the link we need some extra information due to quantum ambiguities in the correlation function (2.8). Notice that we are taking the knots and links to be oriented, and this makes a difference. If $\mathcal{K}^{-1}$ denotes the knot obtained from $\mathcal{K}$ by inverting its orientation, we have that

$$\text{Tr}_R U_{\mathcal{K}^{-1}} = \text{Tr}_R U^{-1}_{\mathcal{K}} = \text{Tr}_R U_{\mathcal{K}}, \quad (2.9)$$

where $\overline{R}$ denotes the conjugate representation.

For further use we notice that, given two linked oriented knots $\mathcal{K}_1$, $\mathcal{K}_2$, one can define an elementary topological invariant, the \textit{linking number}, by

$$\text{lk}(\mathcal{K}_1, \mathcal{K}_2) = \frac{1}{2} \sum_p \epsilon(p), \quad (2.10)$$

where the sum is over all crossing points, and $\epsilon(p) = \pm 1$ is a sign associated to the crossings as indicated in Fig. 2.2. The linking number of a link $\mathcal{L}$ with components $\mathcal{K}_\alpha$, $\alpha = 1, \cdots, L$, is defined by

$$\text{lk}(\mathcal{L}) = \sum_{\alpha < \beta} \text{lk}(\mathcal{K}_\alpha, \mathcal{K}_\beta). \quad (2.11)$$

For example, once an orientation is chosen for the two components of the Hopf link $2_1^2$ shown in Fig. 2.1, one finds two inequivalent oriented links with linking numbers $\pm 1$.

Some of the correlation functions of Wilson loops in Chern–Simons theory turn out to be closely related to important polynomial invariants of knots and
links. For example, one of the most important polynomial invariants of a link $L$ is the HOMFLY polynomial $P_L(q, \lambda)$, which depends on two variables $q$ and $\lambda$ and was introduced by Freyd et al. (1985). This polynomial turns out to be related to the correlation function (2.8) when the gauge group is $U(N)$ and all the components are in the fundamental representation $R_\alpha = \boxtimes$. More precisely, we have

$$W_{\boxtimes \cdots \boxtimes}(L) = \lambda^{lk(L)} \left( \frac{\lambda^{\frac{1}{2}} - \lambda^{-\frac{1}{2}}}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}} \right) P_L(q, \lambda) \quad (2.12)$$

where $lk(L)$ is the linking number of $L$, and the variables $q$ and $\lambda$ are related to the Chern–Simons variables as

$$q = \exp \left( \frac{2\pi i}{k + N} \right), \quad \lambda = q^N. \quad (2.13)$$

When $N = 2$ the HOMFLY polynomial reduces to a one-variable polynomial, the Jones polynomial. When the gauge group of Chern–Simons theory is $SO(N)$, $W_{\boxtimes \cdots \boxtimes}(L)$ is closely related to the Kauffman polynomial. For the mathematical definition and properties of these polynomials, see, for example, Lickorish (1998).

### 2.2 Perturbative approach

#### 2.2.1 Perturbative structure of the partition function

The partition function and correlation functions of Wilson loops in Chern–Simons theory can be computed in a variety of ways. We will here present the basic results of Chern–Simons perturbation theory for the partition function. Since our main interest will be the non-perturbative results of Witten (1989), we will be rather sketchy. For more information on Chern–Simons perturbation theory, we refer the reader to Dijkgraaf (1995) and Labastida (1999) for a physical point of view, and Bar-Natan (1995) and Ohtsuki (2003), for a mathematical perspective.

In the computation of the partition function in perturbation theory, we have first to find the classical solutions of the Chern–Simons equations of motion. If we write $A = \sum_a A^a T_a$, where $T_a$ is a basis of the Lie algebra, we find

$$\frac{\delta S}{\delta A^a_\mu} = \frac{k}{4\pi} \epsilon^{\mu\nu\rho} F^a_{\nu\rho},$$

therefore the classical solutions are just flat connections on $M$. Flat connections are in one-to-one correspondence with group homomorphisms

$$\pi_1(M) \to G. \quad (2.14)$$

For example, if $M = S^3/\mathbb{Z}_p$ is the lens space $L(p, 1)$, one has $\pi_1(L(p, 1)) = \mathbb{Z}_p$, and flat connections are labelled by homomorphisms $\mathbb{Z}_p \to G$. Let us assume that the flat connections on $M$ are a discrete set of points (this happens, for example, if $M$ is a rational homology sphere, since in that case $\pi_1(M)$ is a finite
In that situation, one expresses $Z(M)$ as a sum of terms associated to stationary points:

$$Z(M) = \sum_c Z^{(c)}(M), \quad (2.15)$$

where $c$ labels the different flat connections $A^{(c)}$ on $M$. Each of the $Z^{(c)}(M)$ will be an asymptotic series in $1/k$ of the form

$$Z^{(c)}(M) = Z^{(1)}_{\text{loop}}(M) \exp \left\{ \sum_{\ell=1}^{\infty} S^{(c)}_{\ell} x^\ell \right\}. \quad (2.16)$$

In this equation, $x$ is the effective expansion parameter:

$$x = \frac{2\pi i}{k + y}, \quad (2.17)$$

which takes into account a quantum shift $k \to k + y$ due to finite renormalization effects (Witten, 1989; Álvarez-Gaumé et al., 1990). The one-loop correction $Z^{(c)}_{\text{loop}}(M)$ was first analyzed by Witten (1989), and has been studied in great detail since then (Freed and Gompf, 1991; Jeffrey, 1992; Rozansky, 1995). It has the form

$$Z^{(c)}_{\text{loop}}(M) = \frac{(2\pi x)^{\frac{1}{2}(\dim H^0_c - \dim H^1_c)}}{\text{vol}(H_c)} e^{-\frac{1}{4} S_{\text{CS}}(A^{(c)}) - \frac{i\pi}{4} \sqrt{|\tau^{(c)}_R|}}, \quad (2.18)$$

where $H^{0,1}_c$ are the cohomology groups with values in the Lie algebra of $G$ associated to the flat connection $A^{(c)}$, $\tau^{(c)}_R$ is the Reidemeister–Ray–Singer torsion of $A^{(c)}$, $H_c$ is the isotropy group of $A^{(c)}$, and $\varphi$ is a certain phase. Notice that, for the trivial flat connection $A^{(c)} = 0$, $H_c = G$.

Let us focus on the terms in (2.16) corresponding to the trivial connection, which will be denoted by $S_\ell$. Diagrammatically, the free energy is computed by connected bubble diagrams made out of trivalent vertices (since the interaction in the Chern–Simons action is cubic). We will refer to these diagrams as connected trivalent graphs. $S_\ell$ is the contribution of connected trivalent graphs with $2\ell$ vertices and $\ell + 1$ loops. For each of these graphs we have to compute a group factor and a Feynman integral. However, not all these graphs are independent, since the underlying Lie algebra structure imposes the Jacobi identity:

$$\sum_e \left( f_{abc} f_{edc} + f_{dae} f_{ebc} + f_{ace} f_{edb} \right) = 0. \quad (2.19)$$

This leads to the graph relation known as the IHX relation. Also, anti-symmetry of $f_{abc}$ leads to the so-called AS relation (see, for example, Bar-Natan, 1995; Dijkgraaf, 1995; Ohtsuki, 2002). The existence of these relations suggests to define an equivalence relation in the space of connected trivalent graphs by quotienting by the IHX and the AS relations, and this gives the so-called graph homology. The
space of homology classes of connected diagrams will be denoted by $A(\emptyset)^{\text{conn}}$. This space is graded by half the number of vertices $\ell$, and this number gives the degree of the graph. The space of homology classes of graphs at degree $\ell$ is then denoted by $A(\emptyset)^{\text{conn}}_\ell$. For every $\ell$, this is a finite-dimensional vector space of dimension $d(\ell)$. The dimensions of these spaces are explicitly known for low degrees, see, for example, Bar-Natan (1995), and we have listed some of them in Table 2.1. Given any group $G$, we have a map

$$r_G : A(\emptyset)^{\text{conn}} \rightarrow \mathbb{R}$$ (2.20)

that associates to every graph $\Gamma$ its group theory factor $r_G(\Gamma)$. This map is of course well defined, since different graphs in the same homology class $A(\emptyset)^{\text{conn}}_\ell$ lead by definition to the same group factor. This map is an example of a weight system for $A(\emptyset)^{\text{conn}}_\ell$. Every gauge group gives a weight system for $A(\emptyset)^{\text{conn}}_\ell$, but one may, in principle, find weight systems not associated to gauge groups, although so far the only known example is the one constructed by Rozansky and Witten (1997), which instead uses hyperKähler manifolds. We can now state very

Table 2.1 Dimensions $d(\ell)$ of $A(\emptyset)^{\text{conn}}_\ell$ up to $\ell = 10$.

<table>
<thead>
<tr>
<th>$\ell$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d(\ell)$</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>

precisely what is the structure of the $S_\ell$ appearing in (2.16): since the Feynman diagrams can be grouped into homology classes, we have

$$S_\ell = \sum_{\Gamma \in A(\emptyset)^{\text{conn}}_\ell} r_G(\Gamma) I_\Gamma(M).$$ (2.21)

The factors $I_\Gamma(M)$ appearing in (2.21) are certain sums of integrals of propagators over $M$. It was shown by Axelrod and Singer (1992) that these are differentiable invariants of the three-manifold $M$, and since the dependence on the gauge group has been factored out, they only capture topological information of $M$, in contrast to $Z(M)$, which also depends on the choice of the gauge group. These are the universal perturbative invariants defined by Chern–Simons theory. Notice that, at every order $\ell$ in perturbation theory, there are $d(\ell)$ independent perturbative invariants. Of course, these invariants inherit from $A(\emptyset)^{\text{conn}}_\ell$ the structure of a finite-dimensional vector space, and in particular one can choose a basis of trivalent graphs. A possible choice for $\ell \leq 5$ is the following (Sawon, 2004):
We will denote the graphs with \( k \) circles joined by lines by \( \theta_k \). Therefore, the graph corresponding to \( \ell = 1 \) will be denoted by \( \theta \), the graph corresponding to \( \ell = 2 \) will be denoted \( \theta_2 \), and so on.

Notice that Chern–Simons theory detects the graph homology through the weight system associated to Lie algebras, so in principle it could happen that there is an element of graph homology that is not detected by these weight systems. There is, however, a very elegant mathematical definition of the universal perturbative invariant of a three-manifold that works directly in the graph homology. This is called the LMO invariant (Le et al., 1998) and it is a formal linear combination of homology graphs with rational coefficients:

\[
\omega(M) = \sum_{\Gamma \in \mathcal{A}(\emptyset)^\text{conn}} I^\text{LMO}_\Gamma(M) \Gamma \in \mathcal{A}(\emptyset)^\text{conn}[Q].
\] (2.23)

It is believed that the universal invariants extracted from Chern–Simons perturbation theory agree with the LMO invariant. More precisely, since the LMO invariant \( \omega(M) \) is taken to be 0 for \( S^3 \), we have:

\[
I^\text{LMO}_\Gamma(M) = I_\Gamma(M) - I_\Gamma(S^3),
\] (2.24)

as long as the graph \( \Gamma \) is detected by Lie algebra weight systems. In that sense the LMO invariant is more refined than the universal perturbative invariants extracted from Chern–Simons theory; see Ohtsuki (2002) for a detailed introduction to the LMO invariant and its properties.

As a final remark, we point out that the perturbative evaluation of Wilson loop correlators can also be done using standard procedures. First, one has to expand the holonomy operator as

\[
W^\mathcal{K}_R(A) = \text{Tr}_R \left[ 1 + \oint_{\mathcal{K}} dx^\mu A_\mu(x) + \oint_{\mathcal{K}} dx^\mu \int_x^y dy^\nu A_\nu(y) A_\mu(x) + \cdots \right],
\]

where \( A_\mu = \sum_a A_\mu^a T_a \). Then, after gauge fixing, one can proceed and evaluate the correlation functions in standard perturbation theory. The perturbative study of
Fig. 2.3. Graphic representation of the generator $(T_a)_{ij}$ of a Lie algebra.

Fig. 2.4. Graphic representation of the normalization condition (2.25).

Wilson loops was started by Guadagnini et al. (1990), and a nice review of its development can be found in Labastida (1999). The resulting structure can be formalized in a beautiful way by introducing an algebra of diagrams, as we did for the partition functions, and the corresponding universal perturbative invariants are Vassiliev invariants of knots; see Bar-Natan (1995) and Ohtsuki (2002) for further information.

2.2.2 Group factors

The computation of $S_\ell$ involves the evaluation of group factors of Feynman diagrams, which we have denoted by $r_G(\Gamma)$ above. Here, we give some details about how to evaluate these factors when $G = U(N)$, following the diagrammatic techniques of Cvitanovic (1976) and Bar-Natan (1995). A systematic discussion of these techniques can be found in Cvitanovic (2004).

The basic idea to evaluate group factors is, in fact, very similar to the double-line notation of 't Hooft: express indices in the adjoint representation in terms of indices in the fundamental (and anti-fundamental) representation. In the case of $U(N)$, the adjoint representation is just the tensor product of the fundamental and the anti-fundamental representation. Let us first normalize the trace in the fundamental representation by setting

$$\text{Tr} (T_a T_b) = \delta_{ab}, \quad a, b = 1, \cdots, N^2. \quad (2.25)$$

One can then see that

$$\sum_a (T_a)_{ij} (T_a)_{kl} = \delta_{il} \delta_{kj}. \quad (2.26)$$

If we represent the generator $(T_a)_{ij}$ as in Fig. 2.3, the relation (2.26) can in turn be represented as Fig. 2.5. This is simply the statement that the adjoint representation of $U(N)$ is given by $V_N \otimes V_N^\ast$. Similarly, the normalization condition
(2.25) is graphically represented as Fig. 2.4. The evaluation of group factors of Feynman diagrams involves, of course, the structure constants of the Lie algebra $f_{abc}$, associated to the cubic vertex. By tracing the defining relation of the structure constants we find

\[ f_{abc} = \text{Tr} (T_a T_b T_c) - \text{Tr} (T_b T_a T_c), \]  

(2.27)

which we represent as Fig. 2.6. Putting this together with Fig. 2.5, we obtain the graphical rule represented in Fig. 2.7. We can interpret this as a rule that tells us how to split a single-line Feynman diagram of the $U(N)$ theory into fatgraphs: given a Feynman diagram, we substitute each vertex by the double line vertex without twists, minus the double-line vertex with twists in all edges. If the diagram has $2\ell$ vertices, we will generate $4^\ell$ fatgraphs (some of them may be equal), with a $\pm$ sign, which can be interpreted as Riemann surfaces with holes. The group factor of a fatgraph with $h$ holes is simply $N^h$.

**Example 2.1** Group factors of some simple diagrams. One can use the above rules to compute the group factor of the two-loop Feynman diagram

\[ \circlearrowright. \]  

(2.28)

By resolving the two vertices we obtain 2 different fatgraphs: the graph in Fig. 1.3 with weight 2, and the graph in Fig. 1.4 with weight $-2$. One then finds:

\[ r_{U(N)}(\theta) = 2N(N^2 - 1). \]  

(2.29)
Fig. 2.7. Graphic rule to transform Feynman diagrams into double-line diagrams.

Similarly, the same procedure gives

\[ r_{U(N)}(\theta_2) = 4N^2(N^2 - 1). \]  \hspace{1cm} (2.30)

Using the diagrammatic techniques that we have summarized, it is actually easy to prove that

\[ r_G(\theta) = 2yd, \quad r_G(\theta_2) = 4y^2d \]  \hspace{1cm} (2.31)

for any simple gauge group, where we recall that \( y \) is the dual Coxeter number and \( d \) the dimension of the group. One of the consequences of (2.31), together with (2.4), is that the only universal perturbative invariant that depends on a choice of framing is precisely \( I_{\theta}(M) \). Under a change of framing one has

\[ I_{\theta}(M) \rightarrow I_{\theta}(M) - \frac{n}{48}, \]  \hspace{1cm} (2.32)

while the other universal perturbative invariants remain the same.

It is easy to see from the evaluation of group factors that the perturbative expansion of the free energy of Chern–Simons theory around the trivial connection can be written in the form

\[ F = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} F_{g,h} x^{2g-2+h} N^h, \]  \hspace{1cm} (2.33)

as happened with matrix models. In fact, this structure for the partition function holds for any quantum theory containing only fields in the adjoint representation (’t Hooft, 1974). As in the case of matrix models, we can also reorganize the perturbative series (2.33) as in (1.26). The ’t Hooft coupling of Chern–Simons theory is given by

\[ t = Nx, \]  \hspace{1cm} (2.34)

and one can define the quantity \( F_g(t) \) by summing over all holes keeping the genus \( g \) fixed. We will see later in this chapter how to compute the coefficients \( F_{g,h} \) and the function \( F_g(t) \) for Chern–Simons theory on \( S^3 \).
2.3 Canonical quantization and surgery

As was shown by Witten (1989), Chern–Simons theory is exactly solvable by using non-perturbative methods and the relation to the Wess–Zumino–Witten (WZW) model. In order to present this solution, it is convenient to recall some basic facts about the canonical quantization of the model.

Let $M$ be a three-manifold with boundary given by a Riemann surface $\Sigma$. We can insert a general operator $O$ in $M$, which will, in general, be a product of Wilson loops along different knots and in arbitrary representations of the gauge group. We will consider the case in which the Wilson loops do not intersect the surface $\Sigma$. The path integral over the three-manifold with boundary $M$ gives a wavefunction $\Psi_{M,O}(A)$ that is a functional of the values of the field on $\Sigma$. Schematically, we have:

$$\Psi_{M,O}(A) = \langle A | \Psi_{M,O} \rangle = \int_{A \mid \Sigma = A} DA e^{i S_O}.$$

In fact, associated to the Riemann surface $\Sigma$ we have a Hilbert space $\mathcal{H}(\Sigma)$, which can be obtained by doing canonical quantization of Chern–Simons theory on $\Sigma \times \mathbb{R}$. Before spelling out in detail the structure of these Hilbert spaces, let us make some general considerations about the computation of physical quantities.

In the context of canonical quantization, the partition function can be computed as follows. We first perform a Heegaard splitting of the three-manifold, i.e. we represent it as the connected sum of two three-manifolds $M_1$ and $M_2$ sharing a common boundary $\Sigma$, where $\Sigma$ is a Riemann surface. If $f : \Sigma \rightarrow \Sigma$ is a homeomorphism, we will write $M = M_1 \cup_f M_2$, so that $M$ is obtained by gluing $M_1$ to $M_2$ through their common boundary and using the homeomorphism $f$. This is represented in Fig. 2.8. We can then compute the full path integral (2.3) over $M$ by computing first the path integral over $M_1$ to obtain a state $| \Psi_{M_1} \rangle$ in $\mathcal{H}(\Sigma)$. The boundary of $M_2$ is also $\Sigma$, but with opposite orientation, so its Hilbert space is the dual space $\mathcal{H}^*(\Sigma)$. The path integral over $M_2$ then produces a state $\langle \Psi_{M_2} | \in \mathcal{H}^*(\Sigma)$. The homeomorphism $f : \Sigma \rightarrow \Sigma$ will be represented by an operator acting on $\mathcal{H}(\Sigma)$,

$$U_f : \mathcal{H}(\Sigma) \rightarrow \mathcal{H}(\Sigma).$$

and the partition function can be finally evaluated as

$$Z(M) = \langle \Psi_{M_2} | U_f | \Psi_{M_1} \rangle.$$

Therefore, if we know explicitly what the wavefunctions and the operators associated to homeomorphisms are, we can compute the partition function. The result of the computation is, of course, independent of the particular Heegaard splitting of $M$.

One of the most fundamental results of Witten (1989) is a precise description of $\mathcal{H}(\Sigma)$: it is the space of conformal blocks of a WZW model on $\Sigma$ with gauge
Fig. 2.8. Heegaard splitting of a three-manifold $M$ into two three-manifolds $M_1$ and $M_2$ with a common boundary $\Sigma$.

group $G$ and level $k$ (for an extensive review of the WZW model, see, for example, Di Francesco et al., 1997). In particular, $\mathcal{H}(\Sigma)$ has finite dimension. We will not review here the derivation of this fundamental result. Instead we will use the relevant information from the WZW model in order to solve Chern–Simons theory in some important cases.

The description of the space of conformal blocks on Riemann surfaces can be made very explicit when $\Sigma$ is a sphere or a torus. For $\Sigma = S^2$, the space of conformal blocks is one-dimensional, so $\mathcal{H}(S^2)$ is spanned by a single element. For $\Sigma = T^2$, the space of conformal blocks is in one-to-one correspondence with the integrable representations of the affine Lie algebra associated to $G$ at level $k$. We will use the following notations: the fundamental weights of $G$ will be denoted by $\lambda_i$, and the simple roots by $\alpha_i$, $i = 1, \cdots, r$, where $r$ denotes the rank of $G$. The weight and root lattices of $G$ are denoted by $\Lambda^w$ and $\Lambda^r$, respectively, and $|\Delta_+|$ denotes the number of positive roots. The fundamental chamber $\mathcal{F}_l$ is given by $\Lambda^w/l\Lambda^r$, modded out by the action of the Weyl group. For example, in $SU(N)$ a weight $p = \sum_{i=1}^r p_i \lambda_i$ is in $\mathcal{F}_l$ if

$$\sum_{i=1}^r p_i < l, \quad \text{and} \quad p_i > 0, \ i = 1, \cdots, r.$$  

(2.38)

We recall that a representation given by a highest weight $\Lambda$ is integrable if $\rho + \Lambda$ is in the fundamental chamber $\mathcal{F}_l$, where $l = k + y$ ($\rho$ denotes as usual the Weyl vector, given by the sum of the fundamental weights). In the following, the states in the Hilbert state of the torus $\mathcal{H}(T^2)$ will be denoted by $|p \rangle = |\rho + \Lambda \rangle$ where $\rho + \Lambda \in \mathcal{F}_l$, as we have stated, is an integrable representation of the WZW model at level $k$. We will also denote these states by $|R \rangle$, where $R$ is the representation associated to $\Lambda$. The state $|\rho \rangle$ will be denoted by $|0 \rangle$. The states $|R \rangle$ can be chosen to be orthonormal (Witten, 1989; Elitzur et al., 1989; Labastida and Ramallo, 1989), so we have

$$\langle R | R' \rangle = \delta_{RR'}.$$  

(2.39)

There is a special class of homeomorphisms of $T^2$ that have a simple expression as operators in $\mathcal{H}(T^2)$; these are the $SL(2, \mathbb{Z})$ transformations. Recall that the group $SL(2, \mathbb{Z})$ consists of $2 \times 2$ matrices with integer entries and unit determinant. If
Fig. 2.9. Performing the path integral on a solid torus with a Wilson line in representation $R$ gives the state $|R\rangle$ in $\mathcal{H}(T^2)$.

(1, 0) and (0, 1) denote the two one-cycles of $T^2$, we can specify the action of an $SL(2,\mathbb{Z})$ transformation on the torus by giving its action on this homology basis. The $SL(2,\mathbb{Z})$ group is generated by the transformations $T$ and $S$, which are given by

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \quad (2.40)$$

Notice that the $S$ transformation exchanges the one-cycles of the torus. These transformations can be lifted to $\mathcal{H}(T^2)$, and they have the following matrix elements in the basis of integrable representations:

$$T_{pp'} = \delta_{p,p'}e^{2\pi i(h_p-c/24)},$$

$$S_{pp'} = \frac{i|\Delta_+|}{(k + y)^{r/2}} \left( \frac{\text{Vol } \Lambda^w}{\text{Vol } \Lambda^r} \right)^{\frac{r}{2}} \sum_{w \in W} \epsilon(w) \exp\left( -\frac{2\pi i}{k + y} p \cdot w(p') \right). \quad (2.41)$$

In the first equation, $c$ is the central charge of the WZW model, and $h_p$ is the conformal weight of the primary field associated to $p$:

$$h_p = \frac{p^2 - \rho^2}{2(k + y)}, \quad (2.42)$$

where we recall that $p$ is of the form $\rho + \Lambda$. In the second equation, the sum over $w$ is a sum over the elements of the Weyl group $W$, $\epsilon(w)$ is the signature of the element $w$, and $\text{Vol } \Lambda^w(\text{Vol } \Lambda^r)$ denote, respectively, the volume of the weight (root) lattice. We will often write $S_{RR'}$ for $S_{pp'}$, where $p = \rho + \Lambda$, $p' = \rho + \Lambda'$ and $\Lambda, \Lambda'$ are the highest weights corresponding to the representations $R, R'$.

What is the description of the states $|R\rangle$ in $\mathcal{H}(T^2)$ from the point of view of canonical quantization? Consider the solid torus $T = D \times S^1$, where $D$ is a disc in $\mathbb{R}^2$. This is a three-manifold whose boundary is a $T^2$, and it has a non-contractible cycle given by the $S^1$. Let us now consider the Chern–Simons path integral on the solid torus, with the insertion of the operator $\mathcal{O}_R = \text{Tr}_RU$ given
by a Wilson loop in the representation $R$ around the non-contractible cycle, as shown in Fig. 2.9. In this way, one obtains a state in $\mathcal{H}(T^2)$, and one has

$$|\Psi_{T,C_R}\rangle = |R\rangle.$$ 

(2.43)

In particular, the path integral over the solid torus with no operator insertion gives $|0\rangle$, the ‘vacuum’ state.

These results allow us to compute the partition function of any three-manifold that admits a Heegaard splitting along a torus. Imagine, for example, that we take two solid tori and we glue them along their boundary with the identity map. Since a solid torus is a disc times a circle, $D \times S^1$, by performing this operation we get a manifold that is $S^1$ times the two discs glued together along their boundaries. Therefore, with this surgery we obtain $S^2 \times S^1$, and (2.37) then gives

$$Z(S^2 \times S^1) = \langle 0|0 \rangle = 1.$$ 

(2.44)

If we do the gluing, however, after performing an $S$-transformation on the $T^2$ the resulting manifold is instead $S^3$. To see this, notice that the complement to a solid torus inside $S^3$ is indeed another solid torus whose non-contractible cycle is homologous to the contractible cycle in the first torus. We then find

$$Z(S^3) = \langle 0|S|0 \rangle = S_{00}.$$ 

(2.45)

By using Weyl’s denominator formula,

$$\sum_{w \in W} \epsilon(w)e^{w(\rho)} = \prod_{\alpha > 0} 2 \sinh \alpha / 2,$$ 

(2.46)

where $\alpha > 0$ are positive roots, one finds

$$Z(S^3) = \frac{1}{(k + y)^{r/2}} \left( \frac{\text{Vol} \Lambda^w}{\text{Vol} \Lambda^r} \right)^{1/2} \prod_{\alpha > 0} 2 \sin \left( \frac{\pi (\alpha \cdot \rho)}{k + y} \right).$$ 

(2.47)

The above result can be generalized in order to compute path integrals in $S^3$ with some knots and links. Consider a solid torus where a Wilson line in representation $R$ has been inserted. The corresponding state is $|R\rangle$, as we explained before. If we now glue this to an empty solid torus after an $S$-transformation, we obtain a trivial knot, or unknot, in $S^3$. The path integral with the insertion is then,

$$Z_R = \langle 0|S|R \rangle.$$ 

(2.48)

It follows that the normalized vacuum expectation value for the unknot in $S^3$, in representation $R$, is given by

$$W_R(\text{unknot}) = \frac{S_{0R}}{S_{00}} = \frac{\sum_{w \in W} \epsilon(w)e^{-\frac{2\pi i}{k+y} \rho \cdot w(\Lambda+\rho)}}{\sum_{w \in W} \epsilon(w)e^{-\frac{2\pi i}{k+y} \rho \cdot w(\rho)}}.$$ 

(2.49)
This expression can be written in terms of characters of the group $G$. Remember that the character of the representation $R$, evaluated on an element $a \in \Lambda^w \otimes R$ is defined by

$$\text{ch}_R(a) = \sum_{\mu \in M_R} e^{a \cdot \mu},$$

(2.50)

where $M_R$ is the set of weights associated to the irreducible representation $R$. By using Weyl’s character formula we can write

$$W_R(\text{unknot}) = \text{ch}_R\left[-\frac{2\pi i}{k+y}\right].$$

(2.51)

Moreover, using (2.46), we finally obtain

$$W_R(\text{unknot}) = \prod_{\alpha > 0} \frac{\sin\left(\frac{\pi}{k+y} \alpha \cdot (\Lambda + \rho)\right)}{\sin\left(\frac{\pi}{k+y} \alpha \cdot \rho\right)}.$$

(2.52)

This quantity is often called the quantum dimension of $R$, and it is denoted by $\text{dim}_q R$.

We can also consider a solid torus with a Wilson loop in representation $R$, glued to another solid torus with the representation $R'$ through an $S$-transformation. What we obtain is clearly a link in $S^3$ with two components, which is the Hopf link shown in Fig. 2.1. Carefully taking into account the orientation, we find that this is the Hopf link with linking number $+1$. The path integral with this insertion is:

$$Z_{RR'} = \langle R'|S|R \rangle,$$

(2.53)

so the normalized vacuum expectation value is

$$W_{RR'}(\text{Hopf}^+1) = \frac{S_{RR'}}{S_{00}} = \frac{S_{R'R}^{-1}}{S_{00}},$$

(2.54)

where the superscript $+1$ refers to the linking number. Here, we have used that the bras $\langle R \rangle$ are canonically associated to conjugate representations $\overline{R}$, and that $S_{R'R} = S_{R'R}^{-1}$ (see for example Di Francesco et al., 1997). Therefore, the Chern–Simons invariant of the Hopf link is essentially an $S$-matrix element. In order to obtain the invariant of the Hopf link with linking number $-1$, we notice that the two Hopf links can be related by changing the orientation of one of the components. We then have

$$W_{RR'}(\text{Hopf}^{-1}) = \frac{S_{R'R}}{S_{00}},$$

(2.55)

where we have used the property (2.9).

When we take $G = U(N)$, the above vacuum expectation values for unknots and Hopf links can be evaluated very explicitly in terms of Schur polynomials.
It is well known that the character of the unitary group in the representation $R$ is given by the Schur polynomial $s_R$ (see for example Fulton and Harris, 1991). There is a precise relation between the element $a$ on which one evaluates the character in (2.50) and the variables entering the Schur polynomial. Let $\mu_i$, $i = 1, \ldots, N$, be the weights associated to the fundamental representation of $U(N)$. Notice that, if $R$ is given by a Young tableau whose rows have lengths $l_1 \geq \cdots \geq l_N$, then $\Lambda_R = \sum_i l_i \mu_i$. We also have

$$\rho = \sum_{i=1}^N \frac{1}{2} (N - 2i + 1) \mu_i. \quad (2.56)$$

Let $a \in \Lambda^w \otimes R$ be given by

$$a = \sum_{i=1}^N a_i \mu_i. \quad (2.57)$$

Then,

$$\text{ch}_R[a] = s_R(x_i = e^{a_i}). \quad (2.58)$$

For example, in the case of the quantum dimension, one has $\dim_q R = \dim_{\overline{R}}$, and we find

$$\dim_q R = s_R(x_i = q^{\frac{1}{2}(N-2i+1)}), \quad (2.59)$$

where $q$ is given in (2.13). By using that $s_R$ is homogeneous of degree $\ell(R)$ in the variables $x_i$ we finally obtain

$$\dim_q R = \lambda^{\ell(R)/2} s_R(x_i = q^{-i+\frac{1}{2}})$$

where $\lambda = q^N$ as in (2.13), and there are $N$ variables $x_i$. The quantum dimension can be written very explicitly in terms of the $q$-numbers:

$$[a] = q^{\frac{a}{2}} - q^{-\frac{a}{2}}, \quad [a]_\lambda = \lambda^\frac{a}{2} q^{\frac{a}{2}} - \lambda^{-\frac{a}{2}} q^{-\frac{a}{2}}. \quad (2.60)$$

If $R$ corresponds to a Young tableau with $c_R$ rows of lengths $l_i$, $i = 1, \cdots, c_R$, the quantum dimension is given by:

$$\dim_q R = \prod_{1 \leq i < j \leq c_R} \frac{|l_i - l_j + j - i|}{|j - i|} \prod_{i=1}^{c_R} \prod_{v=i+1}^{c_R} \frac{[v]_\lambda}{[v-i+c_R]} \prod_{i=1}^{c_R} \prod_{v=i+1}^{c_R} [v-i+c_R]. \quad (2.61)$$

It is easy to check that in the limit $k + N \to \infty$ (i.e. in the semi-classical limit) the quantum dimension becomes the dimension of the representation $R$. Notice that the quantum dimension is a rational function of $q^{\pm \frac{1}{2}}, \lambda^{\pm \frac{1}{2}}$. This is a general property of all normalized vacuum expectation values of knots and links in $S^3$. 

CANONICAL QUANTIZATION AND SURGERY 41
The $S$-matrix elements that appear in (2.54) and (2.55) can be evaluated through the explicit expression (2.41), by using the relation between $U(N)$ characters and Schur functions that we explained above. Notice first that
\[
S_{00}^{-1}R_1R_2 = \text{ch}_{R_1} \left[ \frac{2\pi i}{k+y} (\Lambda_{R_2} + \rho) \right] \text{ch}_{R_2} \left[ \frac{2\pi i}{k+y} \rho \right].
\]
(2.62)
If we denote by $l_{i}^{R_2}$, $i = 1, \cdots, c_{R_2}$ the lengths of rows for the Young tableau corresponding to $R_2$, it is easy to see that
\[
W_{R_1R_2}(q, \lambda) = (\lambda q)\frac{|R_1|}{2} s_{R_1}(x_i = q^{i_R^2 - i}) \dim q R_2,
\]
(2.63)
where we set $l_{i}^{R_2} = 0$ for $i > c_{R_2}$. A convenient way to evaluate $s_{R_1}(x_i = q^{i_R^2 - i})$ for a partition $\{l_i^R\}_{i=1,\cdots,c_R}$ associated to $R$ is to use the Jacobi–Trudi formula (A.6). It is easy to show that the generating functional of elementary symmetric functions (A.2) for this specialization is given by
\[
E_{R}(t) = E_{\emptyset}(t) \prod_{j=1}^{c_R} \frac{1 + q^{l_R^2 - j} t}{1 + q^{-j} t},
\]
(2.64)
where
\[
E_{\emptyset}(t) = 1 + \sum_{n=1}^{\infty} a_n t^n,
\]
(2.65)
and the coefficients $a_n$ are defined by
\[
a_n = \prod_{r=1}^{n} \frac{1 - \lambda^{-1} q^{-1}}{q^r - 1}.
\]
(2.66)
The formula (2.63), together with the expressions above for $E_{R}(t)$, provides an explicit expression for (2.54) as a rational function of $q^{\pm \frac{1}{2}}$, $\lambda^{\pm \frac{1}{2}}$, and it was first written down by Morton and Lukac (2003).

Surgery techniques also make it possible to evaluate the partition function of Chern–Simons theory on any manifold $M$ in a purely combinatorial way, in terms of link invariants in $S^3$. By Lickorish's theorem (see, for example, Lickorish, 1998), any three-manifold $M$ can be obtained by surgery on a link $L$ in $S^3$. Let us denote by $K_i$, $i = 1, \cdots, L$, the components of $L$. The surgery operation means that around each of the knots $K_i$ we take a tubular neighbourhood $\text{Tub}(K_i)$ that we remove from $S^3$. This tubular neighbourhood is a solid torus with a contractible cycle $\alpha_i$ and a non-contractible cycle $\beta_i$. We then glue the solid torus back after performing an $\text{SL}(2, \mathbb{Z})$ transformation given by the matrix
\[
U(p_i, q_i) = \begin{pmatrix} p_i & r_i \\ q_i & s_i \end{pmatrix}.
\]
(2.67)
This means that the cycles $p_i \alpha_i + q_i \beta_i$ and $r_i \alpha_i + s_i \beta_i$ on the boundary of the complement of $K_i$ are identified with the cycles $\alpha_i$, $\beta_i$ in $\text{Tub}(K_i)$. In the case
of simply-laced gauge groups, the $SL(2, \mathbb{Z})$ transformation given by $U^{(p,q)}$ has the following matrix elements in the above basis (Jeffrey, 1992; Rozansky, 1996; Hansen and Takata, 2004):

$$U_{\alpha\beta}^{(p,q)} = \frac{[i \text{sign}(q)]^{\Delta_+}}{(l|q|)^{r/2}} \exp \left[ -\frac{id\pi}{12} \Phi(U^{(p,q)}) \right] \left( \frac{\text{Vol} \Lambda^w}{\text{Vol} \Lambda^r} \right)^{1/2} \sum_{n \in \Lambda^r/q \Lambda^r} \sum_{w \in \mathcal{W}} \epsilon(w) \exp \left\{ \frac{i\pi}{lq} \left( p\alpha^2 - 2\alpha(ln + w(\beta)) + s(ln + w(\beta))^2 \right) \right\}. \tag{2.68}$$

In this equation, $\Phi(U^{(p,q)})$ is the Rademacher function:

$$\Phi \left[ \begin{array}{c} p \\ q \\ r \\ s \end{array} \right] = \frac{p + s}{q} - 12s(p, q), \tag{2.69}$$

where $s(p, q)$ is the Dedekind sum

$$s(p, q) = \frac{1}{4q} \sum_{n=1}^{q-1} \cot \left( \frac{\pi n}{q} \right) \cot \left( \frac{\pi np}{q} \right). \tag{2.70}$$

With these data we can already present Witten’s result for the Chern–Simons partition function of $M$. As before, suppose that $M$ is obtained by surgery on a link $L$ in $S^3$. Let us denote by

$$|0, \cdots, 0\rangle \in \mathcal{H}(T^2)^{\otimes L} \tag{2.71}$$

the state corresponding to the linked tubular neighbourhoods that are obtained after removing the $L$ components of the link. The label 0 for each component of the link indicates that they are in the trivial representation (we are not turning on Wilson lines, and the surgery around the link is purely geometric). The state corresponding to the complementary space in $S^3$ will be denoted by $\langle \psi |$. Notice that, if we turn on Wilson lines in representations $R_1, \cdots, R_L$ along the non-contractible cycles of the tori, and we glue the state back, we obtain the (un-normalized) Chern–Simons link invariant of $L$ in representations $R_1, \cdots, R_L$. In other words,

$$\langle \psi | R_1, \cdots, R_L \rangle = Z_{R_1 \cdots R_L}(L). \tag{2.72}$$

If we glue back the tori after performing an $SL(2, \mathbb{Z})$ transformation $U^{(p_i,q_i)}$ on each tori, $i = 1, \cdots, L$, the partition function of Chern–Simons theory on the resulting three-manifold is given by

$$\langle \psi | \prod_{i=1}^{L} U^{(p_i,q_i)} |0, \cdots, 0\rangle. \tag{2.73}$$

But
\[ \prod_{i=1}^{L} \mathcal{U}(p_i, q_i) |0, \ldots, 0\rangle = \sum_{R_1, \ldots, R_L \in \mathcal{F}} \prod_{i=1}^{L} \mathcal{U}_{R_i}^{(p_i, q_i)} |R_1, \ldots, R_L\rangle, \]

therefore the partition function on \( M \) reads

\[ Z(M) = e^{i \phi_{fr}} \sum_{R_1, \ldots, R_L \in \mathcal{F}} \prod_{i=1}^{L} \mathcal{U}_{R_i}^{(p_i, q_i)} Z_{R_1, \ldots, R_L}(\mathcal{L}). \] (2.74)

The phase factor \( e^{i \phi_{fr}} \) is a framing correction that guarantees that the resulting invariant is in the canonical framing for the three-manifold \( M \). Its explicit expression is (Jeffrey, 1992):

\[ \phi_{fr} = \frac{\pi k d}{12l} \left( \sum_{i=1}^{L} \Phi(U^{(p_i, q_i)}) - 3\sigma(\mathcal{L}) \right), \] (2.75)

where \( \sigma(\mathcal{L}) \) is the signature of the linking matrix of \( \mathcal{L} \) (whose entries are the linking numbers of the different components of \( \mathcal{L} \)). We then see that the computation of Chern–Simons invariants of arbitrary manifolds can be reduced in principle to the computation of link invariants in \( S^3 \).

### 2.4 Framing dependence

In the above discussion on the correlation functions of Wilson loops we have glossed over an important ingredient. We already mentioned that, in order to define the partition function of Chern–Simons theory at the quantum level, one has to specify a framing of the three-manifold. It turns out that the evaluation of correlation functions like (2.8) also involves a choice of framing of the knots, as discovered by Witten (1989).

A good starting point to understand the framing is to take Chern–Simons theory with gauge group \( U(1) \). The Abelian Chern–Simons theory turns out to be extremely simple, since the cubic term in (2.1) drops out, and we are left with a Gaussian theory (Polyakov, 1988). \( U(1) \) representations are labelled by integers, and the correlation function (2.8) can be computed exactly. In order to do that, however, one has to choose a framing for each of the knots \( K_\alpha \).

This arises as follows: in evaluating the correlation function, contractions of the holonomies corresponding to different \( K_\alpha \) produce the following integral:

\[ \text{lk}(K_\alpha, K_\beta) = \frac{1}{4\pi} \oint_{K_\alpha} dx^\mu \oint_{K_\beta} dy^\nu \epsilon_{\mu\nu\rho} \frac{(x - y)^\rho}{|x - y|^3}. \] (2.76)

This is a topological invariant, i.e. it is invariant under deformations of the knots \( K_\alpha, K_\beta \), and it is, in fact, the Gauss integral representation of their linking
number $\text{lk}(\mathcal{K}_\alpha, \mathcal{K}_\beta)$ defined in (2.10). On the other hand, contractions of the holonomies corresponding to the same knot $\mathcal{K}$ involve the integral

$$\phi(\mathcal{K}) = \frac{1}{4\pi} \oint_{\mathcal{K}} dx^\mu \oint_{\mathcal{K}} dy^\nu \epsilon_{\mu\nu\rho} \frac{(x - y)^\rho}{|x - y|^3}. \quad (2.77)$$

This integral is well defined and finite (see, for example, Guadagnini et al., 1990), and it is called the cotorsion or writhe of $\mathcal{K}$. It gives the self-linking number of $\mathcal{K}$: if we project $\mathcal{K}$ on a plane, and we denote by $n_+(\mathcal{K})$ the number of positive (negative) crossings as indicated in Fig. 2.2, then we have that

$$\phi(\mathcal{K}) = n_+(\mathcal{K}) - n_-(\mathcal{K}). \quad (2.78)$$

The problem is that the cotorsion is not invariant under deformations of the knot. In order to preserve topological invariance of the correlation function, one has to choose another definition of the composite operator $(\oint_{\mathcal{K}} A)^2$ by means of a framing. A framing of the knot consists of choosing another knot $\mathcal{K}_f$ around $\mathcal{K}$, specified by a normal vector field $n$. The cotorsion $\phi(\mathcal{K})$ then becomes

$$\phi_f(\mathcal{K}) = \frac{1}{4\pi} \oint_{\mathcal{K}} dx^\mu \oint_{\mathcal{K}_f} dy^\nu \epsilon_{\mu\nu\rho} \frac{(x - y)^\rho}{|x - y|^3} = \text{lk}(\mathcal{K}, \mathcal{K}_f). \quad (2.79)$$

The correlation function that we obtain in this way is a topological invariant (since it only involves linking numbers) but the price that we have to pay is that our regularization depends on a set of integers $p_\alpha = \text{lk}(\mathcal{K}_\alpha, \mathcal{K}_f)$ (one for each knot). The correlation function (2.8) can now be computed, after choosing the framings, as follows:

$$\left\langle \prod_\alpha \exp \left( n_\alpha \oint_{\mathcal{K}_\alpha} A \right) \right\rangle = \exp \left\{ \frac{\pi i}{k} \left( \sum_\alpha n_\alpha^2 p_\alpha + \sum_{\alpha \neq \beta} n_\alpha n_\beta \text{lk}(\mathcal{K}_\alpha, \mathcal{K}_\beta) \right) \right\}. \quad (2.80)$$

This regularization is simply the ‘point-splitting’ method familiar in the context of quantum field theory.

Let us now consider Chern–Simons theory with gauge group $U(N)$, and suppose that we are interested in the computation of (2.8), in the context of perturbation theory. It is easy to see that self-contractions of the holonomies lead to the same kind of ambiguities that we found in the Abelian case, i.e. a choice of framing has to be made for each knot $\mathcal{K}_\alpha$. The only difference from the Abelian case is that the self-contraction of $\mathcal{K}_\alpha$ gives a group factor $\text{Tr}_{R_\alpha} (T_a T_a)$, where $T_a$ is a basis of the Lie algebra (see, for example, Guadagnini et al., 1990). The precise result can be better stated as the effect on the correlation function (2.8) under a change of framing, and it says that, under a change of framing of $\mathcal{K}_\alpha$ by $p_\alpha$ units, the vacuum expectation value of the product of Wilson loops changes as follows (Witten, 1989):

$$W_{R_1 \cdots R_L} \rightarrow \exp \left[ 2\pi i \sum_{\alpha=1}^L p_\alpha h_{R_\alpha} \right] W_{R_1 \cdots R_L}. \quad (2.81)$$
In this equation, $h_R$ is the conformal weight of the WZW primary field corresponding to the representation $R$. One can write (2.42) as

$$h_R = \frac{C_R}{2(k + N)}, \quad (2.82)$$

where $C_R = \text{Tr}_R(T_aT_a)$ is the quadratic Casimir in the representation $R$. For $U(N)$ one has

$$C_R = N\ell(R) + \kappa_R, \quad (2.83)$$

where $\ell(R)$ is the total number of boxes in the tableau, and

$$\kappa_R = \ell(R) + \sum_i (l_i^2 - 2il_i). \quad (2.84)$$

In terms of the variables (2.13) the change under framing (2.81) can be written as

$$W_{R_1 \cdots R_L}(L) \to q^{\frac{1}{2} \sum_{\alpha=1}^{L} \kappa_{R\alpha} p_{\alpha} \chi_{\alpha} \sum_{a=1}^L \ell(R_a)p_{\alpha}} W_{R_1 \cdots R_L}. \quad (2.85)$$

Therefore, the evaluation of vacuum expectation values of Wilson loop operators in Chern–Simons theory depends on a choice of framing for knots. It turns out that for knots and links in $S^3$, there is a standard or canonical framing, defined by requiring that the self-linking number is zero. The expressions we have given before for the Chern–Simons invariant of the unknot and the Hopf link are all in the standard framing. Once the value of the invariant is known in the standard framing, the value in any other framing specified by non-zero integers $p_{\alpha}$ can be easily obtained from (2.81).

### 2.5 Results on Wilson loops and knot invariants

In this section, we discuss some useful results for the computation of vacuum expectation values of Wilson loops. We first state some general properties, and then we give more concrete results for particular knots and links. A good reference on the computation and properties of Wilson loop averages in Chern–Simons theory is Guadagnini (1992).

#### 2.5.1 General properties

1) The first property we want to state is the factorization property for the vacuum expectation values of disjoint links, which says the following. Let $\mathcal{L}$ be a link with $L$ components $\mathcal{K}_1, \cdots, \mathcal{K}_L$ that are disjoint knots, and let us attach the representation $R_\alpha$ to the $\alpha$-th component. Then one has

$$W_{R_1 \cdots R_L}(\mathcal{L}) = \prod_{\alpha=1}^L W_{R_\alpha}(\mathcal{K}_\alpha). \quad (2.86)$$

This property is easy to prove in Chern–Simons theory. It only involves some elementary surgery arguments and the fact that $\mathcal{H}(S^2)$ is one-dimensional. A proof can be found in Witten (1989).
2) The second property we will consider is parity symmetry. Chern–Simons theory is a theory of oriented links, and under a parity transformation a link $L$ will transform into its mirror $L^\ast$. The mirror of $L$ is obtained from its planar projection simply by changing undercrossings by overcrossings, and vice versa. On the other hand, parity changes the sign of the Chern–Simons action, in other words $k + N \rightarrow -(k + N)$. We then find that vacuum expectation values transform as

$$W_{R_1 \ldots R_L}(L^\ast, q, \lambda) = W_{R_1 \ldots R_L}(L, q^{-1}, \lambda^{-1}).$$  \hspace{1cm} (2.87)

This is interesting from a knot-theoretic point of view, since it implies that Chern–Simons invariants of links can distinguish in principle a link from its mirror image. As an example of this property, notice for example that the unknot is identical to its mirror image, therefore quantum dimensions satisfy

$$(\dim_q R)(q^{-1}, \lambda^{-1}) = (\dim_q R)(q, \lambda).$$  \hspace{1cm} (2.88)

3) Let us now discuss the simplest example of a fusion rule in Chern–Simons theory. Consider a vacuum expectation value of the form

$$\langle \text{Tr} R_1 U \text{Tr} R_2 U \rangle,$$  \hspace{1cm} (2.89)

where $U$ is the holonomy of the gauge field around a knot $K$. The classical operator $\text{Tr} R_1 U \text{Tr} R_2 U$ can always be written as

$$\text{Tr} R_1 U \text{Tr} R_2 U = \text{Tr} R_1 \otimes R_2 U = \sum_R N_{R_1, R_2}^R \text{Tr} R U,$$  \hspace{1cm} (2.90)

where $R_1 \otimes R_2$ denotes the tensor product, and $N_{R_1, R_2}^R$ are tensor product coefficients. In Chern–Simons theory, the quantum Wilson loop operators satisfy a very similar relation, with the only difference that the coefficients become the fusion coefficients for integrable representations of the WZW model. This can be understood easily if we take into account that the admissible representations that appear in the theory are the integrable ones, so one has to truncate the list of ‘classical’ representations, and this implies in particular that the product rules of classical traces have to be modified. However, in the computation of knot invariants in $U(N)$ Chern–Simons theory it is natural to work in a setting in which both $k$ and $N$ are much larger than any of the representations involved. In that case, the vacuum expectation values of the theory satisfy

$$\langle \text{Tr} R_1 U \text{Tr} R_2 U \rangle = \sum_R N_{R_1, R_2}^R \langle \text{Tr} R U \rangle,$$  \hspace{1cm} (2.91)

where $N_{R_1, R_2}^R$ are the Littlewood-Richardson coefficients of $U(N)$. As a simple application of the fusion rule, imagine that we want to compute $\langle \text{Tr} R_1 U_1 \text{Tr} R_2 U_2 \rangle$, where $U_1, U_2$ are holonomies around disjoint unknots with zero framing. We can take the unknots to be very close, in such a way that the paths along which we
compute the holonomy are the same. In that case, this vacuum expectation value becomes exactly the l.h.s of (2.91). Using also the factorization property (2.86), we deduce the following fusion rule:

$$\dim_q R_1 \dim_q R_2 = \sum_R N^R_{R_1 R_2} \dim_q R.$$  (2.92)

4) Another important property is the behaviour of correlation functions under direct sum. This operation is defined as follows. Let us consider two links $L_1$, $L_2$ with components $K_1, K$ and $K_2, K$, respectively, i.e. the component knot $K$ is the same in $L_1$ and $L_2$. The direct sum $L = L_1 \# L_2$ is a link of three components that is obtained by joining $L_1$ and $L_2$ through $K$. It is not difficult to prove that the Chern–Simons invariant of $L$ is given by (Witten, 1989)

$$W_{R_1 R_2 R}(L) = \frac{W_{R_1 R}(L_1) W_{R_2 R}(L_2)}{W_R(K)}.$$  (2.93)

As an application of this rule, let us consider the three-component link in Fig. 2.10. This link is a direct sum of two Hopf links whose common component is an unknot in representation $R$, and the knots $K_1, K_2$ are unknots in representations $R_1, R_2$. Equation (2.93) expresses the Chern–Simons invariant of $L$ in terms of invariants of Hopf links and quantum dimensions. Notice that the invariant of the link in Fig. 2.10 can also be computed by using the fusion rules. If we fuse the two parallel unknots with representations $R_1, R_2$, we find

$$W_{R_1 R_2 R}(L) = \sum_{R'} N^R_{R_1 R_2} \langle \text{Tr}_{R'} U' \text{Tr}_R U \rangle,$$  (2.94)

where $U$ is the holonomy around the unknot in representation $R$, and $U'$ is the holonomy around the unknot that is obtained by fusing the two parallel unknots in Fig. 2.10. Equation (2.94) expresses the invariant (2.93) in terms of the invariants of a Hopf link with representations $R'$, $R$.

5) Finally, skein rules give relations between vacuum expectation values of Wilson loops. The form of these relations depends on the choice of group and representation, and the simplest case corresponds to $G = U(N)$ and all the representations being the fundamental one. In this case, the skein relation reads...
as follows. Let $\mathcal{L}$ be a link in $S^3$, and let us focus on one of the crossings in its plane projection. The crossing can be an overcrossing, like the one depicted in $L_+$ in Fig. 2.11, or an undercrossing, like the one depicted in $L_-$. If the crossing is $L_+$, we can form two other links either by undoing the crossing (and producing $L_0$ of Fig. 2.11) or by changing $L_+$ into $L_-$. In both cases the rest of the link is left unchanged. Similarly, if the crossing is $L_-$, we form two links by changing $L_-$ into $L_+$ or into $L_0$. The links produced in this way will be in general topologically inequivalent to the original one (they can even have a different number of components). The skein relation says that

$$\lambda^{\frac{1}{2}} W_{\square\cdots\square}(L_+) - \lambda^{-\frac{1}{2}} W_{\square\cdots\square}(L_-) = (q^{\frac{1}{2}} - q^{-\frac{1}{2}}) W_{\square\cdots\square}(L_0)$$

and expresses the Chern–Simons invariant of the original link in terms of the links that are obtained by changing the crossing. By using this relation recursively, one can undo all the crossings in the original link and express $W_{\square\cdots\square}(\mathcal{L})$ in terms of the Chern–Simons invariant of the unknot in the fundamental representation.

As a final comment, it is worth mentioning that the link invariants of Chern–Simons theory do not distinguish all possible links. In particular, they do not distinguish topologically inequivalent links that can be related through an operation called mutation (see, for example, Ramadevi et al., 1995).

2.5.2 Torus knots

Although Chern–Simons theory is exactly solvable, in practice the computation of vacuum expectation values for knots and links can be complicated. For the fundamental representation, the skein relation (2.95) gives, in principle, the answer after finitely many steps, but of course this procedure becomes cumbersome if the number of steps becomes very large. More importantly, although one can write down skein relations for arbitrary representations, they do not determine uniquely the value of the invariant, and other techniques are needed.

A particularly useful framework to compute knot invariants is the formalism of knot operators (Labastida et al., 1991). In this formalism, one constructs operators that ‘create’ knots wrapped around a Riemann surface in the representation $R$ of the gauge group associated to the highest weight $\Lambda$:

$$W^K_{\Lambda} : \mathcal{H}(\Sigma) \rightarrow \mathcal{H}(\Sigma).$$

(2.96)
Notice that the topology of $\Sigma$ restricts the type of knots that one can consider. So far, these operators have been constructed only in the case when $\Sigma = T^2$. Knots that can be put on the surface of a torus are called *torus knots*, and they are labelled by two coprime integers $(n, m)$ that specify the number of times that they wrap the two cycles of the torus. Here, $n$ refers to the winding number around the non-contractible cycle of the solid torus, while $m$ refers to the contractible one. The trefoil knot $3_1$ in Fig. 2.1 is the $(2, 3)$ torus knot, and the knot $5_1$ is the $(2, 5)$ torus knot. The operator that creates the $(n, m)$ torus knot in the representation associated to $\Lambda$ will be denoted by $W^{(n,m)}_{\Lambda}$, and it has a fairly explicit expression (Labastida et al., 1991):

$$W^{(n,m)}_{\Lambda}|p\rangle = e^{2\pi i n m h_{\rho + \Lambda}} \sum_{\mu \in M_{\Lambda}} \exp \left[ i\pi \mu^2 \frac{nm}{k + N} + 2\pi i \frac{m}{k + N} p \cdot \mu \right]|p + n\mu\rangle. \tag{2.97}$$

In this equation, $|p\rangle$ is an arbitrary state in $\mathcal{H}(T^2)$, and $M_{\Lambda}$ is the set of weights corresponding to the irreducible representation with highest weight $\Lambda$. The factor involving the conformal weight $h_{\rho + \Lambda}$ is introduced in order to obtain the invariant in the standard framing. Equation (2.97) allows us to compute the vacuum expectation value of the Wilson loop around a torus knot in $S^3$ as follows: first, one makes a Heegaard splitting of $S^3$ into two solid tori, as we explained before. Then, one puts the torus knot on the surface of one of the solid tori by acting with the knot operator (2.97) on the vacuum state $|\rho\rangle$. Finally, one glues together the tori by performing an $S$-transformation. The normalized vacuum expectation value of the Wilson loop is then given by:

$$\langle W^{(n,m)}_{\Lambda} \rangle = \frac{\langle \rho | SW^{(n,m)}_{\Lambda} |\rho\rangle}{\langle \rho | S |\rho\rangle}. \tag{2.98}$$

One can write (2.98) in a more compact way as follows. When the operator $W^{(n,m)}_{\Lambda}$ acts on the vacuum state $|\rho\rangle$, the r.h.s. of (2.97) is a linear combination of states of the form $|\rho + n\mu\rangle$, where $\mu \in M_{\Lambda}$. The corresponding weights have representatives in the Weyl alcove $\mathcal{F}_l$ that can be obtained by a series of Weyl reflections. In other words, given $n$ and $\mu$, there is a weight $\rho + \xi$ in $\mathcal{F}_l$ and a Weyl reflection $w_\xi \in \mathcal{W}$ such that $\rho + n\mu = w_\xi(\rho + \xi)$. We will denote the set of representatives of the weights $\rho + n\mu$ in $\mathcal{F}_l$ by $\mathcal{M}(n, \Lambda)$. We then conclude that the Chern–Simons invariant of a torus knot $(n, m)$ can be written as:

$$e^{2\pi i n m h_{\rho + \Lambda}} \sum_{\rho + \xi \in \mathcal{M}(n, \Lambda)} \epsilon(w_\xi) \exp \left[ \frac{i\pi m}{n(k + N)} \xi \cdot (\xi + 2\rho) \right] \text{ch}_{\xi} \left[ -\frac{2\pi i}{k + N} \rho \right]. \tag{2.99}$$

Notice that, since the representatives $\rho + \xi$ live in $\mathcal{F}_l$, the weights $\xi$ can be considered as highest weights for a representation, hence the character in (2.99) makes sense.

As an example of this procedure, one can compute the invariant of a torus knot $(n, m)$ in the fundamental representation, where $\Lambda = \lambda_1$. By performing
Weyl reflections, one can show that $\mathcal{M}(n, \lambda_1)$ is given by the following weights (Labastida and Mariño, 1995):

$$\rho + (n - i)\lambda_1 + \lambda_i, \quad i = 1, \ldots, N. \quad (2.100)$$

The computation of the characters is now straightforward (they are just the quantum dimensions of the weights (2.100)), and one finally obtains:

$$W_{(n,m)} = q^{\frac{n}{2}} \frac{\lambda^{-\frac{1}{2}} (\lambda q^{-1})^{(m-1)(n-1)} q^n - 1}{\prod_{i=0}^{n-1} (p+i+1 = n) (-1)^i q^{mi + \frac{1}{4} (p(p+1) - i(i+1)) \prod_{j=-p}^{i} (\lambda - q^j)} [i]! [p]!}. \quad (2.101)$$

If we divide by the vacuum expectation value of the unknot, we find the expression for the HOMFLY polynomial first obtained by Jones (1987). For the trefoil one has, for example:

$$W_{(2,3)} = \frac{1}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}} (-2\lambda^{\frac{1}{2}} + 3\lambda^2 - \lambda^{\frac{5}{2}}) + (q^{\frac{3}{2}} - q^{-\frac{3}{2}}) (-\lambda^{\frac{1}{2}} + \lambda^{\frac{3}{2}}). \quad (2.102)$$

With more effort one can write down formulae for the invariants of torus knots and links in arbitrary representations (Labastida and Mariño, 1995, 2001; Labastida et al., 2000), although they are rather complicated. They afford, however, a systematic computation of the invariants of these knots. For example, for the trefoil in representations with two boxes one finds:

$$W_{(2,3)} = \frac{(\lambda - 1)(\lambda q - 1)}{\lambda (q^{\frac{1}{2}} - q^{-\frac{1}{2}})^2 (1 + q)} \left( (\lambda q^{-1})^2 (1 - \lambda q^2 + q^3 - \lambda q^3 + q^4) \right),$$

$$W_{(2,3)} = \frac{(\lambda - 1)(\lambda - q)}{\lambda (q^{\frac{1}{2}} - q^{-\frac{1}{2}})^2 (1 + q)} \left( (\lambda q^{-2})^2 (1 - \lambda - \lambda q + \lambda^2 q + q^3 - \lambda q^3 + q^4 + q^6) \right). \quad (2.103)$$

2.6 $U(\infty)$ representation theory

As we will see later in this book, the relation between Chern–Simons theory and string theory involves the vacuum expectation values of Wilson loop operators for arbitrary irreducible representations of $U(N)$. This means that $N$ has to be bigger than the number of boxes of any representation under consideration. Formally, we have to consider then $U(\infty)$ representation theory, and it is convenient to develop an appropriate framework for it. This framework is essentially the theory of symmetric polynomials (in an infinite number of variables), which in turn can be translated into the language of bosons and fermions in two dimensions,
and it has appeared before in the study of large-$N$ Yang–Mills theory (Douglas, 1993; Cordes et al., 1994) and in the theory of classical integrable systems (Miwa et al., 2000; Babelon et al., 2003). We will review it here, focusing on the applications to Chern–Simons theory. The basics of symmetric polynomials are reviewed in the Appendix.

Let us consider the set of Wilson loop operators associated to a fixed knot $K$. This set has two natural bases: the basis labelled by representations $R$, which we have already discussed, and the basis labelled by conjugacy classes of the symmetric group. Wilson loop operators in the basis of conjugacy classes are constructed as follows. Let $U$ be the holonomy of the gauge connection around the knot $K$. Let $\vec{k} = (k_1, k_2, \cdots)$ be a vector of infinite entries, almost all of which are zero. This vector defines naturally a conjugacy class $C(\vec{k})$ of the symmetric group $S_\ell$ of $\ell$ elements, with

$$\ell = \sum_j j k_j. \tag{2.104}$$

The conjugacy class $C(\vec{k})$ is simply the class that has $k_j$ cycles of length $j$. We will also use the following quantity:

$$|\vec{k}| = \sum_j k_j. \tag{2.105}$$

We now define the operator

$$\Upsilon_{\vec{k}}(U) = \prod_{j=1}^{\infty} (\text{Tr} U^j)^{k_j}. \tag{2.106}$$

These operators, for all possible $\vec{k}$, give the basis of Wilson loop operators labelled by conjugacy classes. The basis of representations and the basis of conjugacy classes for Wilson loop operators correspond to two well-known bases for symmetric polynomials. This is easily seen by taking $U$ to be a diagonal matrix $U = \text{diag}(x_1, \cdots, x_N)$. It is then an elementary result in the representation theory of the unitary group that $\text{Tr}_R U$ is the Schur polynomial in $x_i$ defined in (A.5):

$$\text{Tr}_R U = s_R(x). \tag{2.107}$$

It is immediately seen that

$$\Upsilon_{\vec{k}}(U) = P_{\vec{k}}(x), \tag{2.108}$$

where the Newton polynomials are defined in (A.7). It follows from the Frobenius formula (A.9) between the two bases of symmetric polynomials that the operators (2.106) are linear combinations of the operators labelled by representations:

$$\Upsilon_{\vec{k}}(U) = \sum_R \chi_R(C(\vec{k})) \text{Tr}_R U, \tag{2.109}$$
where \( \chi_R(C(\vec{k})) \) are the characters of the symmetric group \( S_\ell \) in the representation \( R \) evaluated at the conjugacy class \( C(\vec{k}) \). The above formula can be inverted as

\[
\text{Tr}_R(U) = \sum_{\vec{k}} \frac{\chi_R(C(\vec{k}))}{z_{\vec{k}}} \Upsilon_{\vec{k}}(U),
\]

with

\[
z_{\vec{k}} = \prod_j k_j! j^{k_j}. \tag{2.111}
\]

As we have already pointed out, we are interested in considering symmetric polynomials in an infinite number of variables. This can be formally done, as explained by Macdonald (1995), and allows one to consider arbitrary representations \( R \) and vectors \( \vec{k} \). A particularly useful way of dealing with symmetric polynomials in an infinite number of variables is to consider the set of operators \( \Upsilon_{\vec{k}}(U) \) as a basis for the infinite-dimensional Hilbert space associated to a chiral boson in two dimensions. The Frobenius relation is, in this language, a manifestation of the equivalence between fermions and bosons in two dimensions. Let us explain this in some detail.

The mode expansion of a chiral boson in two dimensions is:

\[
\partial \phi = \sum_{m \in \mathbb{Z}} \alpha_m z^{-m-1}. \tag{2.112}
\]

The modes \( \alpha_m \) satisfy the commutation relations:

\[
[\alpha_n, \alpha_m] = n \delta_{n+m,0}. \tag{2.113}
\]

In order to construct the Hilbert space of states, we first define a vacuum by

\[
\alpha_n |0 \rangle = 0, \quad n > 0. \tag{2.114}
\]

As usual, we will call the operators \( \alpha_{-n} \) creation operators for \( n > 0 \) and annihilation operators for \( n < 0 \).

We can also define the fermion fields \( \psi(z), \psi^*(z) \) in two dimensions by

\[
\psi(z) = \sum_{n \in \mathbb{Z}+1/2} \psi_n z^{-n-1/2},
\psi^*(z) = \sum_{n \in \mathbb{Z}+1/2} \psi_n^* z^{-n-1/2}. \tag{2.115}
\]

The fermionic modes \( \psi_n, \psi_n^* \) satisfy the anti-commutation relations

\[
\{\psi_n, \psi_m\} = \{\psi_n^*, \psi_m^*\} = 0,
\{\psi_n, \psi_m^*\} = \delta_{nm}. \tag{2.116}
\]
The vacuum defined in (2.114) satisfies
\[ \psi_n |0\rangle = \psi^*_n |0\rangle = 0, \quad n > 0. \] (2.117)

It is also useful to define a normal ordering of fermionic modes:
\[ : \psi_i \psi^*_j : = \psi_i \psi^*_j - \langle 0 | \psi_i \psi^*_j | 0 \rangle. \] (2.118)

The equivalence of fermions and bosons in two dimensions can be stated as the following relation:
\[ \partial \phi(z) = : \psi^*(z) \psi(z) :, \] (2.119)

which in terms of modes reads
\[ \alpha_n = \sum_{j \in \mathbb{Z} + 1/2} : \psi_{-j} \psi^*_{j+n} :. \] (2.120)

Let us now consider the Hilbert space \( \mathcal{H} \) obtained by acting with creation operators in the vacuum. A natural basis of this space is labelled by infinite-dimensional vectors \( \vec{k} \):
\[ |\vec{k}\rangle = \prod_{j=1}^{\infty} (\alpha_{-j})^{k_j} |0\rangle. \] (2.121)
This is an orthogonal basis, since
\[ \langle \vec{k} | \vec{k}' \rangle = z_{\vec{k}} \delta_{\vec{k}, \vec{k}'}, \] (2.122)
and we have the completeness relation
\[ \sum_{\vec{k}} \frac{1}{z_{\vec{k}}} |\vec{k}\rangle \langle \vec{k}| = 1_{\mathcal{H}}. \] (2.123)

We can construct another basis of \( \mathcal{H} \) by using fermion operators. Let \( R \) be a Young tableau. Instead of specifying it through the lengths of the rows, as we have been doing so far, we can use the so-called Frobenius notation. This notation works as follows: we cut the tableau along the diagonal, and count the boxes above and below the diagonal, in the horizontal and vertical directions, respectively. This leads to two sets of integers, denoted by \( m_i \) and \( n_i \), respectively. The Young tableau associated to \( R \) will be denoted, in Frobenius notation, by \( (m_i | n_i) \). In terms of lengths of rows \( h_i \) and columns \( v_i \) in the tableau, we have
\[ m_i = h_i - i, \quad n_i = v_i - i. \] (2.124)
For example, the tableau

is written in Frobenius notation as \( (4,1|2,1) \). The state \( |R\rangle \) associated to the Young tableau \( (m_i | n_i) \), \( i = 1, \cdots, r \), is defined by
\[ |R\rangle = \epsilon(n) \prod_{i=1}^{r} \psi_{-m_i-1/2} \psi_{-n_i-1/2} |0\rangle, \quad (2.125) \]

where \( \epsilon(n) \) is a sign given by
\[ \epsilon(n) = (-1)^{\sum_{i=1}^{r} n_i + r(r-1)/2}. \quad (2.126) \]

This basis is orthonormal,
\[ \langle R|R' \rangle = \delta_{RR'}, \quad (2.127) \]

and one can show that it is related to the \( |\vec{k}\rangle \) basis through the Frobenius relation (see, for example, Miwa et al., 2000; Babelon et al., 2003):
\[ |R\rangle = \sum_{\vec{k}} \chi_{R} (C(\vec{k})) \frac{1}{z_{\vec{k}}} |\vec{k}\rangle. \quad (2.128) \]

There is a close relation between the Hilbert space \( \mathcal{H} \) and the space of symmetric polynomials. In order to do that, we define a coherent state in \( \mathcal{H} \) associated to \( \{ t_n \}_{n \geq 1} \) through the equation
\[ |t\rangle = \exp \left( \sum_{n=1}^{\infty} \frac{t_n}{n} \alpha_{-n} \right) |0\rangle. \quad (2.129) \]

The product of two coherent states is given by
\[ \langle s|t \rangle = \exp \left( \sum_{n=1}^{\infty} \frac{t_n s_n}{n} \right). \]

Notice that
\[ |t\rangle = \sum_{\vec{k}} \frac{t_{\vec{k}}}{z_{\vec{k}}} |\vec{k}\rangle, \quad (2.130) \]

where \( t_{\vec{k}} = \prod_j t_j^{k_j} \). We can now define a very explicit map between the Hilbert space \( \mathcal{H} \) and the space of symmetric polynomials in an infinite number of variables \( \mathcal{P} \) by
\[ |\phi\rangle \to \langle t|\phi \rangle, \quad (2.131) \]

where \( |\phi\rangle \in \mathcal{H} \). The r.h.s. is a polynomial in the \( t_n \), but if we further identify
\[ t_j = P_j(x), \quad (2.132) \]

where \( P_j(x) \) is the power sum series defined in (A.8), we find an isomorphism
\[ \mathcal{H} \to \mathcal{P}. \quad (2.133) \]

For example, the basis \( |\vec{k}\rangle \) corresponds to the Newton polynomials, since
\[ \langle t | \vec{k} \rangle = P_{\vec{k}}(x). \]  

Also, using (2.128) one finds

\[ \langle t | R \rangle = s_{R}(t), \]  

where \( s_{R}(t) \) is the Schur polynomial written in terms of Newton polynomials.

The language of bosons and fermions turns out to be extremely useful to analyse various aspects of representation theory that appear in Chern–Simons theory. For example, the total number of boxes of a representation \( \ell(R) \) is the eigenvalue of the operator

\[ \ell = \sum_{n \in \mathbb{Z} + \frac{1}{2}} n : \psi_{-n} \psi_{n}^{*}, \]  

when acting on the state \( |R\rangle \), and the quantity \( \kappa_{R} \) defined in (2.84) is the eigenvalue of the operator

\[ \kappa = \sum_{n \in \mathbb{Z} + \frac{1}{2}} n^2 : \psi_{-n} \psi_{n}^{*}. \]  

This can be shown as follows: in terms of the quantities defined in (2.124), the eigenvalue of the operator \( \kappa \) acting on \( |R\rangle \) is clearly

\[ \sum_{i=1}^{r} \left( m_{i} + \frac{1}{2} \right)^{2} - \sum_{i=1}^{r} \left( n_{i} + \frac{1}{2} \right)^{2}. \]  

Using now the identity (see for example Cordes et al. (1994), Section 4.4.2):

\[ \ell(R) \sum_{i=1}^{r} (h_{i} - i)^{2} - i^{2} = \sum_{i=1}^{r} \left( (h_{i} - i)^{2} - (v_{i} - i + 1)^{2} \right), \]  

it follows that (2.138) equals \( \kappa_{R} \). Finally, notice that the operators (2.136) and (2.137) can be written in terms of bosons, leading to

\[ \ell = \oint \frac{dz}{2\pi i} \partial \phi(z), \quad \kappa = \frac{1}{3} \oint \frac{dz}{2\pi i} : (\partial \phi)^{3} :, \]  

Let us now discuss some structures of the Hilbert space \( \mathcal{H} \). First, it has an automorphism \( \omega \) that, in fermionic language, essentially exchanges \( \psi \) with \( \psi^{*} \):

\[ \omega(\psi_{n}) = (-1)^{n} \psi_{n}^{*}, \quad \omega(\psi_{n}^{*}) = (-1)^{n} \psi_{n}. \]  

It is easy to see that \( \omega \) acts as an involution on \( \mathcal{H} \), i.e. \( \omega^{2} = 1 \) on \( \mathcal{H} \). In the representation basis it maps

\[ |R\rangle \rightarrow |R^{t}\rangle, \]  

where \( R^{t} \) is the transposed Young tableau (where rows and columns are exchanged w.r.t. \( R \)). The involution \( \omega \) corresponds to the involution in the space of symmetric polynomials that exchanges the complete symmetric functions with the elementary symmetric functions (Macdonald, 1995).
Exercise 2.1 Show that the state $|\omega(t)\rangle$, obtained by acting with the involution $\omega$ defined in (2.141) on the coherent state $|t\rangle$, satisfies

$$\omega(t_n) = (-1)^{n+1}t_n.$$ (2.143)

The space $\mathcal{H}$ also has a ring structure given by the tensor product of representations, and defined in the representation basis by

$$|R_1\rangle \otimes |R_2\rangle = \sum_R N_{R_1 R_2}^R |R\rangle,$$ (2.144)

where $N_{R_1 R_2}^R$ are Littlewood–Richardson coefficients. The tensor product of coherent states is particularly simple:

$$|t_1\rangle \otimes |t_2\rangle = |t_1 + t_2\rangle.$$ (2.145)

We also have a coproduct

$$\Delta : \mathcal{H} \rightarrow \mathcal{H} \times \mathcal{H},$$ (2.146)

which in the representation basis decomposes a representation according to

$$\Delta(|R\rangle) = \sum_{R_1 R_2} N_{R_1 R_2}^R |R_1\rangle \otimes |R_2\rangle.$$ (2.147)

For example,

$$\Delta(|00\rangle) = |00\rangle \otimes |00\rangle + |00\rangle \otimes |00\rangle + |00\rangle \otimes |00\rangle.$$ (2.148)

In view of (2.147), it is useful to define the state

$$|R'/R''\rangle = \sum_{R''} N_{R' R''}^R |R\rangle.$$ (2.149)

Under the correspondence (2.131) with symmetric polynomials, the above state is mapped to the skew Schur function $s_{R'/R''}(x)$ defined in (A.11). The coproduct $\Delta$ turns out to be extremely simple in the boson basis, since it maps

$$\alpha_{-n} |0\rangle \rightarrow \alpha_{-n}^{(1)} |0\rangle_1 \otimes |0\rangle_2 + |0\rangle_1 \otimes \alpha_{-n}^{(2)} |0\rangle_2,$$ (2.149)

where we have added subscripts and superscripts to make clear the tensor product structure. From (2.149) it is easy to deduce that the coproduct $\Delta$ acts on coherent states as

$$\Delta(|t\rangle) = |t\rangle_1 \otimes |t\rangle_2.$$ (2.150)

One can actually show that $\mathcal{H}$, endowed with the tensor product (2.144) and with the coproduct $\Delta$, has the structure of a Hopf algebra whose primitive elements are
precisely $\alpha_{-n}|0\rangle$. We refer to Macdonald (1995), I.5, Example 25, for a detailed explanation of this point. Notice that (2.150) leads to the identity

$$\sum_{Q,R,R'} N_{R,R'}^Q s_Q(t)|R\rangle_1 \otimes |R'\rangle_2 = |t\rangle_1 \otimes |t\rangle_2.$$  
(2.151)

The language of CFT in two dimensions allows one to define various useful states for computations. It is easy to show, for example, that the coherent state

$$|D\rangle = \exp\left(\sum_{n=1}^{\infty} \frac{\lambda_n^2 - \lambda_{-n}^2}{n(q^{\frac{1}{2}} - q^{-\frac{1}{2}})} \alpha_{-n}\right)|0\rangle$$  
(2.152)

satisfies

$$\dim_q R = \langle R|D\rangle.$$  
(2.153)

Let us now consider the construction of generating functionals for knot and link invariants. We first define the Ooguri–Vafa operator. Let $U$ be a $U(N)$ holonomy matrix for a knot, and let $V$ be a $U(M)$ matrix (a ‘source’ term). We assume that both $N$ and $M$ are formally very large, so that one can make sense of any representation. The Ooguri–Vafa operator, introduced by Ooguri and Vafa (2000), is given by

$$Z(U, V) = \exp\left[\sum_{n=1}^{\infty} \frac{1}{n} \text{Tr} U^n \text{Tr} V^n\right].$$  
(2.154)

When expanded, this operator can be written in the $k$-basis as follows,

$$Z(U, V) = 1 + \sum_{\vec{k}} \frac{1}{z_{\vec{k}}} \Upsilon_{\vec{k}}(U) \Upsilon_{\vec{k}}(V).$$  
(2.155)

We see that $Z(U, V)$ includes all possible Wilson loop operators $\Upsilon_{\vec{k}}(U)$ associated to a knot $K$. One can also use the Frobenius formula to show that

$$Z(U, V) = \sum_{R} \text{Tr}_R(U) \text{Tr}_R(V),$$  
(2.156)

where the sum over representations starts with the trivial one. From the point of view of symmetric polynomials, the equality of (2.154) and (2.156) can be rewritten as the well-known identity

$$\sum_{R} s_R(x)s_R(y) = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n} P_n(x)P_n(y)\right).$$  
(2.157)

In $Z(U, V)$ we assume that $U$ is the holonomy of a dynamical gauge field and that $V$ is a source. The vacuum expectation value
\[ Z(V) = \langle Z(U, V) \rangle = \sum_R W_R \text{Tr}_R V \] (2.158)

then has information about the vacuum expectation values of the Wilson loop operators in all possible representations, and by taking its logarithm one can define the connected vacuum expectation values \( W^{(c)}_k \):

\[ F(V) = \log Z(V) = \sum_k \frac{1}{z_k} W^{(c)}_k \Upsilon_k(V). \] (2.159)

One has, for example:

\[ W^{(c)}_{(2,0,\ldots)} = \langle (\text{Tr} U)^2 \rangle - \langle \text{Tr} U \rangle^2 = W_{\square} + W_{\boxslash} - W_{\square}. \]

**Example 2.2 Generating functionals for the unknot.** In the case of the unknot at zero framing, it is easy to compute the generating functional \( Z(V) \) and \( F(V) \) if we take into account the fact that quantum dimensions are given by Schur polynomials evaluated at \( x_i = q^{\frac{1}{2} (N - 2i + 1)} \). We then find

\[ Z(V) = \exp \left( \sum_{n=1}^{\infty} \frac{1}{n} \frac{\lambda^{-\frac{n}{2}} - \lambda^{-\frac{n}{2}}}{q^{\frac{n}{2}} - q^{-\frac{n}{2}}} \text{Tr} V^n \right) = \langle t|\mathcal{D} \rangle, \] (2.160)

where we identified \( t_n = \text{Tr} V^n \).

### 2.7 The 1/N expansion in Chern–Simons theory

As we explained in Section 2.2, the perturbative series of Chern–Simons theory around the trivial connection can be re-expressed in terms of fatgraphs. In particular, one should be able to study the free energy of Chern–Simons theory on the three-sphere in the 1/N expansion, i.e. to expand it as in (2.33) and to resum all fatgraphs of fixed genus in this expansion to obtain the quantities \( F_g(t) \). In this section we will obtain closed expressions for \( F_{g,h} \) and \( F_g(t) \) in the case of Chern–Simons theory defined on \( S^3 \), following Gopakumar and Vafa (1998a; 1999). For earlier work on the 1/N expansion of Chern–Simons theory, see Camperi et al. (1990), Periwal (1993) and Correale and Guadagnini (1994).

A direct computation of \( F_{g,h} \) from perturbation theory is difficult, since it involves the evaluation of integrals of products of propagators over the three-sphere. However, in the case of \( S^3 \) we have an exact expression for the partition function and we can expand it in both \( x \) and \( N \) to obtain the coefficients of (2.33). The partition function of CS with gauge group \( U(N) \) on the three-sphere can be obtained from the formula (2.47) for \( SU(N) \) after multiplying it by an overall \( N^{1/2}/(k + N)^{1/2} \), which is the partition function of the \( U(1) \) factor. The final result is

\[ Z = \frac{1}{(k + N)^{N/2}} \prod_{\alpha > 0} 2 \sin \left( \frac{\pi (\alpha \cdot \rho)}{k + N} \right). \] (2.161)
Using the explicit description of the positive roots of SU($N$), one gets

$$F = \log Z = -\frac{N}{2} \log(k + N) + \sum_{j=1}^{N-1} (N - j) \log \left[ 2 \sin \frac{\pi j}{k + N} \right].$$  \hspace{1cm} (2.162)

We can now write the sin as

$$\sin \pi z = \pi z \prod_{n=1}^{\infty} \left( 1 - \frac{z^2}{n^2} \right),$$  \hspace{1cm} (2.163)

and we find that the free energy is the sum of two parts. We will call the first one the \textit{non-perturbative} part:

$$F^{np} = -\frac{N^2}{2} \log(k + N) + \frac{1}{2} N(N - 1) \log 2\pi + \sum_{j=1}^{N-1} (N - j) \log j,$$  \hspace{1cm} (2.164)

and the other part will be called the \textit{perturbative} part:

$$F^p = \sum_{j=1}^{N} (N - j) \sum_{n=1}^{\infty} \log \left[ 1 - \frac{j^2 g_s^2}{4\pi^2 n^2} \right],$$  \hspace{1cm} (2.165)

where we have denoted

$$g_s = \frac{2\pi}{k + N},$$  \hspace{1cm} (2.166)

which, as we will see later, coincides with the open string coupling constant under the gauge/string theory duality.

To see that (2.164) has a non-perturbative origin, we rewrite it as

$$F^{np} = \log \left( \frac{2\pi g_s}{\text{vol}(U(N))} \right)^{\frac{1}{2} N^2},$$  \hspace{1cm} (2.167)

where we used the explicit formula (1.41). This indeed corresponds to the volume of the gauge group in the one-loop contribution (2.18), where $A^{(c)}$ is in this case the trivial flat connection. Therefore, $F^{np}$ is the log of the prefactor of the path integral, which is not captured by Feynman diagrams.

Let us now work out the perturbative part (2.165), following Gopakumar and Vafa (1998a and 1999). By expanding the log, using that $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$ and the formula

$$\sum_{j=1}^{N} j^k = \frac{1}{k + 1} \sum_{l=1}^{k+1} (-1)^{k-l+1} \binom{k + 1}{l} B_{k+1-l} N^l,$$  \hspace{1cm} (2.168)

where $B_n$ are Bernoulli numbers, we find that (2.165) can be written as
\[ F^p = \sum_{g=0}^{\infty} \sum_{h=2}^{\infty} F^p_{g,h} g_s^{2g-2+h} N^h , \] (2.169)

where \( F^p_{g,h} \) is given by:

\[ F^p_{0,h} = - \frac{|B_{h-2}|}{(h-2)!}, \quad h \geq 4, \]

\[ F^p_{1,h} = \frac{1}{12} \frac{|B_h|}{h h!}. \] (2.170)

Notice that \( F^p_{0,h} \) vanishes for \( h \leq 3 \). For \( g \geq 2 \) one obtains

\[ F^p_{g,h} = \frac{\zeta(2g-2+h)}{(2\pi)^{2g-3+h}} \left( \frac{2g-3+h}{h} \right) \frac{B_{2g}}{2g(2g-2)}. \] (2.171)

This gives the contribution of connected diagrams with two loops and beyond to the free energy of Chern–Simons theory on the sphere. The non-perturbative part also admits an asymptotic expansion that can be easily worked out by expanding the Barnes function that appears in the volume factor (Periwal, 1993; Ooguri and Vafa 2002). Using the results in Section 1.1 one readily finds:

\[ F^{\text{np}} = F_G, \] (2.172)

where \( F_G \) is the free energy of the Gaussian matrix model written in (1.45).

In order to find \( F^G(t) \) we have to sum over the holes, as in (1.26). The 't Hooft parameter is given by \( t = xN = ig_s N \), and

\[ F^p_g(t) = \sum_{h=1}^{\infty} F^p_{g,h} (-it)^h. \] (2.173)

Let us first focus on \( g \geq 2 \). To perform the sum explicitly, we again write the \( \zeta \) function as \( \zeta(2g-2+2p) = \sum_{n=1}^{\infty} n^{2-2g-2p} \), and use the binomial series,

\[ \frac{1}{(1-z)^q} = \sum_{n=0}^{\infty} \binom{q+n-1}{n} z^n \] (2.174)

to obtain:

\[ F^p_g(t) = \frac{(-1)^g |B_{2g} B_{2g-2}|}{2g(2g-2)(2g-2)!} + \frac{B_{2g}}{2g(2g-2)} \sum_{n \in \mathbb{Z}}' \frac{1}{(-it + 2\pi n)^{2g-2}}, \] (2.175)

where \( ' \) means that we omit \( n = 0 \). Now we notice that, if we write

\[ F^{\text{np}} = \sum_{g=0}^{\infty} F^{\text{np}}_g(t) g_s^{2g-2}, \] (2.176)
then for, \( g \geq 2 \), \( F_{np}^g(t) = B_{2g}/(2g(2g - 2)(-it)^{2g-2} \), which is precisely the \( n = 0 \) term missing in (2.175). We then define:

\[
F_g(t) = F_p^g(t) + F_{np}^g(t).
\]

Finally, since

\[
\sum_{n \in \mathbb{Z}} \frac{1}{n + z} = \frac{2\pi i}{1 - e^{-2\pi i z}},
\]

by taking derivatives w.r.t. \( z \) we can write

\[
F_g(t) = \frac{(-1)^g|B_{2g}B_{2g-2}|}{2g(2g - 2)(2g - 2)!} + \frac{|B_{2g}|}{2g(2g - 2)!} \text{Li}_{3-2g}(e^{-t}),
\]

again for \( g \geq 2 \). The function \( \text{Li}_j \) appearing in this equation is the polylogarithm of index \( j \), defined by

\[
\text{Li}_j(x) = \sum_{n=1}^{\infty} \frac{x^n}{n^j}.
\]

The computation for \( g = 0, 1 \) is very similar, and one obtains:

\[
F_0(t) = -\frac{t^3}{12} + \frac{\pi^2 t}{6} + \zeta(3) + \text{Li}_3(e^{-t}),
\]

\[
F_1(t) = \frac{t}{24} + \frac{1}{12} \log (1 - e^{-t}).
\]

This gives the resummed functions \( F_g(t) \) for all \( g \geq 0 \). In the next section we will see how to derive this by using the matrix model technology developed in the previous chapter.

So far we have considered the \( 1/N \) expansion of the free energy. The vacuum expectation values of Wilson loops can be also analysed from this point of view. A simple field theory argument shows that the connected vacuum expectation value of \( h \) gauge-invariant operators has an expansion in powers of \( N^{2-2g+h} \) (see, for example, Coleman, 1988). This means that the correlation functions of Wilson loops that have a well-defined behaviour in the \( 1/N \) expansion are the connected vacuum expectation values \( W_{\bar{k}}^{(c)} \) introduced in (2.159). Their expansion is of the form,

\[
W_{\bar{k}}^{(c)} = \sum_{g=0}^{\infty} W_{\bar{k};g}^{(c)}(t)x^{2g-2+|\bar{k}|},
\]

where \( W_{\bar{k};g}^{(c)}(t) \) is a function of the 't Hooft parameter and \( |\bar{k}| \) is defined in (2.105).

### 2.8 Chern–Simons theory as a matrix model

As we have seen, Chern–Simons theory is exactly solvable and on certain three-manifolds one can write very explicit answers for the partition function. It turns
out that for certain rational homology spheres (including $S^3$), this partition function admits a matrix model representation. This was found in Mariño (2004a) based on previous work by Rozansky (1996) and Lawrence and Rozansky (1999), and was further developed in Aganagic et al. (2004). In the case of $S^3$ the easiest way to derive the matrix model representation of the Chern–Simons partition function is through direct computation. Consider the following integral:

$$Z_{CS} = \frac{e^{-\frac{x}{2}N(N^2-1)}}{N!} \int \prod_{i=1}^{N} \frac{d\beta_i}{2\pi} e^{-\sum_i \beta_i^2/2x} \prod_{i<j} \left(2 \sinh \frac{\beta_i - \beta_j}{2}\right)^2.$$

(2.183)

It can easily be seen that this reproduces the partition function of $U(N)$ Chern–Simons theory on $S^3$, given in (2.47), and the derivation is left as an exercise.

**Exercise 2.2** Use the Weyl formula (2.46) to write (2.183) as a Gaussian integral, and show that it reproduces (2.161).

The measure factor in (2.183)

$$\prod_{i<j} \left(2 \sinh \frac{\beta_i - \beta_j}{2}\right)^2$$

(2.184)

is not the standard Vandermonde determinant, although it reduces to it for small separations among the eigenvalues. In fact, for very small $x$, the Gaussian potential in (2.183) will be very narrow, forcing the eigenvalues to be close to each other, and one can expand the sinh in (2.184) in power series. At leading order, we find the usual Gaussian matrix model, while the corrections to it can be evaluated systematically by computing correlators in the Gaussian theory. In this way, one obtains the perturbative expansion of Chern–Simons theory; see Mariño (2004a) for details.

Here we will analyse the model by using the technology developed in Sections 1.2 and 1.3. First, we want to write the above integral as a standard matrix integral with the usual Vandermonde discriminant. This can be achieved with the change of variables (Forrester, 1994; Tierz, 2004)

$$\exp(\beta_i + t) = \lambda_i,$$

(2.185)

where $t = Nx$, as usual. It is easy to see that the above integral becomes, up to a factor $\exp(-N^3x/2)$,

$$Z_{SW} = \frac{1}{N!} \int \prod_{i=1}^{N} \frac{d\lambda_i}{2\pi} \Delta^2(\lambda) \exp\left(-\sum_{i=1}^{N} \left(\log \lambda_i\right)^2/2x\right),$$

(2.186)

therefore we are considering the matrix model

$$Z_{SW} = \frac{1}{\text{vol}(U(N))} \int dM \ e^{-\frac{1}{2x} \text{Tr} \left(\log M\right)^2}.$$

(2.187)

We will call this model the **Stieltjes–Wigert matrix model**, hence the subscript in (2.186) and (2.187). This is because it can be exactly solved with the so-called Stieltjes–Wigert polynomials, as we will explain in a moment.
We want to analyse now the saddle-point approximation to the matrix integral (2.183), or equivalently to (2.186). Since the model in (2.186) has the standard Vandermonde determinant, we can use the techniques of section 1.2. Although the formulae there were obtained for a polynomial potential, some of them generalize to arbitrary polynomials. In particular, to obtain the resolvent $\omega_0(p)$ we can use the formula (1.67) with

$$W'(z) = \frac{\log z}{z}.$$  

(2.188)

Notice that this potential has a minimum at $z = 1$. We then expect a one-cut solution where the endpoints of the interval $a(t), b(t)$ will satisfy $a(0) = b(0) = 1$. In order to compute the integral (1.67) with $W'(z)$ given in (2.188), we deform the integration contour. In the case of polynomial potentials, we picked one residue at $z = p$ and another one at infinity. Here, since the logarithm has a branch cut, we cannot push the contour to infinity. Instead, we deform the contour as indicated in Fig. 2.12: we pick the pole at $z = p$, and then we surround the cut of the logarithm along the negative real axis and the singularity at $z = 0$ with a small circle $C_\epsilon$ of radius $\epsilon$. This kind of situation is typical of the solution of matrix models with the character expansion (Kazakov et al., 1996). The resulting integrals are:

$$\frac{1}{2t} \left\{ - \int_{-\infty}^{-\epsilon} \frac{dz}{z(z-p)\sqrt{(z-a)(z-b)}} + \int_{C_\epsilon} \frac{dz \log z}{z(z-p)\sqrt{(z-a)(z-b)}} \right\}. \quad (2.189)$$

Both are singular as $\epsilon \to 0$, but the singularities cancel, and after some computations one finds for the resolvent:

\[ \text{Fig. 2.12. This shows the deformation of the contour needed to compute the planar resolvent of the Chern–Simons matrix integral. We pick a residue at } z = p, \text{ and we have to encircle the singularity at the origin as well as the branch cut of the logarithm, which on the left-hand side is represented by the dashed lines.} \]
\[
\omega_0(p) = -\frac{1}{2tp} \log \left[ \frac{(\sqrt{a}\sqrt{p-b} - \sqrt{b}\sqrt{p-a})^2}{(\sqrt{p-a} - \sqrt{p-b})^2} \right] + \frac{\sqrt{(p-a)(p-b)}}{2tp\sqrt{ab}} \log \left[ \frac{4ab}{2\sqrt{ab} + a + b} \right].
\]

(2.190)

In order to satisfy the asymptotics (1.64) the second term must vanish, and the first one must go as \(1/p\). This implies

\[
4ab = 2\sqrt{ab} + a + b,
\]

\[
\sqrt{a} + \sqrt{b} = 2e^t,
\]

and from here we obtain the positions of the endpoints of the cut \(a, b\) as a function of the 't Hooft parameter:

\[
a(t) = 2e^{2t} - e^t + 2e^{\frac{3t}{2}}\sqrt{e^t - 1},
b(t) = 2e^{2t} - e^t - 2e^{\frac{3t}{2}}\sqrt{e^t - 1}.
\]

(2.192)

Notice that, for \(t = 0\), \(a(0) = b(0) = 1\), as expected. The final expression for the resolvent is then

\[
\omega_0(p) = -\frac{1}{tp} \log \left[ \frac{1 + e^{-t}p + \sqrt{(1 + e^{-t}p)^2 - 4p}}{2p} \right],
\]

(2.193)

and from here we can easily find the density of eigenvalues

\[
\rho(\lambda) = \frac{1}{\pi t\lambda} \tan^{-1} \left[ \frac{\sqrt{4\lambda - (1 + e^{-t}\lambda)^2}}{1 + e^{-t}\lambda} \right].
\]

(2.194)

If we now define

\[
u(p) = t(1 - p\omega_0(p)) + \pi i,
\]

(2.195)

we see that it solves the equation

\[
e^u + e^v + e^{v-u+t} + 1 = 0,
\]

(2.196)

where we put \(p = e^{t-v}\). This was first found by Aganagic et al. (2004) by a similar analysis. Equation (2.196) is the analogue of (1.77) in the case of polynomial matrix models, and can be regarded as an algebraic equation describing a non-compact Riemann surface.

As we mentioned before, the matrix model (2.187) can be solved exactly with a set of orthogonal polynomials called the Stieltjes–Wigert polynomials. The fact
that the Chern–Simons matrix model is essentially equivalent to the Stieltjes-Wigert matrix model was pointed out by Tierz (2004). The Stieltjes–Wigert polynomials are defined as follows (Szegö, 1959):

\[ p_n(\lambda) = (-1)^n q^{n^2 + \frac{n}{2}} \sum_{\nu=0}^{n} \left( \begin{array}{c} n \\ \nu \end{array} \right) q^{\frac{\nu^2 - \nu^2}{2}} (-q^{-\frac{1}{2}} \lambda)^\nu, \]  

(2.197)

and satisfy the orthogonality condition (1.94) with

\[ d_\mu(\lambda) = e^{-\frac{1}{2\pi} (\log \lambda)^2} \frac{d\lambda}{2\pi} \]  

(2.198)

and

\[ h_n = q^{\frac{3}{4} n (n+1) + \frac{1}{2} \cdot \frac{[n]}{[n]! \cdot [n - m]!}}. \]

(2.199)

The recursion coefficients appearing in (1.99) are, in this case,

\[ r_n = q^{3n} (q^n - 1), \quad s_n = -q^{\frac{1}{2} + n} (q^{n+1} + q^n - 1). \]

One can easily check that the normalized vacuum expectation value of Tr$_R$ $M$ in this ensemble is given by

\[ \langle \text{Tr}_R M \rangle_{SW} = e^{\frac{3\ell(R)}{4} q^{\frac{n}{2}} \text{dim}_q R}, \]  

(2.200)

where $\ell(R)$ is the number of boxes of $R$, and $\kappa_R$ is defined in (2.84).

Notice that, for this set of orthogonal polynomials, the expansion (1.114) is very simple since

\[ R_0(\xi) = e^{4t\xi} (1 - e^{-t\xi}), \quad R_{2s}(\xi) = 0, \quad s > 0, \quad s(\xi) = e^{t\xi} (1 - 2e^{t\xi}). \]

(2.201)

As we pointed out in Section 2.3, $R_0(\xi)$ and $s(\xi)$ can be used to determine the endpoints of the cut in the resolvent through (1.118). It is easy to see that (2.201) indeed lead to (2.192), and that by using (1.117) one obtains (2.194). In fact, it is well known that the expression (2.194) is the density of zeroes of the Stieltjes-Wigert polynomials (Kuijlaars and Van Assche, 1999; Chen and Lawrence, 1998).
We can now use the technology developed in Section 2.3 to compute $F_g(t)$. Since
\[ F_{CS} = F_{SW} - \frac{7}{12} t^3 + \frac{1}{12} t, \]
the formula (1.115) gives
\[ F_{CS}^0(t) = \frac{t^3}{12} - \frac{\pi^2 t}{6} - \text{Li}_3(e^{-t}) + \zeta(3), \]
where the polylogarithm of index $j$ was defined in (2.180). The above result is in precise agreement with the result obtained in the previous section by resumming the perturbative series, up to an overall minus sign that comes from the fact that the expansion variable here is $x$, while in Section 2.7 it was $g_s = -ix$. With some extra work we can also compute $F_g(t)$, for all $g > 0$, starting from (1.119). We just have to compute $f^{(p)}(1) - f^{(p)}(0)$, for $p$ odd, where
\[ f(\xi) = (1 - \xi)\phi(\xi, t), \quad \phi(\xi, t) = \log \frac{1 - e^{-t\xi}}{\xi} + 4t\xi. \]
It is easy to show that
\[ \phi^{(p)}(\xi, t) = (-1)^{p+1} \left\{ \text{Li}_{1-p}(e^{-t\xi}) t^p - \frac{(p-1)!}{\xi^p} \right\}, \]
and by using the expansion
\[ \frac{1}{1 - e^{-t}} = \frac{1}{t} + \sum_{k=0}^{\infty} (-1)^{k+1} B_{k+1} \frac{t^k}{(k+1)!} \]
one gets
\[ \phi^{(p)}(0, t) = \frac{(-1)^p B_g}{p} t^p. \]
Putting everything together, we find for $g > 1$
\[ F_g(t) = \frac{B_{2g} B_{2g-2}}{2g(2g-2)(2g-2)!} + \frac{B_{2g}}{2g(2g-2)!} \text{Li}_{3-2g}(e^{-t}) - \frac{B_{2g}}{2g(2g-2)} t^{2-2g}. \]
Since the last part is the free energy at genus $g$ of the Gaussian model, we conclude that the free energy $F_g^{CS}(t)$ of the Chern–Simons matrix model at genus $g$ is given by
\[ F_g^{CS}(t) = \frac{B_{2g} B_{2g-2}}{2g(2g-2)(2g-2)!} + \frac{B_{2g}}{2g(2g-2)!} \text{Li}_{3-2g}(e^{-t}), \]
which agrees with the resummation (2.179) up to a factor $(-1)^{g-1}$ that comes from the fact that the expansion variable we are considering here is $x$ instead of $g_s$. 
Before closing this chapter, we would like to mention that the matrix model techniques that we introduced in this section can be used to compute the perturbation expansion and large-$N$ behaviour of the partition function of Chern–Simons theory in more complicated three-manifolds (Mariño, 2004a; Aganagic et al., 2004; Halmagyi and Yasnov, 2003).
Part II

Topological strings
This page intentionally left blank
3

TOPOLOGICAL SIGMA MODELS

String theory can be regarded, at the algebraic level, as a two-dimensional conformal field theory coupled to two-dimensional gravity. When the conformal field theory is also a topological field theory (i.e. a theory whose correlation functions do not depend on the metric on the Riemann surface), the resulting string theory turns out to be very simple and in many cases can be completely solved. A string theory that is constructed in this way is called a topological string theory.

The starting point to obtain a topological string theory is therefore a conformal field theory with topological invariance. Such theories are called topological conformal field theories and can be constructed out of $\mathcal{N} = 2$ superconformal field theories in two dimensions by a procedure called twisting. In this book, we will consider a class of topological string theories in which the topological field theory is taken to be a topological sigma model with target space a Calabi–Yau manifold. We will first review the $\mathcal{N} = 2$ supersymmetric sigma model, and then we will introduce the twisting procedure. In two dimensions there are two possible twists, leading respectively to the A-type and the B-type topological sigma models, which we study in detail in this chapter. More details on topological sigma models can be found in the book by Hori et al. (2003).

3.1 The $\mathcal{N} = 2$ supersymmetric sigma model

In this section, we give a brief summary of the $\mathcal{N} = 2$ supersymmetric sigma model. We will follow very closely the presentation of Labastida and Llatas (1992).

Let us first state our notation and conventions. We will be considering two-dimensional quantum field theories, with Euclidean signature, defined on a Riemann surface $\Sigma_g$ with co-ordinates $x^1, x^2$. We will also use complex co-ordinates $z = x^1 + i x^2$, $\bar{z} = x^1 - i x^2$. Locally, we can always find a flat Euclidean metric whose components are given in complex co-ordinates by

$$g_{zz} = g_{\bar{z}\bar{z}} = \frac{1}{2}, \quad g_{z\bar{z}} = g_{\bar{z}z} = 0. \quad (3.1)$$

The epsilon symbol is $\epsilon^{\bar{z}z} = -\epsilon^{z\bar{z}} = 2i$. Our choice of Euclidean Dirac matrices $\gamma^\mu$ is

$$(\gamma^1)_{\alpha\beta} = \sigma^1, \quad (\gamma^2)_{\alpha\beta} = \sigma^2,$$

where $\sigma^1, \sigma^2$ are Pauli matrices. We will denote the spinor indices by $\alpha = +, -, -$ and they are lowered and raised by the matrix $C_{\alpha\beta} = \sigma^1$, so that $(\gamma^1)_{\alpha\beta} = 1_{\alpha\beta}$, $(\gamma^2)_{\alpha\beta} = (\sigma^2)_{\alpha\beta}$. 

71
The generators of $\mathcal{N} = 2$ supersymmetry are denoted by $Q_{\alpha a}$, where $\alpha = +, -$ are Lorentz indices, and $a = +, -$ are R-charge indices. The $\mathcal{N} = 2$ supersymmetry algebra contains the following relations:

$$\{Q_{\alpha +}, Q_{\beta -}\} = \gamma^\mu_{\alpha\beta} P_\mu,$$

$$\{Q_{\alpha \pm}, Q_{\beta \pm}\} = 0,$$

$$[J, Q_{\pm a}] = \pm \frac{1}{2} Q_{\pm a},$$

$$[F_L, Q_{\pm \pm}] = 0,$$

$$[F_R, Q_{\pm \pm}] = 0,$$

$$[F_R, Q_{-\pm}] = \pm \frac{1}{2} Q_{-\pm}.$$  \hfill (3.2)

We have assumed that there are no central charges in the algebra. $J$ is the generator of Lorentz $SO(2)$ transformations and $F_{L,R}$ are the left and right internal $U(1)$ currents, respectively. They combine into vectorial and axial currents $F_V, F_A$ as follows:

$$F_V = F_L + F_R, \quad F_A = F_L - F_R.$$  \hfill (3.3)

We will consider here a theory for an $\mathcal{N} = 2$ chiral multiplet. The most convenient description of this multiplet is by using $\mathcal{N} = 2$ superspace in two dimensions, which is essentially identical to the usual $\mathcal{N} = 1$ superspace in four dimensions (a useful introduction to $\mathcal{N} = 1$ superspace in four dimensions can be found in Wess and Bagger, 1992; a detailed presentation of $\mathcal{N} = 2$ superspace in two dimensions appears in Hori et al., 2003). In $\mathcal{N} = 2$ superspace we have superspace covariant derivatives $D_{\alpha a}$ satisfying the following algebra:

$$\{D_{\alpha +}, D_{\beta -}\} = 2\partial_z,$$

$$\{D_{\alpha -}, D_{\beta +}\} = 2\partial_{\bar{z}},$$  \hfill (3.4)

while all other anti-commutators among the $D_{\alpha a}$ vanish. The chiral multiplet is described by a scalar $\mathcal{N} = 2$ superfield $\Phi$ satisfying the following relations:

$$D_{+, -} \Phi = D_{-, +} \Phi = 0.$$  \hfill (3.5)

Anti-chiral multiplets $\bar{\Phi}$ satisfy, on the contrary,

$$D_{+, +} \Phi = D_{-, +} \Phi = 0.$$  \hfill (3.6)

Let us now consider a collection of $d$ chiral superfields $\Phi^I$, and $d$ anti-chiral superfields $\Phi^T$, where $I, T = 1, \cdots, d$. We can define component fields for these superfields as follows:

$$\Phi^I | = x^I,$$

$$D_{\alpha +} \Phi^I | = \psi^I_{\alpha +},$$

$$D_{\alpha -} \Phi^I | = \psi^I_{\alpha -},$$

$$D_{-, +} D_{+, +} \Phi^I | = F^I_{-+, +},$$

$$D_{+, -} D_{-, -} \Phi^T | = F^T_{+, -, -},$$

where the vertical bar means that we take the component of the superfield with $\theta = 0$. Here, $F^I_{-+, +}$ and $F^T_{+, -, -}$ are auxiliary fields. We can also write down
very easily the supersymmetry transformations of the different component fields under the $\mathcal{N} = 2$ supersymmetry. For an $\mathcal{N} = 2$ superfield the transformation rule takes the form
\[ \delta \Phi = \eta^{\alpha a} Q_{\alpha a} \Phi, \] (3.8)
where $\eta^{\alpha a}$ is a constant $\mathcal{N} = 2$ supersymmetry parameter. Projecting onto components, and using the definitions (3.7), one finds
\[ \delta x^I = \eta^{++} \psi^I_{++,} + \eta^{-+} \psi^I_{-,+}, \]
\[ \delta \psi^I_{++,} = \eta^{++} F^I_{++,++} + 2 \eta^{++} \partial_\bar{z} x^I, \]
\[ \delta \psi^I_{-,+} = -\eta^{++} F^I_{-,++,} + 2 \eta^{-+} \partial_\bar{z} x^I, \]
\[ \delta F^I_{++,++} = 2 \eta^{-+} \partial_\bar{z} \psi^I_{++,+} - 2 \eta^{++} \partial_\bar{z} \psi^I_{-,+}, \]
\[ \delta x^T = \eta^{--} \psi^T_{--,} + \eta^{-+} \psi^T_{-+-}, \]
\[ \delta \psi^T_{--,} = \eta^{--} F^T_{--,--} + 2 \eta^{--} \partial_\bar{z} x^T, \]
\[ \delta \psi^T_{-+-} = -\eta^{--} F^T_{-+-,--} + 2 \eta^{-+} \partial_\bar{z} x^T, \]
\[ \delta F^T_{-+-,--} = 2 \eta^{--} \partial_\bar{z} \psi^T_{-+-,} - 2 \eta^{-+} \partial_\bar{z} \psi^T_{-+-}. \]
(3.9)

The transformations under the $R$-symmetry can be read off from the $U(1)$ indices of the fields.

3.2 Topological twist

The supersymmetric sigma model that we defined in the last section can be twisted in two different ways to produce two inequivalent topological quantum
Table 3.1 Quantum numbers of $Q_{\alpha a}$ under the different $U(1)$ symmetries.

<table>
<thead>
<tr>
<th></th>
<th>$U(1)_V$</th>
<th>$U(1)_A$</th>
<th>$U(1)_E$</th>
<th>A-twist $U(1)_E'$</th>
<th>B-twist $U(1)_E'$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Q_{++,}$</td>
<td>$+1/2$</td>
<td>$+1/2$</td>
<td>$+1/2$</td>
<td>0</td>
<td>$+1$</td>
</tr>
<tr>
<td>$Q_{--,}$</td>
<td>$+1/2$</td>
<td>$-1/2$</td>
<td>$-1/2$</td>
<td>$-1$</td>
<td>$-1$</td>
</tr>
<tr>
<td>$Q_{+-,}$</td>
<td>$-1/2$</td>
<td>$-1/2$</td>
<td>$+1/2$</td>
<td>$+1$</td>
<td>0</td>
</tr>
<tr>
<td>$Q_{-,+}$</td>
<td>$-1/2$</td>
<td>$+1/2$</td>
<td>$-1/2$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Field theories in two dimensions. These two inequivalent twisting procedures are called the A-twist and the B-twist, and they give rise to the topological type-A sigma model and the topological type-B sigma model, respectively. The A-twisting was introduced by Witten (1988), while the B-twisting was introduced by Vafa (1991), Labastida and Llatas (1992) and Witten (1991a).

The twisting procedure amounts to a redefinition of the spin of the fields (equivalently, of the energy-momentum tensor of the theory) by using the internal $F_V$ or $F_A$ currents. In the A-twist, one redefines the spin current as follows:

$$\tilde{J} = J - F_V,$$

while in the B-twist the redefinition is given by

$$\tilde{J} = J + F_A.$$

The choice of sign is made so as to agree with the conventions of Witten (1991a). The above redefinition means that we are replacing the $U(1)_E$ Lorentz symmetry by the diagonal embedding $U(1)_E' \subset U(1)_E \times U(1)_{V,A}$ for the A- and the B-twist, respectively. It is very illuminating to make a table where we write down the quantum numbers of all the components of $Q_{\alpha a}$ under the different symmetries involved. We see that in both cases one gets two scalar supercharges, and one vector-valued supercharge. This suggests defining the following operator, also called the topological charge,

$$\begin{align*}
A - \text{twist} : & \quad Q = Q_{+,+} + Q_{-,-}, \\
B - \text{twist} : & \quad Q = Q_{+-} + Q_{-,+},
\end{align*}$$

which is a scalar in the resulting theories. $Q$ is a Grassmannian, scalar charge, so twisted theories violate the spin-statistics theorem. We will also define a vector charge $G_\mu$ through the following equations:

$$\begin{align*}
A - \text{twist} : & \quad G_z = Q_{+,+}, \quad G_{\bar{z}} = Q_{-,-}, \\
B - \text{twist} : & \quad G_z = Q_{+-}, \quad G_{\bar{z}} = Q_{-,+}.
\end{align*}$$

One can check that, as a consequence of the supersymmetry algebra, the topological charge is nilpotent,
\[ \mathcal{Q}^2 = 0, \]  
(3.16)

and also that

\[ \{ \mathcal{Q}, G_\mu \} = P_\mu. \]  
(3.17)

These are the most important relations characterizing the so-called topological algebra that is obtained by the twist of the \( \mathcal{N} = 2 \) algebra.

The twisting procedure can be interpreted in a more physical way as a gauging of a global \( U(1) \) current: when twisting we are, in fact, taking a \( U(1) \) conserved global current of the model and coupling it to the spin connection on the two-dimensional Riemann surface. In the original, untwisted models, we have two global currents, the vectorial and the axial \( U(1) \) currents, and gauging one or the other leads to type-A or type-B topological theories. To see how this works in more detail, let us consider the \( \mathcal{N} = 2 \) supersymmetric sigma model, which involves \( d \) chiral multiplets \( X^I \) and \( d \) anti-chiral multiplets \( X^I \). In particular, the model contains a bosonic field \( x : \Sigma_g \rightarrow X \) and two Dirac spinors

\[
\psi^{I}_{\pm,+} \in \Gamma(\Sigma_g, S^\pm \otimes \phi^*(TX)), \quad \psi^{\bar{I}}_{\pm,-} \in \Gamma(\Sigma_g, S^\pm \otimes \phi^*(TX)),
\]  
(3.18)

where \( S^\pm \) are the positive and negative chirality spinor bundles, respectively, and \( TX \) is the holomorphic tangent bundle of \( X \). The kinetic fermion term in the action (3.10) is

\[
S_f = \int_{\Sigma_g} d^2 z G_{\bar{I}J}(\psi^{\bar{I}}_{+,+} D_{\bar{z}} \psi^{J}_{+,+} + \psi^{\bar{I}}_{-,+} D_{\bar{z}} \psi^{J}_{-,+}),
\]  
(3.19)

where the covariant derivatives are given in local co-ordinates by

\[
D_{\bar{z}} \psi^{J}_{+,+} = \partial_{\bar{z}} \psi^{J}_{+,+} + \frac{i}{2} \omega_{\bar{z}} \psi^{J}_{+,+} + \Gamma^J_{KL} \partial_{\bar{z}} x^K \psi^{L}_{+,+},
\]

(3.20)

\[
D_{\bar{z}} \psi^{J}_{-,+} = \partial_{\bar{z}} \psi^{J}_{-,+} - \frac{i}{2} \omega_{\bar{z}} \psi^{J}_{-,+} + \Gamma^J_{KL} \partial_{\bar{z}} x^K \psi^{L}_{-,+}.
\]

In this equation, \( \omega_{\bar{z},\bar{z}} \) are the components of the spin connection on \( \Sigma_g \). This theory has a conserved, non-anomalous vector current with components

\[
j_{V}^{\bar{I}} = G_{\bar{I}J} \psi^{\bar{I}}_{-,+} \psi^{J}_{-,+}, \quad j_{V}^{\bar{I}} = 2G_{\bar{I}J} \psi^{\bar{I}}_{+,+} \psi^{J}_{-,+},
\]  
(3.21)

and an anomalous axial current with components

\[
j_{A}^{\bar{I}} = -G_{\bar{I}J} \psi^{\bar{I}}_{-,+} \psi^{J}_{-,+}, \quad j_{A}^{\bar{I}} = G_{\bar{I}J} \psi^{\bar{I}}_{+,+} \psi^{J}_{+,+}.
\]  
(3.22)

The anomaly in the axial current is given by the index of the Dirac operator and reads

\[
\int_{\Sigma_g} x^*(c_1(X)),
\]  
(3.23)

where \( c_1(X) \) is the first Chern class of \( X \). The gauging of the Abelian global symmetry involves promoting it to a worldsheet spacetime symmetry, so we have
to add to the original Lagrangian the coupling of the corresponding currents to
the worldsheet spin connection. For the A-twist we have

\[ S_f - \frac{1}{2} \int_{\Sigma_g} \, d^2 z \omega_\mu J^\mu_A = \int_{\Sigma_g} \, d^2 z G_{IJ} \{ \psi^I_+,-(\partial_\bar{z}\psi^J_+,+) + \Gamma^I_{KL} \partial_\bar{z} x^K \psi^L_+,+ \} \]

\[ \quad + \psi^I_+,-(\partial_\bar{z}\psi^J_+,+ - i \omega_\bar{z}\psi^J_+,+) + \Gamma^I_{KL} \partial_\bar{z} x^K \psi^L_-,+) \}, \quad (3.24) \]

while for the B-twist we have

\[ S_f + \frac{1}{2} \int_{\Sigma_g} \, d^2 z \omega_\mu J^\mu_A = \int_{\Sigma_g} \, d^2 z G_{IJ} \{ \psi^I_+,-(\partial_\bar{z}\psi^J_+,+) + i \omega_\bar{z}\psi^J_+,+ + \Gamma^I_{KL} \partial_\bar{z} x^K \psi^L_+,+ \}
\]

\[ \quad + \psi^I_+,-(\partial_\bar{z}\psi^J_+,+ - i \omega_\bar{z}\psi^J_+,+) + \Gamma^I_{KL} \partial_\bar{z} x^K \psi^L_-,+) \}. \quad (3.25) \]

We see that, in the twisted models, the fermion fields have changed their spin content: in the type-A model, \( \psi^J_+,- \) is a scalar field, while \( \psi^J_-,+ \) is a \((0,1)\)-form.

In the type-B model, on the contrary, \( \psi^J_+,- \) is a \((1,0)\)-form, while \( \psi^J_-,+ \) is a \((0,1)\)-form.

This analysis of the twist as the gauging of a global current also manifests an important issue: if we gauge an anomalous global current, the resulting model will inherit it as a gauge/gravitational anomaly, making the twisted model ill-defined. Since the vector current is non-anomalous, the resulting type-A model is well-defined for an arbitrary Kähler target \( X \). However, in the case of the type-B model the axial current has an anomaly given by (3.23), therefore the resulting model is ill-defined unless \( c_1(X) = 0 \). Kähler manifolds satisfying this condition are called Calabi–Yau manifolds, and they will play a crucial role in the following.

As we will see in a moment, twisted \( N = 2 \) theories are examples of topological quantum field theories of the cohomological or Witten type, in the classification scheme of Birmingham et al. (1991) that we mentioned at the beginning of Chapter 2. A topological quantum field theory of the Witten type is a quantum field theory defined on a manifold \( M \) that has an underlying scalar symmetry \( \delta \) acting on the fields \( \phi \), in such a way that the action of \( \delta \) is a full symmetry.

The symmetry \( \delta \) is usually called a topological symmetry. In (3.27) we have assumed that the symmetry \( \delta \) is not anomalous, so that it is a full symmetry.
of the quantum theory. Moreover, we have ‘integrated by parts’ in field space, therefore we have assumed that there are no contributions coming from boundary terms. In some situations these assumptions do not hold, and the theory is not strictly topological.

In a cohomological theory, the observables are the $\delta$-invariant operators. On the other hand, operators that are $\delta$-exact decouple from the theory, since their correlation functions vanish. We will then restrict the set of observables to the cohomology of $\delta$:

$$\mathcal{O} \in \frac{\text{Ker } \delta}{\text{Im } \delta}. \quad (3.28)$$

In all known examples of cohomological field theories, $\delta$ is a Grassmannian symmetry, and in the examples we will consider

$$\delta^2 = 0. \quad (3.29)$$

In general one can have $\delta^2 = Z$, where $Z$ is a global symmetry of the theory, see for example Labastida and Mariño (1997) for a detailed discussion of this.

The structure of topological quantum field theories of the Witten type leads immediately to a procedure for constructing non-local observables starting from local ones. Let us suppose that we have found an operator $\phi^{(0)}$ which is in the cohomology of $\delta$, as well as a series of operators $\phi^{(n)}$, $n = 1, \cdots, \dim M$, that are differential forms of degree $n$ on $M$ such that,

$$d\phi^{(n)} = \delta\phi^{(n+1)}, \quad n \geq 0. \quad (3.30)$$

In this equation, $d$ denotes the exterior derivative on $M$. The operators $\phi^{(n)}$ are called the topological descendants of $\phi^{(0)}$. It is easy to see that the operator

$$W_{\phi^{(0)}}^{(\gamma_n)} = \int_{\gamma_n} \phi^{(n)}, \quad (3.31)$$

where $\gamma_n \in H_n(M)$, is a topological observable:

$$\delta W_{\phi^{(0)}}^{(\gamma_n)} = \int_{\gamma_n} \delta\phi^{(n)} = \int_{\gamma_n} d\phi^{(n-1)} = \int_{\partial\gamma_n} \phi^{(n-1)} = 0, \quad (3.32)$$

since $\partial\gamma_n = 0$. Similarly, it is easy to show if $\gamma_n$ is trivial in homology (i.e. if it is $\partial$-exact), then $W_{\phi^{(0)}}^{(\gamma_n)} = 0$ is $\delta$-exact. Equations (3.30) are called descent equations. The conclusion of this analysis is that, given a (scalar) topological observable $\phi^{(0)}$ and a solution to the descent equations (3.30), one can construct a family of topological observables:

$$W_{\phi^{(0)}}^{(\gamma_n)} i_n = 1, \cdots, b_n; \quad n = 1, \cdots, \dim M, \quad (3.33)$$

in one-to-one correspondence with the homology classes of the spacetime $M$. 
It is easy to see that in any theory where (3.26) is satisfied there is a simple procedure to construct a solution to (3.30) given a scalar observable \( \phi^{(0)} \). If (3.26) holds, then one has:

\[
P_\mu = T_{0\mu} = \delta G_\mu,
\]

(3.34)

where

\[
G_\mu \equiv G_{0\mu}.
\]

(3.35)

Since \( \delta \) is a Grassmannian symmetry, \( G_\mu \) is an anti-commuting operator and a one-form in spacetime. If we are given a \( \delta \)-invariant operator \( \phi^{(0)}(x) \), we can use (3.35) to construct

\[
\phi^{(n)}_{\mu_1 \mu_2 \cdots \mu_n}(x) = G_{\mu_1} G_{\mu_2} \cdots G_{\mu_n} \phi^{(0)}(x),
\]

(3.36)

where \( n \leq 1, \cdots, \text{dim} M \). On the other hand, since the \( G_{\mu_i} \) anti-commute,

\[
\phi^{(n)} = \frac{1}{n!} \phi^{(n)}_{\mu_1 \mu_2 \cdots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n},
\]

(3.37)

is an \( n \)-form on \( M \). By using (3.34), the \( \delta \)-invariance of \( \phi^{(0)} \), as well as \( P_\mu = \partial_\mu \), one can easily check that these forms satisfy the descent equations (3.30). This solution to (3.30) is usually called the canonical solution to the descent equations.

We can now make a connection between this general structure and the twisting procedure. The nilpotent, scalar, Grassmannian symmetry \( \delta \) is provided by the topological charge \( Q \). In order to prove that the theory is indeed topological of the cohomological type we have to show that (3.26) holds, and this depends on the model under consideration. Notice, however, that the structure of the twisted supersymmetry algebra gives the relation (3.17), which provides a partial fulfillment of (3.26). Indeed, the operator defined in (3.35) also appears in the twisted supersymmetry algebra as the vector-valued Grassmannian charge defined in (3.15). Our conclusion is that twisted \( \mathcal{N} = 2 \) supersymmetric theories in two dimensions have the right ingredients to be topological field theories of the cohomological type. In the next sections we will see that, indeed, for the twisted sigma model, the key relation (3.26) holds, so the resulting theories are topological.

### 3.3 The topological type-A model

The topological type-A model is obtained after doing the A-twist of the \( \mathcal{N} = 2 \) supersymmetric sigma model. After the twisting, the fermionic and auxiliary fields of the supersymmetric sigma model change their spin, and we will rename them as follows:

\[
\chi^I = \psi^I_{+,+}, \quad \rho^I_z = \psi^I_{-,+}, \quad F^I_{-} = F^I_{-+,+-}
\]

\[
\chi^I = \psi^I_{-}, \quad \rho^I_z = \psi^I_{+}, \quad F^I_{+} = F^I_{+-,-}.
\]

(3.38)

We will denote real indices in the tangent space of the target manifold \( X \) by \( i \), with \( i = 1, \cdots, 2d \). The field content of the type-A topological sigma model is
then the following. First, we have a map \( x : \Sigma_g \rightarrow X \) from a Riemann surface of genus \( g \) to a Kähler manifold \( X \) of complex dimension \( d \). We also have Grassmann fields \( \chi \in x^*(TX) \), which are scalars on \( \Sigma_g \), and a Grassmannian one-form \( \rho_\alpha \) with values in \( x^*(TX) \). This last field satisfies a self-duality condition that implies that its only non-zero components are \( \rho^I_\top \in x^*(T^{(1,0)}X) \) and \( \rho^I_\top \in x^*(T^{(0,1)}X) \).

The action for the theory is

\[
S_A = \int_{\Sigma_g} d^2z \left[ G_{IJ}( - F^I_\top F^J_\top - 2 \rho^I_\top D_z \chi^J - 2 \rho^J_\top D_z \chi^I + 4 \partial_x x^I \partial_z x^J ) + \partial_K \partial_Z G_{IJ} \rho^K_\top \rho^K_\top \chi^I \chi^J + \partial_K G_{IJ} \chi^I F^J_\top \rho^K_\top + \partial_K G_{IJ} \chi^J F^I_\top \rho^K_\top \right].
\]

(3.39)

The topological charge \( Q \) acts as follows on the fields:

\[
[Q, x^i] = \chi^i, \quad \{Q, \rho^I_\top \} = 2 \partial_x x^I - F^I_\top, \quad [Q, F^I_\top] = 2 \partial_x \chi^I, \quad [Q, F^I_\top] = 2 \partial_z \chi^I, \quad [Q, F^I_\top] = 2 \partial_z \chi^I.
\]

In the twisted theory we also have a \( U(1) \) ghost number symmetry that comes from the \( U(1)_A \) symmetry of the untwisted model. The ghost numbers of the fields \( x, \chi \) and \( \rho, F \) are 0, 1, \(-1\) and 0, respectively. Notice that the Grassmannian charge \( Q \) then has ghost number 1.

The formulation presented in (3.39) and (3.40) is not very convenient, since the action and transformations are not covariant with respect to the target metric. To improve this we redefine the auxiliary fields as follows:

\[
\tilde{F}^I_\top = F^I_\top - \Gamma^I_{JK} \chi^J \rho^K_\top, \quad \tilde{F}^I_\top = F^I_\top - \Gamma^I_{JK} \chi^J \rho^K_\top.
\]

(3.40)

The action (3.39) then becomes

\[
S = \int_{\Sigma_g} d^2z \left[ G_{IJ}(4 \partial_x x^I \partial_z x^J - 2 \rho^I_\top D_z \chi^J - 2 \rho^J_\top D_z \chi^I - \tilde{F}^I_\top \tilde{F}^I_\top ) + R_{IJKL} \rho^K_\top \rho^K_\top \chi^I \chi^J \chi^L \chi^L \right].
\]

(3.41)

The \( Q \)-transformations now read as follows:

\[
[Q, x^i] = \chi^i, \quad [Q, \chi^i] = 0, \quad [Q, \rho^I_\top] = 2 \partial_x x^I - \tilde{F}^I_\top - \Gamma^I_{JK} \chi^J \rho^K_\top, \quad [Q, \rho^I_\top] = 2 \partial_x x^I - \tilde{F}^I_\top - \Gamma^I_{JK} \chi^J \rho^K_\top,
\]

\[
[Q, \tilde{F}^I_\top] = 2 \partial_x \chi^I - \Gamma^I_{JK} \chi^J \tilde{F}^K_\top + R^I_{JKL} \chi^K \chi^L \rho^K_\top, \quad [Q, \tilde{F}^I_\top] = 2 \partial_x \chi^I - \Gamma^I_{JK} \chi^J \tilde{F}^K_\top + R^I_{JKL} \chi^K \chi^L \rho^K_\top.
\]

(3.42)

Notice that, by construction, one has \( Q^2 = 0 \). Geometrically, if we interpret \( \chi^i \) as the basis \( dx^i \) of differential forms on \( X \), we see that \( Q \) acts on \( x^i, \chi^i \) like...
the de Rham differential operator on the target space $X$. This analogy can be made more precise in the context of the so-called Mathai–Quillen formalism of topological field theory; see Cordes et al. (1994), for a detailed exposition of this formalism and its application to the type-A model.

We will now show that the above action and $Q$-transformations lead, in fact, to a topological field theory. First, we covariantize with respect to the two-dimensional metric to obtain:

$$S_A = \int_{\Sigma_g} d^2z\sqrt{g} \left[ G_{I\bar{J}} (g^{\mu\nu} \partial_{\mu} x^I \partial_{\nu} x^\bar{J} + \frac{i\epsilon^{\mu\nu}}{\sqrt{g}} \partial_{\mu} x^I \partial_{\nu} x^\bar{J} - g^{\mu\nu} \rho^I_\mu D_{\nu} \chi^\bar{J} - g^{\mu\nu} \rho^\bar{J}_\mu D_{\nu} \chi^I - \frac{1}{2} g^{\mu\nu} \rho^J_\mu \rho^I_\nu \chi^\bar{J} + \frac{1}{2} g^{\mu\nu} R_{I\bar{J}KL} \rho^\bar{J}_\mu \rho^I_\nu \chi^K \chi^L \right].$$

(3.43)

The $Q$-transformations can be easily covariantized as well (see Labastida and Llatas, 1991, for explicit formulae), and one can easily check that

$$S_A = \{Q, V\},$$

(3.44)

where

$$V = \frac{1}{2} \int_{\Sigma_g} d^2z\sqrt{g} g^{\mu\nu} G_{I\bar{J}} \left[ \frac{1}{2} \rho^I_\mu \tilde{F}^\bar{J}_\nu + \frac{1}{2} \rho^\bar{J}_\mu \tilde{F}^I_\nu + (\rho^I_\mu \partial_{\nu} x^\bar{J} + \rho^\bar{J}_\mu \partial_{\nu} x^I) \right].$$

(3.45)

In other words, the action is $Q$-exact. Since the action of $Q$ does not depend on the two-dimensional metric on $\Sigma_g$, it immediately follows that the energy-momentum tensor is also $Q$-exact. Therefore, the twisted A model is a topological field theory of the cohomological type. The quantity $V$ is sometimes called the \textit{gauge fermion}. The ‘antighost’ $G_{\mu\nu}$ appearing in (3.26) is given by $G_{\mu\nu} = \delta V/\delta g^{\mu\nu}$ and has ghost number $-1$.

It is interesting to note that the second term in (3.43) can be written as

$$\int_{\Sigma_g} x^* (J),$$

(3.46)

where $J = i G_{I\bar{J}} dx^I \wedge dx^\bar{J}$ is the Kähler class of $X$. This term is a topological invariant characterizing the homotopy type of the map $x : \Sigma_g \rightarrow X$. We can also add a coupling to a $B$-field into the action,

$$\int_{\Sigma_g} x^* (B),$$

(3.47)

which will replace the Kähler form by the complexified Kähler form $\omega = J + iB$.

Since the twisted A model is a cohomological field theory, the relevant operators, as we discussed in the previous section, are the observables, i.e. the operators
that belong to the $\mathcal{Q}$-cohomology. One can easily check that the $\mathcal{Q}$-cohomology is given by operators of the form

$$O_\phi = \phi_{i_1 \cdots i_p} \chi^{i_1} \cdots \chi^{i_p},$$

(3.48)

where $\phi = \phi_{i_1 \cdots i_p} dx^{i_1} \wedge \cdots \wedge dx^{i_p}$ is a closed $p$-form representing a non-trivial class in $H^p(X)$. Therefore, in this case the $\mathcal{Q}$-cohomology is in one-to-one correspondence with the de Rham cohomology of the target manifold $X$. This is in agreement with the fact that $\mathcal{Q}$ can be interpreted as the de Rham differential on $X$. Notice that the degree of the differential form corresponds to the ghost number of the operator. Moreover, one can derive a selection rule for correlation functions of such operators: the vacuum expectation value $\langle O_{\phi_1} \cdots O_{\phi_\ell} \rangle$ vanishes unless

$$\sum_{k=1}^{\ell} \deg(O_{\phi_k}) = 2d(1 - g) + 2 \int_{\Sigma_g} x^*(c_1(X)),$$

(3.49)

where $\deg(O_{\phi_k}) = \deg(\phi_k)$ and $c_1(X)$ is the first Chern class of the Kähler manifold $X$. This selection rule arises as follows: the twisted theory has a $U(1)$ ghost current, a global $U(1)$ symmetry that rotates the twisted fermions. Since $\chi$ and $\rho$ have opposite ghost numbers, this symmetry is anomalous, and the anomaly is given by the r.h.s. of (3.49), which calculates the number of zero modes of the twisted Dirac operator (in other words, the r.h.s. is minus the ghost number of the vacuum). As usual in quantum field theory, the operators with non-trivial vacuum expectation values have to soak up the zero modes associated to the anomaly. It is interesting to note that, for Calabi–Yau threefolds, i.e. Kähler manifolds of complex dimension 3, and such that $c_1(X) = 0$, the last term in (3.49) vanishes and the ghost number anomaly is $6 - 6g$, as in the usual bosonic string. We will see the implications of this fact in the next chapter.

The main goal in any quantum field theory is to evaluate the correlation functions of observables. In topological field theories of the cohomological type, the $\mathcal{Q}$-exactness of the action has an important consequence for this evaluation. Let us introduce a parameter $t$ in front of the action (which plays the role of $1/h$), and consider the unnormalized vacuum expectation value

$$\langle \mathcal{O} \rangle(t) = \int \mathcal{D}\phi \mathcal{O} e^{-tS_A(\phi)},$$

(3.50)

where $\phi$ denotes the set of fields of the theory and $\mathcal{O}$ is an observable. Since the action is $\mathcal{Q}$-exact, and $\mathcal{O}$ is $\mathcal{Q}$-closed, one has

$$\frac{d}{dt} \langle \mathcal{O} \rangle(t) = \pm \langle \{\mathcal{Q}, \mathcal{O}\mathcal{V}\} \rangle = 0,$$

(3.51)

and (3.50) is independent of $t$. In particular, we can evaluate it in the limit of $t$ large, which is the semi-classical approximation. This means that the semi-classical approximation is exact. This is a general property of topological field
theories of cohomological type where the action is $Q$-exact. We therefore can compute the correlation functions exactly by just doing a semi-classical computation. The first step in such a computation is to identify the possible instanton sectors of the theory. In the A model the instantons are holomorphic maps $x : \Sigma_g \to X$. The different instanton sectors are classified topologically by the homology class
\[ \beta = x_*[(\Sigma_g)] \in H_2(X,\mathbb{Z}). \] (3.52)

Sometimes it is useful to introduce a basis $[S_i]$ of $H_2(X,\mathbb{Z})$, where $i = 1, \cdots, b_2(X)$, in such a way that we can expand $\beta$ as $\beta = \sum_i n_i[S_i]$. The instanton sectors are then labelled by $b_2(X)$ integers $n_i$. These instantons are also called worldsheet instantons. A simple analysis (see, for example, Witten, 1988) shows that the contribution of an instanton sector to the path integral reduces to an integration over the moduli space of instantons in that sector.

Let us now present some results for the correlation functions of the A model involving three operators $O_{\phi_i}$ associated to 2-forms $\phi_i$ on a Calabi–Yau threefold $X$. The selection rule (3.49) says that, in genus $g = 0$ (i.e. when the Riemann surface is a sphere) and on a Calabi–Yau threefold, these correlation functions are generically non-vanishing. The computation of a given correlation function involves summing over the different topological sectors of worldsheet instantons. In the trivial sector, i.e. when $\beta = 0$, the image of the sphere is a point in the target. The moduli space of instantons is just the target space $X$, and the correlation function is just the classical intersection number $D_1 \cap D_2 \cap D_3$ of the three divisors $D_i$, $i = 1, 2, 3$, associated to the 2-forms. The non-trivial instanton sectors give an infinite series. The final answer is, schematically,
\[ \langle O_{\phi_1} O_{\phi_2} O_{\phi_3} \rangle = (D_1 \cap D_2 \cap D_3) + \sum_{\beta} I_{0,3,\beta}(\phi_1,\phi_2,\phi_3)Q^\beta. \] (3.53)

The notation is as follows: we have introduced the complexified Kähler parameters
\[ t_i = \int_{S_i} \omega, \quad i = 1, \cdots, b_2(X), \] (3.54)
where $\omega$ is the complexified Kähler form of $X$, and $S_i$ is a basis of $H_2(X)$. We also define $Q_i = e^{-t_i}$, and if $\beta = \sum_i n_i[S_i]$, then $Q^\beta$ denotes $\prod_i Q_i^{n_i}$. The coefficient $I_{0,3,\beta}(\phi_1,\phi_2,\phi_3)$ ‘counts’ in an appropriate sense the number of holomorphic maps from the sphere to the Calabi–Yau that send the point of insertion of $O_{\phi_i}$ to the divisor $D_i$. It can be shown that the coefficients $I_{0,3,\beta}(\phi_1,\phi_2,\phi_3)$ can be written as
\[ I_{0,3,\beta}(\phi_1,\phi_2,\phi_3) = N_{0,\beta} \int_{\beta} \phi_1 \int_{\beta} \phi_2 \int_{\beta} \phi_3, \] (3.55)
in terms of invariants $N_{0,\beta}$ that encode all the information about the three-point functions (3.53) of the topological sigma model. The invariants $N_{0,\beta}$ are our first
example of Gromov–Witten invariants. It is convenient to put all these invariants together in a generating functional called the prepotential:

\[ F_0(t) = \sum_\beta N_{0,\beta} Q^\beta. \]  

(3.56)

This prepotential depends on the \( h^{1,1}(X) \) complexified Kähler parameters of the Calabi–Yau \( X \).

What happens if we go beyond \( g = 0 \)? For \( g = 1 \) and \( c_1(X) = 0 \), the selection rule (3.49) says that the only quantity that may lead to a non-trivial answer is the partition function itself, while for \( g > 1 \) all correlation functions vanish. This corresponds mathematically to the fact that, for a generic metric on the Riemann surface \( \Sigma_g \), there are no holomorphic maps in genus \( g > 1 \). In order to circumvent this problem, we have to consider the topological string theory made out of the topological sigma model, i.e. we have to couple the theory to two-dimensional gravity and to consider all possible metrics on the Riemann surface. We will address this issue in the next chapter.

3.4 The topological type-B model

We now analyse the model obtained after the B twisting. We will assume that the target manifold \( X \) is Calabi–Yau, otherwise the resulting type-B model is ill-defined. The fermionic and auxiliary fields of the supersymmetric sigma model change their spin as prescribed by (3.12), and we obtain the following fields:

\[
\begin{align*}
\rho^I_\pm &= 2\psi^I_\pm, & \chi^T &= \psi^T_{-,-}, & F^I &= 2F^T_{-,-,-}, \\
\rho^I &= 2\psi^I_+ & \bar{\chi}^T &= \psi^T_{-,-}, & \bar{F}^T &= 2F^T_{-,-,-}.
\end{align*}
\]  

(3.57)

It is convenient to perform the following redefinition, suggested by Witten (1991a):

\[
\begin{align*}
\eta^I &= \chi^T + \bar{\chi}^I, \\
\theta_I &= G_{IJ}(\chi^J - \bar{\chi}^J).
\end{align*}
\]  

(3.58)

We then get the following field content: a map \( x : \Sigma_g \to X \), which is a scalar, commuting field, two sets of Grassmann fields \( \eta^I, \theta^I \in x^* (TX) \), which are scalars on \( \Sigma_g \), and a Grassmannian one-form on \( \Sigma_g, \rho^I_\alpha \), with values in \( x^* (TX) \). We also have commuting auxiliary fields \( F^I, \bar{F}^T \). As in the case of the type-A model, it is convenient to redefine the auxiliary fields appropriately (see Labastida and Llatas, 1992, for the details). We will present here the final results after this redefinition (following Labastida and Mariño, 1994). The \( Q \)-transformations read

\[
\begin{align*}
\{ Q, x^I \} &= 0, & \{ Q, \eta^T \} &= 0, \\
\{ Q, x^T \} &= \eta^T, & \{ Q, \theta_I \} &= G_{IJ} F^J, \\
\{ Q, \rho^I_\pm \} &= \partial_\pm x^I, & \{ Q, F^I \} &= D_\pm \rho^I_\pm - D_\pm \rho^I_\mp + R^I_{JLK} \eta^J \rho^J_\pm \rho^K_\pm, \\
\{ Q, \rho^I \} &= \partial_\pm x^I, & \{ Q, \bar{F}^T \} &= -\Gamma_{JK}^T \eta^J F^K,
\end{align*}
\]  

(3.59)
and satisfy $Q^2 = 0$. Notice that $Q$ acts differently on holomorphic and anti-holomorphic co-ordinates. In contrast to what happens in the type-A model, it explicitly depends on the splitting between holomorphic and anti-holomorphic co-ordinates on $X$, in other words, it depends explicitly on the choice of complex structure on $X$. If we interpret $\eta^T_I$ as a basis for anti-holomorphic differential forms on $X$, the action of $Q$ on $x^I$, $x^\bar{I}$ may be interpreted as the Dolbeault anti-holomorphic differential $\bar{\partial}$. The action for the theory is

$$S_B = \int_{\Sigma_g} d^2z \left[ G_{I\bar{J}} (\partial_z x^I \partial_{\bar{z}} x^\bar{J} + \partial_{\bar{z}} x^\bar{J} \partial_z x^I) - \rho_z^I (G_{I\bar{J}} D_z \eta^\bar{J} + D_{\bar{z}} \theta_I) - \rho^\bar{J}_z (G_{I\bar{J}} D_{\bar{z}} \eta^J - D_z \theta_I) - R^I_{\bar{J}JK} \eta_I \eta^K \rho_J^\bar{J} \theta_I - G_{I\bar{J}} F^I F^\bar{J} \right].$$

(3.60)

We can explicitly introduce the metric on $\Sigma_g$ in this action and verify that it is $Q$-exact:

$$S_B = \{Q, V\},$$

(3.61)

where $V$ is now given by

$$V = \int_{\Sigma_g} d^2z \sqrt{g} \left[ G_{I\bar{J}} g^{\mu\nu} \rho^I_{\mu} \partial_\nu x^\bar{J} - F^I \theta_I \right].$$

(3.62)

Finally, we also have a $U(1)$ ghost number symmetry, in which $x$, $\eta$, $\theta$ and $\rho$ have ghost numbers 0, 1, 1, and $-1$, respectively.

Since the action is $Q$-exact, the theory is topological and the semi-classical approximation is exact. In contrast to the type-A model, there are no non-trivial instantons for this theory: the classical configurations are just constant maps $x : \Sigma_g \rightarrow X$. It follows that path integrals in the type-B model reduce to integrals over $X$, as found by Witten (1991a).

What are the observables in this theory? It is easy to see that the operators in the $Q$-cohomology are of the form

$$O_\phi = \phi_{I_1 \cdots I_p}^{J_1 \cdots J_q} \eta^{I_1} \cdots \eta^{I_p} \theta_{J_1} \cdots \theta_{J_q},$$

(3.63)

where

$$\phi = \phi_{I_1 \cdots I_p}^{J_1 \cdots J_q} dx^{I_1} \wedge \cdots \wedge dx^{I_p} \partial_{x^{J_1}} \wedge \cdots \wedge \partial_{x^{J_q}}$$

(3.64)

is an element of $H^p_{Q}(X, \wedge^q TX)$. Therefore, the $Q$-cohomology is in one-to-one correspondence with the twisted Dolbeault cohomology of the target manifold $X$. We can then consider correlation functions of the form

$$\langle \prod_a O_{\phi_a} \rangle.$$  

(3.65)

This correlation function vanishes unless the following selection rule is satisfied:

$$\sum_a p_a = \sum_a q_a = d(1 - g),$$

(3.66)

where $g$ is the genus of the Riemann surface. This selection rule comes from the $U(1)_L \times U(1)_R$ anomalous global current. Due to the arguments presented above,
this correlation function can be computed in the semi-classical limit, where the path integral reduces to an integration over the target $X$. The product of operators in (3.65) corresponds to a form in $H_d^d(X, \wedge^d TX)$. To integrate such a form over $X$ we crucially need the Calabi–Yau condition. This arises as follows. One of the most important properties of Calabi–Yau manifolds (which can actually be taken as their defining feature) is that they have a holomorphic, nowhere-vanishing section $\Omega$ of the canonical bundle $K_X = \Omega^d \otimes \mathbb{O}$. Since the section is nowhere-vanishing, the canonical line bundle is trivial and we recover the condition $c_1(K_X) = c_1(X) = 0$. This means in particular that we have an invertible map

$$
\Omega^{0,p}(\wedge^q TX) \longrightarrow \Omega^{d-q,p}(X)
$$

$$
\phi_{I_1 \ldots I_q} \longrightarrow \Omega_{I_1 \ldots I_q I_{q+1} \ldots I_d} \phi_{J_1 \ldots J_p}
$$

(3.67)

where the $(d,0)$-form $\Omega$ is used to contract the indices. Since $\Omega$ is holomorphic, this descends to the $\partial\bar{\partial}$-cohomology. It then follows that an element in $H_{\partial\bar{\partial}}^d(X, \wedge^d TX)$ maps to an element in $H_{\partial\bar{\partial}}^{0,d}(X)$. After further multiplication by $\Omega$, one can then integrate a $(d,d)$-form over $X$. This is the prescription to compute correlation functions like (3.65). A simple and important example of this procedure is the case of a Calabi–Yau threefold, $d = 3$, and operators associated to forms in $H_{\partial\bar{\partial}}^1(X, TX)$, or by using (3.67), to forms in $H_{\partial\bar{\partial}}^2(X)$. These operators are important since they correspond to infinitesimal deformations of the complex structure of $X$. The selection rule (3.66) says that we have to integrate three of these operators, and the correlation function reads in this case

$$
\langle O_{\phi_1} O_{\phi_2} O_{\phi_3} \rangle = \int_X (\phi_1)^{I_1}_{J_1} (\phi_2)^{I_2}_{J_2} (\phi_3)^{I_3}_{J_3} \Omega_{I_1 I_2 I_3} dz^{J_1} dz^{J_2} dz^{J_3} \wedge \Omega.
$$

(3.68)

It turns out that the full information of the correlators (3.68) in genus zero can be encoded in a single function called the prepotential. We will quickly review here some of the basic results of special geometry and the theory of the prepotential for the topological B model, and we refer the reader to Candelas and de la Ossa (1991) or Hori et al. (2003) for more details. The correlation functions in the B model, like for example (3.68), depend on a choice of complex structure, as we have already emphasized. The different complex structures form a moduli space $\mathcal{M}$ of dimension $h^{2,1}$. A convenient parametrization of $\mathcal{M}$ is the following. Choose first a symplectic basis for $H_3(X)$, denoted by $(A_a, B^a)$, with $a = 0, 1, \ldots, h^{2,1}$, and such that $A_a \cap B^b = \delta_a^b$. We then define the periods of the Calabi–Yau manifold as

$$
z_a = \int_{A_a} \Omega, \quad \mathcal{F}^a = \int_{B^a} \Omega, \quad a = 0, \ldots, h^{2,1}.
$$

(3.69)

Of course, the symplectic group $\text{Sp}(2h^{2,1} + 2, \mathbb{R})$ acts on the vector $(z^a, \mathcal{F}_a)$. A basic result of the theory of deformation of complex structures says that the $z^a$ are (locally) complex projective co-ordinates for $\mathcal{M}$. Inhomogeneous co-ordinates
can be introduced in a local patch where one of the projective co-ordinates, say $z_0$, is different from zero, and taking
\[ t_a = \frac{z_a}{z_0}, \quad a = 1, \ldots, h^{2,1}. \] (3.70)

The co-ordinates $z_a$ are called \textit{special projective co-ordinates}, and since they parametrize $\mathcal{M}$ we deduce that the other set of periods must depend on them, \textit{i.e.} $\mathcal{F}^a = \mathcal{F}^a(z)$. Using the periods (3.69) we can define a function $\mathcal{F}(z)$ by the equation
\[ \mathcal{F} = \frac{1}{2} \sum_{a=0}^{h^{2,1}} z_a \mathcal{F}^a. \] (3.71)

This function satisfies
\[ \mathcal{F}^a(z) = \frac{\partial \mathcal{F}}{\partial z_a} \] (3.72)

and turns out to be homogeneous of degree two in the $z_a$. Therefore, one can rescale it in order to obtain a function of the inhomogeneous co-ordinates $t_a$:
\[ F_0(t_a) = \frac{1}{z_0^2} \mathcal{F}(z_a). \] (3.73)

$F_0(t_a)$ is called the B-model prepotential, and it depends on the $h^{2,1}(X)$ moduli $t_a$ which parametrize the complex structure moduli of $X$. The fact that the co-ordinates $z_a$ are projective is related to the freedom in normalizing the three-form $\Omega$. In order to obtain expressions in terms of the inhomogeneous co-ordinates $t_a$, we simply have to rescale $\Omega \rightarrow \frac{1}{z_0} \Omega$, and the periods $(z_a, \mathcal{F}^a)$ become
\[ (1, t_a, 2F_0 - \sum_{a=1}^{h^{2,1}} t_a \frac{\partial F_0}{\partial t_a}, \frac{\partial F_0}{\partial t_a}). \] (3.74)

One of the key results in special geometry is that the correlation functions (3.68) can be computed in terms of the prepotential $F_0(t_a)$. Given a deformation of the complex structure parametrized by $t_a$, the corresponding tangent vector $\partial/\partial t_a$ is associated to a differential form of type $(2, 1)$. This form leads to an operator $\mathcal{O}_a$, and the three-point functions involving these operators turn out to be given by
\[ \langle \mathcal{O}_a \mathcal{O}_b \mathcal{O}_c \rangle = \frac{\partial^3 F_0}{\partial t_a \partial t_b \partial t_c}. \] (3.75)

The prepotential $F_0(t)$ encodes the relevant information about the B model on the sphere. In the next chapter we will see how to define generalizations of $F_0(t)$ to higher genera.
4 TOPOLOGICAL STRINGS

4.1 Coupling to gravity

In the previous chapter we defined two topological field theories in two dimensions, the type-A and the type-B topological sigma models. Although they contain a lot of information in genus 0, they turn out to be trivial for \( g > 1 \) due to the selection rules (3.49) and (3.66). This is essentially due to the fact that, in order to define these theories, we consider a fixed metric in the Riemann surface. In order to obtain a non-trivial theory in higher genus we have to introduce the degrees of freedom of the two-dimensional metric. This means that we have to couple the topological field theories considered in the last chapter to two-dimensional gravity. The coupling to gravity is done by using the fact pointed out by Dijkgraaf et al. (1991), Witten (1995), and Bershadsky et al. (1994), that the structure of the twisted theory is tantalizingly close to that of the bosonic string. In the bosonic string, there is a nilpotent BRST operator, \( \mathcal{Q}_{\text{BRST}} \), and the energy-momentum tensor turns out to be a \( \mathcal{Q}_{\text{BRST}} \)-commutator:

\[
T(z) = [\mathcal{Q}_{\text{BRST}}, b(z)].
\]

In addition, there is a ghost number with anomaly \( 3\chi(\Sigma_g) = 6 - 6g \), in such a way that \( \mathcal{Q}_{\text{BRST}} \) and \( b(z) \) have ghost number 1 and -1, respectively. This is precisely the same structure that we found in (3.26), and the composite field \( G^{\mu\nu} \) plays the role of an anti-ghost. Furthermore, the anomalies in the ghost current are precisely \( 6g - 6 \) for a Calabi–Yau threefold, and therefore the ghost number symmetry in the twisted sigma models plays exactly the same role as the ghost number symmetry in the bosonic string. One can then just follow the prescription for coupling to gravity of bosonic string theory (see, for example, Polchinski, 1998) and define a genus \( g \geq 1 \) free energy as follows:

\[
F_g = \int_{\mathcal{M}_g} \langle 6g - 6 \prod_{k=1} G, \mu_k \rangle, \quad (4.1)
\]

where

\[
(G, \mu_k) = \int_{\Sigma_g} d^2z (G_{zz}(\mu_k)z \bar{z} + G_{\bar{z}z}(\bar{\mu}_k)\bar{z}z), \quad (4.2)
\]

and \( \mu_k \) are the Beltrami differentials. The vacuum expectation value in (4.1) refers to the path integral over the fields of the topological sigma model under consideration, and gives a differential form on the moduli space of Riemann surfaces of genus \( g \), \( \mathcal{M}_g \), which is then integrated over. The model obtained after coupling to gravity the type-A (B) topological sigma model will be called type-A (B) topological string theory. One can show that, as happened to the prepotential
4.2 Relation to compactifications of type II string theory

The topological string amplitudes $F_g$ that we have considered above, as well as the prepotential $F_0$ that appears already at tree level, have an interpretation in terms of compactifications of type II superstring theory.

Type II superstrings have maximal supersymmetry in ten dimensions, and appear in two varieties: type IIA and type IIB, which at low energies reduce to type IIA/B supergravity in ten dimensions, respectively (see Polchinski, 1998, for a review of these issues). Once this theory is compactified on a Calabi–Yau threefold, the resulting theory in four dimensions has $\mathcal{N} = 2$ supersymmetry, and it is, in fact, $\mathcal{N} = 2$ supergravity in four dimensions. In this theory there are three kinds of multiplets. First, there is the so-called supergravity multiplet, which includes the four-dimensional graviton, two gravitinos, and a $U(1)$ gauge field called the graviphoton. There are also $\mathcal{N} = 2$ matter multiplets: the $\mathcal{N} = 2$ vector multiplet (VM), and the hypermultiplet (HM). In what follows, we will focus on the vector multiplets.

The number of VMs in the effective four-dimensional theory depends on the internal geometry of the Calabi–Yau manifold $X$ through its Hodge numbers. In compactifications of type IIA theory, the number of vector multiplets is $h^{1,1}(X)$, while in type IIB this number is given by $h^{2,1}(X)$. Notice that these are precisely the moduli that determine the prepotential in the type-A and the type-B model, respectively. We will denote the complex scalar field in the $a$-th VM by $t^a$, where $a = 1, \cdots, h^{1,1}(X)$ in the type IIA theory, and $a = 1, \cdots, h^{2,1}(X)$ in the type IIB theory. A vacuum expectation value for this complex scalar field can be interpreted as a choice of moduli of the Calabi–Yau.

The effective action for the $\mathcal{N} = 2$ VMs up to two derivatives is encoded in a single holomorphic function, called the prepotential $F_0(t^a)$. Let us denote by $T_a$ and $W^\alpha_a$, respectively, the $\mathcal{N} = 1$ chiral and gauge multiplets associated to the $a$-th $\mathcal{N} = 2$ VM. The four-dimensional action can be written, in $\mathcal{N} = 1$ superspace, as follows

$$\frac{1}{4\pi} \text{Im} \left[ \int d^4\theta \frac{\partial F_0(T)}{\partial T_a} T_a + \int d^2\theta \frac{1}{2} \frac{\partial^2 F_0}{\partial T_a \partial T_b} W^\alpha_a W_{\alpha \beta} \right].$$

The first term in (4.3) leads to a non-linear sigma model whose target space is Kähler, with Kähler potential given by

$$K = \text{Im} \left( T_a \frac{\partial F_0(T)}{\partial T_a} \right).$$

The metric on the moduli space is given by
\begin{equation}
\mathrm{d}s^2 = g_{a\bar{b}} \, \mathrm{d}t_a \, \mathrm{d}\bar{t}_b = \text{Im} \, \tau_{a\bar{b}} \, \mathrm{d}t_a \, \mathrm{d}\bar{t}_b, \tag{4.5}
\end{equation}

where,
\begin{equation}
\tau_{a\bar{b}} = \frac{\partial^2 F_0}{\partial t_a \partial \bar{t}_b}. \tag{4.6}
\end{equation}

From (4.3) it is easy to obtain the gauge kinetic term for the vector fields
\begin{equation}
\int \mathrm{d}^4x \, \tau_{a\bar{b}}(t) F_{\mu \nu}^a \, F_{\mu \nu}^b, \tag{4.7}
\end{equation}

where $F_{\mu \nu}^a$ is the field strength of the $U(1)$ vector field of the $a$-th VM. Therefore, (4.6) is the matrix of gauge couplings. Since the action (4.3) is completely determined by the prepotential, it is important to determine $F_0(t_a)$. One of the basic results in the theory is that the prepotential of the vector multiplets in type IIA/B superstring compactifications is given by the prepotential of the type-A/B topological models that we considered in the previous chapter.

The four-dimensional effective theory obtained upon compactification of type II superstring theory on $X$ contains more complicated couplings. The graviphoton field strength $T_{\mu \nu}$ is the lowest component of an $\mathcal{N} = 2$ supergravity multiplet
\begin{equation}
W_{\mu \nu}^{ij} = \frac{1}{2} \epsilon_{ij} T_{\mu \nu} - R_{\mu \nu \lambda \rho} \theta^i \sigma^{\lambda \rho} \theta^j + \cdots. \tag{4.8}
\end{equation}

In this equation, $i, j = 1, 2$ are indices of the internal $SU(2)_R$ symmetry of $\mathcal{N} = 2$, $\theta^i$ are fermionic co-ordinates for $\mathcal{N} = 2$ superspace, and $\cdots$ refers to the supersymmetric completion (see Antoniadis et al., 1994, for details). The couplings we are interested in are given, in $\mathcal{N} = 2$ superspace, by
\begin{equation}
\int \mathrm{d}^4x \, \mathrm{d}^4\theta F_g(t_a)(W^2)^{g-1} = \int \mathrm{d}^4x F_g(t_a)(T^2)^{g-1} R^2 + \cdots, \quad g \geq 1, \tag{4.9}
\end{equation}

where the $\cdots$ indicate supersymmetric completion. It was shown by Antoniadis et al. (1994) that the $F_g(t_a)$ agree with the higher-genus topological string amplitudes defined in (4.1). This gives an important relation between topological string amplitudes and physical superstring amplitudes that has many uses and implications, as we will see.

### 4.3 The type-A topological string

In most of this book we will focus on the type-A topological string, and in this section we give a detailed discussion of some of its properties.

#### 4.3.1 Mathematical description

As was the case for the type-A prepotential, the free energies $F_g$, $g \geq 1$, of the type-A topological string can also be evaluated as a sum over instanton sectors, i.e. over holomorphic curves. Therefore they have the structure
where $N_{g,\beta}$ ‘count’ in an appropriate sense the number of holomorphic curves of genus $g$ in the two-homology class $\beta$. We will refer to $N_{g,\beta}$ as the Gromov–Witten invariant of the Calabi–Yau $X$ in genus $g$ and in the class $\beta$. They generalize the Gromov–Witten invariants in genus 0 that were introduced in (3.55).

The Gromov–Witten invariants appearing in (4.10) can be defined in a rigorous mathematical way, and have played an important role in algebraic geometry and symplectic geometry. We will now give a short summary of the main mathematical ideas involved in Gromov–Witten theory.

The coupling of the type-A model to gravity involves the moduli space of Riemann surfaces $M_g$, as we have just seen. In order to construct the Gromov–Witten invariants in full generality we also need the moduli space of possible metrics (or equivalently, complex structures) on a Riemann surface with punctures, which is the famous Deligne-Mumford space $\overline{M}_{g,n}$ of $n$-pointed stable curves (the definition of what stable means can be found for example in Harris and Morrison, 1998). The relevant moduli space in the theory of topological strings $\overline{M}_{g,n}(X,\beta)$ is a generalization of $\overline{M}_{g,n}$, and depends on a choice of a two-homology class $\beta$ in $X$. Very roughly, a point in $\overline{M}_{g,n}(X,\beta)$ can be written as $(f, \Sigma_g, p_1, \cdots, p_n)$ and is given by (a) a point in $\overline{M}_{g,n}$, i.e. a Riemann surface with $n$ punctures, $(\Sigma_g, p_1, \cdots, p_n)$, together with a choice of complex structure on $\Sigma_g$, and (b) a map $f : \Sigma_g \to X$ that is holomorphic with respect to this choice of complex structure and such that $f_*[\Sigma_g] = \beta$. The set of all such points forms a good moduli space provided a certain number of conditions are satisfied (see for example Cox and Katz (1999) and Hori et al. (2003) for a detailed discussion of these issues). $\overline{M}_{g,n}(X,\beta)$ is the basic moduli space we will need in the theory of topological strings. Its complex virtual dimension is given by

$$ (1 - g)(d - 3) + n + \int_{\Sigma_g} f^*(c_1(X)), \quad (4.11) $$

which is given by the r.h.s. of (3.49) plus $3g - 3 + n$, which is the dimension of $\overline{M}_{g,n}$ and takes into account the extra moduli that come from the coupling to two-dimensional gravity. We also have two natural maps

$$ \pi_1 : \overline{M}_{g,n}(X,\beta) \longrightarrow X^n, $$

$$ \pi_2 : \overline{M}_{g,n}(X,\beta) \longrightarrow \overline{M}_{g,n}. \quad (4.12) $$

The first map is easy to define: given a point $(f, \Sigma_g, p_1, \cdots, p_n)$ in $\overline{M}_{g,n}(X,\beta)$, we just compute $(f(p_1), \cdots, f(p_n))$. The second map essentially sends $(f, \Sigma_g, p_1, \cdots, p_n)$ to $(\Sigma_g, p_1, \cdots, p_n)$, i.e. forgets the information about the map and keeps the information about the punctured curve.

We can now formally define the Gromov–Witten invariant $I_{g,n,\beta}$ as follows. Let us consider cohomology classes $\phi_1, \cdots, \phi_n$ in $H^*(X)$. If we pull back their
tensor product to $H^\ast(\overline{M}_{g,n}(X,\beta))$ via $\pi_1$, we get a differential form on the moduli space of maps that we can integrate (as long as there is a well-defined fundamental class for this space):

$$I_{g,n,\beta}(\phi_1, \cdots, \phi_n) = \int_{\overline{M}_{g,n}(X,\beta)} \pi_1^\ast(\phi_1 \otimes \cdots \otimes \phi_n).$$  \hfill (4.13)

The Gromov–Witten invariant $I_{g,n,\beta}(\phi_1, \cdots, \phi_n)$ vanishes unless the degree of the form equals the dimension of the moduli space. Therefore, we have the following constraint:

$$\frac{1}{2} \sum_{i=1}^n \deg(\phi_i) = (1 - g)(d - 3) + n + \int_{\Sigma_g} f^\ast(c_1(X)).$$ \hfill (4.14)

Notice that Calabi–Yau threefolds play a special role in the theory, since for those targets the virtual dimension only depends on the number of punctures, and therefore the above condition is always satisfied if the forms $\phi_i$ have degree 2. These invariants generalize the invariants obtained from topological sigma models. In particular, $I_{0,3,\beta}$ are the invariants involved in the evaluation of correlation functions of the topological sigma model with a Calabi–Yau threefold as its target in (3.53). When $n = 0$, one gets an invariant $I_{g,0,\beta}$ that does not require any insertions. This is precisely the Gromov–Witten invariant $N_{g,\beta}$ that appears in (4.10). Notice that these invariants are in general rational, due to the orbifold character of the moduli spaces involved.

By using the Gysin map $\pi_2!$, one can reduce any integral of the form (4.13) to an integral over the moduli space of curves $\overline{M}_{g,n}$. The resulting integrals involve two types of differential forms. The first type of forms are the Mumford classes $\psi_i$, $i = 1, \cdots, n$, which are constructed as follows. We first define the line bundle $\mathcal{L}_i$ over $\overline{M}_{g,n}$ to be the line bundle whose fibre over each curve $\Sigma_g$ is the cotangent space of $\Sigma_g$ at $p_i$ (where $p_i$ is the $i$-th marked point). We then have,

$$\psi_i = c_1(\mathcal{L}_i), \quad i = 1, \cdots, n.$$ \hfill (4.15)

The second type of differential forms are the Hodge classes $\lambda_j$, $j = 1, \cdots, g$, which are defined as follows. On $\overline{M}_{g,n}$ there is a complex vector bundle $\mathcal{E}$ of rank $g$, called the Hodge bundle, whose fibre at a point $\Sigma_g$ is $H^0(\Sigma_g, K_{\Sigma_g})$ (i.e. the space of holomorphic sections of the canonical line bundle $K_{\Sigma_g}$ of $\Sigma_g$). The Hodge classes are simply the Chern classes of this bundle,

$$\lambda_j = c_j(\mathcal{E}).$$ \hfill (4.16)

The integrals of the $\psi$ classes can be obtained by the results of Witten (1991b) and Kontsevich (1992), while the integrals involving $\psi$ and $\lambda$ classes (the so-called Hodge integrals) can, in principle, be computed by reducing them to pure $\psi$ integrals (Faber, 1999). Explicit formulae for some Hodge integrals can be found,
for example, in Getzler and Pandharipande (1998) and Faber and Pandharipande (2000). As we will see later in this book, one of the outcomes of the string/gauge correspondence for Chern–Simons theory is an explicit formula for a wide class of Hodge integrals.

4.3.2 Integrality properties and Gopakumar–Vafa invariants

The free energies $F_g$ of topological string theory, which contain information about the Gromov–Witten invariants of the Calabi–Yau manifold $X$, play an important role in type IIA string theory, since they capture the prepotential and the $F_g$ couplings in the four-dimensional $\mathcal{N} = 2$ supergravity that is obtained when type IIA theory is compactified on $X$.

The relation between superstrings and topological strings has been a source of insights for both models, and in particular has indicated a hidden integrality structure in the Gromov–Witten invariants $N_{g,\beta}$. In order to make manifest this structure it is useful to introduce a generating functional for the all-genus free energy:

$$F(g_s, t) = \sum_{g=0}^{\infty} F_g(t) g_s^{2g-2}. \tag{4.17}$$

The parameter $g_s$ can be regarded as a formal variable, but in the context of type II strings it is the string coupling constant. Gopakumar and Vafa (1998b) showed that the generating functional (4.17) can be written as a generalized index that counts BPS states in the type IIA superstring theory compactified on $X$. We now sketch this derivation, referring the reader to the original paper (Gopakumar and Vafa, 1998b) or to Hori et al. (2003) for more details.

Consider the compactification of type IIA string theory on $X$, and let us turn on a self-dual, constant graviphoton field strength, in such a way that $T^2 = g_s^2$. Taking into account (4.9), we find that this leads to the following term in the four-dimensional effective action:

$$F(g_s, t) R^2, \tag{4.18}$$

together with its supersymmetric completion. It turns out that this coupling can be computed by integrating out at one loop a certain class of charged particles in the presence of the constant, self-dual graviphoton field strength. These particles come in spin multiplets, i.e. they are labelled by their spin representation of the Lorentz group $SU(2)_L \times SU(2)_R$. Due to self-duality, this correction is only sensitive to the left spin (say) of the particle, since the self-dual graviphoton field strength only couples to the $SU(2)_L$ spin content of the particle, labelled by $j_L$. The contribution of multiplets of mass $m$ and spin $j_L$ to this coupling is given by the Schwinger formula,

$$\int_0^{\infty} \frac{d\tau}{\tau} \frac{e^{-\tau m}}{(2 \sin \frac{\tau g_s}{2})^2} \text{Tr} (-1)^F e^{2i\tau j_L^3} g_s, \tag{4.19}$$
where \( F = 2j^3_L + 2j^3_R \) is the total fermion number. What are the particles that have to be taken into account in this computation? They are BPS particles of the four-dimensional compactification. Since IIA theory contains D2 and D0 branes, one can obtain these BPS particles by wrapping D2 branes bound to D0 branes around Riemann surfaces inside \( X \). The mass of a BPS particle is given by the corresponding central charge of the D2/D0 brane bound state. If the D2 brane wraps a curve in the homology class \( \beta \in H_2(X) \), the central charge is given by the area of the curve, which depends on a choice of Kähler parameters \( t \) and is given by \( t \cdot \beta \). We have also to take into account that, given a wrapped D2 brane, we can form a bound state with any number \( d \) of D0 branes. The central charge of a BPS particle coming from a D2 brane wrapping a curve in the class \( \beta \) bound to \( d \) D0 branes is given by

\[
m = t \cdot \beta + 2\pi id. \tag{4.20}
\]

Let us assume that there are \( n^{(j^L,j^R)}_{\beta} \) particles with spin content \((j^L,j^R)\) and coming from a D2 brane wrapped around a Riemann surface in the class \( \beta \). After tracing over \( j^R \) with the sign \((-1)^{2j^R} \) we obtain

\[
n^{j^L}_{\beta} = \sum_{j^R} (-1)^{2j^R} (2j^R + 1)n^{(j^L,j^R)}_{\beta}. \tag{4.21}
\]

The \( SU(2)_L \) content can be given in many bases. Above, we have chosen the basis given by \( j^L \), but there is another useful basis given by

\[
I_g = \left[ 2(0) + \left( \frac{1}{2} \right) \right] \otimes g. \tag{4.22}
\]

Notice that \( I_1 \) consists of two particles of spin 0 and one particle of spin 1/2. The above set, with \( g = 0, 1, \ldots \), gives a basis for \( SU(2)_L \) representations, and we can write

\[
\sum_{j^L} n^{j^L}_{\beta} [j^L] = \sum_{g \geq 0} n^g I_g. \tag{4.23}
\]

This defines the integers \( n^g_{\beta} \), which characterize the set of multiplets contributing to \( F(g_s,t) \). Notice that

\[
\text{Tr}_{I_g} (-1)^F e^{2\pi j^L_s g/s} = \left( 2 \sin \frac{\tau g_s}{2} \right)^{2g}. \tag{4.24}
\]

We can now put together all these ingredients and compute the total amplitude \( F(g,t) \). We have to sum over all possible BPS states, therefore we have to sum over all possible homology classes \( \beta \), all \( SU(2)_L \) representations labelled by \( g \), and all numbers \( d \) of D0 branes. The result is

\[
F(g_s,t) = \sum_{\beta} \sum_{g=0}^{\infty} \sum_{d=-\infty}^{\infty} n^g_{\beta} \int_0^{\infty} \frac{d\tau}{\tau} e^{-\tau(t \cdot \beta + 2\pi i d)} \left( 2 \sin \frac{\tau g_s}{2} \right)^{2g-2}. \tag{4.25}
\]
To proceed, we first sum over $d$ by using
\begin{equation}
\sum_{d=-\infty}^{\infty} e^{-2\pi i d \tau} = \sum_{d=-\infty}^{\infty} \delta(\tau - d). \tag{4.26}
\end{equation}
It is now trivial to integrate over $\tau$, and the final result is
\begin{equation}
F(g_s, t) = \sum_{g=0}^{\infty} \sum_{\beta} \sum_{d=1}^{\infty} n_{\beta}^g \frac{1}{d} \left( 2 \sin \frac{dg_s}{2} \right)^{2g-2} Q^{d\beta}. \tag{4.27}
\end{equation}
The integers $n_{\beta}^g$ are known as Gopakumar–Vafa invariants. They are true invariants of the Calabi–Yau manifold $X$, in the sense that they do not depend on smooth deformations of the target geometry. This is in contrast to the quantities $n_{(j_L,j_R)}^{(\beta)}$, which do depend on deformations. As usual, by tracing over a non-invariant quantity with signs we obtain an invariant quantity.

The structure result (4.27) implies that Gromov–Witten invariants of closed strings, which are in general rational, can be written in terms of these integer invariants. In fact, by knowing the Gromov–Witten invariants $N_{g,\beta}$ we can explicitly compute the Gopakumar–Vafa invariants from (4.27) (an explicit inversion formula can be found in Bryan and Pandharipande, 2001). By expanding in $g_s$, it is easy to show that the Gopakumar–Vafa formula (4.27) predicts the following expression for $F_g(t)$:
\begin{equation}
F_g(t) = \sum_{\beta} \left( \frac{|B_{2g}| n_{\beta}^g}{2g(2g-2)!} + \frac{2(-1)^g n_{\beta}^2}{(2g-2)!} \pm \cdots - \frac{g-2}{12} \frac{n_{\beta}^{g-1} + n_{\beta}^g}{|B_{g-1}|} \right) \text{Li}_{3-2g}(Q^{\beta}), \tag{4.28}
\end{equation}
where $\text{Li}_j$ is the polylogarithm defined in (2.180). The appearance of the polylogarithm of order $3-2g$ in $F_g$ was first predicted from type IIA/heterotic string duality by Mariño and Moore (1999).

The structure found by Gopakumar and Vafa solves some longstanding issues in the theory of Gromov–Witten invariants, in particular the enumerative meaning of the invariants. Two obstructions to finding obvious enumerative meaning to Gromov–Witten invariants are multicovering and bubbling. Multicovering arises as follows. Suppose one finds a holomorphic map $x : I \rightarrow X$ in genus zero and in the class $\beta$. Then, simply by composing this with a degree $d$ cover $I \rightarrow I$, one can find another holomorphic map in the class $d\beta$. Therefore, at every degree, in order to count the actual number of ‘primitive’ holomorphic curves, one should subtract from the corresponding Gromov–Witten invariant the contributions coming from multicovering of curves with lower degree. Another geometric effect that has to be taken into account is bubbling (see, for example, Bershadsky et al., 1993, 1994). Imagine that one finds a map $x : \Sigma_g \rightarrow X$ from a genus $g$ Riemann surface to a Calabi–Yau threefold. By gluing to $\Sigma_g$ a small Riemann surface of genus $h$, and making it very small, one can find an approximate holomorphic map from a Riemann surface whose genus is topologically $g+h$. This means that ‘primitive’ maps at genus $g$ contribute to all genera.
$g' > g$, and in order to count curves properly one should take this effect into account.

The formula (4.28) gives a precise answer to these questions. Consider, for example, the structure of $F_0$. According to the above formula, the contribution of a BPS state is given by the function $L_i$, i.e. each BPS state contributes

$$\sum_{d=1}^{\infty} \frac{Q^{d\beta}}{d^3}. \quad (4.29)$$

This gives the contribution of all the multicoverings of a given ‘primitive’ curve, where $d$ is the degree of the multicovering. In addition, it says that each cover has a weight $1/d^3$. Therefore, the BPS invariant $n_{\beta}^0$ corresponds to primitive holomorphic maps, and the non-integrality of genus-zero Gromov–Witten invariants is due to the effects of multicovering. The multicovering phenomenon in genus 0 was found experimentally in Candelas et al. (1991) and later derived in the context of Gromov–Witten theory by Aspinwall and Morrison (1993). The structure result of Gopakumar and Vafa also predicts that the multicovering of degree $d$ of a genus $g$ curve contributes with a weight $d^3 - 2g$ (coming from $L_i - 2g$).

Moreover, the formula (4.28) implies that a genus $h < g$ BPS state contributes to $F_g(t)$ with a precise weight, and this corresponds to the bubbling effects we mentioned before. For example, a genus 0 BPS state contributes to $F_g$ with a weight $|B_{2g}|/(2g(2g - 2)!)$.

The Gopakumar–Vafa invariants can also be computed in some circumstances directly in terms of the geometry of embedded curves, and in many cases their computation only involves elementary algebraic geometry (Katz et al., 1999). The basic idea behind this approach is the following. The BPS states associated to D2 branes wrapped around an embedded Riemann surface of genus $g$, $\Sigma_g$, are obtained by quantization of the corresponding moduli space, as is well known from the quantization of solitons. The moduli space associated to a D2 brane has, roughly speaking, two parts. One comes from the geometric deformation moduli of the Riemann surface $\Sigma_g$ inside the Calabi–Yau manifold $X$, and we will denote it by $M_{g,\beta}$. The other part comes from the gauge fields on the D2 brane, more precisely from the moduli of flat $U(1)$ bundles over $\Sigma_g$. This moduli space is simply the Jacobian of $\Sigma_g$, which is a torus $T^{2g}$. The resulting Hilbert space is given by the tensor product of their cohomologies:

$$H = H^*(M_{g,\beta}) \otimes H^*(T^{2g}). \quad (4.30)$$

Since $M_{g,\beta}$ and $T^{2g}$ are Kähler manifolds, their cohomologies have an $SU(2)$ action given by their Lefschetz decomposition (see, for example, Griffiths and Harris, 1977). It turns out that the $SU(2)$s acting on the first and second factors in (4.30) give, respectively, the $SU(2)_R$ and $SU(2)_L$ associated to the BPS states (Witten, 1996; Gopakumar and Vafa, 1998b). From the point of view of the Lefschetz decomposition, one has that

$$H^*(T^{2g}) = I_g. \quad (4.31)$$
The SU(2)\textsubscript{R} states with \((-1)^{2j_3}\) correspond, up to a sign, to computing the Euler characteristic of \(M_{g,\beta}\). We then find a geometric interpretation for the integer invariant \(n^g_{\beta}\):

\[
n^g_{\beta} = (-1)^{\dim(M_{g,\beta})} \chi(M_{g,\beta}).
\] (4.32)

In the above discussion, we have assumed that the genus of the embedded curve \(\Sigma_g\) is constant as we vary the moduli, but this is obviously not the case in general: as we move along the moduli space, the Riemann surface develops nodal singularities, and one has to correct (4.32) to take into account this fact. It was argued by Katz et al. (1999) that the right formula expressing the Gopakumar–Vafa invariants in terms of the geometry of embedded surfaces is

\[
n^g_{\beta - \delta} = (-1)^{\dim(M_{g,\delta,\beta})} \chi(M_{g,\delta,\beta}),
\] (4.33)

where \(M_{g,\delta,\beta}\) is the moduli space of irreducible genus \(g\) curves with \(\delta\) ordinary nodes. In spite of these compelling results, a rigorous mathematical definition of the invariants is not yet known.

There is a contribution of constant maps to \(F_g\) for \(g \geq 2\), given by \(N_{g,0}\), which has not been included in (4.27). It was shown by Bershadsky et al. (1994) (see also Getzler and Pandharipande, 1998) that this contribution can be written as a Hodge integral

\[
N_{g,0} = (-1)^g \frac{\chi(X)}{2} \int_{M_g} c^3_{g-1}(\mathcal{E}), \quad g \geq 2,
\] (4.34)

where \(\chi(X)\) is the Euler characteristic of the Calabi–Yau manifold \(X\). The above integral can be evaluated explicitly to give (Faber and Pandharipande, 2000)

\[
N_{g,0} = \frac{(-1)^g \chi(X) |B_{2g}B_{2g-2}|}{4g(2g-2)(2g-2)!}.
\] (4.35)

This expression for the degree zero Gromov-Witten invariants can be also deduced from the physical picture of Gopakumar and Vafa (1998b) in terms of D branes in type IIA string theory. In this picture, the constant map contribution comes from integrating out D0 branes. The embedded Riemann surfaces are now points, and the moduli space of geometric deformations for the D0 branes is the whole target space \(X\). The contribution can then be obtained from the integral (4.25) with \(\beta = 0\) and \(n^g_{\beta} = -\delta_{g0} \chi(X)/2\), where \(\chi(X)\) is the Euler characteristic of the Calabi–Yau threefold \(X\). This leads to the generating functional for degree zero Gromov-Witten invariants

\[
F_{\beta=0} = \frac{-\chi(X)}{2} \sum_{d=1}^{\infty} \frac{1}{d} \left( \frac{1}{2 \sin \frac{dg_s}{2}} \right)^2.
\] (4.36)

A detailed study of the asymptotic expansion of this function as a power series in \(g_s\) leads to the following expression (Dabholkar et al., 2005)
\[ F_{\beta=0} = -\frac{\chi(X)}{2} \left[ g_s^{-2} \zeta(3) + K - \sum_{g=2}^{\infty} g_s^{2g-2} \frac{(-1)^g|B_{2g}||B_{2g-2}|}{2g(2g-2)(2g-2)!} \right] \]  

(4.37)

where

\[ K = \frac{1}{12} \log \frac{2\pi i}{g_s} - \frac{1}{2\pi^2} \zeta'(2) + \frac{1}{12} \gamma_E; \]  

(4.38)

and \( \gamma_E \) is the Euler-Mascheroni constant. In this way we recover the result (4.35) for \( g \geq 2 \). Notice that the terms with \( g \geq 2 \) in the asymptotic expansion can be obtained by simply expanding (4.36) in powers of \( g_s \) and using the zeta function regularization in the resulting sums over \( d \). The degree zero Gromov-Witten invariants (4.35) can also be obtained from type IIA/heterotic string duality (Mariño and Moore, 1999).

### 4.4 Open topological strings

#### 4.4.1 Type A model

So far we have discussed topological sigma models defined on closed Riemann surfaces and closed topological strings, but both theories can be extended to the open case. The natural starting point is a topological sigma model in which the worldsheet is now a Riemann surface \( \Sigma_{g,h} \) of genus \( g \) with \( h \) holes. Such models were analysed in detail by Witten (1995). The main issue is, of course, to specify boundary conditions for the maps \( f : \Sigma_{g,h} \rightarrow X \). It turns out that the relevant boundary conditions are Dirichlet and given by Lagrangian submanifolds of the Calabi–Yau \( X \). A Lagrangian submanifold \( \mathcal{L} \) is a cycle on which the Kähler form vanishes:

\[ J|_{\mathcal{L}} = 0. \]  

(4.39)

If we denote by \( C_i, i = 1, \cdots, h \), the boundaries of \( \Sigma_{g,h} \) we have to pick a Lagrangian submanifold \( \mathcal{L} \), and consider holomorphic maps such that

\[ f(C_i) \subset \mathcal{L}. \]  

(4.40)

These boundary conditions are a consequence of requiring \( \mathcal{Q} \)-invariance at the boundary. One also has boundary conditions on the Grassmann fields of the topological sigma model, which require that \( \chi \) and \( \psi \) at the boundary \( C_i \) take values in \( f^*(T\mathcal{L}) \).

We can also couple the theory to Chan–Paton degrees of freedom on the boundaries, giving rise to a \( U(N) \) gauge symmetry. The model can then be interpreted as a topological open string theory in the presence of \( N \) topological D-branes wrapping the Lagrangian submanifold \( \mathcal{L} \). The Chan–Paton factors give rise to a boundary term in the presence of a gauge connection. If \( A \) is a \( U(N) \) connection on \( \mathcal{L} \), then the path integral has to be modified by inserting

\[ \prod_i \text{Tr} \mathbb{P} \exp \oint_{C_i} f^*(A), \]  

(4.41)

where we pull back the connection to \( C_i \) through the map \( f \), restricted to the boundary. In contrast to physical D-branes in Calabi–Yau manifolds, which wrap
special Lagrangian submanifolds (Becker et al., 1995; Ooguri et al., 1996), in the topological framework the conditions are relaxed to just Lagrangian.

Once boundary conditions have been specified, one can couple the model to gravity similarly to what we did in the closed case. The resulting theory is the type-A open topological string theory, and mathematically describes holomorphic maps from open Riemann surfaces of genus $g$ and with $h$ holes to the Calabi–Yau $X$, with Dirichlet boundary conditions specified by $\mathcal{L}$. These holomorphic maps are called *open string instantons*, and can also be classified topologically. The topological sector of an open string instanton is given by two different kinds of data: the boundary part and the bulk part. For the bulk part, the topological sector is labelled by relative homology classes, since we are requiring the boundaries of $f_\ast[\Sigma_{g,h}]$ to end on $L$. Therefore, we will set

$$f_\ast[\Sigma_{g,h}] = \beta \in H_2(X, \mathcal{L}). \tag{4.42}$$

To specify the topological sector of the boundary, we will assume that $b_1(\mathcal{L}) = 1$, so that $H_1(\mathcal{L})$ is generated by a non-trivial one-cycle $\gamma$. We then have

$$f_\ast[C_i] = w_i \gamma, \quad w_i \in \mathbb{Z}, \quad i = 1, \ldots, h, \tag{4.43}$$
in other words, $w_i$ is the winding number associated to the map $f$ restricted to $C_i$. We will collect these integers into a single $h$-uple denoted by $w = (w_1, \ldots, w_h)$.

The free energy of type-A open topological string theory at fixed genus and boundary data $w$, which we denote by $F_{w,g}(t)$, can be computed as a sum over open string instantons labelled by the bulk classes:

$$F_{w,g}(t) = \sum_\beta F_{w,g,\beta} Q^\beta. \tag{4.44}$$

In this equation, the sum is over relative homology classes $\beta \in H_2(X, \mathcal{L})$, and the notation is as in (3.53). The quantities $F_{w,g,\beta}$ are *open Gromov–Witten invariants*. They ‘count’ in an appropriate sense the number of holomorphically embedded Riemann surfaces of genus $g$ in $X$ with Lagrangian boundary conditions specified by $\mathcal{L}$, and in the class represented by $\beta, w$. They are in general rational numbers. In contrast to conventional Gromov–Witten invariants, a rigorous theory of open Gromov–Witten invariants is not yet available. However, localization techniques make it possible to compute them in some situations (Katz and Liu, 2002; Li and Song, 2002; Graber and Zaslow, 2002; Mayr, 2002).

In order to consider all topological sectors, we have to introduce the string coupling constant $g_s$, which takes care of the genus, as well as a Hermitian $M \times M$ matrix $V$, which takes care of the different winding numbers $w$. The total free energy is defined by

$$F(V) = \sum_{g=0}^\infty \sum_{h=1}^\infty \sum_{w_1, \ldots, w_h} \frac{i^h}{h!} g_s^{2g-2+h} F_{w,g}(t) \text{Tr} V^{w_1} \cdots \text{Tr} V^{w_h}. \tag{4.45}$$

The factor $i^h$ is introduced for convenience, while $h!$ is a symmetry factor which takes into account that the holes are indistinguishable. Notice that, in order
to distinguish all possible topological sectors, one has to take $V$ to have infinite rank, and formally we can think about the different traces in (4.45) as symmetric functions in an infinite number of variables, like we did in Chapter 2.

We saw before that, for closed topological strings, the relation with type IIA superstrings makes it possible to express the free energies in terms of a set of integer invariants related to an appropriate counting of BPS states. In the open case, a relation of this type also holds, and we can re-express the amplitude (4.45) in terms of another set of integer invariants. The appropriate setting in type IIA string theory is the following. Consider $M$ D4 branes wrapped around a special Lagrangian submanifold $L$. The remaining two directions in the uncompactified dimensions lead to an $N = 2$ supersymmetric gauge theory in $\mathbb{R}^{1,1}$. This theory has $b_1(L)$ chiral multiplets $\Sigma_\alpha$ in two dimensions, $\alpha = 1, \cdots, b_1(L)$, corresponding to the geometric deformations of $L$ inside the Calabi–Yau $X$, and in the adjoint representation of $U(M)$. We will assume for simplicity that $b_1(L) = 1$, so we have just one chiral multiplet $\Phi$. The field $V = e^{i\Phi}$ is then a $U(M)$-valued matrix field. It was shown by Ooguri and Vafa (2000) that the couplings $F_{g,w}(t)$ in (4.45) compute the following terms in the supersymmetric gauge theory on $\mathbb{R}^{1,1}$:

$$\int d^2 x d^2 \theta F_{g,w}(t) \text{Tr} V^w_1 \cdots \text{Tr} V^w_h (W^2)^g (W \cdot v)^{h-1},$$

(4.46)

where $W$ is the graviphoton multiplet, $v^{\mu\nu}$ is a vector orthogonal to $\mathbb{R}^{1,1}$ in $\mathbb{R}^{3,1}$, and $W \cdot v = W_{\mu\nu} v^{\mu\nu}$. Once we have a target space interpretation for the open string amplitudes, we can follow the same arguments that we used for closed string amplitudes in the previous section. We first give a constant vacuum expectation value to the graviphoton field strength, $T = g_s$. The couplings (4.46) add up to compute a correction of the form

$$F(V) \int d^2 x R.$$

(4.47)

This can again be calculated in terms of BPS states, now in two dimensions. The two-dimensional Schwinger computation is similar to (4.19), but it has a single power of sin in the denominator. The BPS states that contribute to these coupling are now obtained from D2 branes wrapping around a Riemann surface of genus $g$ with $\ell$ holes, $\Sigma_{g,\ell}$, which ends on the D4 branes. The D2 branes are in addition bound to D0 branes. These BPS states are characterized topologically by the relative cohomology class of the embedded Riemann surface, $\beta \in H_2(X, L)$, and they have two quantum numbers: the $U(1)$ spin $S$ in two dimensions, and the $U(1)$ R-charge $R$ that appears in the supersymmetry algebra. Out of these two charges, we can form the quantum numbers $S_{L,R} = S \pm R$, which are the analogues of $j_{L,R}^3$. Finally, the fact that there are $M$ D4 branes means that the BPS states will form representations of $U(M)$. The mass $m_{R,\beta}$ of a BPS state with these quantum numbers, bound to $d$ D0 branes, is given by

$$e^{-m_{R,\beta}} = e^{-(\beta \cdot t + 2\pi i d)} \text{Tr}_R V,$$

(4.48)
where $V = e^{i\phi}$ and $\phi$ is the vacuum expectation value of the lowest component of $\Phi$. The contribution of a multiplet of BPS particles in two dimensions with charges $R, \beta$ is

$$
\int_0^\infty \frac{d\tau}{\tau} e^{-\tau m_{R,\beta}} \frac{\tau}{2 \sin \frac{\tau g_s}{2}} \text{Tr}_{\mathcal{H}} (-1)^F e^{i\tau s g_s},
$$

(4.49)

where $F = S_L + S_R$, $s = 2S_L$, and $\mathcal{H}$ is the corresponding Hilbert space of states. After summing over $\beta, R$, and $d$, the total contribution to the open free energy can then be written as

$$
\sum_{\beta, R} \sum_{d=1}^\infty \frac{e^{-d\beta t}}{2d \sin \frac{dg_s}{2}} \text{Tr}_{\mathcal{H}} (-1)^F q^{ds} \text{Tr}_R V^d,
$$

(4.50)

where we have used (4.26) to perform the integral over $\tau$, and $q = e^{ig_s}$.

It remains to compute the trace over the Hilbert space of states in (4.50). In order to do that, we will use the geometric description of this space that arises from the moduli space of embedded D2 branes ending on D4 branes. This moduli space is the product of three factors: the moduli of Abelian gauge fields on the worldvolume of the D2 brane, the moduli of geometric deformations of the D2s in the ambient space, and finally the Chan–Paton factors associated to the boundaries of the D2, which lead to the charges $R$. If the D2s are genus $g$ surfaces with $\ell$ holes in the relative cohomology class labelled by $\beta$, the moduli space of Abelian gauge fields gives rise to the Jacobian $J_{g,\ell} = T^{2g+\ell-1}$. Here, we are assuming that the genus and number of holes of the surface are constant along the moduli. Although this assumption is strictly speaking not true geometrically, it leads to the correct structure of the amplitudes, as we saw in the discussion of Gopakumar–Vafa invariants in the closed case. We will denote the moduli space of geometric deformations by $\mathcal{M}_{g,\ell,\beta}$. Finally, for the Chan–Paton degrees of freedom we get a factor of $F$ (the fundamental representation of $SU(M)$) from each hole. The Hilbert space is obtained by computing the cohomology of these moduli, and we obtain

$$
F^{\otimes \ell} \otimes H^*(J_{g,\ell}) \otimes H^*(\mathcal{M}_{g,\ell,\beta}).
$$

(4.51)

As in the closed string case, the $S_L$ quantum number is identified with the degree of the forms in $H^*(J_{g,\ell})$, and $S_R$ is identified with the degree of the forms in $H^*(\mathcal{M}_{g,\ell,Q})$. An important point is that this Hilbert space is associated with the moduli space of $\ell$ distinguished holes, which is not physical, and we have to mod out by the action of the permutation group $S_\ell$. We can factor out the cohomology of the Jacobian $T^{2g}$ of the ‘bulk’ Riemann surface, $H^*(T^{2g})$, since the permutation of the holes does not act on it. The other factor of the Hilbert space is the projected space

$$
\text{Sym} \left( F^{\otimes \ell} \otimes H^*((S^1)^{\ell-1}) \otimes H^*(\mathcal{M}_{g,\ell,Q}) \right).
$$

(4.52)

This projection can be expressed in terms of the projections of the factors on different representations of $S_\ell$. Using results from the representation theory of
the symmetric group (Fulton and Harris, 1991) we find that (4.52) can be written as a direct sum:

$$\bigoplus_{RR'R''} C_{RR'R''}S_R(F^\otimes \ell) \otimes S_{R'}(H^*((S^1)^{\ell-1})) \otimes S_{R''}(H^*(M_{g,\ell,Q})). \quad (4.53)$$

In this equation, $S_R$ is the Schur functor that projects onto the corresponding subspace, and $C_{RR'R''}$ are the Clebsch–Gordan coefficients of the symmetric group, which can be written in terms of characters as follows

$$C_{RR'R''} = \sum_{\vec{k}} \frac{|C(\vec{k})|}{\ell!} \chi_R(C(\vec{k})) \chi_{R'}(C(\vec{k})) \chi_{R''}(C(\vec{k})). \quad (4.54)$$

The space $S_R(F^\otimes \ell)$ is the vector space underlying the irreducible representation $R$ of $U(M)$. We then find that the trace over the Hilbert space $\mathcal{H}_{R,g,\beta}$ associated to the representation $R$ coming from (4.51) is given by:

$$\text{Tr}_{\mathcal{H}_{R,g,\beta}}(-1)^F q^s = \left(2 \sin \frac{g_s}{2}\right)^{2g} \sum_{R',R''} C_{RR'R''} N_{R',R''} \text{Tr}_{S_{R''}(H_\ell)}(-1)^F q^s, \quad (4.55)$$

where we denoted by $H_\ell = H^*((S^1)^{\ell-1})$ the cohomology of the product of circles, and

$$N_{R,g,\beta} = (-1)^n \chi(S_{R''}(H^*(M_{g,\ell,\beta}))) \quad (4.56)$$

is the analogue of (4.32) for the open case. In this equation,

$$n = \text{dim}(S_{R''}(H^*(M_{g,\ell,\beta}))).$$

To complete the argument, we have to compute the last trace in (4.55), which we denote by

$$S_R(q) = \text{Tr}_{S_R(H_\ell)}(-1)^F q^s. \quad (4.57)$$

In order to do that, we have to be more specific about the action of the permutation group on the cohomology elements. Although the permutation group $S_\ell$ acts in a natural way on a Riemann surface with $\ell$ boundaries, there are only $\ell - 1$ independent one-forms associated to the boundary. This is because the one-forms $d\theta_i, i = 1, \cdots, \ell$, which are Poincaré dual to the holes in the Riemann surface, satisfy $\sum_i d\theta_i = 0$. The procedure to construct the Hilbert space $S_R(H_\ell)$ is then as follows. We consider the Hilbert space $\mathcal{F}_\ell$ generated by $\ell$ fermion fields $\psi_i, i = 1, \cdots, \ell$ acting on the vacuum $|0\rangle$, and we decompose it with respect to the different representations $R$ by using the Young symmetrizers of the corresponding tableaux. In doing that, it is very important to take into account the Grassmann nature of the fermions. Finally, we impose the linear constraint $\sum_i \psi_i = 0$. Let us consider, for example, the simple case $\ell = 2$. The space $\mathcal{F}_2$ is spanned by the four states $|0\rangle, \psi_1|0\rangle$ and $\psi_1\psi_2|0\rangle$. The relevant permutation
group $S_2$ corresponds to permuting $\psi_1 \leftrightarrow \psi_2$. Projecting onto the symmetric and anti-symmetric subspaces, we find:

$$\begin{align*}
\Box : & \quad |0\rangle, \ (\psi_1 + \psi_2)|0\rangle ; \\
\blacklozenge : & \quad (\psi_1 - \psi_2)|0\rangle, \ \psi_1\psi_2|0\rangle. 
\end{align*}$$

(4.58)

Using that $\psi_1 + \psi_2 = 0$, the spectrum turns out to be:

$$\begin{align*}
\Box : & \quad |0\rangle ; \\
\blacklozenge : & \quad \psi_1|0\rangle. 
\end{align*}$$

(4.59)

To assign spins to these states, we first note that $\psi_i$ have $s = 1$. However, this does not fix the spin assignment of all the states, as we have to choose a spin for the ground state. There is, in fact, a natural choice that leads to an average zero spin in a given multiplet. In the case at hand the two states differ in spin by 1, and so the symmetric choice of spin assignment is spin $\mp 1/2$. It is clear now how to proceed to find the spin content of the various representations that arise in this way, when we have more holes. The trivial representation of $S_\ell$, corresponding to a Young tableau with $\ell$ boxes and only one row, is given by the vacuum $|0\rangle$, and we assign it spin $-(\ell - 1)/2$. By acting with one fermion $\psi_i$ on the vacuum, we obtain $\ell$ states, forming a reducible representation $(\ell)$ of $S_\ell$ that decomposes into reducible representations as follows:

$$\begin{align*}
(\ell) = (\ell - 1) \oplus (1). 
\end{align*}$$

(4.60)

The first summand corresponds to the standard representation $V$ of $S_\ell$, with a Young tableau of the form

with $\ell - 1$ boxes in the first row. The second summand corresponds to the trivial representation generated by $$(\sum_i \psi_i)|0\rangle$$, which we are setting to zero. To generate the rest of the spectrum, we have to take the anti-symmetrized tensor products $\wedge^d V$ (since $V$ is fermionic). These are irreducible representations of $S_\ell$ and are called hook representations, since their Young tableau is of the form

with $\ell - d$ boxes in the first row. We have then obtained the spin/representation content of the spectrum: the Hilbert spaces $S_R(H_\ell)$ are non-empty only for hook representations of the form (4.62), and in this case they contain one state of statistics $(-1)^d$ and total spin $-(\ell - 1)/2 + d$, which is equal to the spin of the vacuum plus $d$ units of the $d$ fermionic fields that appear in $\wedge^d V$. The trace (4.57) is

$$S_R(q) = \begin{cases} 
(-1)^d q^{-\frac{\ell - 1}{2} + d}, & \text{if } R \text{ is a hook representation}, \\
0, & \text{otherwise}.
\end{cases}$$

(4.63)
Exercise 4.1 Define the following polynomials:

$$P_{\vec{k}}(q) = \sum_{R} \chi_{R}(C(\vec{k})) S_{R}(q).$$  \hfill (4.64)

Show that

$$P_{\vec{k}}(q) = \prod_{j}(q^{-\frac{j}{2}} - q^\frac{j}{2})^{k_{j}}. \hfill (4.65)$$

Show also that

$$S_{R}(q^{-1}) = (-1)^{\ell-1} S_{R'}(q). \hfill (4.66)$$

We can now write down the final result for the open free energies. They can be written as

$$F(V) = \sum_{R} \sum_{d=1}^{\infty} \frac{1}{d} f_{R}(q^{d}, e^{-d\beta \cdot t}) \text{Tr}_{R} V^{d}, \hfill (4.67)$$

where

$$f_{R}(q, e^{-\beta \cdot t}) = \sum_{g \geq 0} \sum_{R', R''} C_{RR'R''} S_{R''}(q) N_{R'R''gQ} \left( 2 \sin \frac{dg_{s}}{2} \right)^{2g-1} e^{-\beta \cdot t}. \hfill (4.68)$$

Here, \((2 \sin \frac{dg_{s}}{2})^{2g}\) comes from the bulk of the Riemann surface, as in the closed string case, and the extra \(1/(2 \sin \frac{dg_{s}}{2})\) comes from the Schwinger computation in two dimensions.

In this derivation, we have assumed that the winding numbers \(w_{i}\) are all positive, so that the product of traces of \(V\) in (4.45) can be written in terms of \(\text{Tr}_{R} V\) for representations \(R\) with a small number of boxes. In this case, one can label \(w\) in terms of a vector \(\vec{k}\) of non-negative entries. Given an \(h\)-uple \(w = (w_{1}, \cdots, w_{h})\), we define a vector \(\vec{k}\) as follows: the \(i\)-th entry of \(\vec{k}\) is the number of \(w_{j}\)s that take the value \(i\). For example, if \(w = (1, 1, 2)\), the corresponding \(\vec{k}\) is \(\vec{k} = (2, 1, 0, \cdots)\). In terms of \(\vec{k}\), the number of holes and the total winding number are given by

$$h = |\vec{k}|, \quad \ell = \sum_{i} w_{i} = \sum_{j} j k_{j}. \hfill (4.69)$$

Note that a given \(\vec{k}\) will correspond to many \(w\)s that differ by a permutation of their entries. In fact there are \(h! / \prod_{j} k_{j}!\) \(h\)-uples \(w\) that give the same vector \(\vec{k}\) (and the same amplitude). We can then write the total free energy for positive winding numbers as

$$F(V) = \sum_{g=0}^{\infty} \sum_{\vec{k}} \frac{|\vec{k}|^{2g-2+h}}{\prod_{j} k_{j}^{2g_{s}}} F_{\vec{k},g}(t) \Upsilon_{\vec{k}}(V), \hfill (4.70)$$

where \(\Upsilon_{\vec{k}}(V)\) was introduced in (2.106). Define now the following integers:
\[ n_{\vec{k},g,\beta} = \sum_R \chi_R(C(\vec{k})) N_{R,g,\beta}. \]  

(4.71)

Using (4.65), one can show that
\[ \sum_{g=0}^{\infty} g^{2g-2+h} F_{w,g}(t) = \frac{(-1)^{h-1}}{\prod_i w_i} \sum_{g=0}^{\infty} \sum_{\beta} \sum_{d|w} n_{w/d,g,\beta} d^{h-1} \left( 2 \sin \frac{dg_s}{2} \right) \sum_{i=1}^{h} \left( 2 \sin \frac{w_i g_s}{2} \right) e^{-d\beta \cdot t}. \]  

(4.72)

Notice that there is one such identity for each \( w \). In this expression, the sum is over all integers \( d \) that satisfy that \( d|w_i \) for all \( i = 1, \ldots, h \). When this is the case, we define the \( h \)-uple \( w/d \) whose \( i \)-th component is \( w_i/d \). The expression (4.72) can be expanded to give formulae for the different genera. For example, in \( g = 0 \) one simply finds
\[ F_{w,g=0,\beta} = (-1)^{h-1} \sum_{d|w} d^{h-3} n_{w/d,0,\beta/d}, \]  

(4.73)

where the integer \( d \) has to divide the vector \( w \) (in the sense explained above) and it is understood that \( n_{w,d,0,\beta/d} \) is zero if \( \beta/d \) is not a relative homology class. Formulae for higher genera can be easily worked out from (4.72); see Mariño and Vafa (2002) for examples. Notice that integrality of \( N_{R,g,\beta} \) implies integrality of \( n_{\vec{k},g,\beta} \), but not the other way around. In that sense, the integer invariants \( N_{R,g,\beta} \) are the more fundamental ones.

In the derivation of (4.72) we have assumed that all the winding numbers are positive. When there are both positive and negative winding numbers, we can consider that \( V \) and \( V^{-1} \) are independent matrices and introduce two sets of vectors \( \vec{k}^{(1)}, \vec{k}^{(2)} \) associated to the positive and the negative winding numbers, respectively. This leads to integer invariants \( N_{R_1,R_2,g,\beta} \) labelled by two representations, and the above results can be easily generalized. It is easy to check that the formula (4.72) remains valid for arbitrary winding numbers.

We have also assumed that the boundary conditions are specified by a single Lagrangian submanifold with a single non-trivial one-cycle. When there are more one-cycles in the geometry, say \( L \), providing possible boundary conditions for the open strings, the above formalism has to be generalized in an obvious way: one needs to specify \( L \) sets of winding numbers \( w^{(\alpha)} \), and the generating functional (4.70) depends on \( L \) different matrices \( V_\alpha \), \( \alpha = 1, \cdots, L \). The total partition function has the structure
\[ Z(V_i) = \sum_{R_1,\cdots,R_{2L}} Z_{R_1 \cdots R_{2L}}(g_s,t) \prod_{\alpha=1}^{2L} \text{Tr}_{R_\alpha} V_\alpha, \]  

(4.74)

where the \( R_{2\alpha-1}, R_{2\alpha} \) correspond to positive and negative winding numbers, respectively, for the \( \alpha \)-th cycle.
4.4.2 Type-B model

The topological B model can also be formulated for open strings, i.e. when the worldsheet is an open Riemann surface with boundaries $\Sigma_{g,h}$ (Witten, 1995; Ooguri et al., 1996). It turns out that the appropriate boundary conditions for the type-B model are Dirichlet along holomorphic cycles of $X$, which we denote by $S$, and Neumann in the remaining directions. Moreover, one can add Chan–Paton factors to the model, and this is implemented by considering a $U(N)$ holomorphic bundle over the holomorphic cycle $S$. The resulting theory can then be interpreted as a topological B model in the presence of $N$ topological D-branes wrapping $S$. In the case of $N$ branes filling spacetime, i.e. $S = X$, the boundary conditions for the fields are $\theta = 0$ along $\partial \Sigma_{g,h}$ and that the pullback to $\partial \Sigma_{g,h}$ of $\ast \rho$ vanishes, where $\ast$ is the Hodge operator (Witten, 1995). The open topological B model can also be coupled to gravity following the same procedure that was used in the closed case, and one obtains in this way the open type-B topological string propagating along the Calabi–Yau manifold $X$. We will come back to this model in Chapter 7.
This page intentionally left blank
CALABI–YAU GEOMETRIES

So far, we have considered topological string theory on general Calabi–Yau threefolds. In this chapter we will discuss a particular class of Calabi–Yau geometries, which are characterized by being non-compact. In particular, we will discuss in detail non-compact toric Calabi–Yau threefolds. These are threefolds that have the structure of a fibration with torus fibres. The manifolds we will be interested in have the structure of a fibration of $\mathbb{R}^3$ by $T^2 \times \mathbb{R}$. It turns out that the geometry of these threefolds can be packaged in a two-dimensional graph that encodes the information about the degeneration locus of the fibration. We will often call these graphs the toric diagrams of the corresponding Calabi–Yau manifolds. Instead of relying on general ideas of toric geometry (which can be found for example in Cox and Katz, 1999, and in Hori et al., 2003), we will mainly use the approach developed by Leung and Vafa (1998), Aganagic and Vafa (2001), and specially Aganagic et al. (2005). We will give first a general introduction to the construction of non-compact Calabi–Yau geometries, and then we will present the toric approach.

5.1 Non-compact Calabi–Yau geometries: an introduction

One of the main insights in the study of topological string theory on Calabi–Yau threefolds is that the simplest models to study are associated to non-compact Calabi–Yau geometries based on manifolds of lower dimension. To construct these geometries, we start with complex manifolds in one or two complex dimensions, which in general will have a non-zero first Chern class. We then consider vector bundles over them (with the appropriate rank and curvature) that lead to a total three-dimensional space with zero first Chern class. In this way, we obtain Calabi–Yau threefolds whose non-trivial geometry is encoded in a lower-dimensional manifold, and therefore they are easier to study.

Let us first consider non-compact Calabi–Yau manifolds whose building block is a one-dimensional compact manifold. These manifolds will be given by a Riemann surface together with an appropriate bundle over it, and geometrically they can be regarded as the local geometry of an embedded Riemann surface in a general Calabi–Yau space. Indeed, consider a Riemann surface $\Sigma_g$ holomorphically embedded inside a Calabi–Yau threefold $X$, and let us look at the holomorphic tangent bundle of $X$ restricted to $\Sigma_g$. We have

$$TX|_{\Sigma_g} = T\Sigma_g \oplus \mathcal{N}_{\Sigma_g}, \quad (5.1)$$

where $\mathcal{N}_{\Sigma_g}$ is a holomorphic rank-two complex vector bundle over $\Sigma_g$, called the normal bundle of $\Sigma_g$, and the Calabi–Yau condition $c_1(X) = 0$ gives
The Calabi–Yau $X$ ‘near $\Sigma_g$’ then looks like the total space of the bundle

$$N \rightarrow \Sigma_g,$$  

(5.3)

where $N$ is regarded here as a rank-two bundle over $\Sigma_g$ satisfying (5.2). The non-compact space (5.3) is an example of a local Calabi–Yau threefold.

When $g = 0$ and $\Sigma_g = \mathbb{P}^1$ it is possible to be more precise about the bundle $N$. A theorem due to Grothendieck says that any holomorphic bundle over $\mathbb{P}^1$ splits into a direct sum of line bundles (for a proof, see for example Griffiths and Harris, 1977, pp. 516–7). Line bundles over $\mathbb{P}^1$ are all of the form $\mathcal{O}(n)$, where $n \in \mathbb{Z}$. The bundle $\mathcal{O}(n)$ can be easily described in terms of two charts on $\mathbb{P}^1$: the north-pole chart, with co-ordinates $z, \Phi$ for the base and the fibre, respectively, and the south-pole chart, with co-ordinates $z', \Phi'$. The change of co-ordinates is given by

$$z' = \frac{1}{z}, \quad \Phi' = z^{-n}\Phi.$$  

(5.4)

We also have that $c_1(\mathcal{O}(n)) = n$. We then find that local Calabi–Yau manifolds that are made out of a two-sphere together with a bundle over it are all of the form

$$\mathcal{O}(-a) \oplus \mathcal{O}(a-2) \rightarrow \mathbb{P}^1,$$  

(5.5)

since the degrees of the bundles have to sum up to $-2$ due to (5.2). An important case occurs when $a = 1$. The resulting non-compact manifold,

$$\mathcal{O}(-1) \oplus \mathcal{O}(-1) \rightarrow \mathbb{P}^1,$$  

(5.6)

is called the resolved conifold for reasons that will be explained later.

We can also consider non-compact Calabi–Yau threefolds based on compact complex surfaces. Consider a complex surface $S$ embedded in a Calabi–Yau manifold $X$. As before, we can split the tangent bundle as

$$TX|_S = TS \oplus N_S,$$  

(5.7)

where the normal bundle $N_S$ is now of rank one. The Calabi–Yau condition leads to

$$c_1(N_S) = c_1(K_S),$$  

(5.8)

where $K_S$ is the canonical line bundle over $S$, and we used that $c_1(TS) = -c_1(K_S)$. Therefore, we have $N_S = K_S$. The Calabi–Yau $X$ ‘near $S$’ looks like the total space of the bundle

$$K_S \rightarrow S.$$  

(5.9)

This construction gives a whole family of non-compact Calabi–Yau manifolds that are also referred to as local Calabi–Yau manifolds. A well-known example is $S = \mathbb{P}^2$, the two-dimensional projective space, which leads to the Calabi–Yau manifold

$$\mathcal{O}(-3) \rightarrow \mathbb{P}^2,$$  

(5.10)

also known as local $\mathbb{P}^2$. Another important example is $S = \mathbb{P}^1 \times \mathbb{P}^1$, which leads to local $\mathbb{P}^1 \times \mathbb{P}^1$. 

$$c_1(N_{\Sigma_g}) = 2g - 2.$$  

(5.2)
5.2 Constructing toric Calabi–Yau manifolds

Many of the examples of non-compact Calabi–Yau threefolds considered above are toric, i.e. they have the structure of a torus fibration, and can be constructed in a systematic way by a ‘cut and paste’ procedure. In this section we will develop these techniques, following the approach of Aganagic et al. (2005).

5.2.1 \( \mathbb{C}^3 \)

The elementary building block for the technique we want to develop is a very simple non-compact Calabi–Yau threefold, namely \( \mathbb{C}^3 \). We will now exhibit its structure as a \( \mathbb{T}^2 \times \mathbb{R} \) fibration over \( \mathbb{R}^3 \), and we will encode this information in a simple trivalent, planar graph.

Let \( z_i \) be complex co-ordinates on \( \mathbb{C}^3 \), \( i = 1, 2, 3 \). We introduce three functions or Hamiltonians

\[
\begin{align*}
r_\alpha(z) &= |z_1|^2 - |z_3|^2, \\
r_\beta(z) &= |z_2|^2 - |z_3|^2, \\
r_\gamma(z) &= \text{Im}(z_1 z_2 z_3). 
\end{align*}
\]  

(5.11)

These Hamiltonians generate three flows on \( \mathbb{C}^3 \) via the standard symplectic form \( \omega = i \sum_j dz_j \wedge d\bar{z}_j \) on \( \mathbb{C}^3 \) and the Poisson brackets

\[
\partial_\nu z_i = \{r_\nu, z_i\}_\omega, \quad \nu = \alpha, \beta, \gamma. 
\]  

(5.12)

This gives the fibration structure that we were looking for: the base of the fibration, \( \mathbb{R}^3 \), is parameterized by the Hamiltonians (5.11), while the fibre \( \mathbb{T}^2 \times \mathbb{R} \) is parameterized by the flows associated to the Hamiltonians. In particular, the \( \mathbb{T}^2 \) fibre is generated by the circle actions

\[
e^{\alpha r_\alpha + \beta r_\beta} : (z_1, z_2, z_3) \rightarrow (e^{i\alpha} z_1, e^{i\beta} z_2, e^{-i(\alpha+\beta)} z_3),
\]  

(5.13)

while \( r_\gamma \) generates the real line \( \mathbb{R} \). We will call the cycle generated by \( r_\alpha \) the \( (0,1) \) cycle, and the cycle generated by \( r_\beta \) the \( (1,0) \) cycle.

Notice that the \( (0,1) \) cycle degenerates over the subspace of \( \mathbb{C}^3 \) described by \( z_1 = 0 = z_3 \), which is the subspace of the base \( \mathbb{R}^3 \) given by \( r_\alpha = r_\gamma = 0, r_\beta \geq 0 \). Similarly, over \( z_2 = 0 = z_3 \) the \( (1,0) \)-cycle degenerates over the subspace \( r_\beta = r_\gamma = 0 \) and \( r_\alpha \geq 0 \). Finally, the one-cycle parameterized by \( \alpha + \beta \) degenerates over \( z_1 = 0 = z_2 \), where \( r_\alpha - r_\beta = 0 = r_\gamma \) and \( r_\alpha \leq 0 \).

We will represent the \( \mathbb{C}^3 \) geometry by a graph that encodes the degeneration loci in the \( \mathbb{R}^3 \) base. In fact, it is useful to have a planar graph by taking \( r_\gamma = 0 \) and drawing the lines in the \( r_\alpha - r_\beta \) plane. The degeneration locus will then be straight lines described by the equation \( pr_\alpha + qr_\beta = \text{const} \). Over this line the \((-q, p)\) cycle of the \( \mathbb{T}^2 \) degenerates. Therefore we correlate the degenerating cycles unambiguously with the lines in the graph (up to \( (q, p) \rightarrow (-q, -p) \)). This yields the graph in Fig. 5.1, drawn in the \( r_\gamma = 0 \) plane.

There is a symmetry in the \( \mathbb{C}^3 \) geometry that makes it possible to find other representations by different toric graphs. These graphs are characterized by three
Fig. 5.1. This graph represents the degeneration locus of the $T^2 \times \mathbb{R}$ fibration of $\mathbb{C}^3$ in the base $\mathbb{R}^3$ parameterized by $(r_\alpha, r_\beta, r_\gamma)$.

Vectors $v_i$ that are obtained from those in Fig. 5.1 by an $SL(2, \mathbb{Z})$ transformation. The vectors have to satisfy
\[ \sum_i v_i = 0. \] (5.14)

The $SL(2, \mathbb{Z})$ symmetry is inherited from the $SL(2, \mathbb{Z})$ symmetry of $T^2$ that appeared in Chapter 2 in the context of Chern–Simons theory. In the above discussion the generators $H_1(T^2)$ have been chosen to be the one-cycles associated to $r_\alpha$ and $r_\beta$, but there are other choices that differ from this one by an $SL(2, \mathbb{Z})$ transformation on the $T^2$. For example, we can choose $r_\alpha$ to generate a $(p, q)$ one-cycle and $r_\beta$ a $(t, s)$ one-cycle, provided that $ps - qt = 1$. These different choices give different trivalent graphs. As we will see in the examples below, the construction of general toric geometries requires these more general graphs representing $\mathbb{C}^3$.

5.2.2 The general case

The non-compact, toric Calabi–Yau threefolds that we will study can be described as symplectic quotients. Let us consider the complex linear space $\mathbb{C}^{N+3}$, described by $N + 3$ co-ordinates $z_1, \cdots, z_{N+3}$, and let us introduce $N$ real equations of the form
\[ \mu_A = \sum_{j=1}^{N+3} Q_A^j |z_j|^2 = t_A, \quad A = 1, \cdots, N. \] (5.15)

In this equation, $Q_A^j$ are integers satisfying
\[ \sum_{j=1}^{N+3} Q^j_A = 0. \] (5.16)

This condition is equivalent to \( c_1(X) = 0 \), i.e. to the Calabi–Yau condition. We consider the action of the group \( G_N = U(1)^N \) on the zs where the \( A \)-th \( U(1) \) acts on \( z_j \) by

\[ z_j \rightarrow \exp(iQ^j_A \alpha_A)z_j. \]

The space defined by (5.15), quotiented by the group action \( G_N \),

\[ X = \bigcap_{A=1}^{N} \mu_A^{-1}(t_A)/G_N \] (5.17)

turns out to be a Calabi–Yau manifold (it can be seen that the condition (5.16) is equivalent to the Calabi–Yau condition). The \( N \) parameters \( t_A \) are Kähler moduli of the Calabi–Yau. This mathematical description of \( X \) appears in the study of the two-dimensional linear sigma model with \( \mathcal{N} = (2, 2) \) supersymmetry (Witten, 1993). The theory has \( N+3 \) chiral fields, whose lowest components are the zs and are charged under \( N \) vector multiplets with charges \( Q^j_A \). The equations (5.15) are the D-term equations, and after dividing by the \( U(1)^N \) gauge group we obtain the Higgs branch of the theory.

The Calabi–Yau manifold \( X \) defined in (5.17) can be described by \( \mathbb{C}^3 \) geometries glued together in an appropriate way. Since each of these \( \mathbb{C}^3 \)'s is represented by the trivalent vertex depicted in Fig. 5.1, we will be able to encode the geometry of (5.17) into a trivalent graph. In order to provide this description, we must first find a decomposition of the set of all co-ordinates \( \{z_j\}_{j=1}^{N+3} \) into triplets \( U_a = (z_{ia}, z_{ja}, z_{ka}) \) that correspond to the decomposition of \( X \) into \( \mathbb{C}^3 \) patches. We pick one of the patches and we associate to it two Hamiltonians \( r_\alpha, r_\beta \) as we did for \( \mathbb{C}^3 \) before. These two co-ordinates will be global co-ordinates in the base \( \mathbb{R}^3 \), therefore they will generate a globally defined \( \mathbb{T}^2 \) fibre. The third co-ordinate in the base is \( r_\gamma = \text{Im}(\prod_{j=1}^{N+3} z_j) \), which is manifestly gauge invariant and moreover, patch by patch, can be identified with the co-ordinate used in the \( \mathbb{C}^3 \) example above. Equation (5.15) can then be used to find the action of \( r_{\alpha,\beta} \) on the other patches.

We will now exemplify this procedure with two important examples: the resolved conifold and the local \( \mathbb{P}^2 \) geometry, which were introduced before as local Calabi–Yau geometries.

**Example 5.1** The resolved conifold. The resolved conifold (5.6) has a description of the form (5.17), with \( N = 1 \). There is only one constraint given by

\[ |z_1|^2 + |z_4|^2 - |z_2|^2 - |z_3|^2 = t, \] (5.18)

and the \( U(1) \) group acts as

\[ z_1, z_2, z_3, z_4 \rightarrow e^{i\alpha} z_1, e^{-i\alpha} z_2, e^{-i\alpha} z_3, e^{i\alpha} z_4. \] (5.19)
Fig. 5.2. The graph associated to the resolved conifold $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1$. This manifold is made out of two $\mathbb{C}^3$ patches glued through a common edge.

Notice that, for $z_2 = z_3 = 0$, (5.18) describes a $\mathbb{P}^1$ whose area is proportional to $t$. Therefore, $(z_1, z_4)$ can be taken as homogeneous co-ordinates of the $\mathbb{P}^1$ that is the basis of the fibration, while $z_2, z_3$ can be regarded as co-ordinates for the fibres.

Let us now give a description in terms of $\mathbb{C}^3$ patches glued together. The first patch will be defined by $z_4 \neq 0$. Using (5.18) we can solve for the modulus of $z_4$ in terms of the other co-ordinates, and using the $U(1)$ action we can gauge away its phase. Therefore, the patch will be parameterized by $U_4 = (z_1, z_2, z_3)$. The Hamiltonians will be, in this case,

$$r_\alpha(z) = |z_2|^2 - |z_1|^2,$$
$$r_\beta(z) = |z_3|^2 - |z_1|^2,$$

which generate the actions

$$e^{\alpha r_\alpha + \beta r_\beta} : (z_1, z_2, z_3) \to (e^{-i(\alpha + \beta)}z_1, e^{i\alpha}z_2, e^{i\beta}z_3).$$

This patch will be represented by the same graph that we found for $\mathbb{C}^3$. The other patch will be defined by $z_1 \neq 0$, therefore we can write it as $U_1 = (z_4, z_2, z_3)$. However, in this patch $z_1$ is no longer a natural co-ordinate, but we can use (5.18) to rewrite the Hamiltonians as

$$r_\alpha(z) = |z_4|^2 - |z_3|^2 - t,$$
$$r_\beta(z) = |z_4|^2 - |z_2|^2 - t,$$

generating the action
\[ e^{\alpha r + \beta r_{\beta}} : (z_4, z_2, z_3) \rightarrow (e^{i(\alpha + \beta)} z_4, e^{-i\beta} z_2, e^{-i\alpha} z_3). \] (5.23)

The degeneration loci in this patch are the following: i) \( z_4 = 0 = z_2 \), corresponding to the line \( r_{\beta} = -t \) where a \((-1,0)\) cycle degenerates; ii) \( z_4 = 0 = z_3 \), corresponding to the line \( r_{\alpha} = -t \), and with a \((0,1)\) cycle degenerating; iii) finally, \( z_2 = 0 = z_3 \), where \( r_{\alpha} - r_{\beta} = 0 \), and a \((1,1)\) cycle degenerates. This patch is identical to the first one, and they are joined together through the common edge where \( z_2 = 0 = z_3 \). The full construction is represented in Fig. 5.2. Notice that the common edge of the graphs represents the \( \mathbb{P}^1 \) of the resolved conifold: along this edge, one of the \( S^1 \)s of \( T^2 \) has degenerated, while the other only degenerates at the endpoints. An \( S^1 \) fibration of an interval that degenerates at its endpoints is simply a two-sphere. The length of the edge is \( t \), the Kähler parameter associated to the \( \mathbb{P}^1 \).

**Example 5.2 Local \( \mathbb{P}^2 \).** Let us now consider a more complicated example, namely local \( \mathbb{P}^2 \), which is the total space of the bundle (5.10). We can describe it again as in (5.17) with \( N = 1 \). There are four complex variables, \( z_0, \ldots, z_3 \), and the constraint (5.15) now reads

\[ |z_1|^2 + |z_2|^2 + |z_3|^2 - 3|z_0|^2 = t. \] (5.24)

The \( U(1) \) action on the \( z \)'s is

\[ (z_0, z_1, z_2, z_3) \rightarrow (e^{-3i\alpha} z_0, e^{i\alpha} z_1, e^{i\alpha} z_2, e^{i\alpha} z_3). \] (5.25)

Notice that \( z_{1,2,3} \) describe the basis \( \mathbb{P}^2 \), while \( z_0 \) parameterizes the complex direction of the fibre.

Let us now give a description in terms of glued \( \mathbb{C}^3 \) patches. There are three patches \( U_i \) defined by \( z_i \neq 0 \), for \( i = 1, 2, 3 \), since at least one of these three co-ordinates must be non-zero in \( X \). All of these three patches look like \( \mathbb{C}^3 \). For example, for \( z_3 \neq 0 \), we can ‘solve’ again for \( z_3 \) in terms of the other three unconstrained co-ordinates that then parameterize \( \mathbb{C}^3 \): \( U_3 = (z_0, z_1, z_2) \). Similar statements hold for the other two patches. Let us now construct the corresponding degeneration graph. In the \( U_3 = (z_0, z_1, z_2) \) patch we take as our Hamiltonians

\[ r_{\alpha} = |z_1|^2 - |z_0|^2, \quad r_{\beta} = |z_2|^2 - |z_0|^2. \] (5.26)

The graph of the degenerate fibres in the \( r_{\alpha} - r_{\beta} \) plane is the same as in the \( \mathbb{C}^3 \) example, Fig. 5.1. The third direction in the base, \( r_{\gamma} \) is now given by the gauge invariant product \( r_{\gamma} = \text{Im}(z_0 z_1 z_2 z_3) \). The same two Hamiltonians \( r_{\alpha,\beta} \) generate the action in the \( U_2 = (z_0, z_1, z_3) \) patch, and we use the constraint (5.24) to rewrite them as follows: since both \( z_0 \) and \( z_1 \) are co-ordinates of this patch \( r_{\alpha} \) does not change. On the other hand, \( r_{\beta} \) must be rewritten since \( z_2 \) is not a natural co-ordinate here. We then find:
Fig. 5.3. The graph of $O(-3) \to \mathbb{P}^2$. This manifold is built out of three $\mathbb{C}^3$ patches.

$$r_\alpha = |z_1|^2 - |z_0|^2,$$

$$r_\beta = t + 2|z_0|^2 - |z_1|^2 - |z_3|^2,$$  

hence

$$e^{\alpha r_\alpha + \beta r_\beta} : (z_0, z_1, z_3) \to (e^{i(-\alpha+2\beta)}z_0, e^{i(\alpha-\beta)}z_1, e^{-\beta}z_3).$$

We see from the above that the fibres degenerate over three lines: i) $r_\alpha + r_\beta = t$, corresponding to $z_0 = 0 = z_3$, and where a $(-1, 1)$ cycle degenerates; ii) there is a line over which a $(-1, 2)$ cycle degenerates where $z_1 = 0 = z_3$, $2r_\alpha + r_\beta = t$, and finally, iii) there is a line over which $r_\alpha = 0$, and a $(0, 1)$-cycle degenerates. The $U_1$ patch is similar, and we end up with the graph for $O(-3) \to \mathbb{P}^2$ shown in Fig. 5.3.

Example 5.3 Lagrangian submanifolds. In order to consider open string amplitudes in the above Calabi–Yau geometries, we have to construct Lagrangian submanifolds providing boundary conditions, as we explained in section 4.4. Let us start by considering the $\mathbb{C}^3$ geometry discussed above. In this case, one can easily construct Lagrangian submanifolds following the work of Harvey and Lawson (1982). In terms of the Hamiltonians in (5.11), we have three types of them:

$$L_1 : r_\alpha = 0, \quad r_\beta = r_1, \quad r_\gamma \geq 0,$$

$$L_2 : r_\alpha = r_2, \quad r_\beta = 0, \quad r_\gamma \geq 0,$$

$$L_3 : r_\alpha = r_\beta = r_3, \quad r_\gamma \geq 0,$$  

(5.28)

where $r_i, i = 1, 2, 3$, are constants. It is not difficult to check that the above submanifolds are indeed Lagrangian (they turn out to be Special Lagrangian as
In terms of the graph description we developed above, they correspond to points in the edges of the planar graph spanned by \((r_\alpha, r_\beta)\), and they project to semi-infinite straight lines on the basis of the fibration, \(\mathbb{R}^3\), parameterized by \(r_\gamma \geq 0\). Since they are located at the edges, where one of the circles of the fibration degenerates, they have the topology of \(\mathbb{C} \times S^1\).

It is easy to generalize the construction to other toric geometries, like the resolved conifold or local \(\mathbb{P}^2\): Lagrangian submanifolds with the topology of \(\mathbb{C} \times S^1\) are just given by points on the edges of the planar graphs. Such Lagrangian submanifolds were first considered in the context of open topological string theory by Aganagic and Vafa (2000), and further studied by Aganagic et al. (2002).

### 5.3 The conifold transition

In this section, we discuss the conifold geometry and the conifold transition, which will play a crucial role in the following chapters. Consider the following algebraic equation:

\[
\sum_{\mu=1}^{4} \eta_\mu^2 = a. \tag{5.29}
\]

As long as \(a \neq 0\), this is a complex threefold called the deformed conifold. This manifold has the topology of \(T^*S^3\). To see this, let us write \(\eta_\mu = x_\mu + iv_\mu\), where \(x_\mu, v_\mu\) are real co-ordinates, and let us take \(a\) to be real. We find the two equations

\[
\sum_{\mu=1}^{4} (x_\mu^2 - v_\mu^2) = a, \quad \sum_{\mu=1}^{4} x_\mu v_\mu = 0. \tag{5.30}
\]

The first equation indicates that the locus \(v_\mu = 0, \mu = 1, \cdots, 4\), describes a sphere \(S^3\) of radius \(R^2 = a\), and the second equation describes the cotangent space of \(S^3\) at the point \(x_\mu\). Therefore, (5.29) is \(T^*S^3\).

The deformed conifold also has the structure of a toric fibration, and it can be represented in a very useful way as follows. Let us first introduce the following complex co-ordinates:

\[
x = \eta_1 + i\eta_2, \quad v = i(\eta_3 - i\eta_4), \quad u = i(\eta_3 + i\eta_4), \quad y = \eta_1 - i\eta_2. \tag{5.31}
\]

The deformed conifold can now be written as

\[
xy = uv + a. \tag{5.32}
\]

In this parameterization, the geometry has a manifest \(T^2\) fibration

\[
(x, y, u, v) \rightarrow (e^{-i\alpha}x, e^{i\alpha}y, e^{-i\beta}u, e^{i\beta}v), \tag{5.33}
\]

where the \(\alpha\) and \(\beta\) actions above can be taken to generate the \((0, 1)\) and \((1, 0)\) cycles of the \(T^2\), respectively. The \(T^2\) fibre can degenerate to \(S^1\) by collapsing
Fig. 5.4. This figure represents $T^*S^3$, regarded as a $T^2 \times \mathbb{R}$ fibration of $\mathbb{R}^3$.

Two of the directions represent the axes of the two cylinders, and the third direction represents the real axis of the $z$-plane. The dashed line represents the embedded $S^3$.

one of its one-cycles. In (5.33), for example, the $U(1)_{\alpha}$ action fixes $x = 0 = y$ and therefore fails to generate a circle there. In the total space, the locus where this happens, i.e. the $x = 0 = y$ subspace of $X$, is a cylinder $uv = -a$. Similarly, the locus where the other circle collapses, $u = 0 = v$, gives another cylinder $xy = a$. Therefore, we can regard the whole geometry as a $T^2 \times \mathbb{R}$ fibration over $\mathbb{R}^3$: if we define $z = uv$, the $\mathbb{R}^3$ of the base is given by $\text{Re}(z)$ and the axes of the two cylinders. The fibre is given by the circles of the two cylinders, and by $\text{Im}(z)$. The $U(1)_{\alpha}$ fibration degenerates at $z = -a$, while the $U(1)_{\beta}$ fibration degenerates at $z = 0$. This is the same kind of fibration structure that we found when discussing the geometries of the form (5.17).

As we did in the examples above, it is very useful to represent the above geometry by depicting the singular loci of the torus action in the base $\mathbb{R}^3$. The loci where the cycles of the torus collapse, which are cylinders, project to lines in the base space. This is shown in Fig. 5.4. Notice that the $S^3$ of the deformed conifold geometry is realized in this picture as a $T^2$ fibration over an interval $I$. This interval is represented in Fig. 5.4 by a dashed line in the $z$-plane between $z = -a$ (where the $(0,1)$ cycle collapses) and $z = 0$ (where the collapsing cycle is the $(1,0)$). The geometric description of $S^3$ that is obtained in this way is, in fact, equivalent to the description given in Chapter 2 in terms of a Heegaard splitting along solid tori. To see this, let us cut the three-sphere into two pieces by cutting the interval $I$ into two smaller intervals $I_{1,2}$ through its midpoint. Each of the halves is a fibration of $T^2 = S^1 \times S^1_c$ over an interval $I_i$, where $S^1_c$ denotes the collapsing cycle. Of course, the non-trivial part of the fibration comes from the collapsing cycle, so we can see each of the halves as $S^1$ times the fibration
Fig. 5.5. On the left-hand side, we represent $S^3$ as a $T^2$ fibration over the interval. One of the circles of the torus degenerates over one endpoint, while the other circle degenerates over the opposite endpoint. Each of the degenerating circles fibres over half the interval to produce a disc $D$, and on the right-hand side the three-sphere is equivalently realized as two $S^1 \times D$ glued through an $S$ transformation.

of the collapsing cycle over $I_i$, which is simply a disk. In other words, we are constructing the three-sphere by gluing two manifolds of the form $S^1 \times D$. These are of course two solid tori, which are glued after exchanging the two cycles, i.e. after performing an $S$ transformation. This is shown in Fig. 5.5.

When $a = 0$ in (5.29) or in (5.32), the geometry becomes singular. This can be easily understood when we regard the deformed conifold as $T^*S^3$: at $a = 0$ the three-sphere collapses. The singular geometry described by the equation

$$xy = uv$$

(5.34)

is usually called the conifold singularity, or simply the conifold. In algebraic geometry, singularities can be avoided in two ways, in general. The first way is to deform the complex geometry, and in our case this leads to the deformed conifold (5.29). The other way is to resolve the singularity, for example by performing a blow-up, and this, in fact, leads to the resolved conifold geometry (5.6) that we studied further in the previous section. The resolution of the geometry can be explained as follows. Set

$$x = \lambda v, \quad u = \lambda y,$$

(5.35)

where $\lambda$ is an inhomogeneous co-ordinate in $\mathbb{P}^1$. This clearly solves (5.34). On the other hand, we can interpret $(v, y)$ and $(x, u)$ as the north-pole chart and the south-pole chart co-ordinates, respectively, for the bundle $\mathcal{O}(-1) \oplus \mathcal{O}(-1)$ on $\mathbb{P}^1$. Equation (5.35) is simply the change of co-ordinates between these two charts, as follows from (5.4). Therefore, by introducing a $\mathbb{P}^1$, we have transformed the singularity (5.34) into a smooth non-compact Calabi–Yau manifold (5.6). To
make contact with the toric description given in (5.18), we put \( x = z_1 z_3, y = z_2 z_4, u = z_1 z_2 \) and \( v = z_3 z_4 \). We then see that \( \lambda = z_1/z_4 \) is the inhomogeneous co-ordinate for the \( \mathbb{P}^1 \) described in (5.18) by \( |z_1|^2 + |z_4|^2 = t \). We therefore have a conifold transition (see, for example, Candelas and de la Ossa, 1990) in which the three-sphere of the deformed conifold shrinks to zero size as \( a \) goes to zero, and then a two-sphere of size \( t \) grows, giving the resolved conifold. In terms of the co-ordinates \( z_1, \cdots, z_4 \), the \( T^2 \) action (5.33) becomes

\[
\begin{align*}
  z_1, z_2, z_3, z_4 &\rightarrow e^{-i(\alpha+\beta)}z_1, e^{i\alpha}z_2, e^{i\beta}z_3, z_4.
\end{align*}
\]

This \( T^2 \) fibration is precisely (5.21). Notice that the singular locus of the fibration of the resolved conifold, which is encoded in the trivalent graph of Fig. 5.2, is inherited from the singular loci depicted in Fig. 5.4. The transition from the deformed to the resolved conifold can then be represented pictorially as in Fig. 5.6.

5.4 Examples of closed string amplitudes

Now that we have presented some detailed constructions of Calabi–Yau threefolds, we can come back to type-A topological string amplitudes, or equivalently to Gromov–Witten invariants. The Gromov–Witten invariants of Calabi–Yau threefolds can be computed in a variety of ways. A powerful technique that can be made mathematically rigorous is the localization technique pioneered by Kontsevich (1995). For compact Calabi–Yau manifolds, only \( N_{g=0,\beta} \) have been computed in detail, but for non-compact, toric Calabi–Yau manifolds localization techniques make it possible to compute \( N_{g,\beta} \) for arbitrary genus. We will now present some results for the topological string amplitudes \( F_g \) of the geometries we described above.

The resolved conifold \( O(-1) \oplus O(-1) \rightarrow \mathbb{P}^1 \) has a single Kähler parameter \( t \) corresponding to the \( \mathbb{P}^1 \) in the base, and its total free energy is given by

\[
F(g_s, t) = \sum_{d=1}^{\infty} \frac{1}{d^2 (2 \sin \frac{dg_s}{2})^2} Q^d.
\]
where $Q = e^{-t}$. We see that the only non-zero Gopakumar–Vafa invariant occurs at degree one and genus zero and is given by $n_{1,0}^0 = 1$. On the other hand, this model already has an infinite number of non-trivial $N_{g,\beta}$ invariants, which can be obtained by expanding the above expression in powers of $g_s$. The above expression was obtained in Gromov–Witten theory by Faber and Pandharipande (2000).

The space $O(-3) \to \mathbb{P}^2$ also has one single Kähler parameter, corresponding to the hyperplane class of $\mathbb{P}^2$. By using the localization techniques of Kontsevich, adapted to the non-compact case, one finds (Chiang et al., 1999; Klemm and Zaslow, 2001)

\[
F_0(t) = -\frac{t^3}{18} + 3Q - \frac{45 Q^2}{8} + \frac{244 Q^3}{9} - \frac{12333 Q^4}{64} \ldots
\]

\[
F_1(t) = -\frac{t}{12} + \frac{Q}{4} - \frac{3 Q^2}{8} - \frac{23 Q^3}{3} + \frac{3437 Q^4}{16} \ldots
\]

\[
F_2(t) = \frac{\chi(X)}{5720} + \frac{Q}{80} + \frac{3 Q^3}{20} + \frac{514 Q^4}{5} \ldots,
\]

and so on. In (5.38), $t$ is the Kähler class of the manifold, $Q = e^{-t}$, and $\chi(X) = 2$ is the Euler characteristic of local $\mathbb{P}^2$. The first term in $F_0$ is proportional to the intersection number $H^3$ of the hyperplane class, while the first term in $F_1$ is proportional to the intersection number $H \cdot c_2(X)$. The first term in $F_2$ is the contribution of constant maps.

As we explained above, we can express the closed string amplitudes in terms of Gopakumar–Vafa invariants. Let us introduce a generating functional for integer invariants as follows:

\[
f(z, Q) = \sum_{g,\beta} n_{g,\beta} z^g Q^{\beta},
\]

where $z$ is a formal parameter. For local $\mathbb{P}^2$ one finds

\[
f(z, Q) = 3Q - 6Q^2 + (27 - 10z)Q^3 - (192 - 231z + 102z^2 - 15z^3)Q^4 + \mathcal{O}(Q^5).
\]

Some of these Gopakumar–Vafa invariants can indeed be obtained with the geometric method outlined in section 4.3. According to (4.32), we have to consider the geometric moduli space of deformations of the embedded Riemann surface inside local $\mathbb{P}^2$. Consider then a curve of degree $d$ in the base of the fibration, namely $\mathbb{P}^2$. If we denote by $x, y, z$ homogeneous co-ordinates on $\mathbb{P}^2$, this curve is described by a polynomial of the form

\[
\sum_{i+j+k=d} a_{ijk} x^i y^j z^k = 0,
\]

where $a_{ijk}$ are complex coefficients. The moduli space of embedded curves is described by these coefficients, modulo rescalings by $\mathbb{C}^*$. There are in total
\[
\binom{d+2}{d} = \frac{d(d+3)}{2} + 1
\]

coefficients \(a_{ijk}\), therefore the moduli space is the projective space \(\mathbb{P}^{\frac{d(d+3)}{2}}\). By the genus-degree formula, these curves have genus

\[
g = \frac{(d-1)(d-2)}{2}.
\]

Using (4.32), we immediately find

\[
n_d\frac{(d-1)(d-2)}{2} = (-1)^{\frac{d(d+3)}{2}} \frac{(d+1)(d+2)}{2}.
\]

This indeed gives \(n_1^0 = 3\), \(n_2^0 = -6\), \(n_3^1 = -10\), and \(n_4^3 = 15\), in agreement with (5.40). Other invariants can be obtained by looking at curves with nodes, as required by (4.33). Further details can be found in Katz et al. (1999).

It should be mentioned that there is a very powerful method to compute the amplitudes \(F_g\), namely mirror symmetry. In the mirror symmetric computation, the \(F_g\) amplitudes are deeply related to the variation of complex structures on the Calabi–Yau manifold (Kodaira-Spencer theory) and can be computed through the holomorphic anomaly equations of Bershadsky et al. (1993 and 1994). Gromov–Witten invariants of non-compact, toric Calabi–Yau threefolds have been computed with mirror symmetry by Chiang et al. (1999), Klemm and Zaslow (2001) and Katz et al. (1999).
Part III

The topological string/gauge theory correspondence
This page intentionally left blank
STRING THEORY AND GAUGE THEORY

In the first part of this book we have studied some simple gauge theories – matrix models and Chern–Simons theory. In the second part, we have studied open and closed topological string theories. We now want to study the relationship between these gauge theories and topological strings.

The idea that gauge theories and string theories should be related in some way is an old one and has permeated the conceptual development of both fields. After all, string theory was born as an attempt to model the strong interactions. A precise correspondence between string theory and gauge theory was formulated by ’t Hooft (1974), who used as a starting point the double-line formulation of perturbation theory. As we explained in the first part of the book, the perturbative expansion of a gauge theory with gauge group $U(N)$ can be rewritten in terms of an expansion of double-line Feynman diagrams. Although we illustrated this issue in the cases of matrix models and Chern–Simons theory, it is clear that such a reorganization of the expansion can be made for any $U(N)$ gauge theory, as the double-line representation of the graphs only exploits the underlying gauge symmetry. If we denote by $g_s$ the coupling constant of the gauge theory, we can write the free energy as

$$F(g_s, N) = \sum_{g=0}^{\infty} \sum_{h=1}^{\infty} F_{g,h} g_s^{2g-2+h} N^h,$$

which we wrote down in the case of matrix models in (1.25) and in the case of Chern–Simons theory in (2.33). The above sum is over double-line diagrams with the topology of an open Riemann surface $\Sigma_{g,h}$ of genus $g$ with $h$ holes, and $F_{g,h}$ can be computed in terms of the Feynman rules associated to the diagram. On the other hand, one could consider (6.1) as an open string amplitude in which we sum over all possible topologies of the worldsheet $\Sigma_{g,h}$. The factors $N^h$ can be interpreted as Chan-Paton factors associated to the boundary of the open string, carrying a $U(N)$ gauge symmetry, and the coupling $g_s$ can be interpreted as the string coupling constant, which weights the contribution of $\Sigma_{g,h}$ with a factor $g_s^{2g-2+h}$. The quantities $F_{g,h}$ would therefore be interpreted as open string amplitudes on $\Sigma_{g,h}$.

One important question that we will address is the following: Is it possible to make more precise this analogy between $U(N)$ gauge theories and open string theories? In other words, given a $U(N)$ gauge theory, is it possible to find an open string theory in such a way that the quantities $F_{g,h}$ that appear in (6.1) can be interpreted as open string amplitudes, for some open string and some target
manifold? As we will see, in some cases the answer is yes, and involves open topological strings whose target is a Calabi–Yau manifold and whose boundary conditions are specified by topological D-branes. The main tool to provide this identification is string field theory, i.e. a field theory defined on the target of the string that describes its spacetime dynamics. The open string field theory we will need was introduced by Witten (1986). Although it was originally constructed for the open bosonic string theory, it can also be applied to topological string theory, and it turns out that on some particular Calabi–Yau backgrounds the full string field theory of the topological string reduces to a simple $U(N)$ gauge theory, where $g_s$ plays the role of the gauge coupling constant and $N$ is the rank of the gauge group. In particular, the string field reduces in this case to a finite number of gauge fields. We will consider two different topological string theories:

1) The type-A model on a Calabi–Yau of the form $X = T^*M$, where $M$ is a three-manifold, and there are $N$ topological D-branes wrapping $M$. In this case, the gauge theory is Chern–Simons theory on $M$ (Witten, 1995).

2) The B model on a Calabi–Yau manifold $X$ that is obtained as the small resolution of the singularity $y^2 = (W'(x))^2$. If $W'(x)$ has degree $n$, the small resolution produces $n$ two-spheres, and one can wrap $N_i$ topological D-branes around each two-sphere, with $i = 1, \cdots, n$. In this case, the gauge theory is a multicut matrix model with potential $W(x)$ (Dijkgraaf and Vafa, 2002a).

The second case involves a slightly more general situation than the one considered in (6.1), since there are $n$ submanifolds $S_1, \cdots, S_n$ where the strings can end. Equivalently, there are $N_i$ branes wrapping the submanifold $S_i$, for $i = 1, \cdots, n$. In this case, the open string amplitude is of the form $F_{g,h_1,\cdots,h_n}$ and the total free energy is given by

$$ F(g_s,N_i) = \sum_{g=0}^{\infty} \sum_{h_1,\cdots,h_n=1}^{\infty} F_{g,h_1,\cdots,h_n} x^{2g-2-h} N_1^{h_1} \cdots N_n^{h_n}, \quad (6.2) $$

where $h = \sum_i h_i$.

However, this is not the whole story concerning the relation between $U(N)$ gauge theories and string theories. As we saw in detail in the case of matrix models and Chern–Simons theory on $S^3$, the fatgraph expansion of a $U(N)$ gauge theory can be resummed formally by introducing the so-called 't Hooft parameter $t = g_s N$. For example, in the case of the free energy, we can rewrite (6.1) in the form

$$ F(g_s,t) = \sum_{g=0}^{\infty} x^{2g-2} F_g(t) \quad (6.3) $$

by defining

$$ F_g(t) = \sum_{h=1}^{\infty} F_{g,h} t^h. \quad (6.4) $$

The expansions (6.3) and (6.4) are typical of a closed string theory (we saw an example of this in the all-genus free energy (4.17) of type-A topological strings).
The gauge coupling \( g_s \) in (6.3) now plays the role of a closed string coupling constant. This leads to the idea, first suggested by 't Hooft (1974), that the resummation of the fatgraph expansion of a \( U(N) \) gauge theory can be interpreted as a closed string theory. In order to substantiate this idea we have to answer the following questions: once we know that a \( U(N) \) gauge theory can be interpreted in terms of an open string theory, can we give a closed string interpretation of the resummed quantities (6.4)? What is then the relation between the open string theory and the closed string theory? What is the string theory interpretation of the 't Hooft parameter \( t \) that appears as an argument in (6.4)? Answering these questions is, in a sense, the most difficult part in establishing a correspondence between string theory and gauge theory, and there are very few examples where this has been done in detail. A correspondence between a gauge theory and a closed string theory based on the \( 1/N \) expansion of the former is often called a \textit{large-N duality}. Notice that the resummation procedure (6.4) leads to a conjectural relation between an open and a closed string theory, independently of the gauge-theory interpretation, and a relation of this type is often called an \textit{open/closed string duality}.

Perhaps the most important example of a string theory/gauge theory correspondence in the sense defined above is the AdS/CFT correspondence of Maldacena (1998) (see Aharony \textit{et al.}, 2000, for an extensive review). In this correspondence, the gauge theory is \( U(N) \), \( N = 4 \) super-Yang–Mills theory on \( \mathbb{R}^{3,1} \). In this case, finding an open string realization of the fatgraph expansion is easy: it is the point-particle limit of type IIB string theory on \( \mathbb{R}^{9,1} \) with \( N \) D3-branes wrapped around \( \mathbb{R}^{3,1} \), i.e. one considers the limit in which the string tension \( \alpha' \) goes to zero. The \( N = 4 \) Yang–Mills theory arises as the worldvolume theory of the D-branes as \( \alpha' \to 0 \). According to Maldacena, the resummation of the fatgraphs à la 't Hooft leads to a closed string theory, namely type IIB superstring theory on \( \text{AdS}_5 \times S^5 \). This is because, at large \( N \), the presence of the D3-branes can be traded for a deformation of the background geometry. In other words, we can make the branes disappear if we change the background geometry at the same time. The geometry induced by the \( N \) D3-branes is a black hole, but the limit \( \alpha' \to 0 \) that is needed to recover the field theory (without further corrections) is the near-horizon limit of the black-hole geometry, and one finally obtains \( \text{AdS}_5 \times S^5 \) with no D-branes. Moreover, the 't Hooft parameter of the \( U(N) \) gauge theory is related to the common radius of \( \text{AdS}_5 \) and \( S^5 \) (i.e. it becomes a geometric modulus associated to the target of the string theory). We then see that the AdS/CFT correspondence suggests that open/closed string dualities should be associated to \textit{geometric transitions} relating different geometric backgrounds for open and closed string theories. This aspect of the story was emphasized by Gopakumar and Vafa (1999).

As we will see in the following chapters, the idea of geometric transitions makes it possible to find a closed string theory description both for Chern–Simons theory on \( S^3 \) and for matrix models with polynomial potentials. By using string field theory, we will show that these gauge theories can be interpreted
as topological open string theories on certain Calabi–Yau manifolds. We will further show that it is possible to find large-$N$ dualities providing a topological closed string theory description of the resummed $1/N$ expansions (6.4). These dualities are based on geometric transitions relating the Calabi–Yau background $X$ for open topological strings to a different Calabi–Yau background $X'$ for closed topological strings. The structure of the relationships between these gauge theories and open and closed topological strings is depicted in Fig. 6.1.

**Fig. 6.1.** This diagram summarizes the different relations between closed topological strings, open topological strings, and gauge theories.
7

STRING FIELD THEORY AND GAUGE THEORIES

In this chapter, following the program sketched in the previous one, we show
that both Chern–Simons gauge theories and matrix models with polynomial
potentials can be realized as open string theories.

7.1 Open string field theory

As we explained in the previous chapter, our strategy to show that Chern–Simons
theory and matrix models with polynomial potentials are open string theories is
to show that these gauge theories describe the spacetime dynamics of topological
open strings on certain backgrounds, and to do this we will use string field theory.
We briefly summarize here some basic ingredients of the cubic string field theory
introduced by Witten (1986) to describe the spacetime dynamics of open bosonic
strings, since we will use the same model to describe topological strings. For a
detailed presentation of open string field theory, see for example Taylor and
Zwiebach (2003).

In bosonic open string field theory, we consider the worldsheet of the string
to be an infinite strip parameterized by a spatial co-ordinate $0 \leq \sigma \leq \pi$ and a
time co-ordinate $-\infty < \tau < \infty$, and we pick the flat metric $ds^2 = d\sigma^2 + d\tau^2$. We
then consider maps $x : I \to X$, with $I = [0, \pi]$ and $X$ the target of the string.
The string field is a functional of open string configurations $\Psi[x(\sigma)]$, with ghost
number one (although we will not indicate it explicitly, this string functional
depends on the ghost fields as well). Witten (1986) defines two operations on
the space of string functionals. The first one is the integration, which is defined
formally by folding the string around its midpoint and gluing the two halves:

$$
\int \Psi = \int Dx(\sigma) \prod_{0 \leq \sigma \leq \pi/2} \delta[x(\sigma) - x(\pi - \sigma)]\Psi[x(\sigma)].
$$

(7.1)

The integration has ghost number $-3$, which is the ghost number of the vacuum.
This corresponds to the usual fact that in open string theory on the disc one has
to soak up three zero modes. One also defines an associative, non-commutative
star product $\star$ of string functionals through the following equation:

$$
\int \Psi_1 \star \cdots \star \Psi_N = \int \prod_{i=1}^N Dx_i(\sigma) \prod_{i=1}^N \prod_{0 \leq \sigma \leq \pi/2} \delta[x_i(\sigma) - x_{i+1}(\pi - \sigma)]\Psi_i[x_i(\sigma)],
$$

(7.2)

where $x_{N+1} \equiv x_1$. The star product simply glues the strings together by folding
them around their midpoints, and gluing the first half of one with the second
half of the following one (see, for example, the review of Taylor and Zwiebach (2003) for more details), and it doesn’t change the ghost number. In terms of these geometric operations, the string field action is given by

\[ S = \frac{1}{g_s} \int \left( \frac{1}{2} \Psi \star Q_{BRST} \Psi + \frac{1}{3} \Psi \star \Psi \star \Psi \right). \] (7.3)

Notice that the integrand has ghost number 3, while the integration has ghost number \(-3\), so that the action (7.3) has zero ghost number. If we add Chan-Paton factors, the string field is promoted to a \(U(N)\) matrix of string fields, and the integration in (7.3) includes a trace \(\text{Tr}\). The action (7.3) has all the information about the spacetime dynamics of open bosonic strings, with or without D-branes. In particular, one can derive the Born–Infeld action describing the dynamics of D-branes from the above action (Taylor, 2000).

We will not need all the technology of string field theory in order to understand open topological strings. The only piece of relevant information is the following: the string functional is a function of the zero mode of the string (which corresponds to the position of the string midpoint), and of the higher oscillators. If we decouple all the oscillators, the string functional becomes an ordinary function of spacetime, the \(\star\) product becomes the usual product of functions, and the integral is the usual integration of functions. The decoupling of the oscillators is, in fact, the pointlike limit of string theory. As we will see, this is the relevant limit for topological open strings.

### 7.2 Chern–Simons theory as an open string theory

#### 7.2.1 Open topological strings on \(T^* M\)

Let \(M\) be an arbitrary (real) three-dimensional manifold, and consider the six-dimensional space given by the cotangent bundle \(T^* M\) of \(M\). This space is a symplectic manifold. If we pick local co-ordinates \(q^a\) on \(M\), \(a = 1, 2, 3\), and local co-ordinates for the fibre \(p_a\), the symplectic form can be written as

\[ J = \sum_{a=1}^{3} dp_a \wedge dq_a. \] (7.4)

One can find a complex structure on \(T^* M\) such that \(J\) is a Kähler form, so \(T^* M\) can be regarded as a Kähler manifold. Since the curvature of the cotangent bundle exactly cancels the curvature of \(M\), it is Ricci-flat, therefore it is a Calabi–Yau manifold. An important example is the cotangent bundle of the three-sphere, \(T^* S^3\), which as shown in section 3.5, can be described holomorphically as the deformed conifold (5.29).

It is obvious that \(M\) is a Lagrangian submanifold in \(T^* M\), since \(J\) vanishes along \(p_a = \text{const}\). Since we have a Calabi–Yau manifold together with a Lagrangian submanifold in it, we can consider a system of \(N\) topological D-branes wrapping \(M\), thus providing Dirichlet boundary conditions for topological open
strings on $T^*M$. Our goal now is to obtain a spacetime action describing the dynamics of these topological D-branes, and as we will see this action is simply Chern–Simons theory on $M$. This will prove to be the sought-for realization of Chern–Simons theory in terms of open strings.

In order to find the spacetime action, we exploit again the analogy between open topological strings and the open bosonic string that we used to define the coupling of topological sigma models to gravity (i.e. that both have a nilpotent BRST operator and an energy-momentum tensor that is $Q_{\text{BRST}}$-exact). Since both theories have a similar structure, the spacetime dynamics of topological D-branes in $T^*M$ is also governed by (7.3), where $Q_{\text{BRST}}$ is given in this case by the topological charge defined in (3.40), and where the star product and the integration operation are as in the bosonic string. The construction of the cubic string field theory also requires the existence of a ghost-number symmetry, which is also present in the topological sigma model, as we discussed in Chapter 3. Sometimes it is convenient to redefine the ghost number by shifting it by $-d/2$ units with respect to the assignment presented in Chapter 3 (here, $d$ is the dimension of the target). When $d = 3$ this corresponds to the normalization used by Witten (1986) in which the ghost vacuum of the $bc$ system is assigned the ghost number $-1/2$.

In order to provide the string field theory description of open topological strings on $T^*M$, we have to determine the precise content of the string field, the $*$ algebra and the integration of string functionals for this particular model. As in the conventional string field theory of the bosonic string, we have to consider the Hamiltonian description of topological open strings. We then take $\Sigma$ to be an infinite strip and consider maps $x : I \rightarrow T^*M$, with $I = [0, \pi]$, such that $\partial I$ is mapped to $M$. The Grassmann field $\psi$, being a one-form on $\Sigma$, can be split as $\psi = \psi_\sigma d\sigma + \psi_\tau d\tau$, but due to the self-duality condition only one of them, say $\psi_\tau$, is independent. The action is taken to be $tS_A$, where $S_A$ is given in (3.43). As we showed in Chapter 3, due to the $Q$-exactness of the action, the semi-classical computation is exact and one can take the limit $t \rightarrow \infty$ to extract the results.

The canonical commutation relations can be easily read off from this action:

$$\left[ \frac{dx^i}{d\tau}(\sigma), x^j(\sigma') \right] = -\frac{i}{t} G^{ij} \delta(\sigma - \sigma'),$$

$$\{\psi_\tau(\sigma), \chi(\sigma')\} = \frac{1}{t} \delta(\sigma - \sigma').$$

(7.5)

The Hilbert space is made up out of functionals $\Psi[x(\sigma), \cdots]$, where $x$ is a map from the interval as we have just described, and the $\cdots$ refer to the Grassmann fields (which play here the rôle of ghost fields). The Hamiltonian is obtained, as usual in string theory, by

$$H = \int_0^\pi d\sigma T_{00}.$$ 

(7.6)

The bosonic piece of $T_{00}$ is just
\begin{equation}
\frac{dx^i}{d\sigma} \frac{dx^j}{d\sigma} + \frac{dx^i}{d\tau} \frac{dx^j}{d\tau},
\end{equation}
and using the canonical commutation relations we find:
\begin{equation}
H = \int_0^\pi d\sigma \left( -\frac{1}{t} G^{ij} \frac{\delta^2}{\delta x^i(\sigma) \delta x^j(\sigma)} + tG_{ij} \frac{dx^i}{d\sigma} \frac{dx^j}{d\sigma} \right).
\end{equation}

We then see that string functionals with \(dx^i/d\sigma \neq 0\) cannot contribute: as \(t \to \infty\), they get infinitely massive and decouple from the spectrum. Therefore, the map \(x : I \to T^*M\) has to be constant and in particular it must be a point in \(M\). A similar analysis holds for the Grassmann fields as well, and the conclusion is that the string functionals are functions of the commuting and the anti-commuting zero modes. Denoting them by \(q^a, \chi^a\), the string functional reduces to
\begin{equation}
\Psi = A^{(0)}(q) + \sum_{p=1}^3 \chi^{a_1} \cdots \chi^{a_p} A^{(p)}_{a_1 \cdots a_p}.
\end{equation}

These functionals can be interpreted as differential forms on \(M\). A differential form of degree \(p\) will have ghost number \(p\). If we have \(N\) D-branes wrapping \(M\), the above differential forms take values in the adjoint representation of the gauge group (i.e. they are valued in the \(U(N)\) Lie algebra). On these functionals, the \(Q\) symmetry acts as the exterior differential, and \(\{Q, \Psi\} = 0\) if the differential forms are closed. Of course in string field theory we do not restrict ourselves to functionals in the \(Q\)-cohomology. We rather compute the string field action for arbitrary functionals, and then the condition of being in the \(Q\)-cohomology arises as a linearized equation of motion.

We are now ready to write the string field action for topological open strings on \(T^*M\) with Lagrangian boundary conditions specified by \(M\). We have seen that the relevant string functionals are of the form (7.9). Since in string field theory the string field has ghost number one, we see that
\begin{equation}
\Psi = \chi^a A_a(q),
\end{equation}
where \(A_a(q)\) is a Hermitian matrix. In other words, the string field is just a \(U(N)\) gauge connection on \(M\). Since the string field only depends on commuting and anti-commuting zero modes, the integration of string functionals becomes ordinary integration of forms on \(M\), and the star product becomes the usual wedge products of forms. We then have the following dictionary:
\begin{align}
\Psi & \to A, & Q_{\text{BRST}} & \to d \\
\star & \to \wedge, & \int & \to \int_M.
\end{align}

The string field action (7.3) is then the usual Chern–Simons action for \(A\), and by comparing with (2.1) we have the following relation between the string coupling constant and the Chern–Simons coupling:
\[ g_s = \frac{2\pi}{k + N}, \]  
(7.12)

after taking into account the shift \( k \rightarrow k + N \).

This result is certainly remarkable. In the usual open bosonic string, the string field involves an infinite tower of string excitations. For the open topological string, the topological character of the model implies that all excitations decouple, except for the lowest-lying one. In other words, the usual reduction to a finite number of degrees of freedom that occurs in topological theories downsizes the string field to a single excitation. In physical terms, what is happening is that string theory reduces in this context to its pointlike limit, since the only relevant degree of freedom of the string is its zero mode, which describes the motion of a pointlike particle. The string field theory becomes a quantum field theory involving a finite number of fields.

However, as explained by Witten (1995), since open topological string theory is a theory that describes open string instantons with Lagrangian boundary conditions, we should expect to have corrections to the above result due to non-trivial worldsheet instantons. It is easy to see that instantons \( x : \Sigma \rightarrow T^* M \) such that \( x(\partial \Sigma) \subset M \) are necessarily constant. Notice first that \( J = d\rho \), where

\[ \rho = \sum_{a=1}^{3} p_a dq_a, \]  
(7.13)

and \( p_a \) vanishes on \( M \). Since \( x \) is a holomorphic map, the instanton action equals the topological piece \( -\int_{\Sigma} x^*(J) \). This can be evaluated to be

\[ \int_{\Sigma} x^*(J) = \int_{\partial \Sigma} x^*(\rho) = 0, \]  
(7.14)

since \( x(\partial \Sigma) \subset M \). Holomorphic maps with the above boundary conditions are necessarily constant, and there are no worldsheet instantons in the geometry. Therefore, there are no instanton corrections to the Chern–Simons action that we derived above.

One of the immediate consequences of the Chern–Simons spacetime description of open topological strings on \( T^* M \) is that the coefficient \( F_{g,h} \) in the perturbative expansion (1.25) of Chern–Simons theory on \( M \) is given by the free energy of the topological string theory in genus \( g \) with \( h \) holes. What is then the interpretation of the fatgraph associated to \( F_{g,h} \) from the point of view of the topological string theory on \( T^* M \)? Even though there are no ‘honest’ worldsheet instantons in this geometry, there are degenerate instantons of zero area in which the Riemann surface degenerates to a graph in \( M \). It is well known that the moduli space of open Riemann surfaces contains this type of configurations. In the case at hand, the fatgraphs appearing in the \( 1/N \) expansion of Chern–Simons theory on \( M \) are precisely the graphs that describe the degenerate instantons of the geometry. This model then gives a very concrete realization of the string picture of the \( 1/N \) expansion discussed in Chapter 6.
7.2.2 More general Calabi–Yau manifolds

In the previous section, we presented an explicit description of open topological strings on $T^*M$, following Witten (1995). What happens if the target is a more general Calabi–Yau manifold?

Let us consider a Calabi–Yau manifold $X$ together with some Lagrangian submanifolds $M_i \subset X$, with $N_i$ D-branes wrapped over $M_i$. In this case, the spacetime description of topological open strings will have two contributions. First, we have the contributions of degenerate holomorphic curves. These are captured by Chern–Simons theories on the manifolds $M_i$, following the same mechanism that we described for $T^*M$. However, as pointed out by Witten (1995), for a general Calabi–Yau $X$ we may also have honest open string instantons contributing to the spacetime description, which will be embedded holomorphic Riemann surfaces with boundaries ending on the Lagrangian submanifolds $M_i$. An open string instanton $\beta$ will intersect the $M_i$ along one-dimensional curves $K_i(\beta)$, which are in general knots inside $M_i$. We know from (4.41) that the boundary of such an instanton will give a Wilson loop insertion in the spacetime action of the form $\prod_i \text{Tr} U_{K_i(\beta)}$, where $U_{K_i(\beta)}$ is the holonomy of the Chern–Simons connection on $M_i$ along the knot $K_i(\beta)$. In addition, this instanton will be weighted by its area (which corresponds to the closed string background). We can then take into account the contributions of all instantons by including the corresponding Chern–Simons theories $S_{\text{CS}}(A_i)$, which account for the degenerate instantons, coupled in an appropriate way with the ‘honest’ holomorphic instantons. The spacetime action will then have the form

$$S(A_i) = \sum_i S_{\text{CS}}(A_i) + \sum_{\beta} e^{-f_\beta} \omega \prod_i \text{Tr} U_{K_i(\beta)},$$

(7.15)

where $\omega$ is the complexified Kähler form. The second sum is over ‘honest’ holomorphic instantons $\beta$. Notice that all the Chern–Simons theories $S_{\text{CS}}(A_i)$ have the same coupling constant, equal to the string coupling constant. More precisely,

$$\frac{2\pi}{k_i + N_i} = g_s.$$

(7.16)

In the action (7.15), the honest holomorphic instantons are put in ‘by hand’ and in principle one has to solve a non-trivial enumerative problem to find them. Once they are included in the action, the path integral over the Chern–Simons connections will join degenerate instantons to these honest worldsheet instantons: if we have a honest worldsheet instanton ending on a knot $K$, it will give rise to a Wilson loop operator in (7.15), and the $1/N$ evaluation of the vacuum expectation value will generate all possible fatgraphs $\Gamma$ joined to the knot $K$, producing in this way partially degenerate worldsheet instantons (the fatgraphs are interpreted, as before, as degenerate instantons). An example of this situation is depicted in Fig. 7.1. This more complicated scenario was explored by Aganagic and Vafa (2001), Diaconescu et al. (2003a, 2003b), and Aganagic et al. (2004). We will give examples of (7.15) in the following Chapters.
7.3 Matrix model as an open string theory

7.3.1 Holomorphic Chern–Simons theory

The analysis of the string field theory of the B model is very similar to that of the A model. Like we did before, we have to determine the precise content of the string field, the $\star$ algebra and the integration of string functionals. By analysing the Hamiltonian, as we did with (7.8), we again find that string functionals with $dx^i/d\sigma \neq 0$ cannot contribute. Therefore, the map $x : I \to X$ has to be constant and, in particular, it must be a point in $X$. A similar analysis holds for the Grassmann fields as well. Since $\theta = 0$ at the boundary, it follows that string functionals are functions of the commuting zero modes $x^i$ and $\eta^I$, and can be written as

$$\Psi = A^{(0)}(x) + \sum_{p \geq 1} \eta^{T_1} \cdots \eta^{T_p} A^{(p)}_{T_1 \cdots T_p}(x). \quad (7.17)$$

These functionals can be interpreted as a sum of $(0, p)$-forms on $X$. If we have $N$ D-branes wrapping $X$, these forms will be valued in $\text{End}(E)$ (where $E$ is a holomorphic $U(N)$ bundle). The $Q$ symmetry acts on these functionals as the Dolbeault operator $\overline{\partial}$ with values in $\text{End}(E)$. Notice that a differential form of degree $p$ will have ghost number $p$.

We can now write the string field action for topological open type-B strings on $X$ with $N$ spacetime-filling branes. The string functional has the form (7.17), and since the string field has ghost number one, we must have

$$\Psi = \eta^T A_T(x), \quad (7.18)$$

where $A_T(x)$ is a $(0, 1)$-form taking values in the endomorphisms of some holomorphic vector bundle $E$. In other words, the string field is just the $(0, 1)$ piece of a gauge connection on $E$. Since the string field only depends on commuting and anti-commuting zero modes, the star product becomes the wedge product of forms in $\Omega^{(0, p)}(\text{End}(E))$, and the integration of string functionals becomes ordinary integration of forms on $X$ wedged with $\Omega$. We then have the following dictionary:

$$\Psi \to A, \quad Q_{\text{BRST}} \to \overline{\partial}$$

$$\star \to \wedge, \quad \int \to \int_X \Omega \wedge. \quad (7.19)$$
The string field action (7.3) is then given by
\[ S = \frac{1}{2g_s} \int_X \Omega \wedge \text{Tr} \left( A \wedge \bar{A} + \frac{2}{3} A \wedge A \wedge A \right). \] (7.20)

This is the so-called \textit{holomorphic Chern–Simons action}. It is a rather peculiar quantum field theory in six dimensions, but as we will see when we consider D-branes of lower dimension, we will be able to obtain from (7.20) more conventional theories by dimensional reduction.

7.3.2 Type B topological strings and matrix models

We have seen that the spacetime description of the open B model with spacetime-filling branes reduces to a six-dimensional theory (7.20). We will now see that, in some circumstances, this theory simplifies drastically and reduces to a matrix model.

Let us consider the string field theory of type-B open topological strings on the Calabi–Yau manifold (5.5). We will consider a situation where we have Dirichlet boundary conditions associated to \( \mathbb{P}^1 \), in other words, there are \( N \) topological D-branes wrapping \( \mathbb{P}^1 \). Since the normal directions to the D-brane worldvolume are non-compact, the spacetime description can be obtained by considering the dimensional reduction of the original string field theory action (7.20). As usual in D-brane physics, the gauge potential \( A \) splits into a gauge potential on the worldvolume of the brane and Higgs fields describing the motion along the non-compact, transverse directions. In a non-trivial geometric situation like the one here, the Higgs fields are sections of the normal bundle. We then get three different fields:
\[ A, \quad \Phi_0, \quad \Phi_1, \] (7.21)
where \( A \) is a \( U(N) \) \((0,1)\) gauge potential on \( \mathbb{P}^1 \), \( \Phi_0 \) is a section of \( \mathcal{O}(-a) \), and \( \Phi_1 \) is a section of \( \mathcal{O}(a-2) \). Both fields, \( \Phi_0 \) and \( \Phi_1 \), take values in the adjoint representation of \( U(N) \). It is easy to see that the action (7.20) becomes
\[ S = \frac{1}{g_s} \int_{\mathbb{P}^1} \text{Tr} \left( \Phi_0 \overline{D_A} \Phi_1 \right), \] (7.22)

where \( \overline{D_A} = \overline{\partial} + [A, \cdot] \) is the anti-holomorphic covariant derivative. Notice that this theory is essentially a gauged \( \beta \gamma \) system, since \( \Phi_0, \Phi_1 \) are quasi-primary conformal fields of dimensions \( a/2, 1-a/2 \), respectively.

We will now consider a more complicated geometry. We start with the Calabi–Yau manifold (5.5) with \( a = 0 \), i.e.
\[ \mathcal{O}(0) \oplus \mathcal{O}(-2) \rightarrow \mathbb{P}^1. \] (7.23)
In this case, \( \Phi_0 \) is a scalar field on \( \mathbb{P}^1 \), while \( \Phi_1 \) is a \((1,0)\) form (since \( K_{\mathbb{P}^1} = \mathcal{O}(-2) \)). If we cover \( \mathbb{P}^1 \) with two patches with local co-ordinates \( z, z' \) related
by $z' = 1/z$, the fields in the two different patches, $\Phi_0, \Phi_1$, and $\Phi'_0, \Phi'_1$ will be related by

$$\Phi'_0 = \Phi_0, \quad \Phi'_1 = z^2 \Phi_1.$$  

(7.24)

We can regard this geometry as a family of $\mathbb{P}^1$s located at $\Phi'_1 = 0$ (the zero section of the non-trivial line bundle $\mathcal{O}(-2)$) parametrized by $\Phi_0 = \Phi'_0 = x \in \mathbb{C}$. The idea is to obtain a geometry where we get $n$ isolated $\mathbb{P}^1$s at fixed positions of $x$. To do that, we introduce an arbitrary polynomial of degree $n + 1$ on $\Phi_0$, $W(\Phi_0)$, and we modify the gluing rules above as follows (Cachazo et al., 2001):

$$z' = 1/z, \quad \Phi'_0 = \Phi_0, \quad \Phi'_1 = z^2 \Phi_1 + W'(\Phi_0)z.$$  

(7.25)

Before, the $\mathbb{P}^1$ was in a family parameterized by $\Phi_0 \in \mathbb{C}$. Now, we see that there are $n$ isolated $\mathbb{P}^1$s located at fixed positions of $\Phi_0$ given by $W'(\Phi_0) = 0$, since this is the only way to have $\Phi_1 = \Phi'_1 = 0$.

The geometry obtained by imposing the gluing rules (7.25) can be interpreted in yet another way. Let $\Phi_0 = x$ and define the co-ordinates

$$u = 2\Phi'_1, \quad v = 2\Phi_1, \quad y = i(2z'\Phi'_1 - W'(x)).$$  

(7.26)

The last equation in (7.25) can now be written as

$$uv + y^2 + W'(x)^2 = 0.$$  

(7.27)

This is a singular geometry, since there are singularities along the line $u = v = y = 0$ for every $x_*$ such that $W'(x_*) = 0$. For example, if $W'(x) = x$, (7.27) becomes, after writing $u, v \to u - iv, u + iv$

$$u^2 + v^2 + x^2 + y^2 = 0.$$  

(7.28)

This is another way of writing the conifold singularity (5.34). For arbitrary polynomials $W(x)$, (7.27) describes more general, singular Calabi–Yau manifolds. Notice that locally, around the singular points $u = v = y = 0, x = x_*$, the geometry described by (7.27) looks like a conifold (whenever $W''(x_*) = 0$). The manifold described by (7.25) is obtained after blowing up the singularities in (7.27), as we did for the resolved conifold. Since it is a resolved geometry, we will denote it by $X_{\text{res}}^W$. For $W'(x) = x$, the manifold (7.25) is, in fact, the resolved conifold.

We can now consider the dynamics of open type-B topological strings on $X_{\text{res}}^W$. We will consider a situation in which we have in total $N$ D-branes in such a way that $N_i$ D-branes are wrapped around the $i$-th $\mathbb{P}^1$, with $i = 1, \cdots, n$. As before, we have three fields in the adjoint representation of $U(N)$, $\Phi_0, \Phi_1$ and the gauge connection $A$. The action describing the dynamics of the D-branes turns out to be given by

$$S = \frac{1}{g_s} \int_{\mathbb{P}^1} \text{Tr} \left( \Phi_1 D_A \Phi_0 + \omega W(\Phi_0) \right),$$  

(7.29)

where $\omega$ is a Kähler form on $\mathbb{P}^1$ with unit volume. This action was derived by Kachru et al. (2000) and by Dijkgraaf and Vafa (2002a). A quick way to see that
the modification of the gluing rules due to adding the polynomial \( W'(\Phi_0) \) leads to the extra term in (7.29) is to use standard techniques in CFT (Dijkgraaf and Vafa, 2002a). The fields \( \Phi_0, \Phi_1 \) are canonically conjugate and on the conformal plane they satisfy the OPE
\[
\Phi_0(z)\Phi_1(w) \sim \frac{g_s}{z - w}.
\]
(7.30)

Let us now regard the geometry described in (7.25) as two disks (or conformal planes) glued through a cylinder. Since we are in the cylinder, we can absorb the factors of \( z \) in the last equation of (7.25). The operator that implements the transformation of \( \Phi \) is
\[
U = \exp \left( \frac{1}{g_s} \oint \text{Tr} W(\Phi_0(z)) \, dz \right),
\]
(7.31)
since from (7.30) it is easy to obtain
\[
\Phi_1' = U\Phi_1 U^{-1}.
\]
(7.32)

We can also write
\[
U = \exp \left( \frac{1}{g_s} \int_{\mathbb{P}^1} \text{Tr} W(\Phi_0(z)) \omega \right),
\]
(7.33)
where \( \omega \) is localized to a band around the equator of \( \mathbb{P}^1 \) (as we will see immediately, the details of \( \omega \) are unimportant, as long as it integrates to 1 on the two-sphere).

One easy check of the above action is that the equations of motion lead to the geometric picture of D-branes wrapping \( n \) holomorphic \( \mathbb{P}^1 \)s in the geometry. The gauge connection is just a Lagrange multiplier enforcing the condition
\[
[\Phi_0, \Phi_1] = 0,
\]
(7.34)
therefore we can diagonalize \( \Phi_0 \) and \( \Phi_1 \) simultaneously. The equation of motion for \( \Phi_0 \) is simply
\[
\bar{\partial} \Phi_0 = 0,
\]
(7.35)
and since we are on \( \mathbb{P}^1 \), we have that \( \Phi_0 \) is a constant, diagonal matrix. Finally, the equation of motion for \( \Phi_1 \) is
\[
\bar{\partial} \Phi_1 = W'(\Phi_0) \omega,
\]
(7.36)
and for non-singular \( \Phi_1 \) configurations both sides of the equation must vanish simultaneously, as we can see by integrating both sides of the equation over \( \mathbb{P}^1 \).

Therefore, \( \Phi_1 = 0 \) and the constant eigenvalues of \( \Phi_0 \) satisfy
\[
W'(\Phi_0) = 0,
\]
(7.37)
i.e. they must be located at the critical points of \( W(x) \). In general, we will have \( N_i \) eigenvalues of \( \Phi_0 \) at the \( i \)-th critical point, \( i = 1, \cdots , n \), and this is precisely the D-brane configuration we are considering.
What happens in the quantum theory? In order to analyze it, we will use the approach developed by Blau and Thompson (1993) for studying two-dimensional gauge theories. First, we choose the maximally Abelian gauge for $\Phi_0$, i.e. we write
\[
\Phi_0 = \Phi^k_0 + \Phi^t_0,
\]
where $\Phi^t_0$ is the projection on the Cartan subalgebra $t$, and $\Phi^k_0$ is the projection on the complementary part $k$. The maximally Abelian gauge is defined by the condition
\[
\Phi^k_0 = 0,
\]
which means that the non-diagonal entries of $\Phi_0$ are gauge-fixed to be zero. This is, in fact, the same gauge that we used before to write the matrix model in the eigenvalue basis. After fixing the gauge, the usual Faddeev-Popov techniques lead to a ghost functional determinant given by
\[
\frac{1}{N!}\text{Det}_k(\text{ad}(\Phi^t_0))_{\Omega^0(\mathbb{P}^1)},
\]
where the subscript $k$ means that the operator $\Phi^t_0$ acts on the space $k$, and the normalization factor $1/N!$ is the inverse of the order of the residual symmetry group, namely the Weyl group that permutes the $N$ entries of $\Phi^t_0$. The integrand of (7.29) reads, after gauge fixing,
\[
\text{Tr} \left( \Phi^t_1 \partial \Phi^t_0 + W(\Phi^t_0) \right) + 2 \sum_\alpha A^\alpha \Phi^{-\alpha}_1 \alpha(\Phi^t_0),
\]
where the $\alpha$ are roots, $E_\alpha$ is a basis of $k$, and we have expanded $\Phi^k_1 = \sum_\alpha \Phi^\alpha_1 E_\alpha$, $A^k \sum_\alpha A^\alpha E_\alpha$. We can now integrate out the $A^\alpha$ to obtain
\[
\frac{1}{\text{Det}_k(\text{ad}(\Phi^t_0))_{\Omega^1,0(\mathbb{P}^1)}} \prod_{\alpha > 0} \delta(\Phi^\alpha_1).
\]
Here, we have used the functional generalization of the standard formula $\delta(ax) = |a|^{-1}\delta(x)$. The integral over $\Phi^k_1$ is trivial, and the inverse determinant in (7.42) combines with (7.40) to produce
\[
\frac{\text{Det}_k(\text{ad}(\Phi^t_0))_{H^0(\mathbb{P}^1)}}{\text{Det}_k(\text{ad}(\Phi^t_0))_{H^1,0(\mathbb{P}^1)}},
\]
where (as usual) non-zero modes cancel (since they are paired by $\partial$) and one ends with the determinants evaluated on the cohomologies. Similarly, integrating out $\Phi^t_1$ in (7.41) leads to $\partial \Phi^t_0 = 0$, therefore $\Phi^t_0$ must be constant. The quotient of determinants is easy to evaluate in this case, and one finds
\[
\left[ \prod_{i<j} (\lambda_i - \lambda_j)^2 \right]^{h^0(\mathbb{P}^1) - h^1,0(\mathbb{P}^1)},
\]
where $\lambda_i$ are the constant eigenvalues of $\Phi_0^t$. Since $h^0(\mathbb{P}^1) = 1$, $h^{1,0}(\mathbb{P}^1) = 0$, we just get the square of the Vandermonde determinant and the partition function reads:

$$Z = \frac{1}{N!} \int \prod_{i=1}^{N} d\lambda_i \prod_{i<j} (\lambda_i - \lambda_j)^2 e^{-\frac{1}{gs} \sum_{i=1}^{N} W(\lambda_i)}. \quad (7.45)$$

In principle, one has to include a sum over non-trivial topological sectors of the Abelian gauge field $A^t$ in order to implement the gauge fixing (7.39) correctly (Blau and Thompson, 1993). Fortunately, in this case the gauge-fixed action does not depend on $A^t$, and the inclusion of topological sectors is irrelevant. The expression (7.45) is (up to a factor $(2\pi)^N$) the gauge-fixed version of the matrix model

$$Z = \frac{1}{\text{vol}(U(N))} \int d\Phi e^{-\frac{1}{gs} \text{Tr} W(\Phi)}. \quad (7.46)$$

We have then derived a surprising result due to Dijkgraaf and Vafa (2002a): the string field theory action for open topological B strings on the Calabi–Yau manifold described by (7.25) is a matrix model with potential $W(\Phi)$. As in the case of Chern–Simons theory, we have found a topological open string theory realization of matrix models with polynomial potentials. This result can be generalized to more complicated open string backgrounds, which turn out to be described as well by matrix models (Dijkgraaf and Vafa, 2002b, 2002c).

7.3.3 Open string amplitudes and multicut solutions

The total free energy $F(N_i, gs)$ of topological B strings on the Calabi–Yau (7.25) in the background of $N = \sum_i N_i$ branes wrapped around $n$ $\mathbb{P}^1$s is of the form (6.2), and as we have just seen it is given by the free energy of the matrix model (7.46). In particular, the coefficients $F_{g,h_1,\ldots,h_n}$ can be computed perturbatively in the matrix model. We have to be careful, however, to specify the classical vacua around which we are doing perturbation theory. Recall from the analysis of the matrix model that the classical solution that describes the brane configuration is characterized by having $N_i$ eigenvalues of the matrix located at the $i$-th critical point of the potential $W(x)$. In the saddle-point approximation, this means that we have to consider a multicut solution, with eigenvalues ‘condensed’ around all the extrema of the potential. Therefore, in contrast to the multicut solution discussed in section 2.2, we have that (1) all critical points of $W(x)$ have to be considered, and not only the minima, and (2) the number of eigenvalues in each cut is not determined dynamically as in (1.90), but it is rather fixed to be $N_i$ in the $i$-th cut, where $i$ runs from 1 to the number of minima, $n$. In other words, the integral of the density of eigenvalues $\rho(\lambda)$ along each cut equals a fixed filling fraction $\nu_i = N_i/N$:

$$\int_{x_{2i-1}}^{x_{2i-1}} d\lambda \rho(\lambda) = \nu_i, \quad i = 1, \cdots, n, \quad (7.47)$$
where $N = \sum_{i=1}^{n} N_i$ is the total number of eigenvalues. Let us introduce the partial 't Hooft couplings
\[ t_i = g_s N_i = t \nu_i. \] (7.48)
Taking into account (1.65) and (1.75), we can write (7.47) as
\[ t_i = \frac{1}{4 \pi i} \oint_{A_i} y(\lambda) d\lambda, \quad i = 1, \cdots, n, \] (7.49)
where $A_i$ is the closed cycle of the hyperelliptic curve (1.77) that surrounds the cut $C_i$. Assuming for simplicity that all the $t_i$ are different from zero, and taking into account that $\sum_i t_i = t$, we see that (7.49) gives $n - 1$ independent conditions, where $n$ is the number of critical points of $W(x)$. These conditions, together with the $n + 1$ conditions (1.88) (where $s$ is now equal to $n$), determine the positions of the endpoints $x_i$ as functions of the $t_i$ and the coupling constants in $W(x)$. It is clear that the solution obtained in this way is not an equilibrium solution of the matrix model, since cuts can be centred around local maxima and different cuts will have different values of the effective potential. This is not surprising, since we are not considering the matrix model as a quantum-mechanical system per se, but as an effective description of the original brane system. The different choices of filling fractions correspond to different choices of classical vacua for the brane system.

A subtle issue concerning the above matrix model is the following. The matrix field $\Phi$ in (7.46) comes from the B model field $\Phi_0$, which is a holomorphic field. Therefore, the matrix integral (7.45) should be understood as a contour integral, and in order to define the theory a choice of contour should be made. This can be done in perturbation theory, by choosing for example a contour that leads to the usual results for Gaussian integration, and therefore at this level the matrix model is not different from the usual Hermitian matrix model (Dijkgraaf and Vafa, 2002a; Witten, 2003). In some cases, however, regarding (7.46) as a holomorphic matrix model can be clarifying; see Lazaroiu (2003) for an exhaustive discussion.

The above description of the multicut solution refers to the saddle-point approximation. What is the meaning of the multicut solutions from the point of view of perturbation theory? To address this issue, let us consider for simplicity the case of the cubic potential (Klemm et al., 2003):
\[ \frac{1}{g_s} W(\Phi) = \frac{1}{2 g_s} \text{Tr} \Phi^2 + \frac{1}{3 g_s} \text{Tr} \Phi^3. \] (7.50)
This potential has two critical points, $a_1 = 0$ and $a_2 = -1/\beta$. The most general multicut solution will have two cuts. There will be $N_1$ eigenvalues sitting at $\Phi = 0$, and $N_2$ eigenvalues sitting at $\Phi = -1/\beta$. The partition function $Z$ of the matrix model is
\[ Z = \frac{1}{N!} \int \prod_{i=1}^{N} \frac{d\lambda_i}{2\pi} \Delta^2(\lambda)e^{-\frac{1}{2g_s} \sum_i \lambda_i^2 - \frac{1}{g_s} \beta \sum_i \lambda_i^3}, \] (7.51)
where $\Delta(\lambda) = \prod_{i<j}(\lambda_i - \lambda_j)$ is the Vandermonde determinant. We can now expand the integrand around the vacuum with $\lambda_i = 0$ for $i = 1, \ldots, N_1$, and $\lambda_i = -\frac{1}{\beta}$ for $i = N_1 + 1, \ldots, N$. Denoting the fluctuations by $\mu_i$ and $\nu_j$, the Vandermonde determinant becomes

$$\Delta^2(\lambda) = \prod_{1 \leq i_1 < i_2 \leq N_1} (\mu_{i_1} - \mu_{i_2})^2 \prod_{1 \leq j_1 < j_2 \leq N_2} (\nu_{j_1} - \nu_{j_2})^2 \prod_{1 \leq i \leq N_1, 1 \leq j \leq N_2} (\mu_i - \nu_j + \frac{1}{\beta})^2.$$  (7.52)

We also expand the potential around this vacuum and get

$$W = \sum_{i=1}^{N_1} \left( \frac{1}{2g_s} \mu_i^2 + \frac{\beta}{3g_s} \mu_i^3 \right) - \sum_{i=1}^{N_2} \left( \frac{1}{2g_s} \nu_i^2 - \frac{\beta}{3g_s} \nu_i^3 \right) + \frac{1}{6\beta^2 g_s} N_2.$$  (7.53)

Notice that the propagator of the fluctuations around $-1/\beta$ has the ‘wrong’ sign, since we are expanding around a local maximum. The interaction between the two sets of eigenvalues, which is given by the last factor in (7.52), can be exponentiated and included in the action. This generates an interaction term between the two eigenvalue bands

$$W_{\text{int}} = 2N_1 N_2 \log \beta + 2 \sum_{k=1}^{\infty} \frac{1}{k} \beta^k \sum_{i,j} \sum_{p=0}^{k} (-1)^p \binom{k}{p} \mu_i^p \nu_j^{k-p}.$$  (7.54)

By rewriting the partition function in terms of matrices instead of their eigenvalues, we can represent this model as an effective two-matrix model, involving an $N_1 \times N_1$ matrix $\Phi_1$, and an $N_2 \times N_2$ matrix $\Phi_2$:

$$Z = \frac{1}{\text{Vol}(U(N_1)) \times \text{Vol}(U(N_2))} \int d\Phi_1 d\Phi_2 e^{-W_1(\Phi_1) - W_2(\Phi_2) - W(\Phi_1, \Phi_2)},$$  (7.55)

where

$$W_1(\Phi_1) = +\text{Tr} \left( \frac{1}{2g_s} \Phi_1^2 + \frac{\beta}{3g_s} \Phi_1^3 \right),$$

$$W_2(\Phi_2) = -\text{Tr} \left( \frac{1}{2g_s} \Phi_2^2 - \frac{\beta}{3g_s} \Phi_2^3 \right),$$

$$W_{\text{int}}(\Phi_1, \Phi_2) = 2 \sum_{k=1}^{\infty} \frac{\beta^k}{k} \sum_{p=0}^{k} (-1)^p \binom{k}{p} \text{Tr} \Phi_1^p \text{Tr} \Phi_2^{k-p} + N_2 W(a_2) + N_1 W(a_1) - 2N_1 N_2 \ln \beta.$$  (7.56)

Here, $\text{Tr} \Phi_1^0 = N_1$, $\text{Tr} \Phi_2^0 = N_2$, $W(a_1) = 0$ and $W(a_2) = 1/(6g_s\beta^2)$. Although the kinetic term for $\Phi_2$ has the ‘wrong’ sign, we can still make sense of the model in perturbation theory by using formal Gaussian integration, and this can in fact be justified in the framework of holomorphic matrix models (Lazaroiu, 2003). Therefore, the two-cut solution of the cubic matrix model can be formally
represented in terms of an effective two-matrix model. It is now straightforward to compute the free energy $F_{\text{pert}} = \log(Z(\beta)/Z(\beta = 0))$ in perturbation theory. It can be expanded as

$$F_{\text{pert}} = -N_1 W(a_1) - N_2 W(a_2) - 2N_1 N_2 \ln \beta + \sum_{h=1}^{\infty} \sum_{g \geq 0} (g_s \beta^2)^{2g-2+h} F_{g,h}(N_1, N_2),$$

where

$$F_{g,h} = \sum_{h_1 + h_2 = h} F_{g,h_1,h_2} \frac{N_1^{h_1} N_2^{h_2}}{h!}$$

is a homogeneous polynomial in $N_1$ and $N_2$ of degree $h$. One finds, up to fourth order in the coupling constant $\beta$, the following result (Klemm et al., 2003):

$$F_{\text{pert}} = -N_1 W(a_1) - N_2 W(a_2) - 2N_1 N_2 \ln \beta + g_s \beta^2 \left[ \left( \frac{2}{3} N_1^3 - 5N_1^2 N_2 + 5N_1 N_2^2 - \frac{2}{3} N_2^3 \right) + \frac{1}{6}(N_1 - N_2) \right] + g_s^2 \beta^4 \left[ \left( \frac{8}{3} N_1^4 - \frac{91}{3} N_1^3 N_2 + 59N_1^2 N_2^2 - \frac{91}{3} N_1 N_2^3 + \frac{8}{3} N_2^4 \right) \right] + \left( \frac{7}{3} N_1^2 - \frac{31}{3} N_1 N_2 + \frac{7}{3} N_2^2 \right) + \cdots$$

From this explicit perturbative computation one can read off the first few coefficients $F_{g,h_1,h_2}$. Of course, this procedure can be generalized, and the $n$-cut solution can be represented by an effective $n$ matrix model with interactions among the different matrices that come from the expansion of the Vandermonde determinant. These interactions can also be incorporated in terms of ghost fields, as explained in Dijkgraaf et al., 2003. This makes it possible to compute corrections to the saddle-point approximation in perturbation theory. One can also use the multicut solution to the loop equations (Akemann, 1996; Kostov, 1999), with minor modifications, to compute the genus one correction in closed form (Klemm et al., 2003; Dijkgraaf et al., 2004; Chekhov, 2004).
This page intentionally left blank
8

GEOMETRIC TRANSITIONS

8.1 The conifold transition and the large \( N \) duality

After the analysis of the last chapter, we know that Chern–Simons theory on \( S^3 \) is a topological open string theory on \( T^*S^3 \). According to the program presented in Chapter 6 to relate string theory to gauge theory, the next step is to see if there is a closed string theory leading to the resummations (2.179) and (2.181). As shown by Gopakumar and Vafa in an important paper (1999), the answer is yes.

The intuition behind the result of Gopakumar and Vafa is that, as suggested by the AdS/CFT correspondence, and as we discussed in Chapter 6, open/closed string dualities are related to geometric transitions in the background geometry. This suggests that we look for a transition involving the background \( T^*S^3 \). But we studied such a transition in Section 5.4: in the conifold transition, starting with the deformed conifold, we make the \( S^3 \) shrink to zero size so as to obtain the conifold singularity, which we then resolve to obtain the resolved conifold. This was represented in Fig. 5.6. Since Chern–Simons theory is an open topological string on the deformed conifold geometry with \( N \) topological D-branes wrapping the three-sphere, it is natural to conjecture that at large \( N \) the D-branes induce a conifold transition in the background geometry, as in the AdS/CFT correspondence. We then end up with the resolved conifold and no D-branes. But in the absence of D-branes that enforce boundary conditions we just have a theory of closed topological strings. Following this reasoning, Gopakumar and Vafa conjectured that Chern–Simons theory on \( S^3 \) is equivalent to closed topological string theory on the resolved conifold.

What is the relation between the different parameters of the gauge theory and the string theory? By the general arguments presented in Chapter 6, the closed string coupling constant is identified with the open string coupling constant, or equivalently with the Chern–Simons effective coupling constant (2.166). The other parameter on the Chern–Simons side is the 't Hooft coupling

\[
t = ig_s N = xN.
\]  

(8.1)

On the closed topological string side, the only parameter entering the amplitudes \( F_g \) is the Kähler parameter measuring the size of the \( \mathbb{P}^1 \) in the resolved geometry. It is natural to identify the two parameters, and as we will see in a moment this is indeed the case. It is interesting to note that this identification is also natural in view of the AdS/CFT correspondence, where the 't Hooft parameter of the gauge theory is identified with a geometric modulus of the closed string target, namely the radius of \( S^5 \).

143
A non-trivial test of the duality advocated by Gopakumar and Vafa is to verify that the free energy of $U(N)$ Chern–Simons theory on the three-sphere agrees with the free energy of closed topological strings on the resolved conifold. If we compare the result for the resummed amplitude $F_g(t)$ obtained in (2.179) with (4.28), we see that the Chern–Simons amplitudes have precisely the form of the free energy of a closed topological string, with $n_1^0 = 1$, with the rest of the Gopakumar–Vafa invariants being zero. Also, from the first term, which gives the contribution of the constant maps, we find that $\chi(X) = 2$. This is precisely the structure of the free energy of closed topological strings on the resolved conifold, as shown in (5.37), and we confirm that the 't Hooft parameter (8.1) gets identified with the Kähler parameter. This is a remarkable check of the conjecture.

Another piece of evidence for the conjecture comes from the matrix model analysis of Chern–Simons theory on $S^3$ in Section 2.8. There, we found that the resolvent (therefore, the master field configuration controlling the large $N$ expansion) is encoded in the algebraic curve (2.196). This curve turns out to describe the mirror of the resolved conifold geometry (see, for example, Hori and Vafa (2000) and Aganagic et al., 2002), where $t$ is again identified with the Kähler parameter. We then see that the master field of the matrix model formulation of the gauge theory encodes information about the target geometry of the closed string description. This is one of the new insights that appear in the study of large-$N$ dualities involving topological strings, and we will shortly see another example of this.

The conjecture of Gopakumar and Vafa can be justified physically by doing a careful analysis of the worldsheet description of the topological strings involved in the duality (Ooguri and Vafa, 2002). It can also be proved by considering the geometric transition in the context of type II superstring theory (Vafa, 2001a) and lifting it to M-theory (Acharya, 2000; Atiyah et al., 2001).

8.2 Incorporating Wilson loops

As we have extensively discussed in Chapter 2, most of the richness of Chern–Simons theory on $S^3$ is due to the Wilson loop operators along knots. How do we incorporate Wilson loops in the closed string picture that we have just developed? Let us consider the $1/N$ expansion (2.182) of connected vacuum expectation values of Wilson loops. Formally, it can be regarded as an open string expansion, and the $W_{\vec{k},g}(t)$ can be interpreted as amplitudes in an open string theory at genus $g$ and with $h = |\vec{k}|$ holes (summed over all possible bulk classes). According to this interpretation, the Wilson loop creates a one-cycle in the target space where the boundaries of Riemann surfaces can end. The vector $\vec{k}$ specifies the winding numbers for the boundaries as follows: there are $k_j$ boundaries wrapping $j$ times the one-cycle associated to the Wilson loop. The generating functional for connected vacuum expectation values (2.159) is interpreted as the total free energy of an open string. The open strings that are relevant to the string interpretation of Wilson loop amplitudes should not be
confused with the open strings that we associated to the expansion (6.1). The open strings underlying (2.182) should be regarded as an open string sector in the closed string theory associated to the resummed expansion (6.4). This open string interpretation of the Wilson loops is, in fact, valid for any $U(N)$ gauge theory with a closed string description.

In the case of Chern–Simons theory on $S^3$, since the string description involves topological strings, it is natural to assume that Wilson loops are going to be described by open topological strings in the resolved conifold, and this means that we need a Lagrangian submanifold specifying the boundary conditions for the strings. These issues were addressed in an important paper by Ooguri and Vafa (2000). In order to give boundary conditions for the open strings in the resolved conifold, Ooguri and Vafa constructed a natural Lagrangian submanifold $\tilde{\mathcal{C}}_K$ in $T^* S^3$ for any knot $K$ in $S^3$. This construction is rather canonical, and it is called the conormal bundle of $K$. The details are as follows: suppose that a knot $K$ is parameterized by a curve $q(s)$, where $s \in [0, 2\pi)$. The conormal bundle of $K$ is the space

$$\tilde{\mathcal{C}}_K = \left\{ (q(s), p) \in T^* S^3 \mid \sum_a p_a \dot{q}_a = 0, \ 0 \leq s \leq 2\pi \right\} \quad (8.2)$$

where $p_a$ are co-ordinates for the cotangent bundle, and $\dot{q}_a$ denote the derivatives w.r.t. $s$. This space is an $\mathbb{R}^2$-fibration of the knot itself, where the fibre on the point $q(s)$ is given by the two-dimensional subspace of $T_{q(s)}^* S^3$ of planes orthogonal to $\dot{q}(s)$. $\tilde{\mathcal{C}}_K$ has, in fact, the topology of $S^1 \times \mathbb{R}^2$, and intersects $S^3$ along the knot $K$.

One can now consider, together with the $N$ branes wrapping $S^3$, a set of $M$ probe branes wrapping $\tilde{\mathcal{C}}_K$. As usual when we have two sets of D-branes, we have three different types of strings: the strings with both ends on the $N$ branes are described by $U(N)$ Chern–Simons theory on $S^3$, as we argued before. In the same way, the strings with both ends on the $M$ branes are described by $U(M)$ Chern–Simons theory on $\tilde{\mathcal{C}}_K$. But there is a new sector due to strings stretched between the $N$ branes and the $M$ branes. To study these strings, we can make an analysis similar to the one we did in Section 7.2. First, we have to impose again that $dx^i/d\sigma = 0$. Therefore, $x^i$ has to be a constant, and since the endpoints of the string lie on different submanifolds, the only possibility is that $x^i \in S^3 \cap \tilde{\mathcal{C}}_K = K$. A similar analysis holds for the Grassmann fields, and we then find that the string functionals describing the new sector of strings are a function of a single commuting zero mode $q$ parametrizing $K$, and a single anti-commuting zero mode $\chi$. In other words,

$$\mathcal{A} = \phi(q) + \chi \xi(q), \quad (8.3)$$

where $\phi$ is a complex scalar field in the bifundamental representation $(N, \mathcal{M})$, and living in the intersection of the two branes, $\mathcal{K}$. The fact that the scalar is complex is due to the fact that our strings are oriented, and we have to consider
both a real scalar in the representation \((N, \overline{M})\) together with another real scalar in the representation \((\overline{N}, M)\), which we can put together as a complex scalar in one of the representations. The \(Q\) operator is just the exterior differential \(d\) on \(S^1\).

As we explained above, the string field is the part of the above functional with shifted ghost number \(-1/2\). However, now the shifted ghost number assignment is different from the one in \(S^3\), and for a differential form of degree \(p\) it is given by \(p - 1/2\). This is because the target is now \(S^1\) with \(d = 1\). Therefore, the surviving field is in this case the scalar \(\phi(q)\). This is consistent with the fact that, since the spacetime dynamics takes place now on a circle, and since \(Q = d\), the kinetic term for the string field action (7.3) is only non-trivial if the string field is a scalar. The full action for \(\phi(q)\) is simply \(\oint K \overline{\phi} d\phi\). However, there are also two background gauge fields that interact with the Chan-Paton factors at the endpoints of the strings. These are the \(U(N)\) gauge connection \(A\) on \(S^3\), and the \(U(M)\) gauge connection \(\tilde{A}\) on \(\tilde{C}_K\). The complex scalar couples to the gauge fields in the standard way,

\[
\oint \mathcal{K} \text{Tr} (\overline{\phi} A \phi - \phi \tilde{A} \overline{\phi}).
\]

Here, we regard \(\tilde{A}\) as a source. If we now integrate out \(\phi\) we obtain

\[
\exp \left[ - \log \det \left( \frac{d}{ds} + \sum_a (A_a - \tilde{A}_a) \frac{dq_a}{ds} \right) \right],
\]

which can be easily evaluated as

\[
\exp \left[ - \text{Tr} \log (1 - U \otimes V) \right] = \exp \left\{ \sum_{n=1}^{\infty} \frac{1}{n} \text{Tr} U^n \text{Tr} V^n \right\},
\]

where \(U, V^{-1}\) are the holonomies of \(A, \tilde{A}\) around the knot \(K\), and we have dropped an overall constant. In this way, we obtain the effective action for the \(A\) field

\[
S_{CS}(A) + \sum_{n=1}^{\infty} \frac{1}{n} \text{Tr} U^n \text{Tr} V^n,
\]

where \(S_{CS}(A)\) is, of course, the Chern–Simons action for \(A\). Therefore, in the presence of the probe branes, the action involves an insertion of the Ooguri–Vafa operator that was introduced in (2.154). Since we are regarding the \(M\) branes as a probe, the holonomy \(V\) is an arbitrary source. The extra term in (8.7) can be interpreted as coming from an annulus of zero length interpolating between the two sets of D-branes. Later, we will consider a simple generalization of the above for an annulus of finite length.

Let us now follow this system through the geometric transition. The \(N\) branes disappear, and the background geometry becomes the resolved conifold. However,
the $M$ probe branes are still there. It is natural to conjecture that they are
now wrapping a Lagrangian submanifold $\mathcal{C}_K$ of $\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1$ that can
be obtained from $\hat{\mathcal{C}}_K$ through the geometric transition. The final outcome is
the existence of a map between knots in $S^3$ and Lagrangian submanifolds in
$\mathcal{O}(-1) \oplus \mathcal{O}(-1) \to \mathbb{P}^1$:
\[ K \to \mathcal{C}_K. \] (8.8)
Moreover, one has $b_1(\mathcal{C}_K) = 1$. This conjecture is clearly well motivated in the
physics. Ooguri and Vafa (2000) constructed $\mathcal{C}_K$ explicitly when $K$ is the un-
knot, and Labastida et al. (2000) proposed Lagrangian submanifolds for certain
algebraic knots and links (including torus knots). Finally, Taubes (2001) has con-
structed a map from knots to Lagrangian submanifolds in the resolved conifold
for a wide class of knots.

The Lagrangian submanifold $\mathcal{C}_K$ in the resolved geometry gives precisely the
open string sector that is needed in order to extend the large-$N$ duality to Wil-
son loops. According to Ooguri and Vafa (2000), the free energy $F(V)$ of open
topological strings (4.45) with boundary conditions specified by $\mathcal{C}_K$ is identical
to the free energy of the deformed Chern–Simons theory with action (8.7), which
is given by (2.159). We then have the following conjecture:

\[ F(V) = F_{CS}(V). \] (8.9)
In this equation, the l.h.s. denotes the open topological string free energy as-
sociated to $\mathcal{C}_K$, and the r.h.s. denotes the Chern–Simons free energy associated
to the Ooguri–Vafa operator of the knot $\mathcal{K}$. Notice that, since $b_1(\mathcal{C}_K) = 1$, the
topological sectors of maps with positive winding numbers correspond to vectors $\vec{k}$ labelling the connected vacuum expectation values, and one finds
\[ i|\vec{k}| \sum_{g=0}^{\infty} F_{g,\vec{k}}(t) g^{2g-2 + |\vec{k}|} = - \frac{1}{\prod_j j^k} W^{(c)}_{\vec{k}}. \] (8.10)
It is further assumed that there is an analytic continuation of $F(V)$ from negative
to positive winding numbers in such a way that the equality (8.9) holds in general.
Another useful way to state the correspondence (8.9) is to use the total partition
function of topological open strings (4.74) instead of the free energy. The duality
between open string amplitudes and Wilson loop expectation values reads simply
\[ Z_R = W_R(\mathcal{K}), \] (8.11)
where $Z_R$ was introduced in (4.74) and $W_R(\mathcal{K})$ is the Chern–Simons invariant
of the knot $\mathcal{K}$ in representation $R$.

When $\mathcal{K}$ is the unknot in the three-sphere, the conjecture of Ooguri and Vafa
can be tested in full detail (Ooguri and Vafa, 2000; Mariño and Vafa, 2002). We
will describe this test in Chapter 9, in the context of the topological vertex. For
more general knots and links, the open string free energy is not known, but one
can test the duality indirectly by verifying that the Chern–Simons side satisfies
the structural properties of open string amplitudes that we explained in Section
4.4. This aspect is discussed in some detail in Chapter 10.
8.3 Geometric transitions for more general toric manifolds

The duality between Chern–Simons on $S^3$ and closed topological strings on the resolved conifold gives a very nice realization of the gauge/string theory duality. However, from the ‘gravity’ point of view we do not learn much about the closed string geometry, since the resolved conifold is quite simple (recall that it only has one non-trivial Gopakumar–Vafa invariant). It would be very interesting to find a topological gauge theory dual to more complicated geometries, like the ones we discussed in Chapter 5, in such a way that we could use our knowledge of gauge theory to learn about enumerative invariants of closed strings, and about closed strings in general.

The program of extending the geometric transition of Gopakumar and Vafa was started by Aganagic and Vafa (2001). Their basic idea was to construct geometries that locally contain $T^*S^3$s, and then follow the geometric transitions to dual geometries where the deformed conifolds are replaced by resolved conifolds. Remarkably, a large class of non-compact toric manifolds can be realized in this way, as was made clear by Aganagic et al. (2004) and Diaconescu et al. (2003b). In this section we will present some examples where closed string amplitudes can be computed by using this idea.

The geometries that we discussed in Chapter 4 are $T^2 \times \mathbb{R}$ fibrations of $\mathbb{R}^3$, which contain two-spheres (represented by the compact edges of the geometry). In this section we will construct geometries with the same fibration structure that contain three-spheres, and can be related by a geometric transition to some of the toric geometries that we analysed in Chapter 5.

Recall from the discussion in Section 5.4, that the deformed conifold has the structure of a $T^2 \times \mathbb{R}$ fibration of $\mathbb{R}^3$ that can be encoded in a non-planar graph as in Fig. 5.4. The degeneration loci of the cycles of the torus fibre are represented in this graph by straight lines, while the $S^3$ is represented by a dashed line joining these loci. This graphical procedure can be generalized, and it is easy to construct more general $T^2 \times \mathbb{R}$ fibrations of $\mathbb{R}^3$ by specifying degeneration loci in a diagram that represents the $\mathbb{R}^3$ basis. A simple example is shown in Fig. 8.1. This geometry contains two $S^3$s, represented by dashed lines. These three-spheres are also constructed as torus fibrations over the interval, and the cycles that degenerate at the endpoints can be read off from the graph. In fact, both are described by a $T^2$ fibration where the $(0,1)$ cycle collapses at one endpoint, and the $(1,0)$ cycle collapses at the other endpoint. As we explained in Section 5.3, this gives a Heegaard splitting of the three-sphere along solid tori. These tori are glued together through the $S$ transformation that relates one of the collapsing cycles to the other.

We can also construct geometries that contain more general three-manifolds. If a manifold $M$ admits a Heegaard splitting along two solid tori, it will be specified by an $\text{SL}(2,\mathbb{Z})$ matrix $V_M$ mapping the $(p_L,q_L)$ cycle of one $T^2$ to the $(p_R,q_R)$ cycle of the other $T^2$. Equivalently, $M$ can be obtained as a torus fibration over an interval where the $(p_L,q_L)$ and $(p_R,q_R)$ cycles degenerate at the endpoints, as we explained in Chapter 5 in the simple case of the $(1,0)$ and
Fig. 8.1. A Calabi–Yau that is a $\mathbf{T}^2 \times \mathbb{R}$ fibration of $\mathbb{R}^3$. The dashed lines represent $S^3$s.

Fig. 8.2. This figure shows the geometric transition of the Calabi–Yau depicted in Fig. 8.1. In the leftmost geometry there are two three-spheres, represented by dashed lines. The intermediate geometry is singular, and the figure on the right shows the planar graph associated to the smooth toric Calabi–Yau after the transition. It contains three $\mathbb{P}^1$s with Kähler parameters $t_1$, $t_2$ and $t$.

(0, 1) cycles. The local geometry $T^*M$ will be described by two overlapping lines with slopes $-p_L/q_L$ and $-p_R/q_R$. The dashed line joining them will represent the three-manifold $M$.

Given a graph like the one in Fig. 8.1, one can try to use the conifold transition ‘locally,’ as was first explained by Aganagic and Vafa (2001). The above geometry, for example, contains two deformed conifolds with their corresponding three-spheres, therefore there is a geometric transition where the three-spheres shrink to zero size and then the corresponding singularities are blown up to give a resolved geometry. This geometric transition is depicted in Fig. 8.2. The resolved geometry is clearly toric, and it can easily be built up by gluing four trivalent vertices, as we explained in Chapter 5. It has two Kähler classes corresponding to the two blown-up two-spheres, denoted by $t_1$, $t_2$ in Fig. 8.2. It also contains a third two-sphere associated to the intermediate, horizontal leg, with Kähler parameter $t$. Although we have focused on the example depicted in Fig. 8.1, it is clear what the general philosophy is: one considers a ‘deformed’ geometry and performs geometric transitions ‘locally.’ The resulting ‘resolved’ geometry will be a toric Calabi–Yau manifold of the type discussed in Chapter 5. The planar
graph describing the resolved geometry can be easily reconstructed from the non-planar graph describing the deformed geometry.

We will now use the generalized geometric transition that we found in the last subsection in order to compute the topological string amplitudes. Let us first wrap \( N_i \) branes, \( i = 1, 2 \), around the two \( S^3 \)s of the deformed geometry depicted in Fig. 8.1. What is the effective topological action describing the resulting open strings? Since this geometry is not globally of the form \( T^*M \), we are in the situation described in Section 7.2.2: for open strings with both ends on the same \( S^3 \), the dynamics is described by Chern–Simons theory with gauge group \( U(N_i) \), therefore we will have two Chern–Simons theories with groups \( U(N_1) \) and \( U(N_2) \).

However, there is a new sector of open strings stretched between the two three-spheres: these are the non-degenerate instantons that appear in (7.15).

Instead of describing these open string instantons in geometric terms, it is better to use the spacetime physics associated to these strings. A similar situation was considered when we analysed the incorporation of Wilson loops in the geometric transition. There we had two sets of intersecting D-branes, giving a massless complex scalar field living in the intersection and in the bifundamental representation of the gauge groups. In the situation depicted in Fig. 8.1, the same arguments indicate that there is a complex scalar \( \phi \) in the representation \((N_1, N_2)\), corresponding to the bifundamental strings stretched between the two sets of D-branes. The difference from the situation that we were considering before is that this complex scalar is now massive, since the strings have a finite length, and its mass is proportional to the ‘distance’ between the two three-spheres. This length is measured by a complexified Kähler parameter that will be denoted by \( r \). The kinetic term will now have an extra term of the form

\[
-r \oint_{S^1} \bar{\phi} \phi.
\]

We can now integrate out this complex scalar field as we did in (8.6) to obtain the correction to the Chern–Simons actions on the three-spheres due to the presence of the new sector of open strings:

\[
\mathcal{O}(U_1, U_2; r) = \exp \left\{ \sum_{n=1}^{\infty} \frac{e^{-nr}}{n} \Tr U_1^n \Tr U_2^n \right\},
\]

(8.13)

where \( U_{1,2} \) are the holonomies of the corresponding gauge fields around the \( S^1 \) in (8.12). The operator \( \mathcal{O} \) can also be interpreted as the amplitude for a primitive annulus of area \( r \) together with its multicovers, which are labelled by \( n \). This annulus ‘connects’ the two \( S^3 \)s, i.e. one of its boundaries is a circle in one three-sphere, and the other boundary is a circle in the other sphere. The sum over \( n \) in (8.13) is precisely the sum over open string instantons in the second term of (7.15), for this particular geometry.

The problem now is to determine how many configurations like this one contribute to the full amplitude. It turns out that the only contributions come
The only nondegenerate instantons contributing to the geometry of Fig. 8.1 come from an annulus stretching along the degeneracy locus.

From open strings stretching along the degeneracy locus, i.e., along the edges of the graph that represents the geometry. This was found by Diaconescu et al. (2003a, 2003b) by using localization arguments, and derived by Aganagic et al. (2004) by exploiting invariance under deformation of complex structures. This result simplifies the problem enormously, and gives a precise description of all the non-degenerate instantons contributing in this geometry: they are annuli stretching along the fixed lines of the $T^2$ action, together with their multicovertings, and the $S^1$ in (8.12) is the circle that fibres over the edge connecting $M_1$ and $M_2$.

This is illustrated in Fig. 8.3. The action describing the dynamics of topological D-branes in the example we are considering above is then

$$S = S_{CS}(A_1) + S_{CS}(A_2) + \sum_{n=1}^{\infty} \frac{e^{-nr}}{n} \text{Tr} U_1^n \text{Tr} U_2^n,$$

(8.14)

where the $A_i$ are $U(N_i)$ gauge connections on $M_i = S^3$, $i = 1, 2$, and $U_i$ are the corresponding holonomies around the $S^1$. There is a very convenient way to write the free energy of the theory with the above action. First, notice that, by following the same steps that led to (2.156), one can write the operator (8.6) as

$$O(U_1, U_2; r) = \sum_R \text{Tr}_R U_1 e^{-\ell r} \text{Tr}_R U_2;$$

(8.15)

where $\ell$ denotes the number of boxes of the representation $R$. In the situation depicted in Fig. 8.3, we see that the boundaries of the annulus give a knot in $M_1$, and another knot in $M_2$. Therefore, the total free energy can be written as:

$$F = F_{CS}(N_1, g_s) + F_{CS}(N_2, g_s) + \log \sum_R e^{-\ell r} W_R(K_1) W_R(K_2),$$

(8.16)

where $F_{CS}(N_i, g_s)$ denotes the free energy of Chern–Simons theory with gauge group $U(N_i)$. These correspond to the degenerate instantons that come from
Fig. 8.4. The geometry of Fig. 8.1 can be cut into three pieces. The piece that contains the annulus gives by canonical quantization the state (8.17).

Each of the two-spheres. Of course, in order to compute (8.16) we need some extra information: we have to know what the knots $K_i$ are topologically, and also if there is some framing induced by the geometry. It turns out that these questions can be easily answered if we evaluate the path integral by a cut-and-paste procedure. The geometry of the knots is then encoded in the geometry of the degeneracy locus.

The evaluation proceeds as follows: we cut the geometry into three pieces, as indicated in Fig. 8.4, by Heegaard splitting the two three-spheres into solid tori. The first piece comes from a solid torus embedded in the total geometry with no insertion, obtained by splitting $M_1$. This gives the state $\langle 0 |$ in $\mathcal{H}_1^*(\mathbb{T}^2)$, where the subscript 1 refers to the Hilbert space of the $U(N_1)$ Chern–Simons theory on $M_1$. Similarly, the third piece is another solid torus from the splitting of $M_2$, and gives the state $|0\rangle_2$. The path integral with the insertion of $\mathcal{O}(U_1, U_2; r)$ produces the following operator in the canonical formalism:

$$\mathcal{O} = \sum_R |R\rangle_1 e^{-\ell r} \langle R| \in \mathcal{H}_1(\mathbb{T}^2) \otimes \mathcal{H}_2(\mathbb{T}^2), \quad (8.17)$$

where $|R\rangle$ is the Chern–Simons state that we constructed in Chapter 2, and we have introduced subscripts for the labels of the different Hilbert spaces. The gluing is made, as before, through the $S$ transformation on both sides, and the total partition function is then given by $\langle 0|S\mathcal{O}S|0\rangle_2$, so we find

$$Z(g_s, N_{1,2}, r) = \sum_R \langle 0|S|R\rangle_1 e^{-\ell r} \langle R|S|0\rangle_2. \quad (8.18)$$

Comparing with (8.16), we see that

$$W_R(K_i) = \frac{S_{0R}}{S_{00}}(g_s, t_i), \quad i = 1, 2, \quad (8.19)$$

where $g_s$ is the open string coupling constant $2\pi/(k_i + N_i)$ (which is the same for the two Chern–Simons theories, see (7.16)) and $t_i = g_s N_i$ are the ’t Hooft
parameters of the $U(N_i)$ Chern–Simons theories. This means that $K_{1,2}$ are unknots with zero framing in the three-spheres $M_{1,2}$, respectively. Geometrically, each of the boundaries of the annulus in Fig. 8.3 creates a Wilson line along the non-contractible cycle of the solid torus along which we split the three-sphere.

What happens now if we go through the geometric transition of Fig. 8.2? As in the case originally studied by Gopakumar and Vafa, the ’t Hooft parameters become the Kähler parameters $t_1, t_2$ in the toric diagram of Fig. 8.2. There is a third Kähler parameter $t$ in the toric geometry after the transition. It turns out that this parameter is related to the parameter $r$ appearing in (8.18) as follows:

$$t = r - \frac{t_1 + t_2}{2}. \quad (8.20)$$

This relation was first suggested by Diaconescu et al. (2003a). It is clearly needed in order to obtain a free energy of the expected form, with a well-defined limit as $t_1, t_2 \to \infty$. The total free energy of the resulting toric manifold can be obtained from (8.16) and (8.19), and it can be written in closed form as

$$F = \sum_{d=1}^{\infty} \frac{1}{d \left(2 \sin \frac{dq_2}{2}\right)^2} \left\{ e^{-dt_1} + e^{-dt_2} + e^{-dt(1 - e^{-dt_1})(1 - e^{-dt_2})} \right\}. \quad (8.21)$$

From this expression we can read off the Gopakumar–Vafa invariants of the toric manifold. Notice that (8.21) gives the free energy of closed topological strings at all genera. In other words, the non-perturbative solution of Chern–Simons theory (which allows us to compute (8.19) exactly) gives the all-genus answer for the topological string amplitude. This is one of the most important aspects of this approach to topological string theory.

One can consider other non-compact Calabi–Yau manifolds and obtain different closed and open string amplitudes by using these generalized geometric transitions (Aganagic et al., 2004; Diaconescu et al., 2003b). However, this procedure becomes cumbersome, since in some cases one has to take appropriate limits of the amplitudes in order to reproduce the sought-for answers. The underlying problem of this approach is that we are taking as our basic building block for the resolved geometries the tetravalent vertex that corresponds to the resolved conifold. It is clear, however, that the true building block is the trivalent vertex shown in Fig. 5.1, which corresponds to $\mathbb{C}^3$. In the next chapter we will see how one can define an amplitude associated to this trivalent vertex that allows one to recover any open or closed topological string amplitude for non-compact, toric geometries.

### 8.4 Matrix models and geometric transitions

So far, we have analyzed geometric transitions for Chern–Simons theory and type-A topological strings. Let us turn now to matrix models and type-B topological strings. In the previous chapter we have seen that the open topological
string amplitudes on the Calabi–Yau manifold $X_{\text{res}}^W$ are computed by a multicut matrix model whose planar solution (or, equivalently, its master field configuration) is given by a hyperelliptic curve

$$y^2 = W'(x)^2 - R(x).$$

(8.22)

Moreover, we also saw in (7.49) that the partial 't Hooft couplings can be understood as integrals around the $A_i$ cycles of this curve, with $i = 1, \ldots, n$. Let us now compute the variation of the free energy $F_0(t_i)$ when we vary $t_i$. The variation w.r.t. $t_i$ (keeping the $t_j, j \neq i$, fixed) can be obtained by computing the variation in the free energy as we move one eigenvalue from the cut $C_i$ to infinity. This variation is given by (minus) the integral of the force exerted on an eigenvalue, as we move it from the endpoint of the cut to infinity. The path from the endpoint of $C_i$ to infinity, which does not intersect the other cuts $C_j$, will be denoted by $B_i$. Taking into account (1.76), and the fact that $y(p)$ has no discontinuities outside the cuts $C_j$, we find

$$\frac{\partial F_0}{\partial t_i} = \int_{B_i} y(x) dx.$$  

(8.23)

Usually, this integral is divergent, but can easily be regularized by taking $B_i$ to run up to a cutoff point $x = \Lambda$, and subtracting the divergent pieces as the cutoff $\Lambda$ goes to infinity. For example, for the Gaussian matrix model one has

$$\frac{\partial F_0}{\partial t} = \int_{2\sqrt{t}}^{\Lambda} dx \sqrt{x^2 - 4t} = t(\log t - 1) - 2t \log \Lambda + \frac{1}{2} \Lambda^2 + O(1/\Lambda^2).$$

(8.24)

Therefore, the regularized integral gives $t(\log t - 1)$, which is indeed the right result. It is now clear that (7.49) and (8.23) look very much like the relations (3.69) that define the periods (therefore the prepotential) in special geometry. What is the interpretation of the appearance of special geometry?

Recall that our starting point was a Calabi–Yau geometry obtained as a blow-up of the singularity given in (7.27). However, and as we have already seen in the case of the conifold, we can smooth out singularities in algebraic geometry by deforming them rather than by resolving them. In the more general singularity (7.27), there are $n$ singular points that are locally identical to conifold singularities. We want to turn on a deformation that smooths them out by inflating a three-sphere at each of the singularities. It is easy to see that this requires turning on a generic polynomial $R(x)$ of degree $n - 1$. In this way we get the Calabi–Yau manifold

$$u^2 + v^2 + y^2 + W'(x)^2 = R(x).$$

(8.25)

We will call this geometry the deformed manifold $X_{\text{def}}^W$. This is a non-compact Calabi–Yau manifold with holomorphic three-form
\[ \Omega = \frac{1}{2\pi} \frac{dx dy du}{v}. \]  
(8.26)

The three-spheres created by the deformation can be regarded as two-spheres fibred over an interval in the complex \( x \)-plane. To see this, let us consider for simplicity the case of the deformed conifold, with \( W'(x) = x, R(x) = \mu \), where \( \mu \) is a real parameter:

\[ u^2 + v^2 + y^2 + x^2 = \mu. \]  
(8.27)

As we explained in Chapter 5, when we analysed (5.30), this geometry contains a three-sphere that is given by restricting the variables to have real values. If we now consider a fixed, real value of \( x \) in the interval \( -\sqrt{\mu} < x < \sqrt{\mu} \), we get a two-sphere of radius \( \sqrt{\mu - x^2} \). The sphere collapses at the endpoints of the interval, \( x = \pm \sqrt{\mu} \), and the total geometry of the two-sphere together with the interval \( [-\sqrt{\mu}, \sqrt{\mu}] \) is a three-sphere. In the more general case, the curve \( W'(x)^2 - R(x) \) has \( n \) cuts with endpoints \( x_{2i}, x_{2i-1}, i = 1, \cdots, n \), and the \( n \) three-spheres are \( S^2 \) fibrations over these cuts.

Let us now consider closed type-B topological strings propagating on \( X_W^{\text{def}} \). As we saw in Section 3.4, the genus-zero theory is determined by the periods of the three-form \( \Omega \) given in (8.26). We then choose a symplectic basis of three-cycles \( \hat{A}_i, \hat{B}^j \), with \( \hat{A}_i \cap \hat{B}^j = \delta_i^j \). Here, the \( \hat{A}_i \) cycles are the \( n \) three-spheres, and they project to cycles \( A_i \) surrounding the cut \( C_i = [x_{2i}, x_{2i-1}] \) in the \( x \)-plane. The \( \hat{B}_i \) cycles are dual cycles that project in the \( x \)-plane to the \( B_i \) paths (Cachazo et al., 2001). The periods of \( \Omega \) are then given by

\[ t_i = \frac{1}{4\pi} \int_{\hat{A}_i} \Omega, \quad \frac{\partial F_0}{\partial t_i} = \int_{\hat{B}_i} \Omega. \]  
(8.28)

It is easy to see that these periods reduce to the periods (7.49) and (8.23) on the hyperelliptic curve (8.22), respectively. Let us consider again the case of the deformed conifold (8.27), which is simpler since there is only one three-sphere. Let us compute the \( A \)-period over this three-sphere, which is an \( S^2 \) fibration over the cut \( [-\sqrt{\mu}, \sqrt{\mu}] \), by first doing the integral over \( S^2 \), and then doing the integral over the cut. Since \( v = \sqrt{\mu - x^2 - \rho^2} \), where \( \rho^2 = y^2 + u^2 \), the integral of \( \Omega \) over \( S^2 \) is simply

\[ \frac{1}{2\pi} \int_{S^2} \frac{dy dz}{\sqrt{\mu - x^2 - \rho^2}} = \sqrt{\mu - x^2}. \]  
(8.29)

Therefore, the \( A \)-period becomes

\[ t = \frac{1}{2\pi} \int_{-\sqrt{\mu}}^{\sqrt{\mu}} y(x) dx, \]  
(8.30)

where \( y \) is now given by \( y^2 + x^2 = \mu \). This is simply the \( A \)-period (7.48) (up to a redefinition \( y \rightarrow -iy \)). The general case is very similar, and one finally obtains
that the special geometry (8.28) of the deformed Calabi–Yau geometry (8.25) is equivalent to the planar solution of the matrix model, given by the hyperelliptic curve (8.22) and the equations for the partial 't Hooft couplings (7.49) and the planar free energy (8.23).

The physical interpretation of this result is that there is an equivalence between an open type-B topological string theory on the manifold $X^W_{\text{res}}$, with $N$ D-branes wrapping the $n$ spheres obtained by blow-up, and a closed type-B topological string theory on the manifold $X^W_{\text{def}}$, where the $N$ D-branes have disappeared. Moreover, the 't Hooft couplings $t_i$ in the open string theory become geometric periods in the closed string theory. This is another example of an open/closed string duality realized through a geometric transition, and proposed by Cachazo et al. (2001). The geometric transition also provides a closed string theory interpretation for the 't Hooft resummation of the matrix model: it corresponds to closed type-B topological string theory on $X^W_{\text{def}}$. Notice that, as happened with Chern–Simons theory on $S^3$, the master field controlling the planar limit (which is encoded in the planar resolvent, or equivalently in the quantity $y(\lambda)$) leads to an algebraic equation – the hyperelliptic curve (8.22) – that describes very precisely the target of the closed string theory dual.

This geometric transition can be generalized to more complicated situations, see, for example, Cachazo et al. (2001 and 2002), and Ferrari (2004).
In this chapter we put together the cut-and-paste approach to toric Calabi–Yau manifolds developed in Chapter 5 with the large-$N$ duality relating Chern–Simons theory and topological strings, to find a building block for topological string amplitudes on those geometries. This building block is called the topological vertex.

9.1 Framing of topological open string amplitudes

As we will see, the topological vertex is an open string amplitude, and in order to understand it properly we have to discuss one aspect of open string amplitudes that we have not addressed yet: the framing ambiguity. The framing ambiguity was discovered by Aganagic et al. (2002). They realized that when the boundary conditions are specified by non-compact Lagrangian submanifolds like the ones described in (5.28), the corresponding topological open string amplitudes are not unambiguously defined: they depend on a choice of an integer (more precisely, one integer for each boundary). For the Lagrangian submanifolds studied in Example 5.3, the framing ambiguity can be specified by modifying the geometry in an appropriate way. These Lagrangian submanifolds simply correspond to points in the edges of the trivalent graphs. Their geometry can be modified by introducing additional locations in the base $\mathbb{R}^3$ where the $T^2$ fibre degenerates, as we did before when we considered general deformed geometries. In this way, the Lagrangian submanifolds become compact $S^3$ cycles in the geometry, exactly as in Fig. 5.4. The additional lines are labelled by a vector $f = (p, q)$ where the $(-q, p)$ cycle degenerates. This procedure is illustrated in Fig. 9.1. It is useful to introduce the symplectic product of two vectors $v = (v_1, v_2)$ and $w = (w_1, w_2)$ as

$$v \wedge w = v_1 w_2 - v_2 w_1.$$  \hspace{1cm} (9.1)

This product is invariant under SL(2, $\mathbb{Z}$) transformations. If the original Lagrangian submanifold is located at an edge $v$, the condition for the compactified cycle to be a non-degenerate $S^3$ is

$$f \wedge v = 1.$$  \hspace{1cm} (9.2)

Clearly, if $f$ satisfies (9.2), so does $f - nv$ for any integer $n$. The choice of the integer $n$ is precisely the framing ambiguity found by Aganagic et al. (2002). In the case of the Lagrangian submanifolds of $\mathbb{C}^3$ that we constructed in Example 5.3, a particular choice of compactification (therefore of framing) that will be very important in the following is shown in Fig. 9.2.
What is the effect of a change of framing on open topological string amplitudes? A proposal for this was made by Aganagic et al. (2002) and further studied by Mariño and Vafa (2002), based on the duality with Chern–Simons theory. As we explained in Section 8.2, and stated in (8.11), vacuum expectation values of Wilson loops in Chern–Simons theory on $S^3$ compute open string amplitudes. On the other hand, we explained in Section 2.4, that Wilson loop correlation functions depend on a choice of framing. This suggests that the framing ambiguity of Chern–Simons theory corresponds to the ambiguity of topological open string amplitudes that we have just described. This also leads to a very precise prescription to compute the effect of a change of framing for open string amplitudes. Let us consider for simplicity an open string amplitude involving a single Lagrangian submanifold, computed for a framing $f$. If we now consider the framing $f - nv$, the coefficients $Z_R$ of the total partition function (4.74) change as follows

$$Z_R \to (-1)^{n\ell(R)} q^{n_R} Z_R,$$

(9.3)

where $\kappa_R$ was defined in (2.84), and $q = e^{ig}$. This is essentially the behaviour of Chern–Simons invariants under change of framing spelled out in (2.85). The extra sign in (9.3) is crucial to guarantee integrality of the resulting amplitudes, as was verified in Aganagic et al. (2002) and Mariño and Vafa (2002). If the open string amplitudes involves $L$ boundaries, one has to specify $L$ different framings $n_\alpha$, and (9.3) is generalized to

$$Z_{R_1 \cdots R_L} \to (-1)^{\sum_{\alpha=1}^L n_\alpha \ell(R_\alpha)} q^{\sum_{\alpha=1}^L n_\alpha \kappa_{R_\alpha}/2} Z_{R_1 \cdots R_L}.$$

(9.4)

### 9.2 Definition of the topological vertex

In Example 5.3, we showed that we can construct one Lagrangian submanifold in each of the vertices of the toric diagram of $\mathbb{C}^3$. Since each of these submanifolds has the topology of $\mathbb{C} \times S^1$, we can consider the topological open string amplitude associated to this geometry. The total open string partition function will be given by
DEFINITION OF THE TOPOLOGICAL VERTEX 159

\[ Z(V_i) = \sum_{R_1, R_2, R_3} C_{R_1 R_2 R_3} \prod_{i=1}^{3} \text{Tr}_{R_i} V_i, \]  
\hspace{1cm} (9.5)

where \( V_i \) is a matrix source associated to the \( i \)-th Lagrangian submanifold. The amplitude \( C_{R_1 R_2 R_3} \) is a function of the string coupling constant \( g_s \) and, in the genus expansion, it contains information about maps from Riemann surfaces of arbitrary genera into \( \mathbb{C}^3 \) with boundaries on \( L_i \). This open string amplitude is called the topological vertex, and it is the basic object from which, by gluing, one can obtain closed and open string amplitudes on arbitrary toric geometries. Since the vertex is an open string amplitude, it will depend on a choice of three different framings. As we explained in the previous section, this choice will be given by three different vectors \( f_1, f_2 \) and \( f_3 \) that specify extra degeneration loci and lead to a compactification of the \( L_i \). Let us see how to introduce this choice.

We saw in Section 5.2.1, that the \( \mathbb{C}^3 \) geometry can be represented by graphs involving three vectors \( v_i \). These vectors can be obtained from the set in Fig. 5.1 by an SL(2, \( \mathbb{Z} \)) transformation, and satisfy (5.14). We will then introduce a topological vertex amplitude \( C(v_i, f_i)_{R_1 R_2 R_3} \) that depends on both a choice of three vectors \( v_i \) for the edges and a choice of three vectors \( f_i \) for the framings. Due to (9.2) we require

\[ f_i \wedge v_i = 1. \]

We will orient the edges \( v_i \) in a clockwise way. Since wedge products are preserved by SL(2, \( \mathbb{Z} \)), we also have

\[ v_2 \wedge v_1 = v_3 \wedge v_2 = v_1 \wedge v_3 = 1. \]  
\hspace{1cm} (9.6)

However, not all of these choices give independent amplitudes. First of all, there is an underlying SL(2, \( \mathbb{Z} \)) symmetry relating the choices: if \( g \in \text{SL}(2, \mathbb{Z}) \), then the amplitudes are invariant under

\[ (f_i, v_i) \rightarrow (g \cdot f_i, g \cdot v_i). \]

Moreover, if the topological vertex amplitude \( C_{R_1 R_2 R_3}^{(v_i, f_i)} \) is known for a set of framings \( f_i \), then it can be obtained for any set of the form \( f_i - n_i v_i \), and it is given by the general rule (9.4)

\[ C_{R_1 R_2 R_3}^{(v_i, f_i - n_i v_i)} = (-1)^{\sum_i n_i \ell_i(R_i)} \sum_i n_i \kappa_{R_i} / 2 C_{R_1 R_2 R_3}^{(v_i, f_i)} \]  
\hspace{1cm} (9.7)

for all admissible choices of the vectors \( v_i \). Since any two choices of framing can be related through (9.7), it is useful to pick a convenient set of \( f_i \) for any given choice of \( v_i \), which we will define as the canonical framing of the topological vertex. This canonical framing turns out to be

\[ (f_1, f_2, f_3) = (v_2, v_3, v_1). \]

Due to the SL(2, \( \mathbb{Z} \)) symmetry and the transformation rule (9.7), any topological vertex amplitude can be obtained from the amplitude computed for a fixed choice.
of $v_i$ in the canonical framing. A useful choice of the $v_i$ is $v_1 = (-1, -1), v_2 = (0, 1), v_3 = (1, 0)$, as in Fig. 5.1. The vertex amplitude for the canonical choice of $v_i$ and in the canonical framing will be simply denoted by $C_{R_1 R_2 R_3}$. Any other choice of framing will be characterized by framing vectors of the form $f_i - n_i v_i$, and the corresponding vertex amplitude will be denoted by $C_{n_1 n_2 n_3}^{R_1 R_2 R_3}$.

Notice that $n_i = f_i \wedge v_i + 1$ (where $i$ runs mod 3).

One of the most important properties of $C_{R_1 R_2 R_3}$ is its cyclic symmetry. To see this, notice that the SL(2, $\mathbb{Z}$) transformation $g = TS^{-1}$ takes

$$(v_i, f_i) \rightarrow (v_{i+1}, f_{i+1}),$$

where again $i$ runs mod 3. It then follows that

$$C_{R_1 R_2 R_3} = C_{R_3 R_1 R_2} = C_{R_2 R_3 R_1}. \quad (9.8)$$

Finally, it will sometimes be useful to consider the vertex in the basis of conjugacy classes $C_{\vec{k}(1) \vec{k}(2) \vec{k}(3)}$, which is obtained from $C_{R_1 R_2 R_3}$ by

$$C_{\vec{k}(1) \vec{k}(2) \vec{k}(3)} = \sum_{R_i} \prod_{i=1}^{3} \chi_{R_i}(C(\vec{k}^{(i)})) C_{R_1 R_2 R_3}. \quad (9.9)$$

9.3 Gluing rules

We saw in Chapter 5 that any non-compact toric geometry can be encoded in a planar graph that can be obtained by gluing trivalent vertices. It is then natural to expect that the string amplitudes associated to such a diagram can be computed by gluing the open topological string amplitudes associated to the trivalent
vertices, in the same way that one computes amplitudes in perturbative quantum field theory by gluing vertices through propagators. This idea was suggested by Aganagic et al. (2004) and Iqbal (2002), and was developed into a complete set of rules by Aganagic et al. (2005). The gluing rules for the topological vertex turn out to be quite simple. Here we will state three rules (for a change of orientation in one edge, for the propagator, and for the matching of framings in the gluing) which make it possible to compute any closed string amplitude on toric, non-compact Calabi–Yau threefolds. They also make it possible to compute open string amplitudes for Lagrangian submanifolds on edges that go to infinity. The case of Lagrangian submanifolds on inner edges is also very easy to analyze, but we refer the reader to the paper by Aganagic, Klemm, Mariño, and Vafa (2005) for the details. A mathematical point of view on the gluing rules can be found in Diaconescu and Florea (2005) and Li et al. (2004).

1) Orientation. Trivalent vertices are glued along their edges, and this corresponds to gluing curves with holes along their boundaries. In order to do that, the boundaries must have opposite orientations. This change of orientation will be represented as an inversion of the edge vector, therefore in gluing the vertices we will have an outgoing edge on one side, say $v_1$, and an ingoing edge on the other side, $-v_1$. What is the corresponding effect on the amplitude $C_{\vec{k}(1)\vec{k}(2)\vec{k}(3)}$? Changing the orientation of $h$ boundaries along the first edge gives rise to a relative factor $(-1)^h$, where $h = |\vec{k}(1)|$. In the language of topological D-branes, this means that we are gluing branes to anti-branes (Vafa, 2001b). If we denote by $Q^t$ the representation whose Young tableau is the transpose of the Young tableau of $Q$ (i.e. is obtained by exchanging rows and columns), then one has the following relation between characters

$$\chi_{Q^t}(C(\vec{k})) = (-1)^{|\vec{k}|+\ell(Q)}\chi_Q(C(\vec{k})), \quad (9.10)$$

and from here one can easily deduce

$$C_{R_1R_2R_3} \to (-1)^{\ell(R_1)}C_{R_1^*R_2R_3},$$

as we invert the orientation of $v_1$. Of course, a similar equation follows for the other $v_i$.

2) Propagator. Since gluing the edges corresponds to gluing curves with holes along their boundaries, we must have matching of the number of holes and winding numbers along the edge. Therefore, the propagator must be diagonal in the $\vec{k}$ basis. After taking into account the change of orientation discussed above, and after dividing by the order of the automorphism group associated to $\vec{k}$ (which is $z_{\vec{k}}$), we find that the propagator for gluing edges with representations $R_1$, $R_2$ is given by

$$(-1)^{\ell(R_1)}e^{-\ell(R_1)t}\delta_{R_1R_2^*}, \quad (9.11)$$

where $t$ is the Kähler parameter that corresponds to the $\mathbb{P}^1$ represented by the gluing edge.
3) Framing. When gluing two vertices, the framings of the two edges involved in the gluing have to match. This means that, in general, we will have to change the framing of one of the vertices. Let us consider the case in which we glue together two vertices with outgoing vectors \((v_i, v_j, v_k)\) and \((v'_i, v'_j, v'_k)\), respectively, and let us assume that we glue them through the vectors \(v_i, v'_i = -v_i\). We also assume that both vertices are canonically framed, so that \(f_i = v_j, f'_i = v'_j\). In order to match the framings we have to change the framing of, say, \(v'_i\), so that the new framing is \(-f_i\) (the opposite sign is again due to the change of orientation).

There is an integer \(n_i\) such that \(f'_i - n_i v'_i = -f_i\) (since \(f_i \wedge v_i = f'_i \wedge v'_i = 1\), \(f_i + f'_i\) is parallel to \(v_i\)), and it is easy to check that

\[ n_i = v'_j \wedge v_j. \]

The gluing of the two vertex amplitudes is then given by

\[
\sum_{R_i} C_{R_i} R_k R_i e^{-\ell(R_i) t_i} (-1)^{(n_i+1)\ell(R_i)} q^{-n_i \kappa R_i / 2} C_{R'_i} R'_j R'_k, \tag{9.12}
\]

where we have taken into account the change of orientation in the \(v'_i, v'_j, v'_k\) to perform the gluing, and \(t_i\) is the Kähler parameter associated to the edge.

Given then a planar trivalent graph representing a non-compact Calabi–Yau manifold without D-branes, we can compute the closed string amplitude as follows: we give a presentation of the graph in terms of vertices glued together, as we did in Chapter 5. We associate the appropriate amplitude to each trivalent vertex (labelled by representations), and we use the above gluing rules. The edges that go to infinity carry the trivial representation, and we finally sum over all possible representations along the inner edges. The resulting quantity is the total partition function \(Z_{\text{closed}} = e^F\) for closed string amplitudes. We can slightly modify this rule to compute open string amplitudes associated to D-branes, in the simple case in which the Lagrangian submanifolds are located at the outer edges of the graph (i.e. the edges that go to infinity). In this case, we compute the amplitude by associating the representations \(R_1, \cdots, R_L\) to the outer edges with D-branes (in this case, only positive winding numbers contribute). The result is \(Z_{\text{closed}} Z_{R_1 \cdots R_L}\), where \(Z_{R_1 \cdots R_L}\) is the total open string partition function for fixed representations.

We will present some examples of this procedure in a moment. Before doing that, we will derive an explicit expression for the topological vertex amplitude by using a geometric transition.

### 9.4 Derivation of the topological vertex

In order to derive the expression for the vertex, we will consider the configuration drawn in the first picture in Fig. 9.3, which represents a geometry with an \(S^3\) together with three Lagrangian submanifolds \(L_1, L_2\) and \(L_3\). We also make a choice of framing for these Lagrangian submanifolds, indicated by arrows. The worldvolumes of the \(S^3\) and of \(L_1, L_3\) are parallel, and we consider topological
The configuration used to derive the topological vertex amplitude. The figure on the left shows the ‘deformed’ geometry. The figure on the right shows the ‘resolved’ geometry obtained by geometric transition. It contains a resolved conifold and a $\mathbb{P}^1$ of size $t$. We have also depicted in the figure on the left the open strings stretched among the different branes that contribute to the amplitude.

D-branes wrapped on $S^3$ and the $L_i$. The branes wrapping the $L_i$ are probe (spectator) branes, and the large-$N$ transition of the three-sphere leads to a geometry with a resolved conifold and three framed Lagrangian submanifolds. As we have seen in the examples above, the Kähler parameter of the $\mathbb{P}^1$ of the conifold $t$ is the 't Hooft parameter of the Chern–Simons theory on $S^3$. The resulting configuration is shown in the second picture of Fig. 9.3, and can be related to the topological vertex of Fig. 9.2 by (a) taking the Kähler parameter $t$ of the $\mathbb{P}^1$ in the resolved conifold to infinity (so that the extra trivalent vertex disappears), and by (b) moving the Lagrangian submanifold $L_1$ to the outgoing edge along the direction $(-1, -1)$. We will first compute the total open string amplitude by using the geometric transition, and then we will implement (a) and (b).

The open string theory on the $S^3$ is $U(N)$ Chern–Simons theory with some matter fields coming from the three non-compact Lagrangian submanifolds $L_i$. As we discussed in Section 8.3, there are strings stretching between the $S^3$ and $L_{1,2,3}$, and also strings between $L_1$ and $L_3$. These stretched strings are annuli along the degeneracy locus, and they are depicted in Fig. 9.3. The only spacetime excitation associated to these strings is a matter field in the bifundamental representation, and integrating it out corresponds to inserting an annulus operator like (8.13). When the two branes intersect on a circle (like the branes considered in Section 8.2, or like the $S^3$ and $L_2$ in this situation) the matter field is a boson (a scalar field). When the branes are parallel, however, it is a fermion. This is because we can turn the two parallel branes into a brane and an anti-brane intersecting along a circle. This leads to a Grassmann field, as explained in Vafa (2001b), and the resulting operator turns out to be
where \( r \) is the area of the annulus. In Fig. 9.3 the probe branes associated to \( L_2 \) and the dynamical branes on \( S^3 \) intersect on a circle, while the probe branes associated to \( L_1, L_3 \) are parallel to each other and to the dynamical branes. We then have the following operators:

\[
\begin{align*}
\sum_{Q_1} \text{Tr}_{Q_1'} U_1 e^{-\ell(Q_1) r_1} (-1)^{\ell(Q_1)} \text{Tr}_{Q_1} \hat{V}_1, \\
\sum_{R} \text{Tr}_{Q_2} U_2 e^{-\ell(Q_2) r_2} \text{Tr}_{Q_2} V_2, \\
\sum_{Q_3} \text{Tr}_{Q_3'} U_3 e^{-\ell(Q_3) r_3} (-1)^{\ell(Q_3)} \text{Tr}_{Q_3} V_3, \\
\sum_{Q} \text{Tr}_{Q'} V_1 e^{-\ell(Q) r} (-1)^{\ell(Q)} \text{Tr}_{Q} V_3,
\end{align*}
\]

which correspond to the annuli labelled with representations \( Q_1, Q_2, Q_3 \) and \( Q \) in Fig. 9.3. The parameters \( r, r_i, i = 1, 2, 3 \), are the lengths of the annuli, as in Section 8.3. The matrices \( V_2 \) and \( V_3 \) are sources corresponding to D-branes wrapping \( L_2, L_3 \), while \( \hat{V}_1 \), \( V_1 \) are sources for branes wrapping \( L_1 \) with opposite orientations, and represent Chan-Paton factors for open strings ending on opposite sides of \( L_1 \). \( U_1, U_2 \) are holonomies of the gauge connection on \( S^3 \) around the boundaries of the annuli with representations \( Q_1 \) and \( Q_2 \) (the boundary of the annulus carrying the representation \( Q_3 \) is geometrically identical to the boundary of the annulus associated to \( Q_1 \), and it gives the holonomy \( U_1 \) as well). Putting all these ingredients together, we find that the open string amplitude on the deformed geometry is given by

\[
Z(V_1, V_2, V_3) = \frac{1}{S_0} \sum_{Q_1, Q_2, Q_3, Q} (-1)^{\ell(Q_1) + \ell(Q_2) + \ell(Q_3)} \langle \text{Tr}_{Q_2} U_2 \text{Tr}_{Q_1'} U_1 \text{Tr}_{Q_3} U_1 \rangle \\
\times \text{Tr}_{Q_1} \hat{V}_1 \text{Tr}_{Q_1'} V_1 \text{Tr}_{Q_2} V_2 \text{Tr}_{Q_3} V_3, \tag{9.15}
\]

where we have factored out \( 1/S_0 \), the partition function of \( O(-1) \oplus O(-1) \rightarrow \mathbb{P}^1 \), and the factors involving \( r \) and \( r_i, i = 1, 2, 3 \), have been absorbed in the sources. The amplitude (9.15) is an open string amplitude with three boundaries, and \( V_i \) are the corresponding sources. Notice that the annuli that carry the representations \( Q_1, Q_3 \) are supported on the horizontal edge, while the annulus connecting \( L_2 \) to \( S^3 \) lies on the vertical edge. The horizontal and the vertical edge are related by an \( S \) transformation, therefore

\[
\langle \text{Tr}_{Q_2} U_2 \text{Tr}_{Q_1'} U_1 \text{Tr}_{Q_3} U_1 \rangle = \sum_{Q'} N_{Q_1' Q_3'}^{Q} \langle Q_2 | S | Q' \rangle = \sum_{Q'} N_{Q_1' Q_3'}^{Q} S_{Q_2 Q'}^{-1}, \tag{9.16}
\]
where we have fused together the $U_1$ holonomies. From a geometric point of view, this means that the boundaries of the annuli give a link in $S^3$ with the topology depicted in Fig. 2.10 (where the representations $R, R_1, R_2$ in Fig. 2.10 are now $Q_2, Q_3$ and $Q_1$, respectively), and the above expression is simply (2.94). We can also use the direct sum formula (2.93) for this invariant, and we finally arrive at the following expression for (9.15):

$$Z(V_1, V_2, V_3) = \sum_{Q_1, Q_2, Q_3, Q} (-1)^{\ell(Q_1) + \ell(Q_3) + \ell(Q)} \frac{W_{Q_1 Q_2} W_{Q_2 Q_3}}{W_{Q_2}} \times \text{Tr}_{Q_1} \hat{V}_1 \text{Tr}_{Q_2} V_1 \text{Tr}_{Q_3} V_2 \text{Tr}_{Q_2} V_3,$$

where $W_{R_1 R_2}$ is the Hopf link invariant defined in (2.55) and evaluated in (2.63). (9.17) gives the answer for the open topological string amplitude on the geometry depicted on the left in Fig. 9.3. We now incorporate the two modifications that are needed in order to obtain the topological vertex. First, we have to take $t \to \infty$. As we pointed out in Section 8.3, in order to have a well-defined limit it is crucial to renormalize the Chern–Simons expectation values. The relation (8.20) suggests the definition

$$W_{R_1 R_2} = \lim_{t \to \infty} e^{-\frac{\ell(R_1) + \ell(R_2)}{2} t} W_{R_1 R_2}.$$

This limit exists, since $W_{R_1 R_2}$ is of the form $\lambda^\frac{\ell(R_1) + \ell(R_2)}{2} W_{R_1 R_2} + O(e^{-t})$ (recall that $\lambda = e^t$). The quantity $W_{R_1 R_2}$, which is the ‘leading’ coefficient of the Hopf link invariant (2.55), is the building block of the topological vertex amplitude. It is a rational function of $q^{\pm \frac{1}{2}}$, therefore it only depends on the string coupling constant. We will also denote $W_R = W_{R0}$. The limit (9.18) was first considered by Aganagic et al. (2004).

In order to implement the second modification, we have to understand what is the effect on the amplitude of moving $L_1$ to the outgoing edge along $(−1, −1)$. In order to do that, we consider the simplified situation depicted in Fig. 9.4 where we only have two stacks of D-branes wrapping $L_{1,2}$. The amplitude can be easily computed following the arguments that led to (9.17), and one immediately obtains

$$Z(V_1, V_2) = \sum_{Q_1, Q_2} W_{Q_2 Q_1} (-1)^{\ell(Q_1)} \text{Tr}_{Q_1} V_1 \text{Tr}_{Q_2} V_2.$$

On the other hand, this is a particular case of the topological vertex amplitude with $R_1$ trivial, $f_2 = v_3$ and $f_3 = (0, −1)$, so there is a non-canonical framing on $v_3$ that corresponds to $n = −1$. We deduce

$$C_{0R_2 R_1} = W_{R_2 R_1} q^{k_{R_1}/2}.$$

The amplitude on the right-hand side of Fig. 9.4 is the canonically framed vertex $C_{R_1 R_2 0}$, but by cyclic symmetry this is equal to (9.19) with $R_2 \leftrightarrow R_1$. We
Fig. 9.4. Moving the Lagrangian submanifold with representation $Q_1$ to the outgoing edge.

conclude that in going from the left- to the right-hand side of Fig. 9.4 we must replace

$$(-1)^{ℓ(1)} W_{Q_{2}Q_{1}} \text{Tr}_{Q_{1}} \hat{V}_{1} \text{Tr}_{Q_{2}} V_{2} \rightarrow W_{Q_{2}Q_{1}} q^{\kappa_{Q_{2}}/2} \text{Tr}_{Q_{1}} V_{1} \text{Tr}_{Q_{2}} V_{2}. \quad (9.20)$$

After moving $L_{1}$ to the outgoing edge, all strings end on the same side of the corresponding branes, and this explains why we have replaced $\hat{V}_{1}$ by $V_{1}$ in the above formula. Collecting the coefficient of $\text{Tr}_{R_{1}} V_{1} \text{Tr}_{R_{2}} V_{2} \text{Tr}_{R_{3}} V_{3}$ in the partition function we compute $C_{R_{1},R_{2},R_{3}}^{0,0,-1}$. We then get the following expression for the topological vertex amplitude in the canonical framing:

$$C_{R_{1},R_{2},R_{3}} = q^{\kappa_{R_{2}}+\kappa_{R_{3}}} \sum_{Q_{1},Q_{3},Q} N_{Q_{1}Q_{2}} N_{Q_{2}Q_{3}} W_{R_{1}Q_{1}} W_{R_{2}Q_{2}} W_{R_{3}Q_{3}}. \quad (9.21)$$

This is the final expression for the topological vertex amplitude.

### 9.5 Useful formulae for the vertex

The basic ingredient in the explicit expression (9.21) for the topological vertex is the quantity $W_{R_{1}R_{2}}$ defined in (9.18). Using (2.63) it is possible to give an explicit expression for $W_{R_{1}R_{2}}$ that is useful in computations. It is easy to see that the leading coefficient of $\lambda$ in (2.63) is obtained by taking the leading coefficient of $\lambda$ in $\dim_{q} R_{2}$ and the $\lambda$-independent piece in (2.66). The generating function of elementary symmetric polynomials (2.64) then becomes

$$S(t) = \prod_{j=1}^{c_{R}} \frac{1+q^{j-1}t}{1+q^{1}t}. \quad (9.22)$$

where
\[
S(t) = \prod_{j=1}^{\infty} (1 + q^{-j} t) = 1 + \sum_{r=1}^{\infty} q^{-r(r+1)/2} t^r \prod_{m=1}^{r} m!
\] (9.23)

Notice that (9.22) is the generating function of elementary symmetric polynomials for an infinite number of variables given by \(x_j = q^{R_1 - j}, j = 1, 2, \ldots\). One then deduces that the \(\lambda \to \infty\) limit of \(q^{\ell(R_1)/2} s_{R_1}(x_i = q^{R_2 - i})\) is given by the Schur polynomial \(s_{R_1}(x_i = q^{R_2 - i + i + 1/2})\), which now involves an infinite number of variables \(x_i\). This finally leads to the following expression for \(W_{R_1 R_2}\):

\[
W_{R_1 R_2}(q) = s_{R_1}(x_i = q^{R_2 - i + 1/2}) s_{R_2}(x_i = q^{-i + 1/2}).
\] (9.24)

We will also write this as

\[
W_{R_1 R_2}(q) = s_{R_1}(q^{\rho + i R_2}) s_{R_2}(q^{\rho^\rho}),
\] (9.25)

where the arguments of the Schur functions indicate the above values for the polynomial variables \(x_i\). Using (9.25) one can write (9.21) in terms of skew Schur polynomials (Okounkov et al., 2003):

\[
C_{R_1 R_2 R_3} = q^{1/2(\kappa_{R_2} + \kappa_{R_3})} s_{R_1}^{(q^\rho)} \sum_Q s_{R_1/Q}(q^{\ell(R_2)+\rho}) s_{R_2/Q}(q^{\ell(R_2)+\rho}).
\] (9.26)

We can also write some of these quantities in the operator formalism developed in Section 2.6. Consider the coherent state

\[
|D\rangle = \exp\left(\sum_{n=1}^{\infty} \frac{1}{n(q^{n/2} - q^{-n/2}) \alpha_n - n}\right) |0\rangle.
\] (9.27)

Then, one has

\[
W_R = \langle R|D\rangle,
\] (9.28)

which is a simplified version of (2.153).

**Exercise 9.1** Show that \(W_R\) can be written as

\[
W_R = q^{\kappa_R/4} \prod_{\Box \in R} (q^{h(\Box)/2} - q^{-h(\Box)/2}),
\] (9.29)

where the product in the denominator is over all squares \(\Box\) in the Young tableau of \(R\), and \(h(\Box)\) is the corresponding hook length of the square (for help, see Macdonald (1995), I.3, Example 1). Deduce, as a corollary of (9.29), that

\[
W_{R^t}(q) = q^{-\kappa_R/2} W_R(q) = (-1)^{\ell(R)} W_R(q^{-1}).
\] (9.30)
As \( x \to 0 \) (where \( x \) is defined in (2.17)), we can use the hook-length formula for dimensions of the symmetric group (see for example Fulton and Harris, 1991) to obtain

\[
W_R \to \frac{x^{-\ell(R)}}{\ell(R)!} d_R,
\]

where \( d_R \) is the dimension of \( R \) as a representation of the symmetric group \( S_{\ell(R)} \).

In the case of \( W_{R_1 R_2} \), one has the following equality:

\[
W_{R_1 R_2} = q^{(\kappa_{R_1} + \kappa_{R_2})/2} \sum_R W_{R_1'R_2/R'} \frac{C_{R_1'R_2}^{R_1} (-1)^{\ell(R)}}{n(q_{R_1}^2 - q_{R_2}^2)} (\alpha_n + \alpha_{-n}).
\]

Exercise 9.2 Use (2.150) to prove that

\[
\sum_{R_1, R_2} q^{-\frac{1}{2}(\kappa_{R_1} + \kappa_{R_2})} W_{R_1 R_2} s_{R_1}(t_1) s_{R_2}(t_2) = \langle t_1 | t_2 \rangle \langle \omega(t_1) | D \rangle \langle \omega(t_2) | D \rangle,
\]

where \( \omega(t) \) is defined in (2.143), \( |D\rangle \) is the state defined in (9.27), and \( s_{R}(t) \) is the Schur polynomial. Deduce (9.34) from this.

9.6 Some applications

We will now present some examples of computation of topological string amplitudes by using the topological vertex.

Example 9.1 Resolved conifold. The toric diagram for the resolved conifold geometry is depicted in Fig. 5.2. Our rules give immediately:

\[
Z_{\mathbb{P}^1} = \sum_R C_{00R} (-1)^{\ell(R)} e^{-\ell(R)t} C_{R00}.
\]

Since \( C_{R00} = W_R = s_R(x_i = q^{-i+\frac{1}{2}}) \), we can use (2.157) and (9.30) to obtain

\[
Z_{\mathbb{P}^1} = \exp \left\{ -\sum_{d=1}^{\infty} \frac{e^{-dt}}{d(q^{\frac{d}{2}} - q^{-\frac{d}{2}})^2} \right\},
\]

in agreement with the known result (5.37).
**Example 9.2 Framed unknot.** Let us now consider an open string amplitude, corresponding to the resolved conifold with a Lagrangian brane in one of the external legs, and in arbitrary framing $p$. The open string amplitude is given by

$$Z(V,p) = \sum_Q Z_Q(p) \text{Tr}_Q V,$$  

(9.39)

where

$$Z_Q(p) = \frac{1}{Z_{\mathbb{P}^1}} (-1)^{\ell(Q)p} q^{\frac{w_Q}{2}} \sum_R C_{0QR} (-e^{-t})^{\ell(R)} C_{R00}.$$  

(9.40)

This can be explicitly evaluated (see Exercise 10.2) to give

$$Z_Q(p) = (-1)^{\ell(Q)p} q^{\kappa_Q p/2} e^{-\ell(Q)t/2} \text{dim}_q Q.$$  

(9.41)

The r.h.s. is essentially the Chern–Simons invariant of the unknot. The open string free energy is given by $F(V,p) = \log Z(V,p)$, which can be written as in (4.45). It turns out that the leading term of $F_{w,g}(p,t)$ as $t \to \infty$ (which we will simply denote by $F_{w,g}(p)$) can be computed in open Gromov–Witten theory by using Hodge integrals (Katz and Liu, 2002; see also Li and Song, 2002). The result is:

$$F_{w,g}(p) = (-1)^{p\ell+1} (p(p+1))^{h-1} \left( \prod_{i=1}^{h} \frac{\prod_{j=1}^{w_i-1} (j + w_ip)}{(w_i - 1)!} \right) \times \text{Res}_{u=0} \int_{\overline{M}_{g,h}} c_g(\mathcal{E}^\vee(u)) c_g(\mathcal{E}^\vee((-p - 1)u)) c_g(\mathcal{E}^\vee(pd)) u^{2h-4} \prod_{i=1}^{h} (u - w_i \psi_i).$$  

(9.42)

In this formula, $\overline{M}_{g,h}$ is the Deligne–Mumford moduli space, $\psi_i$ are the Mumford classes defined in (4.15), $\mathcal{E}$ is the Hodge bundle over $\overline{M}_{g,h}$, and its dual is denoted by $\mathcal{E}^\vee$. We have also written

$$c_g(\mathcal{E}^\vee(u)) = \sum_{i=0}^{g} c_{g-i}(\mathcal{E}^\vee)u^i,$$  

(9.43)

where $c_j(\mathcal{E}^\vee)$ are Chern classes, and similarly for the other two factors. On the other hand, the $t \to \infty$ limit of $Z_Q$ is

$$(-1)^{\ell(Q)p} q^{\kappa_Q p/2} W_Q.$$  

(9.44)

By equating the open Gromov–Witten result with the Chern–Simons result one finds a highly non-trivial identity that expresses the Hodge integrals appearing in (9.42) in terms of the $W_Q$s, as first noticed by Mariño and Vafa (2002). The explicit expression that one finds is
\[
\sum_{g=0}^{\infty} F_{k,g}^2 g^{2g-2+|\vec{k}|} = \\
(-1)^{p\ell_1-|\vec{k}|-\ell} \prod_j k_j! \sum_{n \geq 1} \frac{(-1)^n}{n} \sum_{\vec{k}_1, \ldots, \vec{k}_n} \delta_{\sum_{\sigma=1}^n \vec{k}_\sigma, \vec{k}} \prod_{R_s} \chi_{R_s}(C(\vec{k}_\sigma)) \sum_{n} \frac{1}{z_{\vec{k}_\sigma}}
\]

\[
\times e^{(p+\frac{1}{2})\kappa_{R_s} g_s/2} \prod_{1 \leq i < j \leq c_{R_s}} \sin \left[ \left( l_i^\sigma - l_j^\sigma + j - i \right) g_s/2 \right] \sin \left[ \left( j - i \right) g_s/2 \right] \]

\[
\times \prod_{i=1}^{c_{R_s}} \prod_{v=1}^{l_i^\sigma} \frac{1}{2 \sin \left[ (v - i + c_{R_s}) g_s/2 \right]},
\]

where we have relabelled \( w \to \vec{k} \) for positive winding numbers, as explained in Section 4.4.1, and \( \ell = \sum_j jk_j \). This equality is a very explicit mathematical prediction of the duality between Chern–Simons theory and topological string theory. Notice that the generating functional for the quantities (9.44)

\[
R(V, p) = \sum_Q q^{\kappa Q p/2} W_Q \text{Tr}_Q V,
\]

where we have not included the sign \((-1)^{\ell p}\), can be written as

\[
R(t, p) = \langle t| q^{\kappa p/2} |D \rangle,
\]

where the \( t_n \) for the coherent state are given by \( t_n = \text{Tr} V^n \), and \( \kappa \) is the operator defined in (2.137). If we now use (2.140), and take into account that

\[
\alpha_n|t\rangle = n \frac{\partial}{\partial t_n}|t\rangle, \quad \alpha_n|t\rangle = t_n|t\rangle,
\]

we find the following differential equation:

\[
\frac{\partial R(t, p)}{\partial p} = \frac{p}{2} \left( \sum_{n,m \geq 1} (n + m) t_n t_m \frac{\partial R(t, p)}{\partial t_n, t_m} + mnt_m + n^2 \frac{\partial^2 R(t, p)}{\partial t_n \partial t_m} \right).
\]

This is known as the cut and join equation, and together with the initial condition

\[
R(t, 0) = \exp \left( \sum_{n=1}^{\infty} \frac{t_n}{n\left( q^{\frac{p}{2}} - q^{-\frac{p}{2}} \right)} \right)
\]

it determines completely \( R(t, p) \).

The prediction (9.45) has been rigorously proved by Liu et al. (2003a) and by Okounkov and Pandharipande (2004), and has been shown to have many applications in Gromov–Witten theory (Liu et al., 2003b; Zhou, 2003b). The strategy of Liu, Liu and Zhou is precisely to show that the generating functional of the Gromov–Witten invariants (9.42) also satisfies the cut and join equation and the same initial condition (9.49).
Example 9.3 Local $\mathbb{P}^2$. The toric diagram is depicted in Fig. 5.3. Using again the rules explained above, we find the total partition function

$$Z_{\mathbb{P}^2} = \sum_{R_1,R_2,R_3} (-1)^{\ell(R_1)} e^{\sum_i \ell(R_i)} q^{\sum_i \kappa R_i} C_{0R_2^1R_3} C_{0R_1^2R_2} C_{0R_3^1R_1}, \quad (9.50)$$

where $t$ is the Kähler parameter corresponding to the hyperplane class in $\mathbb{P}^2$. Using that $C_{0R_2^1R_3} = W_{R_2} q^{-\kappa R_3/2}$ one recovers the expression for $Z_{\mathbb{P}^2}$ first obtained by Aganagic et al. (2004) by using the method of geometric transition explained in Section 8.3. Notice that the free energy has the structure

$$F_{\mathbb{P}^2} = \log \left\{ 1 + \sum_{\ell=1}^{\infty} a_\ell(q) e^{-\ell t} \right\} = \sum_{\ell=1}^{\infty} a_\ell^{(c)}(q) e^{-\ell t}. \quad (9.51)$$

The coefficients $a_\ell(q), a_\ell^{(c)}(q)$ can be easily obtained in terms of $W_{R_1R_2}$. One finds, for example,

$$a_1^{(c)}(q) = a_1(q) = -\frac{3}{(q^{1/2} - q^{-1/2})^2},$$

$$a_2^{(c)}(q) = \frac{6}{(q^{1/2} - q^{-1/2})^2} + \frac{1}{2} a_1(q^2). \quad (9.52)$$

If we compare to (4.27) and take into account the effects of multicovering, we find the following values for the Gopakumar–Vafa invariants of $O(-3) \rightarrow \mathbb{P}^2$:

$$n_1^0 = 3, \quad n_1^g = 0 \text{ for } g > 0,$$

$$n_2^0 = -6, \quad n_2^g = 0 \text{ for } g > 0, \quad (9.53)$$

in agreement with the results listed in (5.40). In fact, one can go much further with this method and compute the Gopakumar–Vafa invariants to high degree. We again see that the use of exact results in Chern–Simons theory leads to the topological string amplitudes to all genera. A complete listing of the Gopakumar–Vafa invariants up to degree 12 can be found in Aganagic et al. (2004). The partition function (9.50) can also be computed in Gromov–Witten theory by using localization techniques, and one finds indeed the same result (Zhou, 2003b).

Example 9.4 Local $\mathbb{P}^1 \times \mathbb{P}^1$. The local $\mathbb{P}^1 \times \mathbb{P}^1$ geometry is the non-compact Calabi–Yau manifold given by the four-manifold $\mathbb{P}^1 \times \mathbb{P}^1$ together with its anticanonical bundle. It also admits a symplectic quotient description of the form (5.17), this time with $N = 2$ and two Kähler parameters $t_1, t_2$. The charges $Q_j^{1,2}, j = 1, \cdots, 5$ can be grouped into two vectors

$$Q_1 = (-2, 1, 1, 0, 0),$$

$$Q_2 = (-2, 0, 0, 1, 1). \quad (9.54)$$
The toric diagram for this geometry can be easily worked out from this description, and it is represented in Fig. 9.5. Using the gluing rules we find the closed string partition function

$$Z_{\mathbb{P}^1 \times \mathbb{P}^1} = \sum_{R_i} e^{-(\ell(R_1)+\ell(R_3))t_1-(\ell(R_2)+\ell(R_4))t_2} q^{\sum_i \kappa_{R_i}/2} \times C_{0R_4R_1} C_{0R_1R_2} C_{0R_2R_3} C_{0R_3R_4}. \quad (9.55)$$

This amplitude can be written as

$$Z_{\mathbb{P}^1 \times \mathbb{P}^1} = \sum_{R_i} e^{-(\ell(R_1)+\ell(R_3))t_1-(\ell(R_2)+\ell(R_4))t_2} \times W_{R_4R_1} W_{R_1R_2} W_{R_2R_3} W_{R_3R_4}. \quad (9.56)$$

This is the expression first obtained by Aganagic et al. (2004). It has been shown to agree with Gromov–Witten theory by Zhou (2003b).

**Example 9.5** The closed topological vertex. Consider the Calabi–Yau geometry whose toric diagram is depicted in Fig. 9.6. It contains three $\mathbb{P}^1$ touching at a single point. The local Gromov–Witten theory of this geometry was studied by
FURTHER PROPERTIES OF THE TOPOLOGICAL VERTEX 173

Bryan and Karp (2005), who called it the closed topological vertex, and also by Karp et al. (2005). The vertex rules give the following expression for the total partition function:

$$Z(t_1, t_2, t_3) = \sum_{R_1, R_2, R_3} C_{R_1 R_2 R_3} W_{R_1}^1 W_{R_2}^2 W_{R_3}^3 (-1)^{\ell(R_1) + \ell(R_2) + \ell(R_3)} e^{-\sum_{i=1}^{3} \ell(R_i) t_i}.$$  \hspace{1cm} (9.57)

It turns out that this can be evaluated in closed form. By using (2.150) and (9.30) one can rewrite (9.57) as

$$\sum_{R_2} e^{-\ell(R_2) t_2} (-1)^{\ell(R_2)} \langle \omega(s^1) | O | R_2 \rangle \langle R_2^t | O | \omega(s^3) \rangle,$$  \hspace{1cm} (9.58)

where the coherent states $|s^i\rangle$, $i = 1, 3$, are defined by

$$s^i_n = \frac{e^{-n t_i}}{q^{n/2} - q^{-n/2}}, \quad i = 1, 3,$$

and the operator $O$ has been defined in (9.35). We also have the completeness relation

$$\sum_{R} (-1)^{\ell(R)} e^{-\ell(R) t} | R \rangle \langle R^t | = \exp \left( - \sum_{d=1}^{\infty} \frac{1}{d} \alpha_d \right) |0\rangle \langle 0|,$$  \hspace{1cm} (9.59)

where it is understood that the creation operators act on the vacuum $|0\rangle$, while the annihilation operators act on the vacuum $\langle 0|$. By using this relation, the evaluation of (9.58) reduces to products of coherent states. After a short calculation, one obtains (Karp et al., 2005)

$$Z(t_1, t_2, t_3) = \exp \left( - \sum_{d=1}^{\infty} \frac{1}{d(q^{d/2} - q^{-d/2})} \left( e^{-dt_1} + e^{-dt_2} + e^{-dt_3} - e^{-d(t_1 + t_2)} - e^{-d(t_1 + t_3)} - e^{-d(t_2 + t_3)} + e^{-d(t_1 + t_2 + t_3)} \right) \right),$$  \hspace{1cm} (9.60)

in agreement with the algebro-geometric results of Bryan and Karp (2005). Notice from the above expression that there is only a finite number of non-vanishing Gopakumar–Vafa invariants for the above geometry.

9.7 Further properties of the topological vertex

Since it was first introduced by Aganagic et al. (2005), the topological vertex has been shown to satisfy three remarkable properties: it has an underlying integrable structure (Aganagic et al., 2003), it has a natural combinatorial interpretation in terms of counting tridimensional Young tableaux (Okounkov et al., 2003), and it can also be reinterpreted in terms of an appropriate counting of sheaves.
on $\mathbb{C}^3$ (Iqbal et al., 2003; Maulik et al., 2003). We briefly review each of these properties.

1) Integrable structure. If we put $\text{Tr} V_i^n = t_i^n/n$ in (9.5), the resulting function of three infinite sets of ‘times’ $Z(t_i^n)$ turns out to be a tau function of the 3-KP hierarchy as constructed for example by Kac and van de Leur (2003). This integrability property is better understood in the context of mirror symmetry, where the computation of the vertex can be seen to reduce to a theory of free fermions in a Riemann surface with three punctures (Aganagic et al., 2003).

2) Combinatorial interpretation. Consider the problem of enumerating three-dimensional Young tableaux $\pi$ (also called plane partitions) with the following boundary condition: along the edges $x$, $y$, $z$, they end up in two-dimensional Young tableaux with the shapes $R_1$, $R_2$ and $R_3$, respectively. Let us introduce the partition function

$$C_{R_1 R_2 R_3} = \sum_{\pi} q^{\mid \pi \mid},$$

(9.61)

where $\mid \pi \mid$ is the number of boxes in $\pi$, and the sum is over plane partitions satisfying the above boundary conditions. It can be shown that, up to an overall factor independent of the $R_i$, the above partition function equals the topological vertex $C_{R_1 R_2 R_3}$ (Okounkov et al., 2003). This combinatorial interpretation of the topological vertex makes it possible to establish a precise correspondence between quantum topological strings on local, toric Calabi–Yau manifolds and the classical statistical mechanics of melting crystals (Okounkov et al., 2003; Iqbal et al., 2003; Saulina and Vafa, 2004).

3) Relation to the counting of ideal sheaves. Let $X$ be a Calabi–Yau threefold. An ideal sheaf $I$ defines a closed subscheme $Y$ through $O_Y = O_X/I$. This means, roughly speaking, that there is a subvariety $Y$ of $X$ defined by the zero locus of the equations that generate the ideal $I$. Given a two-homology class $\beta$, one can consider the moduli space of ideal sheaves $I_n(X, \beta)$ such that the holomorphic Euler characteristic of $Y$ is $n$ and with $O_Y$ supported on curves in the homology class $\beta$. This is a space of virtual dimension zero, and by counting the number of points with appropriate signs one can define the so-called Donaldson–Thomas invariant $\tilde{N}_{n,\beta}$. The Donaldson–Thomas partition function is given by

$$Z_{DT}(X) = \sum_{\beta} \sum_{n \in \mathbb{Z}} \tilde{N}_{n,\beta} Q^{\beta} (-q)^n,$$

(9.62)

where $q$ is interpreted here as a formal expansion parameter, and the notation for $Q^\beta$ is identical to the one in (3.53). Maulik et al. (2003) have shown that, when $X = \mathbb{C}^3$ (so that $Z_{DT}$ only depends on $q$), the Donaldson–Thomas partition function naturally depends on three sets of representations, and indeed agrees with the topological vertex amplitude $C_{R_1 R_2 R_3}(q)$, where $q$ in (9.62) is identified with $e^{i\varphi}$. They have also shown that the Donaldson–Thomas partition function satisfies the same gluing rules as the topological vertex, leading to the identification of the Donaldson–Thomas partition function $Z_{DT}(X)$ with the
topological string all-genus partition function $Z(X) = e^F$ for all non-compact, toric Calabi–Yau manifolds $X$. They also conjecture that the equality holds for all Calabi–Yau threefolds. Iqbal et al. (2003) rephrase the Donaldson–Thomas partition function for $\mathbb{C}^3$ (which computes the topological vertex) in terms of the counting of non-commutative $U(1)$ instantons in six dimensions. These developments seem to indicate that the topological vertex, apart from providing a powerful computational tool, plays a central logical role in the theory of Gromov–Witten invariants and opens the way to connections to other moduli problems in algebraic geometry.
This page intentionally left blank
In this final chapter we consider two applications of the results presented in this book. The first application is to Chern–Simons knot and link invariants, and the second application is to supersymmetric gauge theories in four dimensions.

10.1 Applications to knot invariants

We have seen in Chapter 4 that Gromov–Witten invariants can be written in terms of integer, or BPS invariants, and in Chapter 8 we have seen that knot invariants can be interpreted as open string amplitudes. Putting these two things together, one finds that Chern–Simons knot invariants can be expressed in terms of new integer invariants, and this leads to some surprising structure results for the Chern–Simons invariants of knots. In this section we will make these structure results more precise and we will give some examples.

In (4.67) the total free energy of open topological strings is written in terms of functions $f_R$ that have some hidden structure explained in (4.68). Given a knot $K$, the corresponding Chern–Simons free energy

$$F_{CS}(V) = \log Z_{CS}(V), \quad Z_{CS}(V) = \sum_{R} W_{R}(K) \text{Tr}_{R} V$$

is an open string generating functional, as explained in Section 8.2. We recall that $W_{R}(K)$ is the Chern–Simons invariant of the knot in representation $R$. We define the functions $f_{R}(q, \lambda)$ through the equation

$$F_{CS}(V) = \sum_{d=1}^{\infty} \sum_{R} \frac{1}{d} f_{R}(q^d, \lambda^d) \text{Tr}_R V^d.$$

This is simply (4.68), specialized to the case in which the open string generating functional is given by Chern–Simons knot invariants. It is easy to see that the $f_{R}$ polynomials are completely determined by this equation in terms of the Chern–Simons invariants $W_{R}$ (Labastida and Mariño, 2001, 2002). One can even derive the explicit formula:

$$f_{R}(q, \lambda) = \sum_{d,m=1}^{\infty} (-1)^{m-1} \frac{\mu(d)}{dm} \sum_{\vec{k}_1, \ldots, \vec{k}_m} \sum_{R_1, \ldots, R_m} \chi_{R}(C(\vec{k}_j)^{(d)})$$

$$\times \prod_{j=1}^{m} \frac{\chi_{R_2}f_{j}(C(\vec{k}_j))}{z_{k_j}} W_{R_2}(q^d, \lambda^d).$$

(10.3)
In this equation, we have introduced the following notation: given a vector $\vec{k}$ and an integer $d$, the vector $\vec{k}(d)$ is defined as follows: the $d\cdot i$ entry of $\vec{k}(d)$, $k_i(d)$, is given by $i$-th entry of $\vec{k}$, $k_i$, while the other entries are zero. Therefore, if $\vec{k} = (k_1, k_2, \ldots)$, then $\vec{k}(d) = (0, \ldots, 0, k_1, 0, \ldots, 0, k_2, 0, \ldots)$, where $k_1$ is in the $d$-th entry, $k_2$ is in the $2d$-th entry, and so on. The sum over $\vec{k}_1, \ldots, \vec{k}_m$ is over all vectors with $|\vec{k}_j| > 0$. In (10.3), $\mu(d)$ denotes the Möbius function. This function is defined as follows: if $d$ has the prime decomposition $d = \prod_{i=1}^{a} p_i^{m_i}$, then $\mu(d) = 0$ if any of the $m_i$ is greater than one. If all $m_i = 1$ (i.e. $d$ is square-free) then $\mu(d) = (-1)^a$. Some examples of (10.3) are the following:

\[
\begin{align*}
 f_{\Box}(q, \lambda) &= W_{\Box}(q, \lambda), \\
 f_{\Box \Box}(q, \lambda) &= W_{\Box \Box}(q, \lambda) - \frac{1}{2} (W_{\Box}(q, \lambda)^2 + W_{\Box}(q^2, \lambda^2)), \\
 f_{\Box}(q, \lambda) &= W_{\Box}(q, \lambda) - \frac{1}{2} (W_{\Box}(q, \lambda)^2 - W_{\Box}(q^2, \lambda^2)).
\end{align*}
\]

Therefore, given a representation $R$ with $\ell$ boxes, the function $f_R$ is given by $W_R$, plus some ‘lower order corrections’ that involve $W'_R$, where $R'$ has $\ell' < \ell$ boxes. One can then easily compute these functions starting from the results for vacuum expectation values of Wilson loops in Chern–Simons theory.

We know from (4.68) that the functions $f_R$ have a non-trivial structure, which can be summarized as follows:

\[
f_R = \sum_{R'} M_{RR'} \hat{f}_{R'},
\]

where the square matrix $M_{RR'}$ is given by

\[
M_{RR'} = \sum_{R''} C_{RR'R''} S_{R''}(q),
\]

and $\hat{f}_R$ are generating functionals for the integer invariants $N_{R,g,\beta}$:

\[
\hat{f}_R(q, \lambda) = \sum_{g \geq 0} \sum_{Q} N_{R,g,Q} (q^{-\frac{1}{2}} - q^{\frac{1}{2}})^{2g-1} \lambda^\beta.
\]

The square matrix $M_{RR'}$ that relates $f_R$ to $\hat{f}_R$ is invertible, therefore one can obtain the polynomials $\hat{f}_R$ from the $f_R$. This gives a very precise way to compute the BPS invariants $N_{R,g,Q}$ from Chern–Simons theory: compute the usual vacuum expectation values $W_R$, extract $f_R$ through the relation (10.3), compute $\hat{f}_R$, and expand them as in (10.7).
Table 10.1 BPS invariants for the trefoil knot in the symmetric representation.

<table>
<thead>
<tr>
<th>$g$</th>
<th>$Q = 1$</th>
<th>$2$</th>
<th>$3$</th>
<th>$4$</th>
<th>$5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-2$</td>
<td>$8$</td>
<td>$-12$</td>
<td>$8$</td>
<td>$-2$</td>
</tr>
<tr>
<td>1</td>
<td>$-1$</td>
<td>$6$</td>
<td>$-10$</td>
<td>$6$</td>
<td>$-1$</td>
</tr>
<tr>
<td>2</td>
<td>$0$</td>
<td>$1$</td>
<td>$-2$</td>
<td>$1$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Table 10.2 BPS invariants for the trefoil knot in the antisymmetric representation.

<table>
<thead>
<tr>
<th>$g$</th>
<th>$Q = 1$</th>
<th>$2$</th>
<th>$3$</th>
<th>$4$</th>
<th>$5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$-4$</td>
<td>$16$</td>
<td>$-24$</td>
<td>$16$</td>
<td>$-4$</td>
</tr>
<tr>
<td>1</td>
<td>$-4$</td>
<td>$20$</td>
<td>$-32$</td>
<td>$20$</td>
<td>$-4$</td>
</tr>
<tr>
<td>2</td>
<td>$-1$</td>
<td>$8$</td>
<td>$-14$</td>
<td>$8$</td>
<td>$-1$</td>
</tr>
<tr>
<td>3</td>
<td>$0$</td>
<td>$1$</td>
<td>$-2$</td>
<td>$1$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Exercise 10.1 Show that

$$M^{-1}_{RR'} = \sum_{\vec{k}} \frac{1}{z_{\vec{k}}^2} \chi_R(C(\vec{k})) \chi_{R'}(C(\vec{k}))(1/P_{\vec{k}}(q)), \quad (10.8)$$

where the polynomials $P_{\vec{k}}(q)$ are defined in (4.64).

The fact that one can extract the integer invariants $N_{R,g,Q}$ in the way we have just described is not obvious from the point of view of Chern–Simons theory, and it gives further support to the large-$N$ duality between Chern–Simons theory and topological strings. The structure result for the $f_R$ functions are a consequence of the BPS structure of open topological string amplitudes, and the fact that vacuum expectation values of Wilson loops satisfy this property is one of the most important checks we have of the topological string/gauge theory correspondence.

Example 10.1 BPS invariants of the trefoil knot. Let us consider the trefoil knot. By using the known values for the Chern–Simons invariants (2.103), and the defining relations for the $f$-polynomials (10.4), one can easily obtain:

$$f_{\square}(q, \lambda) = \frac{q^{-\frac{1}{2}}(q + \lambda^2)}{q^{\frac{1}{2}} - q^{-\frac{1}{2}}} \frac{\lambda(\lambda - 1)^2}{(1 + q^2)(q + \lambda^2 q - \lambda (1 + q^2))},$$

$$f_{\Box}(q, \lambda) = -\frac{1}{q^3} f_{\square}(q, \lambda). \quad (10.9)$$

We can now extract $N_{\square,g,Q}$, $N_{\Box,g,Q}$ from (10.9), by using (10.5). The results are presented in Table 10.1 and Table 10.2, respectively.

All the results above have been stated for knot invariants in the canonical framing. The situation for arbitrary framing was analysed in detail by Mariño and Vafa (2002). Suppose that we consider a knot in $S^3$ in the framing labelled
by an integer \( p \) (the canonical framing corresponds to \( p = 0 \)). Then, the integer invariants \( N_{R,g,Q}(p) \) are obtained from (10.3) but with the vacuum expectation values
\[
W_R^{(p)}(q, \lambda) = (-1)^{\ell(R)p} q^{\frac{1}{2}p\kappa_R} W_R(q, \lambda),
\]
(10.10)
where \( \kappa_R \) is defined in (2.84), and \( \ell(R) \) is the number of boxes in \( R \). One has, for example,
\[
\begin{align*}
    f^{(p)}_{\square}(q, \lambda) &= (-1)^p W_{\square}(q, \lambda), \\
    f^{(p)}_{\Box}(q, \lambda) &= q^p W_{\Box}(q, \lambda) - \frac{1}{2} (W_{\Box}(q, \lambda)^2 + (-1)^p W_{\Box}(q^2, \lambda^2)), \\
    f^{(p)}_{\bigcirc}(q, \lambda) &= q^{-p} W_{\bigcirc}(q, \lambda) - \frac{1}{2} (W_{\bigcirc}(q, \lambda)^2 - (-1)^p W_{\bigcirc}(q^2, \lambda^2)),
\end{align*}
\]
(10.11)
and so on. The rule (10.10) is, of course, the same one that we used for the vertex (9.3), and it is crucial for integrality of \( N_{R,g,Q}(p) \). Some integer invariants for the trefoil knot in arbitrary framing are listed in Mariño and Vafa (2002). Results for the BPS invariants of other knots and links can be found in Labastida et al. (2000) and Ramadevi and Sarkar (2001).

All the above results on the \( f \) functions, integer invariant structure, etc., can be extended to links of \( L \) components.

10.2 Applications to \( \mathcal{N} = 2 \) supersymmetric gauge theory

As we explained in Section 4.2, the prepotential of type-A topological strings on a Calabi–Yau manifold \( X \) gives the prepotential \( F_0(t) \) of the \( \mathcal{N} = 2 \) supergravity theory obtained after the compactification of type IIA string theory on \( X \). The supergravity theory has \( h^{1,1}(X) \) vector multiplets, and the low-energy action is described in \( \mathcal{N} = 1 \) superspace by the effective action (4.3).

The low-energy effective action describing vector multiplets of \( \mathcal{N} = 2 \) supersymmetric gauge theories in four dimensions is also described at leading order by a prepotential and an effective action of the form (4.3). As shown by Seiberg and Witten (1994), this prepotential can be obtained by using holomorphy and various consistency conditions. We will now briefly summarize some aspects of the Seiberg–Witten solution. More detailed accounts can be found in Álvarez-Gaumé and Hassan (1997) and Lerche (1997). Let us consider for simplicity the case of pure \( \mathcal{N} = 2 \) supersymmetric Yang–Mills theory with gauge group \( SU(2) \).

The \( SU(2) \) vector multiplet contains a complex scalar field with potential
\[
V = \frac{1}{2g^2} \text{Tr}[\phi^\dagger, \phi]^2,
\]
where \( g^2 \) is the coupling constant. The classical vacua are determined by \( V = 0 \), i.e. \([\phi, \phi^\dagger] = 0\). This means that after a gauge rotation one can set \( \phi = a\sigma_3 \), and the gauge symmetry is broken to \( U(1) \). There are two \( U(1) \) vector multiplets leading to \( W_\pm \) bosons that become massive through the Higgs mechanism. Classically, the only light degree of freedom is a \( U(1) \), \( \mathcal{N} = 2 \) vector multiplet that
we will denote by $A$. The complex scalar field of this multiplet will be denoted also by $a$. The masses of the $W^\pm$ bosons are $\sim |a|$. The low-energy theory obtained after integrating out the massive degrees of freedom is therefore described by a prepotential $F(a)$. This prepotential has perturbative and nonperturbative corrections. Due to $\mathcal{N} = 2$ supersymmetry, the perturbative correction can be obtained by a one-loop computation, and it is given by

$$F_{1\text{-loop}}(a) = \frac{i}{2\pi} a^2 \ln \frac{a^2}{\Lambda^2},$$

where $\Lambda$ is the dynamically generated scale of the theory (which is asymptotically free). The non-perturbative effects are due to instantons, and a $k$-instanton effect is weighted by $e^{-8\pi^2 k/g^2}$. By using the beta function and asymptotic freedom, this can be written as

$$e^{-8\pi^2 k/g^2} = \left(\frac{\Lambda}{a}\right)^{4k}.$$ (10.12)

Therefore, instanton corrections lead to an infinite series of the form

$$F_{\text{inst}}(a) = \frac{i}{2\pi} a^2 \ln \frac{a^2}{\Lambda^2} + \sum_{k=1}^{\infty} F_k \left(\frac{\Lambda}{a}\right)^{4k} a^2,$$ (10.13)

where $k$ is the instanton number. The coefficients $F_k$ are constants (this is due to the fact that, in a supersymmetric theory, instantons contribute to the path integral only through zero-modes), and the computation of the low-energy effective action amounts to computing $F_k$ for all $k$. The Seiberg–Witten solution provides an ansatz for the solution, encoded in an elliptic curve, which allows one to derive explicit expressions for $F_k$. On the other hand, these coefficients can also be computed directly by instanton calculus, and this line of work culminated in the work of Nekrasov (2003), who was able to give closed expressions for the instanton correction to the prepotential for many gauge groups and matter contents.

One natural question is then the following: is it possible to find a compactification of type IIA theory with the property that, after decoupling the gravity degrees of freedom, the resulting four-dimensional theory describes the low-energy effective action of a $\mathcal{N} = 2$ supersymmetric gauge theory? Since the prepotential of type IIA theory can be computed in many cases, this may lead to an independent method of computing instanton effects in gauge theory. This question was answered in the affirmative by Kachru et al. (1996) and further elaborated by Katz et al. (1997). We will focus here on the type IIA configuration which leads to $SU(2)$, $\mathcal{N} = 2$ Yang–Mills theory.

Consider type IIA theory on the non-compact Calabi–Yau manifold given by local $\mathbb{P}^1 \times \mathbb{P}^1$. We can regard this manifold as an $A_1$ resolved singularity in four dimensions fibred over a two-sphere. One of the $\mathbb{P}^1$s is the collapsing sphere of the $A_1$ singularity, while the other $\mathbb{P}^1$ is the two-sphere that gives the base
of the fibration. We will denote them by $\mathbb{P}^{1}_{f,b}$, respectively. The corresponding complexified Kähler parameters will be denoted by $t_f, t_b$. After compactification on the $A_1$ singularity, one obtains an $SU(2)$ gauge theory in six dimensions (plus gravity modes), with a massless $U(1)$ multiplet represented by $t_f$, and two massive $W^\pm$ multiplets of mass $t_f$. After further compactification on the $\mathbb{P}^{1}_b$ one obtains $\mathcal{N} = 2$ super-Yang–Mills, with a four-dimensional gauge coupling constant given by

$$\frac{1}{g^2} \sim t_b.$$  

(10.14)

In this theory, the mass of the $W^\pm$ multiplets in string units has to be identified with the $a$ modulus of $SU(2)$, $\mathcal{N} = 2$ super-Yang–Mills theory. We therefore have

$$t_f = 2\epsilon a,$$  

(10.15)

where $\epsilon = (\alpha')^{-1/2}$, and $\alpha'$ is the string tension. On the other hand, by integrating out the massive string modes at one-loop one obtains

$$e^{-1/g^2} = e^{-tb} = \left(\frac{\epsilon \Lambda}{2}\right)^4.$$  

(10.16)

The factors of 2 in (10.15) and (10.16) are chosen in order to agree with the conventions in Iqbal and Kashani-Poor (2003). Decoupling the gravity and string modes means taking $\alpha' \to 0$. We therefore find that the gauge theory limit appears when we take $\epsilon \to 0$ in (10.15) and (10.16). The Kähler parameters $t_f, t_b$ behave in this limit as

$$t_b \to \infty, \quad t_f \to 0,$$  

(10.17)

while $\Lambda, a$ are finite. Therefore, in order to obtain the Seiberg–Witten prepotential, one has to insert (10.15) and (10.16) in the prepotential $F_0(t_b, t_f)$ of local $\mathbb{P}^1 \times \mathbb{P}^1$ and take $\epsilon \to 0$.

In (9.56) we obtained a closed formula for all the functions $F_g$ of this geometry. Unfortunately, as it stands, this expression is not appropriate to take the limit above. In order to do that, we have to resum the expansion in $e^{-t_f}$, as noted by Klemm et al. (2003). This can be done systematically in the context of the vertex computations, as first shown by Iqbal and Kashani-Poor (2003). The starting point of the computation is the quantity

$$K_{R_1R_2}(Q) \equiv \sum_{S} W_{R_1S} W_{SR_2} Q^{\ell(S)},$$  

(10.18)

where the sum is over all representations $S$. By using (9.25) and (2.157), one easily finds (Iqbal and Kashani-Poor, 2004; Eguchi and Kanno, 2003):

$$K_{R_1R_2}(Q) = W_{R_1} W_{R_2} \exp\left(\sum_{n=1}^{\infty} \frac{Q^n}{n} f_{R_1R_2}(q^n)\right),$$  

(10.19)
where

\[ f_{R_1R_2}(q) = q e_{R_1}(q) e_{R_2}(q), \]  

and

\[ e_R(q) = \sum_{i=1}^{\infty} q^{R_i - i} = \sum_{i=1}^{\ell(R)} (q^{R_i - i} - q^{-i}) + \frac{1}{q - 1}. \]  

Another useful way to write this result is in terms of an infinite product,

\[ K_{R_1R_2}(Q) = \frac{1}{Z_{\mathbb{P}^1}} W_{R_1} W_{R_2} \prod_{k=1}^{\infty} (1 - q^k Q) C_k(R_1, R_2), \]  

where \( Z_{\mathbb{P}^1} \) is given in (9.38), and the coefficients \( C_k(R_1, R_2) \) are defined by

\[ \sum_k C_k(R_1, R_2) q^k = q e_{R_1}(q) e_{R_2}(q) - \frac{1}{(q^{\frac{1}{2}} - q^{-\frac{1}{2}})^2}. \]  

**Exercise 10.2** Prove that

\[ \sum_k C_k(R_1, R_2) = \ell(R_1) + \ell(R_2), \quad \sum_k k C_k(R_1, R_2) = \frac{1}{2} (\kappa_{R_1} + \kappa_{R_2}). \]  

Derive also the expression (9.41) from (9.40), following steps similar to those used in the derivation of (10.22) from (10.18).

Using (10.22), we find the following expression for (9.56):

\[ Z_{\mathbb{P}^1 \times \mathbb{P}^1} = \frac{1}{Z_{\mathbb{P}^1}} \sum_{R_1, R_2} Q_b^{\ell(R_1) + \ell(R_2)} \frac{W_{R_1}^2 W_{R_2}^2}{\prod_k (1 - q^k Q_f)^{2C_k(R_1, R_2)}}, \]  

where \( Q_{f,b} = e^{-t_{f,b}} \). We can now enforce the gauge theory limit in order to recover the Seiberg–Witten prepotential. This means setting

\[ Q_b = \left( \frac{\epsilon \Lambda}{2} \right)^4, \quad Q_f = e^{2\epsilon a} \]

and taking \( \epsilon \to 0 \). The expression (10.25) also depends on the string coupling constant \( x = ig_s \) through \( q = e^x \). In order to obtain an appropriate limit for all \( F_g(t) \), one also has to take

\[ x = -\epsilon \hbar. \]  

This is due to the fact that each \( F_g(t) \) scales with an overall factor \( \epsilon^{2-2g} \) in the Seiberg–Witten limit, and by redefining \( x \) in this way we cancel it. Therefore, in the limit \( \epsilon \to 0 \) the total partition function has the structure

\[ Z_{\mathbb{P}^1 \times \mathbb{P}^1}(t_f, t_b) \to Z_{SU(2)}(a, \Lambda, \hbar) = \exp \left( \sum_{g=0}^{\infty} F_g(a, \Lambda) \hbar^{2g-2} \right), \]  

where $F_0(a, \Lambda)$ is the Seiberg–Witten prepotential. In the limit one finds,

$$\frac{K_{R_1, R_2}(Q_f)}{K_{00}(Q_f)} \to \frac{1}{\epsilon^{2\ell(R_1)+2\ell(R_2)}} \left( \frac{d_{R_1} d_{R_2} (-1)^{\ell(R_1)+\ell(R_2)}}{\ell(R_1)! \ell(R_2)! \hbar^{\ell(R_1)+\ell(R_2)}} \prod_k (2a + \hbar)^{-C_k(R_1, R_2)} + O(\epsilon) \right),$$

where we have used (9.31) and (10.24). We then get, finally

$$Z_{SU(2)}(a, \Lambda, \hbar) = \sum_{R_1, R_2} \left( \frac{\Lambda^2}{4\hbar} \right)^{2(\ell(R_1)+\ell(R_2))} \frac{d_{R_1}^2 d_{R_2}^2}{(\ell(R_1)!)(\ell(R_2)!)} \prod_k (2a + \hbar)^{2C_k(R_1, R_2)}. \quad (10.29)$$

From here we can extract $F_0(a, \Lambda)$, which should be equal to the Seiberg–Witten $SU(2)$ prepotential. To check that this is indeed the case, one can write (10.29) in yet another way,

$$Z_{SU(2)}(a, \Lambda, \hbar) = \sum_{R_1, R_2} \left( \frac{\Lambda}{2} \right)^{4(\ell(R_1)+\ell(R_2))} \prod_{i, n=1,2} \prod_{i, j=1}^{\infty} \frac{a_{ln} + \hbar(l_i^{R_1} - l_j^{R_2} + j - i)}{a_{ln} + \hbar(j - i)}. \quad (10.30)$$

In this formula, we have defined $a_{12} = -a_{21} = 2a$, and $l^R$ is the vector of row lengths in the Young tableau associated to $R$. This expression can be derived from (10.25) or (10.29) by combinatorics, and we refer the reader to Iqbal and Kashani-Poor (2004) and Eguchi and Kanno (2003) for a detailed proof. The expression (10.30) is exactly the one obtained by Nekrasov (2003) by direct instanton computation, and shown to agree with the Seiberg–Witten solution by Nakajima and Yoshioka (2003) and Nekrasov and Okounkov (2003). We then see that topological strings and the topological vertex make possible the computation of gauge theory quantities in a rather simple way. The expressions (10.29) or (10.30) contain, in addition, an infinite number of so-called gravitational corrections to the prepotential, $F_0(a, \Lambda)$, which can also be computed by instanton methods with the techniques of Nekrasov.

The above computation can be extended in many ways: by considering $A_{N-1}$ fibrations over $\mathbb{P}^1$ one can obtain type IIA compactifications that lead (in the appropriate $\alpha' \to 0$ limit) to $\mathcal{N} = 2$ supersymmetric $SU(N)$ super-Yang–Mills theory in four dimensions, and the resulting partition function leads exactly to Nekrasov’s expression obtained from instanton counting (Iqbal and Kashani-Poor, 2003). One can also use these techniques to analyse gauge theories in four and five dimensions (Hollowood et al., 2003).
APPENDIX A

SYMMETRIC POLYNOMIALS

In this brief Appendix we summarize some useful ingredients of the elementary theory of symmetric functions. A standard reference is Macdonald (1995).

Let $x_1, \cdots, x_N$ denote a set of $N$ variables. The \textit{elementary symmetric polynomials} in these variables, $e_m(x)$, are defined as:

$$e_m(x) = \sum_{i_1 < \cdots < i_m} x_{i_1} \cdots x_{i_m}. \quad (A.1)$$

The generating function of these polynomials is given by

$$E(t) = \sum_{m \geq 0} e_m(x) t^m = \prod_{i=1}^{N} (1 + x_i t). \quad (A.2)$$

The \textit{complete symmetric function} $h_m$ can be defined in terms of its generating function

$$H(t) = \sum_{m \geq 0} h_m t^m = \prod_{i=1}^{N} (1 - x_i t)^{-1}, \quad (A.3)$$

and one has

$$E(t) H(-t) = 1. \quad (A.4)$$

The products of elementary symmetric polynomials and of complete symmetric functions provide two different basis for the symmetric functions of $N$ variables.

Another basis is given by the \textit{Schur polynomials}, $s_R(x)$, which are labelled by representations $R$. We will always express these representations in terms of Young tableaux, so $R$ is given by a partition $(l_1, l_2, \cdots, l_{c_R})$, where $l_i$ is the number of boxes of the $i$-th row of the tableau, and we have $l_1 \geq l_2 \geq \cdots \geq l_{c_R}$. The total number of boxes of a tableau will be denoted by $\ell(R) = \sum_i l_i$. The Schur polynomials are defined as quotients of determinants,

$$s_R(x) = \frac{\det x_j^{l_i + N - i}}{\det x_j^{N - i}}. \quad (A.5)$$

They can be written in terms of the symmetric polynomials $e_i(x_1, \cdots, x_N)$, $i \geq 1$, as follows:

$$s_R = \det M_R, \quad (A.6)$$

where
$M_{ij}^{R} = (e_{i} + j - i)$.  

$M_{R}$ is an $r \times r$ matrix, with $r = c_{R^t}$, and $R^t$ denotes the transposed Young tableau with row lengths $l^t_i$. To evaluate $s_{R}$ we put $e_0 = 1$, $e_k = 0$ for $k < 0$. The expression (A.6) is known as the Jacobi–Trudi identity.

A third set of symmetric functions is given by the Newton polynomials $P_{\vec{k}}(x)$. These are labelled by vectors $\vec{k} = (k_1, k_2, \cdots)$, where the $k_j$ are non-negative integers, and are defined as

$$P_{\vec{k}}(x) = \prod_j P_{k_j}^j(x),$$  \hspace{1cm} (A.7)

where

$$P_j(x) = \sum_{i=1}^{N} x_i^j,$$  \hspace{1cm} (A.8)

are power sums. The Newton polynomials are homogeneous of degree $\ell = \sum_j jk_j$ and give a basis for the symmetric functions in $x_1, \cdots, x_N$ with rational coefficients. They are related to the Schur polynomials through the Frobenius formula

$$P_{\vec{k}}(x) = \sum_{R} \chi_R(C(\vec{k})) s_{R}(x),$$  \hspace{1cm} (A.9)

where the sum is over all tableaux such that $\ell(R) = \ell$. By using orthogonality of the characters, we can invert the above formula as

$$s_{R}(x) = \sum_{\vec{k}} \frac{\chi_R(C(\vec{k}))}{z_{\vec{k}}} P_{\vec{k}}(x),$$  \hspace{1cm} (A.10)

where $z_{\vec{k}}$ is given in (2.111).

Let us now list some useful formulae involving Schur functions. First, define the skew Schur function $s_{R/R'}(x)$ by

$$s_{R/R'}(x) = \sum_{Q} N_{R'Q}^R s_{Q}(x).$$  \hspace{1cm} (A.11)

Then, one has the identities:

$$\sum_{R} s_{R/R_1}(x)s_{R/R_2}(y) = \prod_{i,j \geq 1} (1 - x_i y_j)^{-1} \sum_{Q} s_{R_2/Q}(x)s_{R_1/Q(y)},$$

$$\sum_{R} s_{R/R_1}(x)s_{R'/R_2}(y) = \prod_{i,j \geq 1} (1 + x_i y_j) \sum_{Q} s_{R_2'/Q}(x)s_{R_1'/Q}(y).$$  \hspace{1cm} (A.12)
REFERENCES


REFERENCES


REFERENCES


INDEX

BPS invariants, 93–96, 99–104, 177–180
BPS states, 92–104
Calabi-Yau manifold, 76, 81, 85, 106–119
conifold,
  deformed, 115–118, 128, 155
  resolved, 108, 111–113, 118–119, 168
  transition, 116–117
D branes, 92–100, 125,
  topological, 97, 105, 124, 128–129
descent equations, 77–78
diagram,
  double-line representation, 5, 34, 123
  planar, 6
  non-planar, 7
flat connection, 29
framing, 26, 44–46, 157–158, 161
gauge fermion, 80
ghost number, 79, 127–129
Gopakumar-Vafa invariants, 94–97,
  119–120
Gopakumar-Vafa transition, 143–144
graph, 30
Gromov-Witten invariants, 82, 90–91,
  96–97
group factor, 33–35
HOMFLY polynomial, 29
Lagrangian submanifold, 97
  associated to a knot, 145–147
large-\(N\) duality, 125–126
linking number, 28
multicovering, 94–95
matrix model,
  saddle-point analysis, 11–18
  Gaussian, 4, 7, 16, 19, 22
  multicut solution, 17, 124, 138–141
  one-cut solution, 12
\(\mathcal{N} = 2\) supergravity, 88
Ooguri–Vafa operator, 58, 145
orthogonal polynomials, 18–24
  Hermite polynomials, 19
  Stieltjes–Wigert polynomials, 63, 65–67
prepotential, 82, 85–86, 88–89, 178–179,
  181–182
resolvent, 13–17, 65
Seiberg–Witten solution, 178–182
skew relations, 49
special geometry, 85–86, 154–156
't Hooft coupling, 6, 35, 124, 142
topological charge, 74
topological field theory, 25, 76–78

Wilson loop, 27, 46–51, 145–147, 177–180